applied sciences

MDPI

Article

Real-Time Application for Generating Multiple Experiences from 360° Panoramic Video by Tracking Arbitrary Objects and Viewer's Orientations

Syed Hammad Hussain Shah, Kyungjin Han and Jong Weon Lee *

Department of Software Convergence, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea; hammad.shah38@gmail.com (S.H.H.S.); kjinn.han@gmail.com (K.H.)

* Correspondence: jwlee@sejong.ac.kr; Tel.: +82-10-9195-0423

Received: 24 February 2020; Accepted: 18 March 2020; Published: 26 March 2020

Featured Application: The proposed system specifically targets applications related to 360° multimedia content in virtual reality. Overall, this platform supports the generation of multiple interesting experiences or multimedia contents from a single 360° video, an easy to use authoring system for the 360° content providers, the efficient transferring of the generated 360° experiences to the prospective viewers in virtual reality, and enabling the viewers to share their personal experiences of the 360° multimedia content in virtual reality with the other viewers.

Abstract: We propose a novel authoring and viewing system for generating multiple experiences with a single 360° video and efficiently transferring these experiences to the user. An immersive video contains much more interesting information within the 360° environment than normal videos. There can be multiple interesting areas within a 360° frame at the same time. Due to the narrow field of view in virtual reality head-mounted displays, a user can only view a limited area of a 360° video. Hence, our system is aimed at generating multiple experiences based on interesting information in different regions of a 360° video and efficient transferring of these experiences to prospective users. The proposed system generates experiences by using two approaches: (1) Recording of the user's experience when the user watches a panoramic video using a virtual reality head-mounted display, and (2) tracking of an arbitrary interesting object in a 360° video selected by the user. For tracking of an arbitrary interesting object, we have developed a pipeline around an existing simple object tracker to adapt it for 360° videos. This tracking algorithm was performed in real time on a CPU with high precision. Moreover, to the best of our knowledge, there is no such existing system that can generate a variety of different experiences from a single 360° video and enable the viewer to watch one 360° visual content from various interesting perspectives in immersive virtual reality. Furthermore, we have provided an adaptive focus assistance technique for efficient transferring of the generated experiences to other users in virtual reality. In this study, technical evaluation of the system along with a detailed user study has been performed to assess the system's application. Findings from evaluation of the system showed that a single 360° multimedia content has the capability of generating multiple experiences and transfers among users. Moreover, sharing of the 360° experiences enabled viewers to watch multiple interesting contents with less effort.

Keywords: authoring system; 360° video experiences; object tracking; virtual reality; experience transfer; focus assistance; visualization; human–computer interaction (HCI)

1. Introduction

Virtual reality (VR) and 360° videos mutually deliver an immersive experience of visual content to users. As it represents the whole world, 360° images hold more information than normal images.

VR has increased user interest in 360° visual content by giving them a feeling of being present in the content [1]. Social media platforms such as Facebook and YouTube are also now open to uploads of 360° content [2]. Hence, consumer influence on entertainment through 360° videos is increasing day by day. One main reason is that they get a lot of interesting information to explore in a single visual content source.

Promotion of 360° visual content is rising, but at the same time many challenges are faced while watching a 360° video in VR. One of these challenges is the limited field of view (FOV) in head-mounted displays (HMD). In VR, a user can only focus on a specific region at a time due to limited FOV. Consequently, there is a huge possibility of missing interesting and important information while watching 360° visual content in VR. To overcome this problem, there is a concept of providing a narrative for watching a 360° video in VR. The narrative consists of information about a region of interest (ROI) during the whole 360° video. The narrative for the 360° video shapes the user's actual experience as it guides the user towards an ROI. An ROI depends on some interesting object or specific area in a video. There can be multiple ROIs at a time in a 360° video. Each viewer can focus on different ROIs in VR based on personal interest. Hence, there is a huge possibility of watching a single 360° video in different ways, resulting in various experiences based on the visual information watched in VR. By recording the users' experiences in VR, multiple narratives for a single 360° visual content source can be generated. Later, these experiences can be transferred to new viewers using focus assistance techniques in VR. It helps users focus on ROI at the right time, which saves them from missing important information in visual content. Moreover, 360° video content holds many interesting objects in it. Based on personal interest, viewers can choose various objects for watching in a panoramic video. Therefore, it is possible to generate a narrative for each interesting object by tracking it in the whole 360° video. To create a narrative based on an interesting object, it is required to track that object in the rest of the frames of a panoramic video. It is very difficult to select an object manually in all the frames of a video. Therefore, a fast and robust tracker is required to track the object of the user's interest in the video. There are many existing object trackers for normal videos. When these trackers are applied to a 360° panoramic video, many difficulties arise. The most noticeable problem is the movement of the object out of one side of the horizontal margin of the panoramic frame and reappearing on the other side (Figure 1a,b). Moreover, deformation of the object's shape occurs in panoramic images, which results in the loss of the object. In such scenarios, existing object trackers were not robust enough to re-identify and track objects in 360° panoramic videos. Moreover, the author of the narrative can be interested in any kind of object within a 360° video. Therefore, it is important to enable a user to create a narrative based on tracking of an arbitrary/unknown object. In the case of unknown/arbitrary object tracking in a 360° video, model-based and offline methods were unable to perform well because of a lack of prior information about each desired object due to deformations in the object's shape during the video [3].

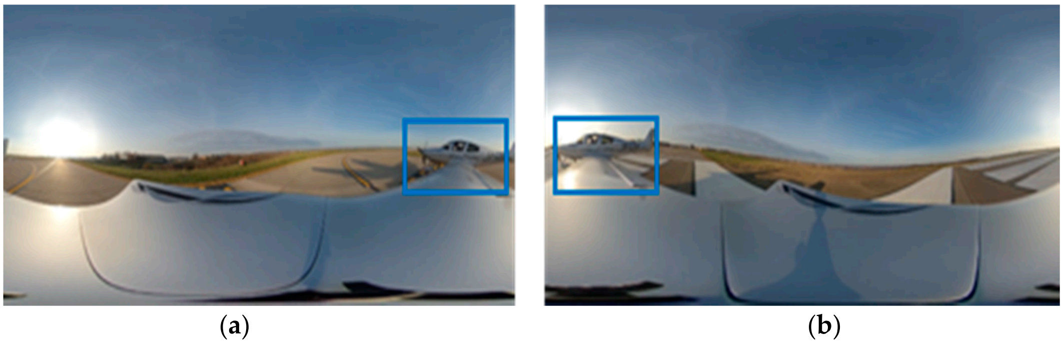

Figure 1. (**a**) Plane cockpit at far right of panoramic frame. (**b**) Plane cockpit reappearing on left side of panoramic frame.

In this paper, we propose an authoring system for generating and watching various experiences of a single 360° video through experience transferring in VR. It mainly consists of two major parts; a creator part and a viewer part as shown in Figure 2.

