

# Shot Orientation Controls for Interactive Cinematography with 360° Video

Amy Pavel  
UC Berkeley  
amypavel@berkeley.edu

Björn Hartmann  
UC Berkeley  
bjoern@eecs.berkeley.edu

Maneesh Agrawala  
Stanford University  
maneesh@cs.stanford.edu

## ABSTRACT

Virtual reality filmmakers creating 360° video currently rely on cinematography techniques that were developed for traditional narrow field of view film. They typically edit together a sequence of shots so that they appear at a **fixed orientation** irrespective of the viewer's field of view. But because viewers set their own camera orientation they may miss important story content while looking in the wrong direction. We present new interactive shot orientation techniques that are designed to help viewers see all of the important content in 360° video stories. Our **viewpoint-oriented** technique reorients the shot at each cut so that the most important content lies in the the viewer's current field of view. Our **active reorientation** technique lets the viewer press a button to immediately reorient the shot so that important content lies in their field of view. We present a 360° video player which implements these techniques and conduct a user study which finds that users spend 5.2-9.5% more time viewing (manually labeled) important points of the scene with our techniques compared to the traditional fixed-orientation cuts. In practice, 360° video creators may label important content, but we also provide an automatic method for determining important content in existing 360° videos.

## ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI):  
Miscellaneous

## Author Keywords

360° video; cinematography

## INTRODUCTION

Panoramic, 360° videos are becoming increasingly widespread. A YouTube search for the “360° video” tag returns over 500,000 videos. Journalists, TV/Film studios, advertisers, independent artists, and amateurs use such 360° videos to tell stories. But, because 360° video is a new medium for storytelling, creators do not yet have the set of conventions that exists in traditional, narrow field of view (NFOV) cinematography. As a result, video creators are

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

UIST 2017 October 22–25, 2017, Quebec City, QC, Canada

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4981-9/17/10.

DOI: <https://doi.org/10.1145/3126594.3126636>

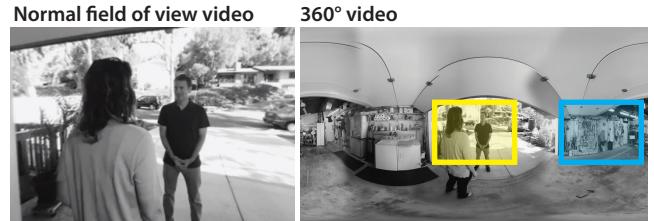
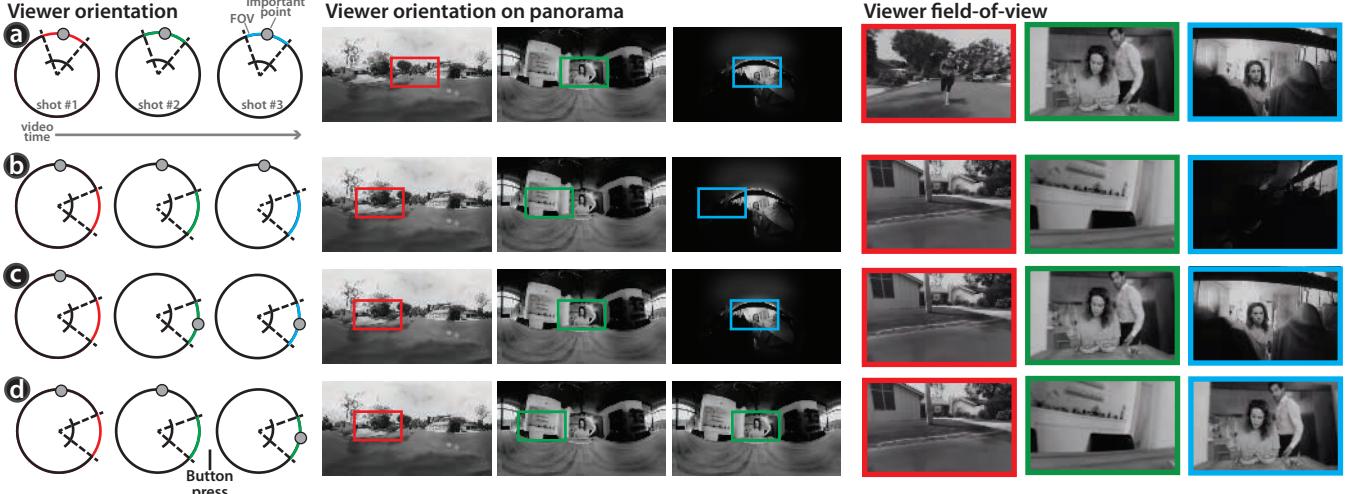


Figure 1. With traditional, narrow field of view video, video creators can guarantee the interesting content will be in the viewer's field of view. With 360° video, the viewer can explore the scene. So, while the viewer's field of view may contain the important content (yellow), it may instead contain only background content (blue). Video by Adam Cusco [14].

using methods from traditional cinematography to combine 360° shots in a sequence to tell a story.

The primary distinguishing feature of 360° video compared to NFOV video is that viewers control the camera orientation and can turn their head at any time to see the surrounding environment. Watching such video in a headset player or a mobile device, viewers can rotate the view to examine different parts of the scene. In a desktop-based player, viewers can drag the video horizontally and vertically with their cursor to rotate the camera. However, the viewer's ability to actively change their field of view creates a new challenge for 360° video creators. Unlike NFOV video, the video creator cannot guarantee that the important story elements (e.g. people, actions, objects) will be within the viewer's field of view when it is most critical for the viewer to see them (Figure 1).

