

---

# VisPhoto: Photography for People with Visual Impairment as Post-Production of Omni-Directional Camera Image

**Masakazu Iwamura**

Osaka Prefecture University  
1-1 Gakuencho, Naka, Sakai,  
Osaka 599-8531, Japan  
masa@cs.osakafu-u.ac.jp

**Naoki Hirabayashi**

Osaka Prefecture University  
1-1 Gakuencho, Naka, Sakai,  
Osaka 599-8531, Japan  
hirabayashi@m.cs.osakafu-  
u.ac.jp

**Zheng Cheng**

Osaka Prefecture University  
1-1 Gakuencho, Naka, Sakai,  
Osaka 599-8531, Japan  
zheng386@m.cs.osakafu-u.ac.jp

**Kazunori Minatani**

National Center for University  
Entrance Examinations  
2-19-23 Komaba, Meguro-ku,  
Tokyo 153-8501 Japan  
minatani@rd.dnc.ac.jp

**Koichi Kise**

Osaka Prefecture University  
1-1 Gakuencho, Naka, Sakai,  
Osaka 599-8531, Japan  
kise@cs.osakafu-u.ac.jp

**Abstract**

It is known that many people with visual impairments desire to take photos. In taking photos, however, they often face difficulty in pointing the camera at the subject. In this paper, we address this problem by proposing a novel photo-taking system named VisPhoto. Unlike existing ones, the proposed system *generates a photo* as post-production of an omni-directional camera image. That is, when the shutter button is pressed, the proposed system captures an omni-directional camera image. In the post-production, from the objects detected in the image, the user selects the ones that should be included and excluded. Finally, as a “photo,” the system outputs a cropped region satisfying the user’s preference and looking beautiful from an aesthetic point of view.

**Author Keywords**

Omni-directional camera; object detection; sound recording

**CCS Concepts**

•Human-centered computing → Accessibility systems and tools;

---

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).

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

© 2020 Copyright is held by the author/owner(s).

ACM ISBN 978-1-4503-6819-3/20/04.

DOI: <https://doi.org/10.1145/3334480.3382983>

## Introduction

It is known that many people with visual impairment<sup>1</sup> desire to take photos as sighted people do. Their purposes, which are of a wide variety, include visually preserving an impression of a particular experience and sharing their memories and experiences with their sighted counterparts in the form of a photograph [16, 14, 32]

Despite their demands, taking a satisfactory photograph needs to overcome some challenges. Harada et al. have raised issues of framing, lighting and focus [14]. Balata et al. have focused on composition [5]. In reality, however, people with visual impairment face even more fundamental problem; that is, how to include the subject in the camera frame. For addressing this problem, much effort has been spent to develop real-time applications that provide the user with sound (voice) and vibration feedback to aim at the subject [33, 16, 30, 31, 5]. It is also known that VoiceOver for iOS tells the number of people within the frame before taking a photo. However, this kind of approach requires time to aim at the subject. Hence, they are applicable only when the subjects are static or moving slow enough.

In this paper, we address the problem by proposing a novel photo-taking system that employs an omni-directional camera, which is called VisPhoto. Intuitively speaking, as an omni-directional camera can capture an image of all surroundings of the user at once, the user does not have to aim at the subject with spending time. Once the omni-directional camera image is obtained, later the user can crop a specific region in a certain manner. That is, *cropping* corresponds to *taking a photo* in the standard photography. In other words, the proposed system *generates a photo* as

---

<sup>1</sup>Following [14], in this paper, we use the term “people with visual impairment” to refer to people who experience significant difficulty in visually inspecting and identifying the content of photographs.

the post-production<sup>2</sup> of an omni-directional camera image.

The advantages of the proposed system are as follows.

1. As mentioned above, the user is free from the process of aiming at the subject.
2. The rotation of the camera can be automatically corrected. Hence, the user does not have to take care of the camera angle when the user takes a photo.
3. In principle, the user can take a photo of moving objects, which are not easy in the current mainstream approach [33, 16, 30, 31, 5].
4. Unlike the mainstream approach, feedback to the user is not required to be in real time. In other words, a computationally heavy process can run before cropping. In this paper, we apply an object detection method to the omni-directional camera image. The object detection does not take long if it is run on a proper device but takes some time on a portable device such as a smartphone.
5. We can employ an algorithm that predicts the quality score of a cropped region from an aesthetic point of view, so that the copped region is expected to be beautiful.