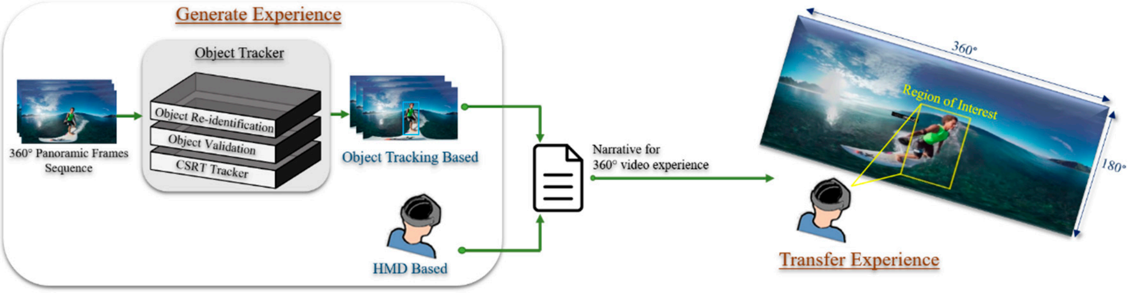

Figure 2. Our framework supports the generation of multiple experiences from a single 360° panoramic video. The first part generates a narrative for the 360° video experience. Later, the experience is transferred to prospective users using adaptive focus assistance.

The creator part focuses on generating the narratives for guiding experiences of 360° videos in VR, while the viewer part encompasses an adaptive focus assistance technique for efficient transferring of generated experiences to prospective users in VR. The creator part is further divided into two main parts. The first part is an HMD-based method. It generates experiences by recording the users' experiences of immersive videos with a VR device. In the second part, we propose a method for generating a narrative (in terms of viewing angle in 360° VR) by fast and robust tracking of arbitrary/unknown objects in panoramic videos. The user can change the selection and track different objects during a video. It helps in generating effective experiences for group performances. The main purpose of tracking an arbitrary object is that an author of experience can be interested in any kind of object within a 360° video. The contribution in the proposed work focuses on the tracking of an arbitrary object in a 360° video by using a recent object tracking approach called a discriminative correlation filter tracker with channel and spatial reliability (DCF-CSR) [4] as part of a whole system. The proposed methodology introduces additional components to the DCF-CSR tracker to adapt it for solving the object tracking problem in 360° panoramic videos. These components encompass the tracking validation and re-identification of an object in case of loss of object and occlusion handling in a 360° panoramic image sequence. A detailed description and implications of these components will be discussed in coming sections. The major contributions of the proposed system are as follows:

1. The sharing of users' personal experiences of 360° videos in VR among each other.
2. An authoring system for generating multiple experiences of a single 360° visual content source.
3. A real-time object tracker for 360° panoramic videos.
4. The provision of adaptive focus assistance for efficient transferring of 360° experiences in VR.

The structure of the paper is as follows. Section 2 provides an overview of existing systems related to our work. Section 3 states the research questions (RQs) and a brief description of the research methods used in this study. The proposed framework and its modules are explained in Section 4. Next, Section 5 presents the experimental results and evaluation of the proposed methodology. In Section 6, an overall evaluation of the system and answers to the RQs related to this study are discussed. Finally, Section 7 concludes the paper with limitations and future directions.

2. Related Work

Many researchers have worked on providing narratives and visual guidance in 360° video to focus on the ROI at the right time [2,5]. From a wide field of view in VR, a user can have an immersive experience of being in specific ROI in 360° video based on their interest [6]. Facebook launched a guide [7] to let the providers of 360° video content set narratives by highlighting points of interest over the course of complete visual content. In this way, it is possible to direct a user's attention towards an ROI in 360° VR. It is very imperative to show only interesting information from a wide view of 360° video to the viewer. However, a difficult aspect is to decide which information is more interesting, since this is very subjective. In 2016 and 2017, Yu-Chuan et al. proposed methods to generate a normal field of view (NFOV) video from 360° video based on regions of interesting information in panoramic frames [8,9]. The first issue is that the NFOV videos restrict immersiveness and the viewer's interaction with the visual content in 360° VR. Moreover, these methods generated a single summarized NFOV video based on interesting information in a panoramic video decided by their learned deep model. However, our system makes it possible to generate various interesting experiences from a single 360° video. Another issue in those approaches was that the user could not give any input in generating a NFOV video. It was similar to a black box to them. However, we enable users to generate 360° video experiences interactively based on their interests. Furthermore, those approaches were deep learning based and required high processing units to operate and were not in real time. In contrast, our system uses a very small amount of processing power and runs in real-time on a CPU, as shown by Figure 11 presented in the section on experimental results, and so is capable of running on devices with hardware constraints.

Panoramic cameras are rapidly gaining in popularity, resulting in the researchers' focus on image processing in 360° videos, e.g., Hu and Lin et al. [10] who designed an agent to control the viewing angle in panoramic sports videos. In 2013, Marcus and Werner proposed a methodology for real-time person tracking in panoramic videos [11]. Some other human detection and tracking methods are also in the literature [12–14], which show that object tracking is currently a hot research topic. A variety of techniques have been used for object tracking. Cui et al. [15] used differences in background and radial profile in dual camera systems for tracking objects. In 2016, Ahmad et al. [16] presented a method for tracking a polar object in 360° polar images. Furthermore, in 2017 they presented another framework for robust and fast object tracking in 360° polar videos [17]. However, the processing speed of their framework did not meet the requirement of being in real time. In 2018, Ahmad et al. made refinements to their previous work presented in [16] and designed an enhanced polar model [18] for fast object tracking in polar sequences. Still, with a speed of 9 fps, they could not get much closer to real time. Perhaps the work presented by Ahmed et al. in [3] is most relevant to our work. The method presented in this work focused on tracking of unknown objects in 360° polar sequences. After manual selection of the desired object, it used an online training method for tracking the object in the rest of the frames. However, the processing speed of all the object tracking algorithms described above was not in real time. There are multiple other object tracking algorithms including the Siamese network for object tracking presented by Luca Bertinetto et al. [19] and multi-domain convolutional neural networks (MDNet) for visual tracking by Hyeonseob Nam et al. [20]. Both presented deep learning-based object tracking algorithms with good accuracy and processing speed on GPU. However, their speed on a CPU did not meet the criteria of being in real time. The goal of our system was to be in real time on a CPU, which could be used by a wide range of users in resource constraint hardware systems. Therefore, we used a simple correlation filter-based tracker with additional modules to achieve efficient object tracking in 360° at high frame speed on a CPU. Moreover, in [21], the results proved that in achieving good accuracy at high framerates, simple correlation filter-based trackers are able to compete with complicated deep architecture-based trackers. Moreover, another problem in using deep learning-based trackers for tracking arbitrary objects is a lack of extensive 360° video data. Deformation of the object occurs in distorted stitched areas of panoramic frames, which makes it difficult to train a network for various appearances of an object.

Focus assistance plays a very important part in storytelling and guiding experiences in VR. There are various focus assistance techniques proposed to follow the ROI including [2,5]. In 2018, Ahmed Elmezeny et al. discussed the important immersion factors that influence effective storytelling in 360° videos [22]. However, there are various existing issues in these techniques. There is a high probability that the viewer may not properly follow the visual guidance and lag behind the actual 360° experience.

Consequently, they lose interesting information or the understanding of the overall story. Existing focus assistance techniques did not provide any mechanism to overcome this problem. However, we make our proposed focus assistance technique more adaptive to the user by providing the adjustment of frame rate and transparency of visual indicators based on the angular distance between viewer's head pose and angle required to view the ROI. In this way, a viewer can easily synchronize with the actual 360° experience with minimum loss of visual information.