Today, 360° video editors rely on a traditional cinematic approach, designing 360° videos so that the sequence of shots appear at a **fixed orientation** irrespective of the viewer's field of view. As a result it can be difficult for viewers to find important story content, particularly right after a shot change, as they may be looking in the wrong direction. Recently 360° video creators have acknowledged this problem, and have proposed best practices for creating more understandable 360° videos [11, 41]. For instance, Brillhart [11] proposes aligning the important content across shot boundaries to increase the probability that the viewer sees the important content in the following shot. As long as viewers orient towards the important content in the first shot and then maintain this field of view over the course of the video, such alignment ensures that they will see the most important content in each shot (Figure 2a). However, if viewers change their field of view before a shot change, they may end up at an irrelevant point in the new scene (Figure 2b).



**Figure 2.** (a) fixed-orientation cuts when the viewer is looking in the expected direction, (b) fixed-orientation cuts when the viewer looks away from the expected direction, (c) viewpoint-oriented cuts – when the viewer looks away from the expected direction on the first shot, the subsequent shots reorient to the viewer’s view point, (d) active reorientation – when the viewer looks away from the expected direction, the viewer can press a button to reorient to the important point. We note that the field of view varies for 360° viewers. Desktop viewers typically range from 75-110°, Google Cardboard and other phone-based viewers have a field of view of 60° and the Oculus Rift has a field of view of 110°. Video by Adam Cusco [14].

Recognizing that many videos contain shots with multiple important points that can attract the viewer’s attention, 360° video creators have also suggested aligning secondary important points across a sequence of shots. While this approach increases the likelihood that viewers will see a secondary important point [10], it is not always possible to line up multiple important points across shot boundaries. Another suggestion is to slow down the timing of the video so that the viewer has more time to explore the scene and hope that they return to the important story content [41]. However, none of these strategies guarantee that viewers will be looking at the important content when the shot changes.

In this paper, we explore new interactive shot orientation techniques that are designed to help viewers see all of the important content 360° videos. Our techniques take two inputs: (1) the viewer’s orientation/field of view; and (2) labels (generated either manually or automatically) that mark the location of important point(s) in the video.

#### *Viewpoint-oriented cuts*

The first technique, **viewpoint-oriented cuts**, automatically reorients the shot at each cut (or shot boundary) so that the most important content lies in the the viewers current field of view (Figure 2c). In contrast to standard fixed-orientation cuts, these viewpoint-oriented cuts let viewers freely explore the scene, but at each shot change viewers are guaranteed to initially see the most important content in the shot, since it is placed directly in their field of view. Thus, viewers are far less likely to miss story elements searching for important content after a shot change. However, this method only helps viewers locate one important point per shot, and does not assist viewers in finding secondary important points.

#### *Active reorientation*

The second technique, **active reorientation**, lets viewers press a button to immediately reorient the shot so that important content lies in their field of view (Figure 2d). This approach lets



**Figure 3.** Active reorientation with multiple important points. Viewers can press a button to toggle between important points. Video by GoPro [1].

viewers interactively reorient the shot to an important point at any time (e.g., return to the narrator in the shot). This approach can handle multiple important points within a shot (e.g., an interviewer on one side of the camera and the interviewee on the opposite side), viewers can press the button to cycle between these important points. Active reorientation lets viewers choose when they wish to orient the shot to important content and reduces the time spent searching for important content within the shot.

We develop a new 360° video player which implements fixed-orientation cuts, viewpoint-based cuts and active reorientation. We use this player to run a study comparing the three techniques and find that users view important content 5.2-9.5% longer using our interactive shot orientation controls. Users rated (5-highest, 1-lowest) active reorientation ( $\mu = 4.07$ ) and viewpoint-oriented cuts ( $\mu = 4.00$ ) as significantly preferable compared to fixed-orientation cuts ( $\mu = 2.14$ ). Based on our study results, we create an automatic method to detect important points and a hybrid technique combining viewpoint-based cuts with active reorientation.

## RELATED WORK

Our approach to developing a 360° video player with interactive shot orientation controls builds on three main areas of related work:

### Editing 360° videos

With the widespread availability of 360° cameras and VR viewers, producers have begun experimenting with cinematography techniques that ensure viewers see important aspects of the story. For example Brillhart [11] describes how she manually aligns important points in shots to help viewers see important story content. However, traditional video editing systems like Adobe Premiere do not provide tools for rotating 360° video, making it difficult to produce such alignments. Nguyen et al. [29] provide an in-headset 360° video editor that is explicitly designed to assist video creators create such alignments of the important content from shot-to-shot. Both Brillhart’s manual editing approach and Nguyen et al.’s system output a single static 360° video file with fixed-orientation cuts. In contrast, our player aligns the important points in the video to the viewer’s field of view in real-time during playback.

### Showing viewers important points in videos

Researchers have developed a variety of techniques for retargeting normal field of view (NFOV) video to different aspect ratios such that the most important content remains in the field of view during playback on a smaller screen [27, 32, 24]. More recently in Pano2Vid, Su et al. [36, 35] have proposed AutoCam. AutoCam creates NFOV video from 360° video using low-level models of saliency to identify important content and then building NFOV camera trajectories that capture the most important content over the course of the shot. Our work differs from AutoCam in two key ways. First we focus on improving playback for already-edited 360° videos rather than creating NFOV videos from unedited 360° content. Second, AutoCam is best-suited to desktop playback because on a desktop interaction with 360° content may be tedious, and on a headset automatic camera movement can induce simulator sickness [23]. Similar to AutoCam, Facebook introduced a new publisher tool for 360° video called Guide that automatically guides the camera between creator specified important points [9]. Using Guide, users may also reorient to important points after rotating away from the guided view. Our technique viewpoint-oriented cuts changes shot orientation imperceptibly on shot boundaries, rather than producing camera movement during a shot. In addition, active reorientation lets users produce cuts by selecting when to move between important points, rather than following a preguided path.

Instead of automatically moving the camera, SwiVRChair [20] introduces a motorized chair to physically rotate and block viewer’s movements to direct their attention. In contrast, we develop a 360° video player that provides software-based techniques for aiding users in viewing important content in 360° videos.