By contrast, since generating a photo is postponed until the post-production, the user has to remember the subject and situation at the moment of taking a photo. As it increases the user’s load, it can be regarded as a disadvantage. To compensate for it, we implemented an audio recording functionality for recording a voice memo like [14, 3]. The

---

<sup>2</sup>Post-production includes all stages of production occurring after taking a raw photo [23].

recorded audio is provided in the post-production along with the date and time the omni-directional camera image was taken.

## Related Work

### *Photo-taking assistance*

Probably, the most used application that assists people with visual impairment in taking a photo is VoiceOver for iOS. It speaks the number of people within the frame before taking a photo. Similarly, the current mainstream approach in existing researches is to provide the user with sound (voice) and vibration feedback to aim at the subject [33, 16, 30, 31, 5]. Table 1 of [5] well summarizes the differences among representative methods with regard to feedback modalities and aiming assistance directions. EasySnap [33, 16] is an iPhone application assisting in aiming at the subject. It provides the user with audio feedback about blur and darkness of the camera view, camera tilt and the size and location of a face and an object. PortraitFramer [16] is similar to EasySnap but for a group portrait containing multiple faces. It tells how many faces are included in a photo, and the size and location of faces. Vazquez et al. [31] have proposed functionality to automatically correct the camera roll. Balata et al. [5] have proposed an aiming assistance application for central and golden-ratio compositions implemented running on a smartphone.

### *Assistance in managing, browsing and sharing photos*

Frohlich and Tallyn [11, 12] have explored the value to combine a photograph with audio recording. They found that ambient sounds as well as voice annotations are valuable for most users. While the research does not focus on people with visual impairment, it is suggestive. EasySnap [16] presented above also realizes an accessible photo album. The location the photo is taken and the content of the photo recognized by a recognition engine are associated with

each photo, so that people with visual impairment can review and share pictures non-visually. Harada et al. [14] have proposed a mobile application that allows capturing photos along with audio recordings for the ambient sound and memo description that can be used for browsing the photos. Adams et al. [4, 3] have proposed an iPhone app named VisSnap that assists people with visual impairment in organizing and browsing a photo library without sight. VizSnap records ambient audio while the user is aiming the camera and a voice memo as well as capturing time, date, and location the photo is taken.

### *Photo-related services*

Applications such as VizWiz [7], Be My Eyes [6], Envision AI [10] and TapTapSee [29] help people with visual impairment get the information on the content of a photo. For the sake of that, the user is required to take a photo of items in question. However, it is not always easy for people with visual impairment to take a photo. Gurari et al. [13] have revealed that  $\sim 28\%$  of visual questions of VizWiz are deemed unanswerable. A part of them is due to the quality of the photos. This shows the difficulty of taking pictures by people with visual impairment. While the proposed system is intended to generate photos to share with sighted people, it can also be used for such assistive technologies.

Some other services are also provided. OrCam MyEye 2 [1] is a wearable device that can read a text, and recognize a face and a product. vOICe [26] is also a wearable device (software) that can recognize the scene in front of the user.

### *Projection of 360-degree videos into normal field-of-view*

A related research field to our proposed method is to project a 360-degree video into a normal field-of-view [28, 27, 15, 19, 17]. Their goals are to generate an informative normal field-of-view video, which looks like taken with a standard camera, from a 360-degree video captured with an omni-