3. Research Questions and Research Method

Our goal was to study the capability of 360° multimedia content in providing different interesting experiences to users. During the research, we developed different research questions (RQs), which are as follows:

- RQ1: Is it possible to generate multiple experiences from a single 360° video?
- RQ2: Does the provision of multiple experiences help viewers experience various interesting content in a single 360° video?
- RQ3: Does sharing of the users' personal experiences among each other help them easily find the ROIs while watching a 360° video?
- RQ4: Do existing simple object trackers perform well on panoramic 360° videos?
- RQ5: Does object tracking make authoring easy and more productive for the 360° video content providers?
- RQ6: Does the user adaptive focus assistance minimize the loss of visual information and efficiently direct the viewer towards an ROI in 360° VR?

Our system, whose functional modules are shown in Figure 2, aims to answer the RQs above. We applied different research methods in order to find the answers to these RQs. This study includes both qualitative and the quantitative methods to assess the system's behavior. A detailed discussion about the research methods is presented in Section 5. Moreover, the answers to the RQs stated above are discussed in Section 6.

4. Proposed Authoring System

Our proposed work was divided into two main parts. The first part focused on generating experiences of 360° video while the second part presented a focus assistance technique to transfer the experiences and enable the users to watch generated experiences efficiently.

4.1. Creator Part

This part generated the narrative for a 360° video experience and consisted of two main modules described as follows:

4.1.1. HMD Based

In this part, a user watched the 360° video in VR using an HMD. During the 360° video in VR, we continuously tracked the user's head orientation using a gyroscope sensor, which gives the viewing direction/user's viewpoint. Our system saved information on those viewpoints during a 360° video every 100 ms. At the end of watching a video, users could preview their own created experience (discussed in Section 4.2). After previewing, they could edit their experience by rewatching the 360° video. It was also possible for them to edit part of or their whole 360° video experience by rewatching it. They could reach a specific video part with the help of the skip or rewind operation of a video player. Finally, they could save their experience to the system after previewing and editing it until they got the desired 360° experience that they wanted to create. The saved experience would act as a narrative for shaping the experience of other viewers for that 360° video. Before saving the experience, a user had to describe it in a sentence by using a virtual keyboard and a controller. The description was provided to assist prospective users in selecting and watching a 360° video experience of their interest.

4.1.2. Object Tracker

This was based on the idea of selecting an arbitrary object or ROI of a user in the first frame and tracking it in the rest of a 360° panoramic video. Based on that object's location in each 2D panoramic frame, we calculated the viewing angle/gaze direction in terms of pitch and yaw to find the position of that object with respect to the 360° world space, which resulted in a narrative for watching the 360° video experience in VR. We developed a pipeline around the DCF-CSR tracker to make it robust for tracking objects in the sequence of 360° panoramic frames. As discussed in Section 1, while tracking the object of interest in panoramic videos, the most noticeable issues were movement of the object across the extreme edges of frames, loss of an object due to occlusions, shape deformation, and unexpected movement that occurred in stitched areas of panoramic frames. In our proposed framework for object tracking, we developed two main modules on top of the DCF-CSR tracker to make it robust for object tracking in 360° panoramic frame sequences, as shown in Figure 3. The first module focused on tracking and validation to ensure the success of the DCF-CSR tracker, while the second module mainly targeted object re-identification in case of the loss of the object. Another purpose of it was to handle the problem of occlusion where another object covered a desired object. Finally, the output of the tracking algorithm in terms of an object's location in the frame was treated as input to the final module of the viewing angle generator for 360° virtual reality. It generated the final narrative for 360° videos, which held the viewing angle in terms of pitch and yaw with respect to time over the complete course of an entire 360° video.

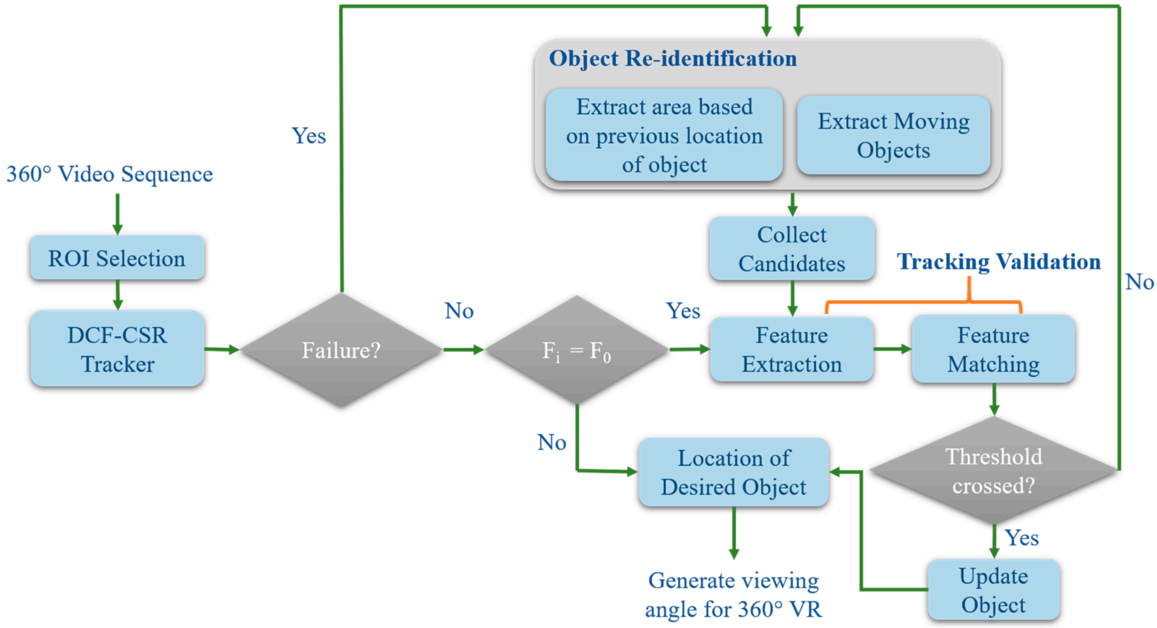

Figure 3. Flow chart of the proposed system for generating viewing angle through object tracking in 360° video.

A detailed description of all building blocks of the object tracker is as follows:

4.1.2.1. DCF-CSR Tracker

This is a short-term tracking algorithm that uses discriminative correlation filters (DCF) with two novelties: Channel reliability and spatial reliability [4]. Spatial reliability improved the filter's discriminative power for a better understanding of the background and channel reliability helped in the localization of a tracked region. This tracking algorithm used standard feature sets of histograms of oriented gradients (HoGs) and color names. It achieved a state-of-the-art performance on major datasets and ran close to real time on a CPU. We tested this tracker on various 360° videos. It performed well for normal scenarios when the object's motion speed was normal, there was no abrupt motion and the object did not move across boundaries of the 360° frames. It also performed well against occlusions.

4.1.2.2. Tracking Validation

While applying the DCF-CSR tracker in 360° videos, many failures occurred due to object shape deformation, abrupt motions in stitched areas of the 360° video and the object moving out of one side of a frame and appearing on the other side, as presented in the results section. This module checked whether the tracker was tracking the right object or if it was lost. It increased the reliability of the tracker by ensuring that the object was tracked correctly. Tracking validation took place after every 10 frames during the whole 360° video, as shown in Figure 3 using Equation (1). In Equation (1), 'F_i' denotes the current frame, 'F_0' represents the frame when validation had to be applied again, 'F_v' represents the frame when validation was last applied, and 'thre' represents the threshold value of frames to be skipped. We tested the threshold value of 'thre' each from 1 to 20 for our system and found the best performance at 'thre = 10' as the threshold value for frame skip count in our scenario. Hence, in our proposed method, the predefined threshold value for 'thre' was set to 10. Validation was applied if and only if the condition shown in Equation (1) became true.

$$F_i = F_0, \text{ where } F_0 = F_v + \text{thre} \tag{1}$$

This means a tracked object validated three times in a second. We decided to skip these frames because an object's appearance does not usually change much within a second. Moreover, applying validation after frames skip incurred less computational cost.