### Navigation in 360° video

As our techniques change the camera orientation in the scene, our work relates to prior work in virtual cinematography which studies virtual camera control in rendered scenes [22, 16],

and automatic real-world camera control for use in remote meetings or lectures [30, 17]. Our problem differs in that we consider already edited 360° videos.

In concurrent work, Outside-In also seeks to let users view important points in 360° video [26] by providing picture-in-picture previews to let users preview and navigate to important content in 360° videos with more than one important point. In other concurrent work, Serrano et al. [33] study how edits with varying shot alignment are understood by viewers. Our work focuses on providing shot orientation controls to the user, either based on their current viewpoint (viewpoint-oriented cuts) or by letting them press a button (active reorientation).

## INTERACTIVE 360° VIDEO PLAYER

We have developed a 360° video player that can play back videos using fixed-orientation cuts, viewpoint-oriented cuts and active reorientation. The web-based player works on both laptop and desktop displays (click and drag rotation), mobile phones (rotation-based orientation change and tap-to-drag rotation), and mobile phones with a Google cardboard viewer.

The input to our video player is a spherical 360° video along with a specification file containing the cut times for each shot boundary and the location in the panorama of one or more important point within each shot. Cut times and important points can be automatically determined, or manually labeled (e.g., by the video creator). In addition, if a shot contains multiple important points, video creators can manually assign a priority ranking to each one (Figure 3).

Given the important points within each shot our player presents the 360° video in any one of three modes:

**Fixed-orientation cuts.** In this mode our player simply renders the input 360° video and assumes that the filmmaker has chosen how to align the important points across shot boundaries. In practice for all of our example videos the filmmaker aligned the most important points across shot boundaries.

**Viewpoint-oriented cuts.** In this mode our player automatically reorients after each cut to ensure that the most important point (i.e. the only important point or the important point with the highest priority ranking) within it is centered horizontally inside the viewer’s field of view immediately. For viewpoint-oriented cuts, we automatically reorient the shot in the horizontal direction only because resetting the shot orientation in the vertical direction can lead to difficult rotation configurations on head-mounted displays. For instance, suppose we have two landscape scenes in shot #1 and shot #2 such that the ground is parallel to the user’s eyes when the user is in a neutral head position (a common setup for a 360° camera). If we reoriented shot #2 on the cut to show the important point while the user is looking at the ground, it would leave the user in an uncomfortable and confusing position for viewing the scene. The viewer would need to look further towards their feet to view the rest of the scene along the horizon, which can be uncomfortable.

**Active reorientation.** In this mode, our player lets viewers actively reorient the shot so that an important point is centered horizontally within their field of view by either clicking a button (desktop), tapping the middle, top of the phone (mobile), or pressing the Google Cardboard button (cardboard). If a shot contains more than one important point (e.g., the main surfer closest to the camera, and a partner surfing in the distance in Figure 3) viewers can press the button to cycle between them in their priority rank order. As in the viewpoint-oriented cuts mode, our player only reorients the shot in the horizontal direction.

## USER STUDY

We conducted a user study to evaluate the effects of our interactive shot orientation techniques (viewpoint-oriented cuts and active reorientation) in comparison to fixed-orientation cuts with respect to 360° video viewing behavior and user preference.

For the study, we selected nine 360° videos (Table 1) to represent a variety of domains (e.g., news, action, fictional stories), editing techniques (e.g., frequent cuts, few cuts), and sources (e.g., independent, GoPro, NYTimes). We extracted 30-75s clips from each of these videos for the study. We manually labeled shot boundaries and important points in each shot for each video. We labeled important points according to cinematography principles (e.g., speaking characters are important). For shots with multiple important points in the same shot, we assigned a priority rank to each point.

## Method

We recruited 14 users from university mailing lists to evaluate our techniques (8 female, 6 male). Their ages ranged from 18 to 34 years and they listed their occupation as undergraduate student, graduate student, or software engineer. All users except one had used a headset to view 360° content in the past.

We gave each user a 5 minute tutorial in which we explained the three techniques (fixed-orientation cuts, viewpoint-oriented cuts, and active reorientation) and let them watch a 30s video clip using each one. After the tutorial, each viewer saw 9 unique videos in total; 3 videos in each of the 3 conditions (fixed orientation, viewpoint oriented, active reorientation). Between users, we varied the technique assigned to each video. We controlled for ordering effects by randomizing the order that the videos appeared, and in-turn randomizing the order user saw each technique. Before each video, we told the user the video title, and the technique assigned to the video. Users viewed all videos using a Google Cardboard hand-held headset which includes a physical button for interaction.

## Measures

While users viewed each video, we recorded head orientation, and button presses on the Google Cardboard along with the corresponding time in the video. We calculate the following metrics for each video:

**Percent time viewing important points.** For each shot, we compute the percent of time spent viewing each of the important points in the shot. Specifically we compute how long each important point in a shot is within 18° of the user’s field

| Video name  | Domain       | Author    | Time per shot (s) | Duration (s) | Avg # pts per shot | Source |
|-------------|--------------|-----------|-------------------|--------------|--------------------|--------|
| dining      | news short   | NYTimes   | 15.73             | 125          | 2.3                | [38]   |
| ice-art     | news short   | NYTimes   | 10.51             | 84           | 2.3                | [37]   |
| hpo-preview | news trailer | HuffPost  | 4.19              | 101          | 1.8                | [2]    |
| snowboard   | sports       | GoPro     | 16.87             | 191          | 1.7                | [1]    |
| surfing     | sports       | GoPro     | 25.71             | 205          | 2.0                | [1]    |
| trees       | adventure    | Nat. Geo. | 20.54             | 204          | 1.4                | [3]    |
| volcano     | adventure    | Red Bull  | 32.64             | 293          | 2.0                | [12]   |
| knives      | fiction      | Indie     | 2.89              | 35           | 1.1                | [14]   |
| invasion    | fiction      | Indie     | 27.18             | 244          | 1.4                | [34]   |

**Table 1.** Set of videos shown in user study representing a variety of domains, authors, shot frequencies, and average numbers of important points per shot. 360 video editors typically employ a longer average shot length (ASL) than traditional film. As a result, the median ASL of our selected videos is 16.87s, which is much longer than the ASL of recent traditional films (~3-5s) [15].

of view - i.e. within macular or central vision, as opposed to peripheral vision [28].