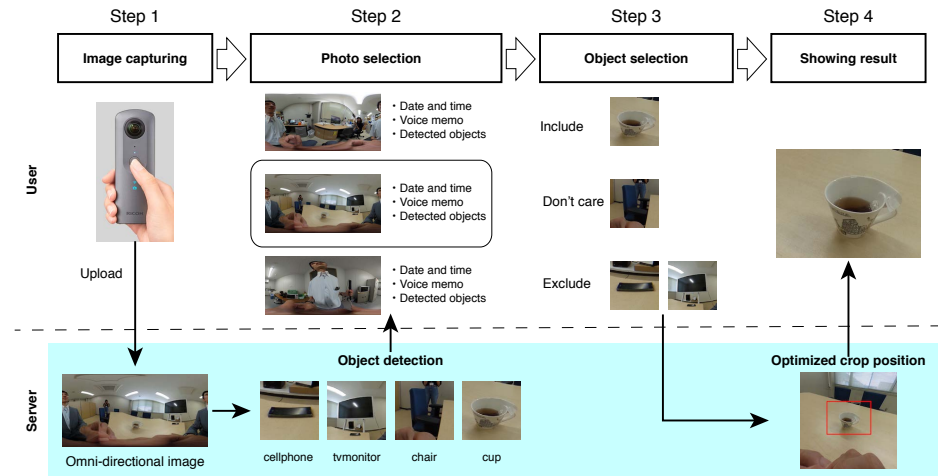


Figure 1: An overview of the proposed system.

directional camera. In particular, [15, 17] take into account objects detected by an object detector. However, they intend not to crop the region the user wants but to crop an informative region.

#### *Light field camera*

The process of the proposed framework is similar to that of light field camera [21] with regard to the capability to generate images the user likes after the scene is captured. A representative vendor of consumer products was Lytro [20], which shut down its operation in 2018. The light field camera of Lytro consists of a microlens array. It can preserve the “light field” at a moment and refocus pictures afterward. For example, assume a photo contains a person and a dog. The user can generate an image a person is in focus and that a dog is in focus as well as an image both are not in focus afterward.

Compared to the proposed framework, a light field camera corresponds to an omni-directional camera and a refocused image corresponds to a cropped image.

## Proposed System

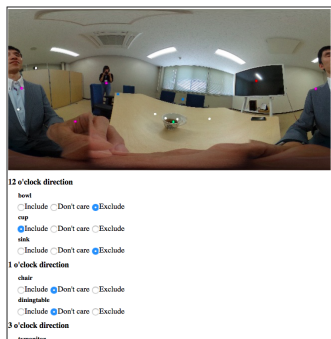
### *Overview*

An overview of the proposed method is illustrated in Fig. 1. The process of generating a photo is divided into four steps: image capturing, photo selection, object selection, and showing the result. In step 1, the user takes an omni-directional camera image followed by recording a voice memo. In steps 2 to 4, the user generates an image reflecting the user’s preference using the web user interface with help of a screen reader if needed. The screenshots of the web user interface in steps 2 to 4 are shown in Fig. 2.

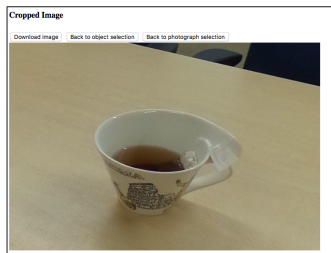
In step 1, in image capturing, we use an omni-directional camera. As the camera is equipped with two fish-eye cam-



(a) Photo selection (Step 2 of Fig. 1).



(b) Object selection (Step 3 of Fig. 1).



(c) Showing result (Step 4 of Fig. 1).

**Figure 2:** Screenshots of the web user interface.

eras, it can take a photo of the surrounding scenery of the user. When the shutter button is pressed, an omnidirectional camera image is taken. At the same time, audio recording starts and lasts until the button is released. If the button is released in less than ten seconds, the recording ends in ten seconds. That is, recording lasts at least ten seconds. This is designed for encouraging the user to speak something that helps remind the user of the subject and situation at the moment of taking a photo. Captured omnidirectional camera images and recorded audios are automatically uploaded to the VisPhoto server when the omnidirectional camera is connected to the Internet.

In step 2 (see Fig. 2(a) for screenshot), the user selects one image among all uploaded images using the VisPhoto web interface which can be accessed with the user's comfortable device such as the user's iPhone and PC with a screen reader. As supplementary information of an image, date and time the image is taken, the recorded audio and the list of detected objects are provided. The list of detected objects is obtained by applying an object detection method to uploaded omnidirectional camera images on the VisPhoto server. Object detection is a well-defined task in computer vision (e.g., [25, 24]); given an image, positions of registered objects are outputted. That is, it finds out object regions associated with registered object names. Of course, it is not perfect. For example, we experienced a case that an unknown stuffed toy is detected as a teddy bear. Even such a wrong detection in the formal definition is often beneficial, as it may retrieve objects which are not registered in the object detector. As the supplementary information, GPS information is not provided because the omnidirectional camera does not record it.