The tracking validation module consisted of two main components: Feature extraction and feature matching. Both components worked mutually. We used the oriented fast and rotated BRIEF (ORB) feature descriptor, which is a fusion of the FAST key point detector [23] and BRIEF descriptor [24]. ORB was faster than the SIFT and SURF descriptors, as well as being better in matching performance [25]. We used ORB feature descriptors due to its fast performance and it incurred less computational cost to our proposed algorithm in feature matching. Moreover, for feature matching, we used the fast library for approximate nearest neighbors (FLANN)-based descriptor matcher [26]. It finds the nearest neighbors based on feature descriptors provided from two images. After performing feature matching, we kept only good matches based on the matching distance of the features between two images, as shown in Figure 4. The threshold for matching distance was 0.8 out of a maximum value of 1. We kept this value high to make the validation accurate and sensitive to avoid wrong object selection.

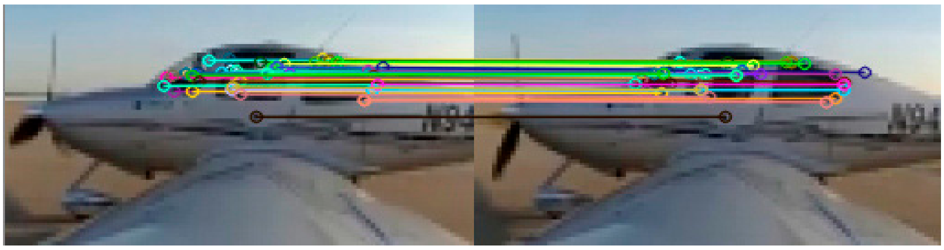

Figure 4. Fast library for approximate nearest neighbors (FLANN)-based matching of oriented fast and rotated BRIEF (ORB) features.

The two images used for feature extraction and feature matching were decided based on the location of an object given by the DCF-CSR tracker. The first image was of the last saved appearance of the tracked object, while the second image corresponded to the latest bounding box given by the tracker. If the feature matching met the threshold, the second image would take the place of the first image and became the latest appearance of an object for future matching in 360° video frames. This process took place continuously with a difference of 10 frames, as shown in Equation (1). If the feature matching did not meet the threshold, we kept performing feature matching with the object's last appearance for the next 60 frames with a skip of 10 frames. If it was still unable to meet threshold, the algorithm gave control to the object re-identification module with an input of the image with the object's last appearance.

4.1.2.3. Object Re-Identification

This is a popular topic in the field of computer vision and its applications. Many algorithms have been proposed for object re-identification, such as [27]. Due to the 360° view, an object stays inside the frame, unlike normal NFOV videos, unless and until it goes out of the sight of the 360° camera. This module executed when the DCF-CSR tracker gave a failure message or when the tracking validation module rejected the tracked region in a 360° video based on knowledge about the object's last appearance. The DCF-CSR tracker mainly lost the object when it went out and appeared on the opposite side of the panoramic frame. Since 360° panoramic videos normally have very high frame resolution, it took a very long time to scan a whole frame for object re-identification. As a result, we used two main techniques to make object re-identification faster for 360° videos.

The first technique was based on segmentation of the moving regions in a frame sequence. Whenever tracking failure occurred, we extracted the regions from the video where movements appeared by using one of the best techniques for real-time segmentation, called a MOG background subtractor [28]. It gave the moving regions as white pixels, whereas unchanged regions were black. The output image of the background subtractor had noise in it due to minute pixel changes. Therefore, noise removal was required to get only meaningful information from the resultant image. To remove the noise from the image and extract the moving objects from the 360° frame, we applied morphological operations. First, we applied erosion to the image with a very small kernel of size (1, 1). Erosion removed the white noise from the image. As erosion removed the white pixel noise, it also made the white regions very thin by decreasing the size of the foreground. Here, we needed to increase the thickness of the white pixels to get the exact moving regions with more information. For this purpose, we applied a dilate operation to the binary image to increase the white pixel area, as shown in Figure 5a. After applying the morphological operations, we drew contours around white pixels to extract the objects from the binary image. After getting the contours from the binary image, we had to extract moving objects from the original colored 360° frame. For this purpose, we applied a bitwise AND operation using the original image and binary image mask. In this way, we got the output image with moving objects as colored regions in it while the remaining area remained black (Figure 5b). After this step, we cropped the areas of moving regions based on the location of contours in the image (Figure 5c). These were considered as initial candidates for object selection.

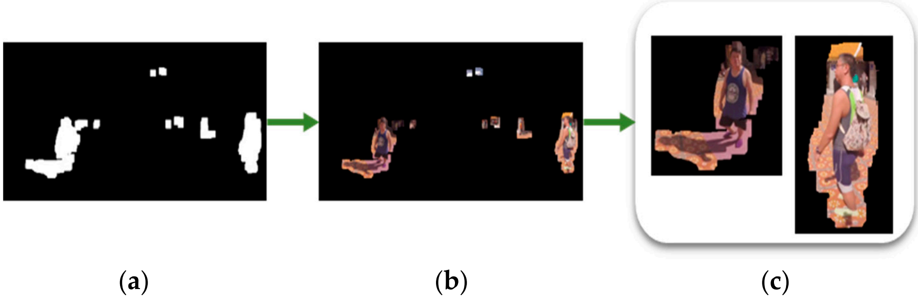

Figure 5. (**a**) Background subtraction and morphological operations, (**b**) bitwise AND, (**c**) initial candidate objects.

Based on the size of the tracked object in the last frame before losing track, we considered this size as a threshold for selecting the candidates for object selection. We determined the size of initial candidate objects based on their contour areas. Candidates that did not meet the size threshold were rejected at this stage. Then, the candidates from the section of extraction of moving objects were finalized to forward them to the next module.

The second technique for object re-identification was to restrict the search area within the 360° frame. This technique also produced candidates for object selection. Here, we restricted the area for object searching around the previous location of the object. We cropped a window around the central pixel coordinate of the object's previous location and performed searching for the desired object only in this area. One of the benefits of this technique was that it reduced the computational cost and kept the algorithm's processing fast. Objects did not move very quickly in successive frames, so it did not affect the tracking performance. Ahmed et al. also used this technique in their work to reduce the computational cost [3]. This restricted area was also considered as the candidate for object selection. Next, all the candidates from the moving object extraction technique and search area restriction technique were given as input to the validation module to find and select an object holding the highest matching value with the target object for further tracking. The validation module again applied feature matching between all the candidates and the target object to find the best exact match that crossed the threshold, as well as having the highest matching value. In the case of not meeting the threshold, this process is repeatedly applied after every 10 frames until the target object is re-identified. In case of failure in re-identification, the user is given control to reselect the ROI manually.