**Percent of shot traversed (angular).** For each shot, we compute the percent of the shot traversed by constructing a histogram of 360 buckets in 1 degree increments along the horizontal axis. We count the bucket as traversed if it fell within an 18° window surrounding the center of the user’s field of view.

After viewing the videos, participants filled out a survey that included preference ratings on a 5 point likert scale, and ranking of the techniques on three facets: preference, level of disorientation while watching the video, and likelihood of viewing important content. We also asked them to compare the advantages and disadvantages of each technique in a free-text response.

## Hypotheses

We consider four hypotheses:

**H1.** Users will spend a greater time during the shot viewing the main (i.e. highest priority) important point using the proposed techniques (viewpoint-oriented cuts and active reorientation) in comparison to the standard fixed-orientation technique, because viewpoint-oriented cuts and active reorientation orient the shot so that the important point falls within the user’s field of view.

**H2.** Users will spend a greater percentage of time during the shot viewing the secondary (i.e. second highest priority) important points using active reorientation than they will using fixed-orientation and viewpoint-oriented cuts, because active reorientation lets users actively orient the shot so that a secondary point falls within the user’s field of view.

**H3.** Users will traverse less of the scene using viewpoint-oriented cuts and active reorientation than they will using the fixed-orientation cuts, because with fixed-orientation cuts users need to rotate more to search for important points than they do with the proposed techniques.

**H4.** Users will prefer viewing videos using the proposed techniques (viewpoint-oriented cuts and active reorientation) over standard fixed-orientation cuts, because users will be able to focus on more interesting content in the 360 videos.

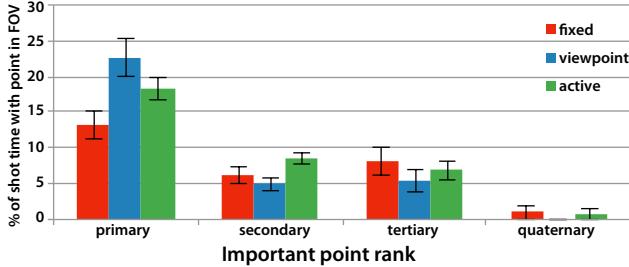


Figure 4. Percent of time per shot spent viewing the important points on average. Error bars show the standard error.

## Results

We computed our measures over the study data. To balance the effect of any given video in considering viewing data, we randomly sampled the video views to have four users for each video-condition pair.

**Percent time viewing important point(s).** We find that users spend a higher percentage of total viewing time examining the single most important point using viewpoint-oriented cuts ( $\mu = 22.6\%$ ,  $\sigma = 15.4$ ) and active reorientation ( $\mu = 18.3\%$ ,  $\sigma = 9.85$ ) than when using fixed-orientation cuts ( $\mu = 13.1\%$ ,  $\sigma = 11.8$ ) (Figure 4). Using a Friedman test we find that viewing mode has a significant effect ( $p < 0.001$ ,  $\chi^2 = 19.06$ ). Post-hoc pairwise Mann-Whitney U tests with Bonferroni correction find significant differences between viewpoint-oriented cuts and fixed-orientation cuts ( $p < 0.01$ ,  $U = 369$ ,  $Z = 3.14$ ,  $r = 0.52$ ) as well as active reorientation cuts and fixed-orientation cuts ( $p < 0.05$ ,  $U = 394$ ,  $Z = 2.86$ ,  $r = 0.48$ ), but not for active reorientation and viewpoint-oriented cuts.

We also consider how active reorientation affects the amount of time users spent viewing the secondary important point. For the 8 of 9 video clips with at least one secondary point of interest, we average over all shots with a secondary point the percent time spent viewing a secondary point of interest. We find that users spend the highest percent of time viewing a secondary point of interest using active reorientation ( $\mu = 8.51\%$ ,  $\sigma = 4.86$ ), followed by fixed-orientation cuts ( $\mu = 6.14\%$ ,  $\sigma = 6.74$ ) and viewpoint-oriented cuts ( $\mu = 4.94\%$ ,  $\sigma = 5.17$ ). A Friedman test shows that viewing mode has a significant effect ( $p < 0.05$ ,  $\chi^2 = 7.94$ ). We find the difference between active reorientation and viewpoint-oriented cuts to be significant ( $p < 0.01$ ,  $U = 281$ ,  $Z = -3.1$ ,  $r = -0.55$ ) and we find the difference between active reorientation and fixed-orientation to be significant ( $p < 0.05$ ,  $U = 326$ ,  $Z = 2.5$ ,  $r = 0.44$ ). As expected (H2), we do not find a significant difference between viewpoint-oriented cuts and fixed-orientation cuts. Finally, few video clips contain at least one shot with a third (6 of 9) or fourth (1 of 9) important point. A Friedman test does not show viewing mode has a significant effect on the amount of time spent viewing tertiary or quaternary points (Figure 4).

