Accessible Image Description Using Sample Example Cues

Dhruba Dahal
Romerike IT Consulting AS
Oslo, Norway
Email: dhruba.dahal03@gmail.com

Raju Shrestha
Oslo Metropolitan University
Oslo, Norway
Email: raju.shrestha@oslomet.no

***Abstract* - Image accessibility on the web is the main focus of this study. Due to the lack of proper image descriptions, it has been difficult to access intended information available on the informative images for the people with sight loss and who use assistive technologies, such as screen reader while surfing websites or web applications. This study defines the lack of effective solutions to author image description as an existing gap, and explores the possibility of helping to write a better image description with the use of different types of sample examples. Results suggest that it is effective to have a similar sample example description to write accessible image descriptions.**

Keywords - image description; accessibility; NCAM guidelines

I. INTRODUCTION

Tim Berners-Lee, the inventor of the World Wide Web, states that the power of the web is in its universality. Accessibility for everyone regardless of disability is an essential aspect. The statement reflects the significance of web accessibility which is about the fundamental design of web for all people regardless of their hardware, software, language, culture, location, and physical or mental ability and the fulfillment of this goal results in an accessible web with a diverse range of sight, hearing, movement, and cognitive ability [1].

This study emphasizes image accessibility on the web. The images on the web might be of several types: informative images, decorative images, functional images, images of text, complex images, groups of images, and image maps. Accessibility of these images simply means if the intended information given in it is accessible to the people including disabled people, such as visually impaired people. It is possible to make images accessible through the text description which is readable by assistive technologies such as a screen reader [2].

If we look at the real-world scenarios, there are massive number of images in the Internet which do not have text descriptions, and many of them that have are not appropriate and not good enough to convey necessary information [3]-[5]. This clearly indicates the lack of availability of descriptive summary of the images on the web intended for image accessibility. Literatures suggest that the main reason behind this may be the negligence of web authors, complexity of writing image description, and lack of time and motivation to read the accessibility guidelines having long text, it would be more effective if instead of traditional textual guidelines, a real-time guidance is provided.

In this work, we have investigated the possibility of encouraging and improving image description for better accessibility by providing example images with sample descriptions, which we call it as sample cues. The reset of the paper is organized as follows. Section II describes the related research. Section III presents the proposed sample cue-based method. Section IV presents experiments and results. Finally, we conclude the paper in Section V.

II. RELATED METHODS USED FOR IMAGE DESCRIPTION

In general, there are two broad categories of methods or authoring techniques used for describing an image. The first one is human powered authoring and the other one is computer algorithm-based authoring.

The system called VizWiz lets blind people take a picture, asks questions, and receive answers from distant workers almost in real-time [6]. TapTapSee, a mobile application developed particularly for blind and visually impaired users, takes a picture of any two- or three-dimensional object and tells the user audibly by identifying the objects within seconds [7]. Splendiani and Ribera [8] suggested to use a decision tree that may reduce ambiguity and enhance the relevance of alternative texts. Likewise, Morash et. al [9] compared two methods, Queried Image Description (QID) method and Free-Response Image Description (FRID) method, for novice Web workers to produce image descriptions for graph images based on National Centre for Accessible Media (NCAM) guidelines [10]. Although there are several human powered systems available, Wu et. al [11] claimed that all these systems so far are constrained by scalability, latency, cost, and privacy concerns.

On the other hand, Cundiff [12] developed a browser extension that adds descriptions to images on the web for blind people. After getting a user click on an image, the extension sends the image URL to the cloud sight API and gets the resulting description to the image. Similarly, Ramnath, Baker, and Vanderwende [13] introduced a system allowing smartphone users to generate captions for their photos. The system is based on a cloud service and the combined outcomes of the different modules result in a large set of candidate captions which are provided to the phone. Several computer algorithm-based solutions are available, which are intended for social media. Automatic alt-text (AAT) [11] is an example, which identifies objects, faces, and themes from photos and generate alternative text for screen reader users on Facebook. However, Morris et al. [14] found that currently available computer-generated

 ISBN: 978-1-61208-692-7

captioning solutions are not robust enough to meet image accessibility requirements. They investigated accessibility of Twitter, which has traditionally been thought of as the most accessible social media platform for blind users, and found that image-based tweets are diverse, largely inaccessible.

III. PROPOSED SAMPLE CUE-BASED IMAGE DESCRIPTION

The literature suggested that the time-consuming accessibility guidelines are not so effective for having useful descriptions to the images uploaded on the web. Therefore, we have proposed a new sample example cue-based method to assist in writing image description to improve image accessibility. Similar example image(s) with description is provided as a sample cue in order to help writing a description for the given image. Description of these sample images are written by accessibility experts by following fourteen NCAM accessibility guidelines that have been developed based on several studies incorporating disabilities. The fourteen guidelines used are listed below.