4.1.2.4. Viewing Angle Generator for 360° VR

The main purpose of this system was to generate experiences of 360° videos for VR. Therefore, the final module of the object tracking pipeline was to generate the viewing angle for a tracked object's location for 360° VR. The viewing angle depends on head rotation in VR. This rotation measure consists of roll, pitch, and yaw, as shown in Figure 6. The viewing angle decides the target area to focus on by the viewer in VR. Thus, after tracking the location of a desired object in the 360° video frame, we calculated the required viewing angle to focus on and watch this object in VR. For viewing angle, we calculated pitch and yaw from x and y coordinates in the panoramic frame. The x and y coordinates used were the central pixel coordinates of the object's location in the panoramic frame. We calculated the viewing angle for the target object over the course of a whole 360° video and stored it in an external file with time information. At the end of the video, those stored values of viewing angle within 360° resulted in a narrative for watching the panoramic video in VR. The generated narrative guided and provided an experience of the 360° video to the viewer in VR.

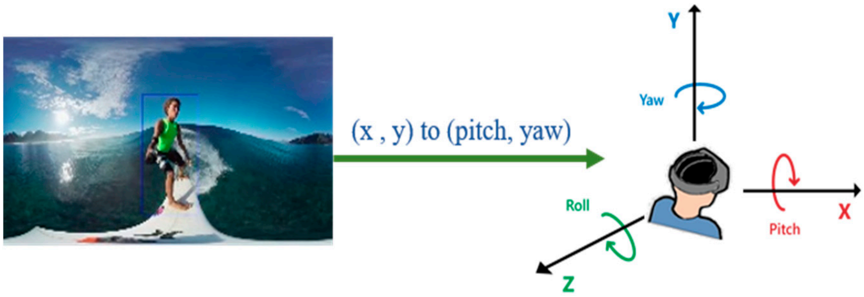

Figure 6. Calculation of viewing angle for 360° virtual reality (VR).

4.2. Viewer Part

This part of the system also held great importance because it ensured the efficient transferring of experience to the users based on the narrative provided for 360° video. In this part of system, we developed a highly adaptive focus assistance technique for VR. The purpose of the focus assistance was to guide the user's viewing direction towards the target ROI. The implemented technique was visual guidance over the top of a 360° VR video player for watching the generated experiences. There were many focus assistance techniques including autorotation. We gave priority to visual guidance over autorotation to avoid motion sickness, as discussed by [29]. With the help of the viewer part of the system, a user could select the desired 360° video experience to watch in VR. Based on the selected experience, our focus assistance technique guided the viewer towards the intended target by using visual indicators. There were many challenges while providing focus assistance in VR for transferring experiences. One major challenge was the synchronization of the viewer with visual guidance. Hence, to make better synchronization between the viewer and 360° video experience, we implemented some methods that made our focus assistance more adaptive to a user and better than the existing techniques. The descriptions of these methods are as follows:

(a) We decreased the opacity of the visual indicator as the viewer's angular distance decreased between the center of the current FOV and intended target. It gave awareness to the viewer about how close he/she was to the intended target. We considered an intended target as one clearly visible by the viewer based on the angular distance (15° for pitch and 20° for yaw) between current head rotation and intended head rotation. We represent the current head pose as 'H_1', the head pose required to focus on an intended target as 'H_2', and the angular distance between them as 'θ' (Figure 7a). If the opacity of the visual indicator is denoted by 'O', the direct relation between 'O' and 'θ' can be represented as in Equation (2).

$$O \propto \theta \tag{2}$$

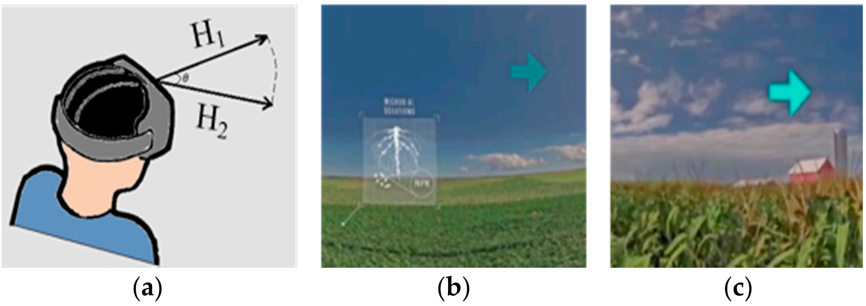

Figure 7. (**a**) This represents the angular distance between current head pose and head pose required to focus on an intended target. The visual indicator has: (**b**) Low opacity due to small angular distance, (**c**) high opacity due to large angular distance.

(b) The second technique focused on the minimization of the loss of visual information, which could occur when a user lagged behind the focus assistance. In this method, we adjusted the playback rate of the 360° panoramic video player based on the angular distance. The maximum angular distance between two points could be 180° for pitch and 360° for yaw. We declined the normal playback rate of the video player by 25% with each rise of 60° in angular displacement. With the decline in playback rate, a user easily felt that he/she was lagging behind the actual experience. It led them to get synchronized with the experience easily and follow the visual guidance efficiently.

4.3. Multiple Experiences from Single 360° Video

The main purpose of the system was to generate multiple experiences from a single 360° video. Users could be interested in different information within a single panoramic frame during a video. The variety of users' interests could lead to the generation of multiple experiences from one visual content source. These various experiences could be saved by using the 'Creator Part' presented in Section 4.1. Figure 8 shows the different ROIs in a frame, focused by different users while watching a 360° video in VR.

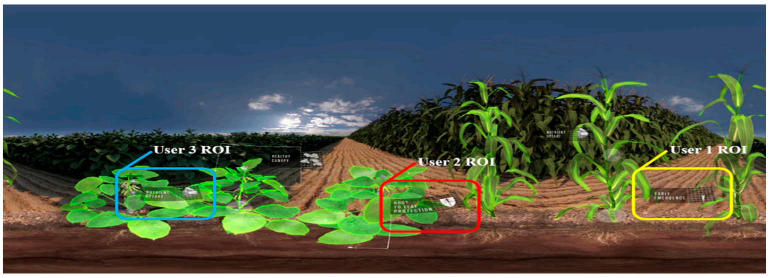

Figure 8. Multiple region of interest (ROIs) in a single 360° video frame based on users' personal interests: (a) User 1 region of interest, (b) user 2 region of interest, and (c) user 3 region of interest. For a single 360° video, there are three different regions of interest based on a user's perception.

5. Experimental Results

To test the overall performance of our system, we divided our experiments into two main parts. The first one was based on a user survey, while the second comprehended the accuracy of the pipeline developed for arbitrary object tracking. Details of both experiments are as follows:

5.1. User Survey

The main purpose of this user survey was to estimate the accuracy of transferring generated experiences to the users and to get an idea about user satisfaction with the proposed authoring system. It mainly tested the HMD-based approach for generating 360° experiences and efficiency of transferring of experience using the proposed focus assistant. In this test, we asked the users to create experiences from 360° video using the HMD-based approach (Section 4.1.1). After creating an experience, the user was asked to follow their own created experience with the help of focus assistance. After creating and viewing their own experiences, we requested users to fill out a designed questionnaire to report the performance of our proposed method. For the user survey, we recruited 25 paid participants. The youngest one was 23 years old, while the eldest user was 31 years old. Thirteen of them rarely used VR devices, four were frequent users and eight had no experience with VR. Participants received nonmonetary compensation of a coupon for a meal in the university's cafe (worth 10 USD). The designed questionnaire was influenced by previous research work and a simulator sickness questionnaire (SSQ) [30]. We categorized our questionnaire into three main categories, usability, efficiency, and satisfaction. The score distribution for each question was from 1 for strongly disagree to 5 for strongly agree. Finally, we calculated the overall score given by the users in each category using the mean opinion score (MOS) technique [31]. The results for each category are as follows:

5.1.1. Usability

This category of the survey mainly focused on the evaluation of convenience in using the system, as well as its learnability. Of the total participants, 53% strongly agreed that the proposed authoring system was very easy to understand, while 41% just agreed, and 6% were neutral (mean (M) = 4.47, standard deviation (SD) = 0.082).