**Percent of shot traversed.** We find that users traverse the lowest percentage of the shot using active reorientation ( $\mu = 71.6\%$ ,  $\sigma = 17.3$ ), more of the shot using viewpoint-oriented cuts ( $\mu = 76.2\%$ ,  $\sigma = 16.1$ ), and the greatest amount

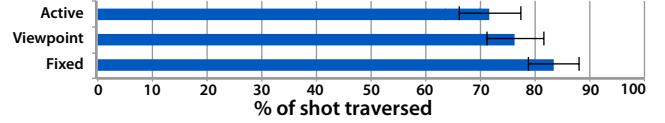


Figure 5. Percent of shot traversed on average for each condition. Error bars show the standard error.

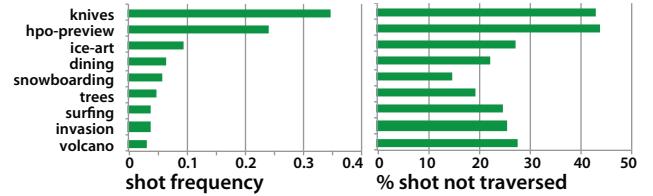


Figure 6. Shot frequency (shots per second) for each video and average % of the shot not traversed. The highest shot frequency videos (“knives” and “hpo-preview”) also have the highest average percentage of the shot not traversed.

of the shot using fixed-orientation cuts ( $\mu = 83.3\%$ ,  $\sigma = 14.2$ ). Using a Friedman test we find significant effect of viewing mode on percent of shot traversed ( $p < 0.05$ ,  $\chi^2 = 7.17$ ). A Mann Whitney U test with Bonferroni correction finds that the only pairwise difference that is significant is between active reorientation and fixed-orientation cuts ( $p < 0.05$ ,  $U = 871$ ,  $Z = -2.51$ ,  $r = -0.42$ ).

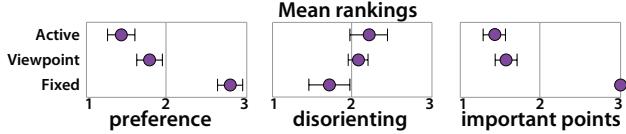
**Subjective ratings.** Users ranked active reorientation ( $\mu = 1.43$ ,  $\sigma = 0.646$ ) and viewpoint-oriented cuts ( $\mu = 1.79$ ,  $\sigma = 0.579$ ) as preferable (1-most preferable, 3-least preferable) to fixed-orientation cuts ( $\mu = 2.79$ ,  $\sigma = 0.579$ ) (Figure 7). We find a significant effect of viewing mode on rankings using Friedman’s nonparametric test ( $p < 0.001$ ,  $\chi^2 = 13.86$ ). Pairwise comparisons using Wilcoxon signed rank test with Bonferroni correction find a significant preference for viewpoint-oriented cuts over fixed-orientation cuts ( $p < 0.05$ ,  $W = 94$ ,  $Z = 3.71$ ,  $r = 0.99$ ) and for active reorientation over fixed-orientation cuts ( $p < 0.05$ ,  $W = 95$ ,  $Z = -3.96$ ,  $r = -1.06$ ).

Users also ranked the techniques on level of disorientation (from 1-most disorienting to 3-least disorienting) and ranked fixed-orientation cuts as most disorienting ( $\mu = 1.71$ ,  $\sigma = 0.99$ ), viewpoint-oriented cuts as second most disorienting ( $\mu = 2.07$ ,  $\sigma = 0.47$ ) and active reorientation to be least disorienting ( $\mu = 2.21$ ,  $\sigma = 0.89$ ) but a Friedman’s test does not show that the differences in ranks are significant.

Finally, users ranked the techniques on their perceived likelihood of viewing important content from most likely to least likely. Users rated viewpoint-oriented cuts ( $\mu = 1.57$ ,  $\sigma = 0.51$ ) and active reorientation ( $\mu = 1.42$ ,  $\sigma = 0.51$ ) over fixed-orientation cuts ( $\mu = 3.00$ ,  $\sigma = 0.00$ ). A Friedman’s test ( $p < 0.001$ ,  $\chi^2 = 21.14$ ) finds the differences to be significant. Pairwise comparisons using the Wilcoxon signed rank test find significant preference for viewpoint-oriented cuts over fixed-orientation cuts ( $p < 0.01$ ,  $W = 105$ ,  $Z = 4.9$ ,  $r = 1.31$ ) and for active reorientation over fixed cuts ( $p < 0.01$ ,  $W = 105$ ,  $Z = -4.9$ ,  $r = -1.31$ ).

## Qualitative feedback

We collected qualitative feedback on the advantages and disadvantages of each technique.



**Figure 7.** Mean rankings for preference (1-most prefer, 3-least prefer), level of disorientation (1-most disorienting, 3-least disorienting), and likelihood of viewing important points (1-most likely, 3-least likely). Error bars show the standard error.

Most (9 of 14) users ranked active reorientation as their most preferred technique, and mentioned the following benefits of active reorientation: they could (1) reorient to interesting content after exploring in a long scene, (2) quickly jump to the important point(s) in a short scene, and (3) toggle between important points. U3 explains “It was really cool to be able to quickly toggle between the interesting points without having to search around the scene constantly to make sure I wasn’t missing anything. I felt like I could still look around and explore the video, but had an option to quickly focus on the important part of a scene when I wanted to.”

The majority of users (9 of 14) preferred viewpoint-oriented cuts second most of the three techniques. Users who preferred viewpoint-oriented cuts over active reorientation mentioned that they didn’t want to have to click a button to see the important point. U9 mentioned that viewpoint-oriented cuts are preferable and that “I don’t know the reason that you wouldn’t want to center on an explicit point on the scene change. Otherwise you could just be randomly looking into dead space”. Only one user (U8) ranked viewpoint-oriented cuts the lowest of the three, explaining that viewpoint-oriented cuts “feel the most passive, I wasn’t as interested in looking around because I was always sure they’d show me whatever they wanted me to see”. Finally, 12 of 14 users rated fixed-orientation cuts the lowest.