In step 3 (see Fig. 2(b) for screenshot), from the objects detected in the selected image, the user selects the ones

that should be included and excluded. The web interface provides the list of the detected objects arranged using the clock position. In the web user interface, for each object, three radio buttons corresponding to “include,” “don't care” and “exclude” are provided.

In step 4 (see Fig. 2(c) for screenshot), as a “photo,” the system outputs a cropped region satisfying the user's preference and looking beautiful from an aesthetic point of view. The photo can be downloaded. The filename of the photo contains the objects that are detected in the cropped image, so that the user can know what objects are contained.

#### Implementation

As the omnidirectional camera, we used Ricoh Theta series. Two models Theta V and Z1 can add new functionalities by installing plug-ins [22]. As their OSs are Android, the plug-in can be developed as an Android app. We implemented the photo-taking app of VisPhoto. The app is fully audio guided. For example, after the app boots, it says “welcome to VisPhoto.” It also speaks “start recording,” “stop recording,” “ready,” “VisPhoto is shutting down” and so on. However, Theta itself is almost free from sounds. Hence, the user needs to boot the VisPhoto app without any sound feedback. This is a drawback of the current implementation. A workaround is to teach the blind user how to use the VisPhoto app alone, e.g., “to boot the device, press the power button for 2 sec. and wait for 16 sec. And then,...”. We need to ask the vendor to improve this.

In the server, as soon as the images are uploaded, object detection is applied to them. As the object detection method, we used a tensorflow implementation [9] of *you only look once* (YOLO) version 1 [24] trained on COCO dataset [18] consisting of 80 object categories. One problem here is that a captured omnidirectional camera image is in the equirectangular format. However, YOLO assumes

a perspective image, which is a standard format that can be taken with a standard digital camera. In order to solve this problem, we projected an omni-directional camera image onto eight perspective images and YOLO was applied to them. The information on the detected objects from eight perspective images are projected back to the original omni-directional camera image and merged. Then, the detected objects are arranged using the clock position.

Once the preference of the user is given in the form of “which objects should be included and excluded in the cropped image,” the desired image is generated by cropping the omni-directional camera image. The process begins with generating a new projected image whose center is equal to the center of the centers of the objects marked as *included*. That is, the coordinate  $(x^C, y^C)$  of the center of the new projected image is given by

$$(x^C, y^C) = \frac{1}{|I_{\text{in}}|} \sum_{i \in I_{\text{in}}} (x_i^C, y_i^C), \quad (1)$$

where  $(x_i^C, y_i^C)$  is the coordinate of the center of the  $i$ -th object, and  $I_{\text{in}}$  is the set of indices of the objects marked as *included*. Then, we define an energy function  $E(\theta)$  as a function of a rectangular region  $\theta$ , and crop the region that minimizes the energy function. The rectangular region  $\theta$  is defined by  $\theta = (x_1, y_1, x_2, y_2)$ , where  $(x_1, y_1)$  and  $(x_2, y_2)$  are the coordinates of the upper-left and lower-right corners of the region, respectively. Similar to  $I_{\text{in}}$ , let  $I_{\text{out}}$  be the set of indices of the objects marked as *excluded*. As is illustrated in Fig. 3, let  $R_i$  be the area ratio of the intersection of the  $i$ -th object and the cropping region to the whole object.

Then, the energy function  $E(\theta)$  is given as follows.

$$E(\theta) = \frac{E_{\text{const}}(\theta) + 1}{E_{\text{aes}}(\theta)} \quad (2)$$

$$E_{\text{const}}(\theta) = \sum_{i \in I_{\text{out}}} R_i + \sum_{i \in I_{\text{in}}} (1 - R_i), \quad (3)$$

where  $E_{\text{aes}}(\theta)$  is the output score of view finding network (VFN) [8, 2]. VFN is an aesthetics-aware deep neural network that finds the best sub-views in an image. A larger score means more beautiful from an aesthetic point of view. Eq. (2) is minimized by using Nelder-Mead method, which finds the minimum of the objective function.