5.1.2. Efficiency

This emphasized evaluating the overall performance of the authoring system in creating and transferring 360° video experiences. It mainly focused on assessing our focus assistance technique. The aim of this was to find out the percentage of actual experience successfully transferred to the user by the proposed method. More than 90% similarity between their actual experience and the experience assisted by the proposed focus assistance technique was reported by 72% of participants, while 28% of participants reported more than 75% similarity. In evaluating the efficiency of the system in generating and transferring 360° video experiences using an HMD device, 68% of the participants strongly agreed that the system was highly efficient, 24% agreed, and 8% remained neutral (M = 4.6, SD = 0.12).

5.1.3. Satisfaction

This quantified satisfaction with the developed authoring system and interest in using it in the future. The positive point of our proposed system was that it required a minimal amount of knowledge to use it. Therefore, a good percentage (72%) of participants were very satisfied (Strongly Agree) while 28% agreed (M = 4.78, SD = 0.06).

Details of the questionnaire and results of the user survey are shown in Table 1.

Table 1. Overall results of user survey.

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Average Score
Score	**1**	**2**	**3**	**4**	**5**	**-**
Usability						
The interface of the system to parse video is easy to understand	0	0	1	13	11	4.4
Tasks such as previewing and editing the recorded experience are easy to perform	0	0	1	11	13	4.48
Convenient to use	0	0	2	11	12	4.4
It does not take a very long time to learn the system and perform operations	0	0	2	6	17	4.6
Overall Results for Usability	0%	0%	6%	41%	53%	-
Efficiency						
When you previewed your experience, the recorded experience was similar to the actual one	0	0	0	7	18	4.72
Knowing that your experience is being recorded while watching a 360° video does not restrict enjoyment	0	0	5	4	16	4.44
The experience with the system was smooth overall and performing operations such as parsing a video did not cause long delays in watching the video	0	0	1	7	17	4.64
Overall Results for Efficiency	0%	0%	8%	24%	68%	-
Satisfaction						
You can record your 360° video experience with very little knowledge about creating 360° video experiences	0	0	0	4	21	4.84
You will recommend this system to others	0	0	0	7	18	4.72
Overall Results for Satisfaction	0%	0%	0%	22%	78%	-
Overall Result						4.58 (91.6%)

5.2. Object Tracker Performance

Object tracking is also an important part of our proposed authoring system. It generated narratives from 360° video based on arbitrary object tracking for delivering the experiences in VR. We computed the precision P; recall R, and F-measure to evaluate the performance of the tracker. The calculation formula for F-measure is F = 2PR/(P + R). Recall is the total true positives divided by the number of occurrences, which should have been detected, while precision is calculated as the number of true positives divided by the number of all responses given by the tracker [32]. The desired object within the frame was considered correctly detected if the detected region of the frame covers 50% of the target object [32].

For validation of our proposed method, we used eight 360° panoramic videos, which consisted of more than 15,000 frames. These videos were taken from multiple open source video channels on YouTube. The videos included multiple objects, which were desirable for tracking, such as a pedestrian, snorkeler, snowboarder, airplane cockpit, etc. as shown in Figure 9, with correctly detected objects and successful tracking.

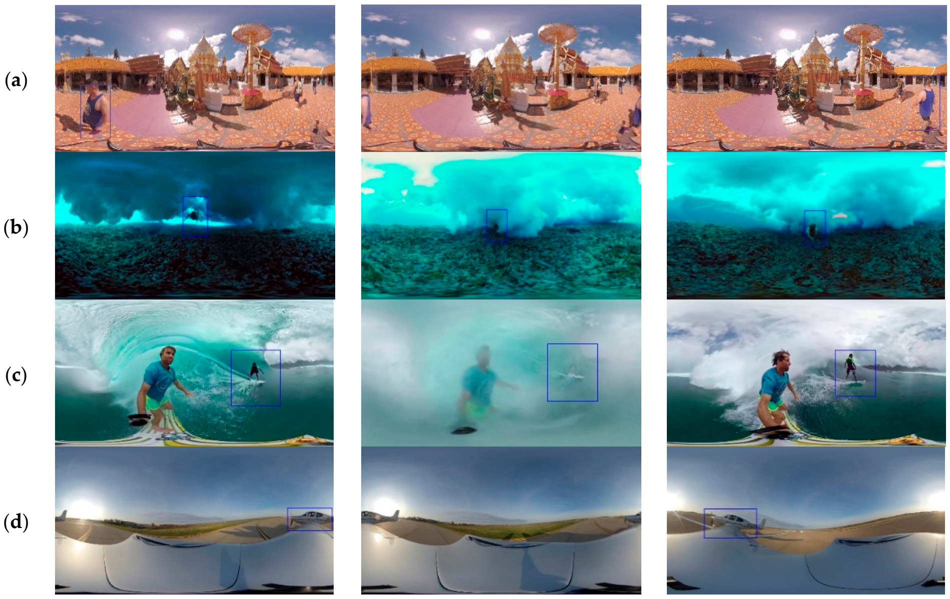

Figure 9. Visual results of proposed object tracking method on 360° videos showing correctly detected desired object and successful object tracking. Three frames from each video, (**a**–**d**) represent different challenges faced during tracking

Figure 9a shows the example of successful object re-identification in the case of multiple candidate objects; Figure 9b shows that tracking is successful even with a small part of the object visible; Figure 9c represents successful object tracking in case of occlusion; and Figure 9d shows the tracker's performance for object re-identification when the object exits from one side and appears on the other side of the 360° panoramic frame.

Table 2 shows the tracker's performance in terms of precision, recall, and F-measure for each 360° panoramic video.

Table 2. Evaluation results in terms of recall, precision, and F-measure.

Video	Desired Object	Frames	Recall	Precision	F-Measure
Street1	Pedestrian	500	0.94	1.0	0.97
Snorkeling	Snorkeler	962	1.0	1.0	1.0
Waterskating1	Front water skater	944	1.0	1.0	1.0
Waterskating2	Water skater in background	1092	0.92	0.94	0.93
Cartoon	Doll	2122	0.89	0.91	0.9
Snowboarding	Snowboarder	4796	0.85	0.88	0.86
Airplane Flight	Airplane cockpit	4598	0.79	0.9	0.84
Street2	Playing child	389	1.0	1.0	1.0
Mean	-	-	0.92	0.95	0.93

If we analyze the information provided in Table 1, the precision measure was higher than the recall measure. The reason for the high precision measure was the strong validation module presented in Section 4.1.2.2. We made the validation module very sensitive, due to which it rejected candidates with less feature matching. Consequently, in the case of object re-identification, it had been rejecting the candidates as soon as it got a candidate object with high matching. This resulted in a slight decrease in the recall measure, as seen in the videos "Snowboarding" and "Airplane Flight." Another reason for the least recall value for the three videos is the division of an object into two parts for a long time while it went out of the margin of the 360° frame and appeared on the other side. If we reduce the sensitivity of the validation module towards matching, there is a possibility of an increase in the recall measure. Figure 10 presents a clear picture about recall and precision for object tracking results in all 360° panoramic videos.