## Discussion

Users spend a higher percentage of time viewing the primary important point using viewpoint-oriented cuts and active reorientation than they do when using fixed-orientation cuts (H1). Also, users spend a higher percentage of time viewing the second-ranked important point using active reorientation than they do using fixed-orientation cuts (H2). This metric only measures the amount of time spent viewing the pre-selected important points labelled by video creators or an automatic method.

Users traverse less of the shot using active reorientation, more of the shot using viewpoint-oriented cuts, and the greatest amount of the shot using fixed-orientation cuts. But we only found a significant difference between active reorientation and fixed-orientation cuts (H3). Active reorientation likely has the lowest percent of the shot traversed because users can reorient to important points when they want to by pressing a button, rather than rotating to the point. In addition, the percent of shot traversed differs between videos with very frequent shots (“hpo-preview”, and “knives”) and the rest of the videos with less frequent shots (Figure 6), because users have less time to explore during short shots. In some cases, video creators may not want to minimize the percent of shot traversed. For

instance, video creators can build suspense in a horror scene by letting users traverse the scene to find the important points.

We find users prefer viewpoint-oriented cuts and active reorientation over the fixed-orientation cuts (H4) with users split between preferring the active reorientation to the viewpoint-oriented cuts (9/14 users) and vice versa (5/14 users). Based on this preference and user comments, we combined the two interfaces to create a hybrid technique.

**Hybrid technique.** In this mode, as in viewpoint-oriented cuts, the player automatically reorients each shot such that the most important point is centered horizontally inside the viewer’s field of view. The viewer may also actively reorient the shot by clicking a button, tapping the phone, or pressing the Google Cardboard button as in active reorientation.

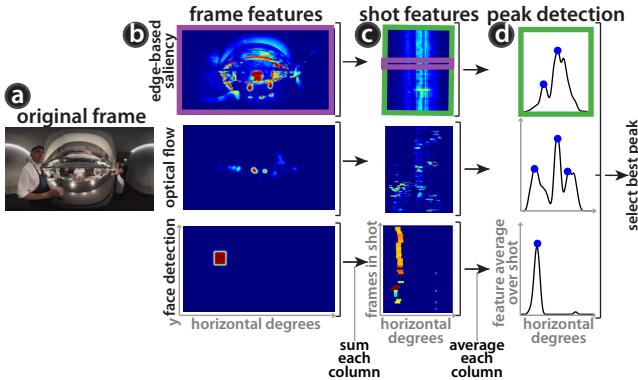
**Design implications.** Overall, our study suggests three main design implications. First, viewpoint-oriented cuts and active reorientation generally work well for a variety of domains and shot lengths. Second, viewpoint-oriented cuts and active reorientation are particularly useful in certain circumstances. Viewpoint-oriented cuts are helpful for videos with frequent shots where users do not have time to reorient, and active reorientation is useful for shots with multiple important points (without active reorientation, users cannot quickly switch between these points). Third, viewpoint-oriented cuts and active reorientation let users efficiently find important points, but video creators may want to prevent such efficiency for certain shots.

**Limitations.** In the study, we selected video clips (Table 1) to represent a variety of domains with differing average shot lengths. We did not investigate the effect of fatigue on user behavior in long videos ( $> 10$  minutes) as many current 360 videos tend to be short (e.g., 95s median length for all NY-Times 360 videos [4], 206s median length for YouTube’s “best of 360” [5]). However, fatigue while watching longer videos may change user behavior as users could traverse less of each shot, or prefer the active reorientation method (pressing a button) over manual reorientation (turning their head). This study also only considers user behavior while viewing a video for the first time. Repeated viewing may change viewer behavior because users will be familiar with the main storyline.

## AUTOMATIC IMPORTANCE DETECTION

For our study, we assume the 360° video creator labels important points for viewpoint-oriented cuts and active reorientation. We present a method for automatically detecting important points in a 360° video. By automatically detecting important points, viewers can enable viewpoint-oriented cuts and active reorientation for existing 360° videos.

Viewpoint-oriented cuts and active reorientation each use the horizontal orientation for each important point in a shot to reorient the shot in the horizontal direction, without affecting the user’s vertical direction. Our automatic important point detection method works by (a) computing per-frame salience maps that account for different types of features (e.g., face detection, optical flow), (b) detecting shot boundaries, (c) finding local maxima in the per-shot feature vectors, and (d) selecting the best local maxima as the important points.



**Figure 8.** Method for selecting important points. We (a) separate the video into single frames, (b) compute feature maps for each frame (edge-based, optical flow, and face detection depicted) and (c) sum each feature map column to find a 1d feature vector for each frame. We average together the feature vectors of all frames in a shot, then (d) find local maxima for each feature before selecting the best local maxima as important points.

**Feature maps.** For videos in our dataset, the important point typically occurs on a key character, object, or title rather than background scenery. 25.7% of the important points in our manually-labeled dataset are also within  $1^\circ$  of the origin (i.e. center of the equirectangular projection). Important points occur close to the origin because video creators line up important points across shots, and they align the first important point with the origin so that the point will initially fall within the viewer’s field of view. To predict the location of important points, we split the video into frames (Figure 8a) then compute the following feature maps for the rectilinear projection of each frame (Figure 8b):

- Face detection (Openface [8])
- Optical flow (Lucas-Kanade [7])
- Edge-based image saliency (Minimum barrier [43])
- Distance from the origin (i.e. frame center)

For face detection and edge-based image saliency we use recent off-the-shelf methods [8, 43] to produce feature maps for each frame. To create a feature map of motion in the frame, we use Harris corners to detect important pixels to track [21] then track the pixels across frames using optical flow [7]. We recalculate pixels to track every 0.3s, as optical flow will lose tracked pixels across shot boundaries. We generate the feature map by setting the value of each pixel tracked, and the surrounding pixels (within 5 pixels of the tracked pixel), to the distance traveled by that pixel. We smooth this feature map with a Gaussian kernel.