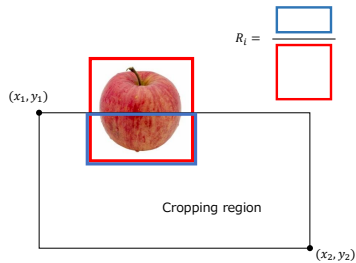
## Conclusion

In this paper, we have contributed to propose a novel photo-taking system that employs an omni-directional camera, which is called VisPhoto. Unlike the mainstream researches which provide the user with feedback to aim at the subject in real time, the proposed system does not need to give feedback when an omni-directional camera image is taken. It is because the omni-directional camera captures an image of all surroundings of the user at once and the user can crop a good region in the post-production. To select a good region, we used the information of detected objects and the quality score of a cropped region from an aesthetic point of view.

The future work includes the evaluation of the proposed method. For evaluating the usability, advantage and disadvantage of the proposed system, we will conduct a user study with people with visual impairment. We will also conduct a user study to evaluate the quality of the generated photos by the proposed system. In the study, we will hire sighted people because the quality of photos is not able to be evaluated by people with visual impairment.

## Acknowledgments

This work is financially supported by The Telecommunication Advancement Foundation.



**Figure 3:** Definition of  $R_i$ . The red rectangle represents the bounding box of an object (apple). The blue rectangle represents the intersection of the object region and the cropping region. Then,  $R_i$  is defined by the ratio of the area of the red rectangle to the area of the blue rectangle.

## REFERENCES

- [1] OrCam MyEye 2. Retrieved Sept. 18, 2019 from <https://www.orcam.com/en/myeye2/>
- [2] View Finding Network. Retrieved Sept. 18, 2019 from <https://modeldepot.io/yilingchen/view-finding-network>
- [3] Dustin Adams, Sri Kurniawan, Cynthia Herrera, Veronica Kang, and Natalie Friedman. 2016. Blind Photographers and VizSnap. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '16*. ACM Press. DOI : <http://dx.doi.org/10.1145/2982142.2982169>
- [4] Dustin Adams, Lourdes Morales, and Sri Kurniawan. 2013. A qualitative study to support a blind photography mobile application. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments - PETRA '13*. ACM Press. DOI : <http://dx.doi.org/10.1145/2504335.2504360>
- [5] Jan Balata, Zdenek Mikovec, and Lukas Neoproud. 2015. BlindCamera: Central and Golden-ratio Composition for Blind Photographers. In *Proceedings of the Multimedia, Interaction, Design and Innovation on ZZZ - MIDI '15*. ACM Press. DOI : <http://dx.doi.org/10.1145/2814464.2814472>
- [6] Be My Eyes. Retrieved Sept. 20, 2019 from <https://www.bemyeyes.com/>
- [7] Jeffrey P. Bigham, Samuel White, Tom Yeh, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, and Brandyn White. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology - UIST '10*. ACM Press. DOI : <http://dx.doi.org/10.1145/1866029.1866080>
- [8] Yi-Ling Chen, Jan Klopp, Min Sun, Shao-Yi Chien, and Kwan-Liu Ma. 2017. Learning to Compose with Professional Photographs on the Web. In *Proceedings of the 2017 ACM on Multimedia Conference - MM '17*. ACM Press. DOI : <http://dx.doi.org/10.1145/3123266.3123274>
- [9] darkflow — tensorflow implementation of YOLO version 1 and 2. Retrieved Sept. 20, 2019 from <https://github.com/thtrieu/darkflow>
- [10] Envision AI. Retrieved Feb. 12, 2020 from <https://www.letsenvision.com/>
- [11] David Frohlich and Ella Tallyn. 1999. Audiophotography: practice and prospects. In *CHI '99 extended abstracts on Human factors in computing systems - CHI '99*. ACM Press. DOI : <http://dx.doi.org/10.1145/632716.632897>
- [12] David M. Frohlich. 2004. *Audiophotography*. Springer Netherlands. DOI : <http://dx.doi.org/10.1007/978-1-4020-2210-4>
- [13] Danna Gurari, Qing Li, Abigale J. Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P. Bigham. 2018. VizWiz Grand Challenge: Answering Visual Questions from Blind People. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE. DOI : <http://dx.doi.org/10.1109/cvpr.2018.00380>