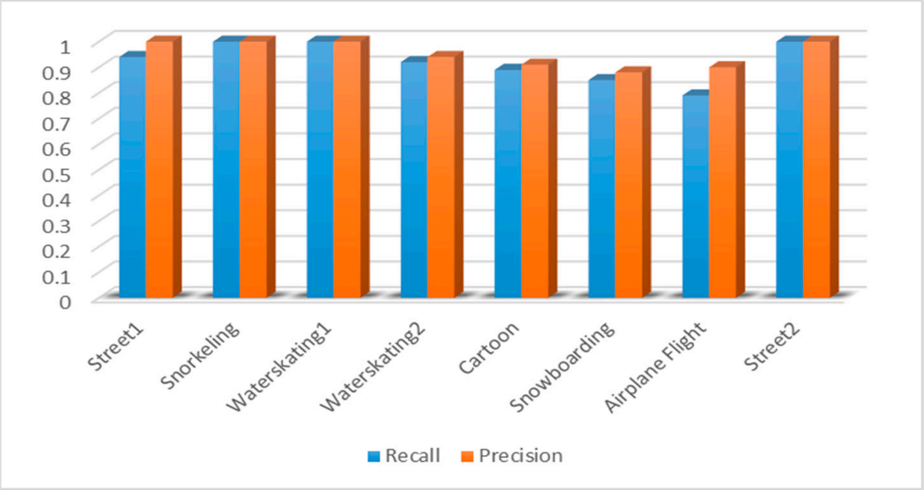

Figure 10. Bar chart of recall and precision values for object tracking in each 360° panoramic video.

5.3. Speed Evaluation

As we developed a system for real-time use, one of our focuses was to keep the processing speed high. The proposed method for object tracking in 360° panoramic videos was implemented on an Intel Core i7-7700k CPU @ 4.20 GHz. We implemented the code of this algorithm in Python along with OpenCV functions. This code was nonoptimized, and by running on a standard machine, it achieved approximately real-time performance with an average speed of 24.1 fps. Resolution of the immersive videos was very high, which resulted in slow computation speed. To further enhance the computation speed, we down-sampled the frame size and tracked the object at this resolution. Instead of down-sampling to a fixed size of frame, a down-sampling rate (25%) was used, which overcame the loss of the content's important information. After tracking the object's location in a down-sampled 360° frame, we remapped the location of the object in the 360° frame with the original resolution. In this way, it boosted the overall speed of the algorithm with no effect on tracking performance. Figure 11 shows a comparison of our method's speed performance with the performances presented by Ahmad et al. [18]. These methods also presented simple online arbitrary object trackers by using the conventional approaches in their techniques for fast object tracking in 360° videos. Figure 11 shows information on computation time per frame for each method along with information on speed in fps.

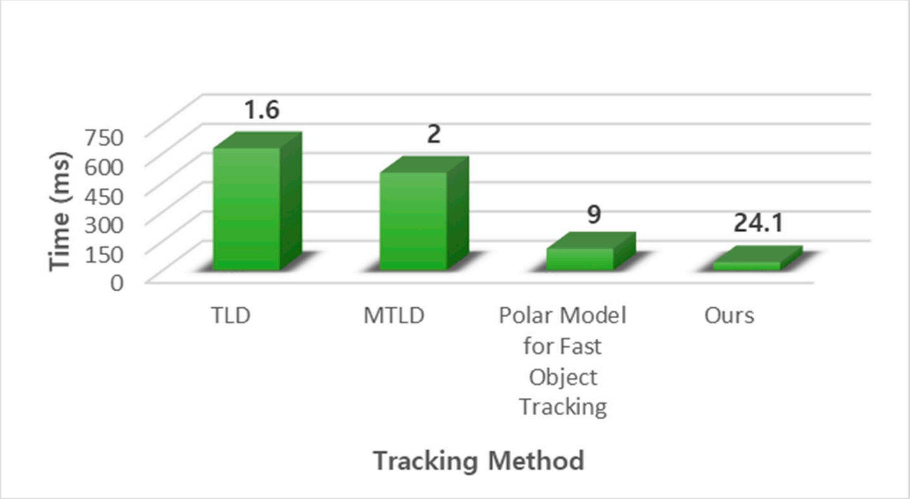

Figure 11. Comparison of tracking methods in terms of average computation time and frames per second (on top of the bars).

6. Discussion

Continuous advances have been made in the multimedia industry. Revolutionary developments in multimedia content have been bringing new experiences to the audience, and 360° multimedia has provided a completely new experience to viewers. In this study, an authoring system was proposed to increase the throughput of 360° multimedia content in terms of providing interesting experiences. The detailed description of the proposed system and experimental results presented in the previous sections proved the potential of a 360° video in providing multiple interesting experiences to viewers. This study supported that unlike the existing 360° cinematography approaches in the literature review, the proposed system offered the generation of multiple experiences from a single 360° video. This aspect of the system provides an answer to RQ1 and RQ2 that multiple experiences can be generated from a single 360° video. Here, an important point to notice is that the number of experiences generated depends on the content of a video. In response to RQ3, a user survey proved that users found it easy to discover interesting information in a 360° environment by following the experience of others. In support of RQ4, it was observed that the simple object tracker with some additional modules proved to have good performance in tracking objects in 360° videos. The tracking of objects turned out to be effective in reducing the effort for authoring 360° video tours, which provided evidence in favor of RQ5. As supported by the results of the user survey, transferring of the 360° experiences in VR using the adaptive focus assistance technique ensured minimum loss of information while directing the viewer towards an ROI, which provides sufficient proof of RQ6.

7. Conclusions and Future Work

We presented an authoring system with the idea of generating multiple experiences from a single 360° video. We introduced arbitrary object tracking in 360° videos and an HMD-based system for authoring multiple experiences in immersive VR. Moreover, we developed an adaptive focus assistance technique to guide the user efficiently in VR for transferring of 360° video experience. Our proposed system ensured the generation of a variety of experiences from a single 360° video. Overall, the proposed system performed very well in generating and transferring 360° video experiences. At the end of the user tests and survey, we analyzed each user's experience that was produced based on the visual content watched by that user. It was noticed by users that the experiences generated from a 360° video held different interesting information. Therefore, it resulted in multiple interesting experiences from one 360° visual content source. The possible number of experiences solely depends on the content presented in a 360° video. This system provided efficient transferring of the experiences to prospective users by using our focus assistance technique. The proposed arbitrary object tracking technique for panoramic videos also produced the best results with average precision of 0.95. The responses obtained from the user survey were very positive. They showed that overall, the audience wants to use our system for getting multiple experiences from one 360° visual content source. In future work, we will enhance the object re-identification module of our object tracker for timely re-identification of the desired object in case of loss of tracking. We will also create a better mechanism to track the partial object at boundaries of the 360° frame. Furthermore, we will work on further enhancement of our focus assistance method for transferring 360° experiences in an even more efficient way. We will also investigate how to record a better 360° video to make it capable of generating multiple interesting experiences from a single content source.