We sum along the columns of each feature map so that for each frame we create 4 feature vectors, one for each feature map (Figure 8c).

**Shot detection.** Building on prior work, we use the per-frame optical flow feature vectors to compute shot boundaries [25, 7]. We use Harris corners [21] to find important pixels to track, and optical flow to track these pixels across frames. Within a single shot, a pixel and its corresponding object will move a small distance ( $\sim < 1\%$  of the image height/width) between subsequent frames. However, when the shot changes,

optical flow can not accurately track an object as it may have disappeared. Such inaccuracies return improbable distances for object movement ( $\sim 25\text{--}50\%$  of the image height/width). We determine a shot change has occurred if the sum of tracked movement falls above a threshold.

**Peak detection.** For each feature, we create a shot feature vector by averaging together the corresponding feature vector for all frames in the shot then smooth the shot feature vector using a Gaussian kernel (Figure 8d). To detect the most important horizontal orientation in the shot, we find the local maxima for each feature that are at least  $60^\circ$  apart such that the field of view around important points will be non-overlapping. We rank the selected local maxima using the clarity of the peak (i.e. area under the peak / area under the rest of the curve), and the feature type from semantically meaningful features (face detection) to low-level image features (optical flow, edge-based saliency). When peaks have low clarity, we select the origin as the predicted point.

## Evaluation

We evaluated our method using live action videos in our dataset (Table 1) along with four new videos [19, 40, 18, 39] for a total of 11 videos. As viewpoint-oriented cuts require shot boundary detection in order to reorient on shot change, we first evaluate the shot boundary detection. The optical-flow based shot boundary detection achieves precision of 89% and recall of 82% for an F1 score of 85%. We prefer the precision to be higher than the recall, as a missed shot boundary will appear as a fixed-orientation cut, but an extra shot boundary will introduce a new cut.

Given correct shot boundaries, we found that our detected most important point is on average  $16.7^\circ$  (less than  $18^\circ$ ) away from one of the ground truth important points as opposed to  $57.7^\circ$  on average for a randomly selected point. In particular, for our method, the predicted point is exactly at one of the ground truth points an average of 30% of the time, less than  $18^\circ$  away from a ground truth important point 75% of the time, and within the field of view or less than  $30^\circ$  away 80% of the time. In comparison, a randomly selected point is less than or equal to  $0^\circ$ ,  $18^\circ$  or  $30^\circ$  degrees away, on average 2%, 18%, and 30% of the time respectively. In the future we will study how well our automatic importance detection technique works in the context of our viewer. In addition, we will incorporate additional features into our method such as text detection [42], person detection [13], and object detection [31].

## FUTURE WORK

Currently we only consider shot orientation control in interactive  $360^\circ$  cinematography. In the future, we will investigate more types of interactive cinematography. For instance, we may change cut timing based on when a viewer reaches a particular point, or when the viewer has explored the entire scene once. In addition, to ensure the viewer sees the main storyline, we could play the main story content only when it falls within the viewer’s field of view. When the viewer looks away from the main content, we could pause the main content, and loop background video (e.g., trees swaying, water moving) using prior work [6].

Finally, we focus on the viewer's experience using interactive cinematography in 360° videos. However, video creators may want to express more control over viewer's playback of the video. For instance, the video creator may want to prevent viewpoint-oriented shots in cases where the viewer's physical location is important (e.g. sitting in a chair that matches the virtual chair). In the future, we would like to consider how the video creator may direct and edit the 360° video when incorporating interactive cinematography.

## ACKNOWLEDGEMENTS

We thank Steve Rubin for creating our video. Our research is supported by an NDSEG fellowship and the Brown Institute for Media Innovation.