- [14] Susumu Harada, Daisuke Sato, Dustin W. Adams, Sri Kurniawan, Hironobu Takagi, and Chieko Asakawa. 2013. Accessible photo album: enhancing the photo sharing experience for people with visual impairment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. ACM Press. DOI: <http://dx.doi.org/10.1145/2470654.2481292>
- [15] Hou-Ning Hu, Yen-Chen Lin, Ming-Yu Liu, Hsien-Tzu Cheng, Yung-Ju Chang, and Min Sun. 2017. Deep 360 Pilot: Learning a Deep Agent for Piloting through 360° Sports Videos. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: <http://dx.doi.org/10.1109/cvpr.2017.153>
- [16] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting blind photography. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '11*. ACM Press. DOI: <http://dx.doi.org/10.1145/2049536.2049573>
- [17] Wei-Sheng Lai, Yujia Huang, Neel Joshi, Christopher Buehler, Ming-Hsuan Yang, and Sing Bing Kang. 2018. Semantic-Driven Generation of Hyperlapse from 360 Degree Video. *IEEE Transactions on Visualization and Computer Graphics* 24, 9 (Sept. 2018), 2610–2621. DOI: <http://dx.doi.org/10.1109/tvcg.2017.2750671>
- [18] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. 2014. Microsoft COCO: Common Objects in Context. *arXiv preprint arXiv:1405.0312* (2014). <https://arxiv.org/abs/1405.0312>
- [19] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell Me Where to Look: Investigating Ways for Assisting Focus in 360° Video. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press. DOI: <http://dx.doi.org/10.1145/3025453.3025757>
- [20] Lytro — Wikipedia. Retrieved Sept. 20, 2019 from <https://en.wikipedia.org/wiki/Lytro>
- [21] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. 2005. *Light Field Photography with a Hand-Held Plenoptic Camera*. Technical Report. Stanford University Computer Science. <http://graphics.stanford.edu/papers/lfcamera/>
- [22] Plug-in store — Ricoh Theta. Retrieved Sept. 20, 2019 from <https://pluginstore.theta360.com/>
- [23] Post-production — Wikipedia. Retrieved Sept. 19, 2019 from <https://en.wikipedia.org/wiki/Post-production>
- [24] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. DOI: <http://dx.doi.org/10.1109/cvpr.2016.91>
- [25] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 91–99.
- [26] Seeing with Sound — The vOICE. Retrieved Sept. 18, 2019 from <https://www.seeingwithsound.com/>



- [27] Yu-Chuan Su and Kristen Grauman. 2017. Making 360° Video Watchable in 2D: Learning Videography for Click Free Viewing. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI:<http://dx.doi.org/10.1109/cvpr.2017.150>
- [28] Yu-Chuan Su, Dinesh Jayaraman, and Kristen Grauman. 2016. Pano2Vid: Automatic Cinematography for Watching 360° Videos. In *Proc. 13th Asian Conference on Computer Vision (ACCV)*. DOI:  
[http://dx.doi.org/10.1007/978-3-319-54190-7\\_10](http://dx.doi.org/10.1007/978-3-319-54190-7_10)
- [29] TapTapSee. Retrieved Sept. 20, 2019 from <https://www.bemyeyes.com/>
- [30] Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12*. ACM Press. DOI:  
<http://dx.doi.org/10.1145/2384916.2384934>
- [31] Marynel Vázquez and Aaron Steinfeld. 2014. An Assisted Photography Framework to Help Visually Impaired Users Properly Aim a Camera. *ACM Transactions on Computer-Human Interaction* 21, 5 (Nov. 2014), 1–29. DOI:  
<http://dx.doi.org/10.1145/2651380>
- [32] Violeta Voykinska, Shiri Azenkot, Shaomei Wu, and Gilly Leshed. 2016. How Blind People Interact with Visual Content on Social Networking Services. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*. ACM Press. DOI:  
<http://dx.doi.org/10.1145/2818048.2820013>
- [33] Samuel White, Hanjie Ji, and Jeffrey P. Bigham. 2010. EasySnap: real-time audio feedback for blind photography. In *Adjunct proceedings of the 23rd annual ACM symposium on User interface software and technology - UIST '10*. ACM Press. DOI:  
<http://dx.doi.org/10.1145/1866218.1866244>