Author Contributions: Conceptualization, J.W.L., S.H.H.S., and K.H.; methodology, S.H.H.S.; investigation, S.H.H.S.; formal analysis, S.H.H.S. and K.H.; validation, S.H.H.S. and K.H.; writing—original draft preparation, S.H.H.S.; writing—review and editing, S.H.H.S., K.H., and J.W.L.; supervision, J.W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2016-0-00312) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cummings, J.J.; Bailenson, J.N. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychol.* **2016**, *19*, 272–309. [CrossRef]
2. Lin, Y.C.; Chang, Y.J.; Hu, H.N.; Cheng, H.T.; Huang, C.W.; Sun, M. Tell me Where to Look: Investigating Ways for Assisting Focus in 360 Video. In Proceedings of the ACM 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 2535–2545.
3. Delforouzi, A.; Tabatabaei, S.A.H.; Shirahama, K.; Grzegorzek, M. Unknown Object Tracking in 360-Degree Camera Images. In Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–6 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1798–1803.
4. Lukezic, A.; Vojir, T.; Cehovin Zajc, L.; Matas, J.; Kristan, M. Discriminative Correlation Filter With Channel and Spatial Reliability. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 6309–6318.
5. Lin, Y.T.; Liao, Y.C.; Teng, S.Y.; Chung, Y.J.; Chan, L.; Chen, B.Y. Outside-in: Visualizing Out-of-Sight Regions-of-Interest in a 360 Video Using Spatial Picture-in-Picture Previews. In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, Quebec City, QC, Canada, 22–25 October 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 255–265.
6. Liu, T.M.; Ju, C.C.; Huang, Y.H.; Chang, T.S.; Yang, K.M.; Lin, Y.T. A 360-Degree 4K x 2K Panoramic Video Processing Over Smart-Phones. In Proceedings of the 2017 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 8–10 January 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 247–249.
7. Facebook. New Publisher Tools for 360 Video. Available online: https://media.fb.com/2016/08/10/new-publisher-toolsfor-360-video/ (accessed on 13 May 2019).
8. Su, Y.C.; Jayaraman, D.; Grauman, K. Pano2Vid: Automatic Cinematography for Watching 360° videos. In Proceedings of the 3rd. Asian Conference on Computer Vision (ACCV), Taipei, Taiwan, 20–24 November 2016; p. 1.
9. Su, Y.C.; Grauman, K. Making 360 Video Watchable in 2d: Learning Videography for Click Free Viewing. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1368–1376.
10. Hu, H.N.; Lin, Y.C.; Liu, M.Y.; Cheng, H.T.; Chang, Y.J.; Sun, M. Deep 360 Pilot: Learning a Deep Agent for Piloting Through 360 Sports Videos. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1396–1405.
11. Thaler, M.; Bailer, W. Real-Time Person Detection and Tracking in Panoramic Video. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 1027–1032.
12. Lin, Z.; Doermann, D.; DeMenthon, D. Hierarchical Part-Template Matching for Human Detection and Segmentation. In Proceedings of the IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brasil, 14–21 October 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 1–8.
13. Wang, L.; Ching Yung, N.H. Three-Dimensional Model-Based Human Detection in Crowded Scenes. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 691–706. [CrossRef]
14. Leibe, B.; Seeman, E.; Schiele, B. Pedestrian Detection in Crowded Scenes. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR05), Providence, RI, USA, 16–21 June 2005; IEEE: Piscataway, NJ, USA, 2005; pp. 878–885.
15. Yuntao, C.; Samarasckera, S.; Qian, H.; Greiffenhagen, M. Indoor Monitoring via the Collaboration Between a Peripheral Sensor and a Foveal Sensor. In Proceedings of the 1998 IEEE Workshop on Visual Surveillance, Bombay, India, 2 January 1998; IEEE: Piscataway, NJ, USA, 1998; pp. 2–9.
16. Delforouzi, A.; Tabatabaei, S.A.H.; Shirahama, K.; Grzegorzek, M. Polar Object Tracking in 360-Degree Camera Images. In Proceedings of the IEEE International Symposium on Multimedia (ISM), San Jose, CA, USA, 11–13 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 347–352.
17. Delforouzi, A.; Grzegorzek, M. Robust and Fast Object Tracking for Challenging 360-degree Videos. In Proceedings of the IEEE International Symposium on Multimedia (ISM), Taichung, Taiwan, 11–13 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 274–277.

18. Delforouzi, A.; Tabatabaei, S.A.H.; Shirahama, K.; Grzegorzek, M. A Polar Model for Fast Object Tracking in 360-degree Camera Images. *Multimed. Tools Appl.* **2018**, *78*, 9275–9297. [CrossRef]
19. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-Convolutional Siamese Networks for Object Tracking. In *Computer Vision—ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science*; Hua, G., Jégou, H., Eds.; Springer: Cham, Switzerland, 2016.
20. Nam, H.; Han, B. Learning Multi-Domain Convolutional Neural Networks for Visual Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 4293–4302.
21. Kiani Galoogahi, H.; Fagg, A.; Huang, C.; Ramanan, D.; Lucey, S. Need for Speed: A Benchmark for Higher Frame Rate Object Tracking. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1134–1143.
22. Elmezeny, A.; Edenhofer, N.; Wimmer, J. Immersive Storytelling in 360-degree Videos: An Analysis of Interplay Between Narrative and Technical Immersion. *J. Virtual Worlds Res.* **2018**, *11*. [CrossRef]
23. Rosten, E.; Drummond, T. Machine Learning for High-Speed Corner Detection. In *Computer Vision—ECCV 2006. ECCV 2006. Lecture Notes in Computer Science*; Leonardis, A., Bischof, H., Pinz, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2006.
24. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. Brief: Binary Robust Independent Elementary Features. In *Computer Vision—ECCV 2010. ECCV 2010. Lecture Notes in Computer Science*; Daniilidis, K., Maragos, P., Paragios, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2010.
25. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G.R. ORB: An Efficient Alternative to SIFT or SURF. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1564–1571.
26. Muja, M.; Lowe, D. *Flann-Fast Library for Approximate Nearest Neighbors User Manual*; Computer Science Department, University of British Columbia: Vancouver, BC, Canada, 2009.
27. Li, X.; Qi, Y.; Wang, Z.; Chen, K.; Liu, Z.; Shi, J.; Luo, P.; Tang, X.; Loy, C.C. Video Object Segmentation With Re-Identification. *arXiv* **2017**, arXiv:1708.00197.
28. KaewTraKulPong, P.; Bowden, R. An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection. In *Video-Based Surveillance Systems*; Remagnino, P., Jones, G.A., Paragios, N., Regazzoni, C.S., Eds.; Springer: Boston, MA, USA, 2002; pp. 135–144.
29. Gugenheimer, J.; Wolf, D.; Haas, G.; Krebs, S.; Rukzio, E. Swivrchair: A Motorized Swivel Chair to Nudge Users' Orientation for 360 Degree Storytelling in Virtual Reality. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 1996–2000.
30. Kennedy, R.S.; Lane, N.E.; Berbaum, K.S.; Lilienthal, M.G. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *Int. J. Aviat. Psychol.* **1993**, *3*, 203–220. [CrossRef]
31. Streijl, R.; Winkler, S.; Hands, D. Mean opinion score (MOS) revisited: Methods and applications, limitations and alternatives. *Multimed. Syst.* **2016**, *22*, 213–227. [CrossRef]
32. Kalal, Z.; Mikolajczyk, K.; Matas, J. Tracking-learning-detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *37*, 1409–1422. [CrossRef] [PubMed]