## REFERENCES

1. 2017. GoPro Youtube Channel. (2017). <https://www.youtube.com/user/GoProCamera>.
2. 2017. Huffington Post Youtube Channel. (2017). <https://www.youtube.com/user/HuffingtonPost>.
3. 2017. National Geographic Youtube Channel. (2017). <https://www.youtube.com/user/NationalGeographic>.
4. 2017. The New York Times YouTube channel, 360 VR Playlist. (June 2017). [https://www.youtube.com/playlist?list=PL4CGYNsoW2iCGZa3\\_Pes8LP\\_jQ\\_GPTW8w](https://www.youtube.com/playlist?list=PL4CGYNsoW2iCGZa3_Pes8LP_jQ_GPTW8w).
5. 2017. YouTube Virtual Reality Channel, Best of 360 Playlist. (June 2017). [https://www.youtube.com/playlist?list=PLU8wpH\\_LfhmsSVRA8bSkn04-2wXvYXS4C](https://www.youtube.com/playlist?list=PLU8wpH_LfhmsSVRA8bSkn04-2wXvYXS4C).
6. Aseem Agarwala, Ke Colin Zheng, Chris Pal, Maneesh Agrawala, Michael Cohen, Brian Curless, David Salesin, and Richard Szeliski. 2005. Panoramic video textures. In *Proc. TOG'05*. ACM.
7. Simon Baker and Iain Matthews. 2004. Lucas-kanade 20 years on: A unifying framework. In *International Journal of Computer Vision*. Springer.
8. Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In *Proc. WACV'16*. IEEE, 1–10.
9. Paul Beddoe-Stephens. 2016. New Publisher Tools for 360 Video. (August 2016). <https://media.fb.com/2016/08/10/new-publisher-tools-for-360-video/>.
10. Jessica Brillhart. 2016a. In the Blink of a Mind: Engagement, Part 1. (2016). <https://medium.com/the-language-of-vr/in-the-blink-of-a-mind-engagement-part-1-eda16ee3c0d8>.
11. Jessica Brillhart. 2016b. The Language of VR - Blog. (2016). <https://medium.com/the-language-of-vr>.
12. Red Bull. 2016. Expedition to the Heart of an Active Volcano | 360 Video. (2016). <https://www.youtube.com/watch?v=OBp2EWPjotk>.
13. Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. In *Proc. CVPR'17*. IEEE.
14. Adam Cusco. 2017. Knives. (2017). <http://adamskosco.com/blog/2016/10/6/bec9b1qspab3enlckac7ij6ij82d5a?rq=knives>.
15. James E Cutting, Jordan E DeLong, and Christine E Nothelfer. 2010. Attention and the evolution of Hollywood film. In *Psychological Science*, Vol. 21. Sage Publications, 432–439.
16. David K Elson and Mark O Riedl. 2007. A Lightweight Intelligent Virtual Cinematography System for Machinima Production.. In *Proc. AIIDE'07*. 8–13.
17. Jonathan Foote and Don Kimber. 2000. Flycam: Practical panoramic video and automatic camera control. In *Proc. ICME'00*. IEEE.
18. GoPro. 2016. GoPro VR: Mentos and Coke experiment with Untamed Science. (2016). <https://www.youtube.com/watch?v=qIKqCWNeIBo>.
19. GoPro. 2017. GoPro Fusion: Relive Reality. (2017). <https://www.youtube.com/watch?v=PygsKZXpYrI>.
20. Jan Gugenheimer, Dennis Wolf, Gabriel Haas, Sebastian Krebs, and Enrico Rukzio. 2016. Swivrchair: A motorized swivel chair to nudge users' orientation for 360 degree storytelling in virtual reality. In *Proc. CHI'16*. ACM.
21. Chris Harris and Mike Stephens. 1988. A combined corner and edge detector.. In *Proc. Alvey Vision Conference*, Vol. 15. Manchester, UK, 10–5244.
22. Li-wei He, Michael F. Cohen, and David H. Salesin. The Virtual Cinematographer: A Paradigm for Automatic Real-time Camera Control and Directing. In *SIGGRAPH '96*. ACM, 217–224.
23. Eugenia M Kolasinski. 1995. *Simulator Sickness in Virtual Environments*. Technical Report. DTIC Document.
24. Philipp Krähenbühl, Manuel Lang, Alexander Hornung, and Markus Gross. 2009. A system for retargeting of streaming video. In *Proc. TOG'09*, Vol. 28. ACM, 126.
25. Rainer Lienhart. 2001. Reliable transition detection in videos: A survey and practitioner's guide. 1, 03 (2001), 469–486.
26. Yung-Ta Lin, Yi-Chi Liao, Shan-Yuan Teng, Yi-Ju Chung, Liwei Chan, and Bing-Yu Chen. 2017. Outside-In: Visualizing Out-of-Sight Regions-of-Interest in a 360 Video Using Spatial Picture-in-Picture Previews. In *Proc. UIST'17*. ACM.
27. Feng Liu and Michael Gleicher. 2006. Video retargeting: automating pan and scan. In *Proc. MM'06*. ACM, 241–250.
28. Christopher G Morris. 1992. *Academic Press dictionary of science and technology*.
29. Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. Vremiere: In-headset Virtual Reality Video Editing. In *Proc. CHI'17*.

30. Abhishek Ranjan, Rorik Henrikson, Jeremy Birnholtz, Ravin Balakrishnan, and Dana Lee. 2010. Automatic camera control using unobtrusive vision and audio tracking. In *Proc. GI'10*. Canadian Information Processing Society, 47–54.
31. Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proc. CVPR'16*. IEEE.
32. Michael Rubinstein, Ariel Shamir, and Shai Avidan. 2008. Improved seam carving for video retargeting. In *Proc. TOG'08*, Vol. 27. ACM, 16.
33. Ana Serrano, Vincent Sitzmann, Jaime Ruiz-Borau, Gordon Wetzstein, Diego Gutierrez, and Belen Masia. 2017. Movie editing and cognitive event segmentation in virtual reality video. In *Proc. SIGGRAPH'17*.
34. Bobab Studios. 2017. Invasion. (2017).  
<http://www.baobabstudios.com/>.
35. Yu-Chuan Su and Kristen Grauman. 2017. Making 360° Video Watchable in 2D: Learning Videography for Click Free Viewing. In *arXiv preprint arXiv:1703.00495*.
36. Yu-Chuan Su, Dinesh Jayaraman, and Kristen Grauman. 2016. Pano2Vid: Automatic Cinematography for Watching 360° Videos. In *arXiv preprint arXiv:1612.02335*.
37. New York Times. 2017a. A Chilly Walk Amid China's Ice Art. (2017). <https://www.nytimes.com/video/world/asia/100000004868768/a-chilly-walk-amid-chinas-ice-art.html>.
38. New York Times. 2017b. Dining at the Met. (2017). <https://www.nytimes.com/video/dining/100000004855665/dining-in-at-the-met-breuer.html>.
39. The New York Times. 2017c. 36 Hours: Tokyo | The Daily 360. (2017).  
<https://www.youtube.com/watch?v=S6dYU0yx880>.
40. The New York Times. 2017d. 52 Places to Go: Grand Teton | The Daily 360. (2017).  
<https://www.youtube.com/watch?v=y19GY195Qnc>.
41. Oculus Story Studio Blog Saschka Unseld. 2015. 5 Lessons Learned While Making Lost. (2015).  
<https://www.oculus.com/story-studio/blog/5-lessons-learned-while-making-lost/>.
42. Tao Wang, David J Wu, Adam Coates, and Andrew Y Ng. 2012. End-to-end text recognition with convolutional neural networks. In *ICPR'12*. IEEE.
43. Jianming Zhang, Stan Sclaroff, Zhe Lin, Xiaohui Shen, Brian Price, and Radomir Mech. 2015. Minimum barrier salient object detection at 80 fps. In *Proc. ICCV'15*, 1404–1412.