NCAM image accessibility guidelines:

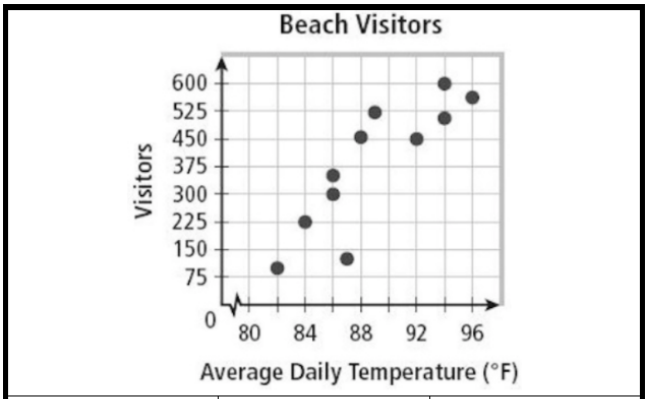
1. The description should be succinct.
2. Colors should not be specified unless it is significant.
3. The new concept or terms should not be introduced.
4. The description should be started with high level context and drilled down to details to enhance understanding.
5. The active verbs in the present tense should be used.
6. Spelling, grammar, and punctuation should be correct.
7. Symbols should be written out properly.
8. The description vocabulary should be added which adds meaning for example, "map" instead of an image.
9. The title and axis labels should be provided.
10. The image should be identified as a scatter plot and be focused on the change of concentration.
11. The central teaching point should be focused to determine if borders, region shapes, and bodies of water are important.
12. The description should be organized using number lists and pull the most important information in the beginning.
13. Physical appearance and actions should be explained rather than emotions and possible intentions.
14. The material should not be interpreted or analyzed, instead, the readers should be allowed to form their own opinions.

Among these fourteen guidelines, the first 8 guidelines are common to all types of images, while guidelines 9 and 10 are specific to graph images, guidelines 11 and 12 are specific to map images, and guidelines 13 and 14 are specific to natural images.

Modern Artificial Intelligence (AI) based algorithms have shown successful classification of images, even beating human intelligence. These algorithms can be used to find a similar example image for a given image to be described. Therefore, the proposed method could be a viable and effective solution for accessible image description.

IV. EXPERIMENTS AND RESULTS

In order to evaluate the effectiveness of the proposed method, this study conducts an online experiment to compare results of different sample cues on image descriptions. A custom web application software was developed for this. Sixty-five participants took part in the experiment who wrote text descriptions for given images with and without sample cues. We have limited our study to three different types of images: graph, map, and natural photos. The participants were asked to write descriptions, first without any sample example description (No cue), then by providing a random image with a description (Random cue), and finally by providing a similar image with a description (Similar cue). Figure 1 shows an example image description written by a participant for a graph image. Sample example images (cues) were selected randomly from the set of pre-classified images (graph, map and natural) with descriptions.



No cue	Random cue	Similar cue
Number of beach visitors with respect to the average daily temperature.	More the daily temperature more the beach visitors.	This graph is a scatter plot of daily visitors in a beach, which shows the increase of visitors in the beach with respect to increase in daily temperatures.

Figure 1 An example image description written by a participant while having no cue, a random cue, and a similar cue.

Effectiveness of the proposed method has been evaluated based on the compliance of the 500 image descriptions entered by the participants to the 14 NCAM guidelines as evaluated by the six experts who have a good knowledge on image accessibility.

As suggested by Allen and Seaman [15] and Boone and Boone [16], compliance of the image descriptions to the NCAM guidelines is measured in a Likert type rating scale from 1 to 4 (1 – strongly disagree, 2 – disagree, 3 – agree, and 4 – strongly agree). Figure 2 shows resulting compliance of the image descriptions (in percentage) to the overall 14 NCAM guidelines in three different cases with no cue, random cue and similar cue. The plots in the figure also shows standard errors in cases of all the four rating scales.

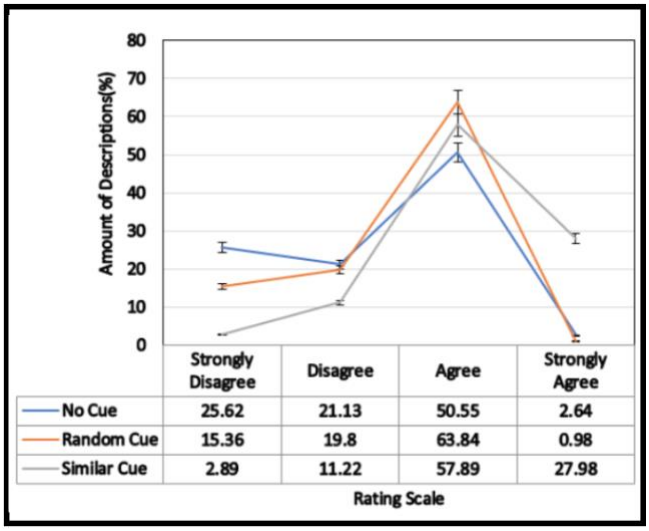


Figure 2. Compliance of image descriptions with overall guidelines in 3 different cases.

From the figure, we see that, when no cues were provided, almost 53% of the image descriptions complied (which includes both 'agree' and 'strongly agree') to the overall guidelines. Compared to this, the compliance percentage increased significantly 12% when random cues were provided. The compliance percentage increased even more by 33% when similar cues were provided.

To determine statistical significance of these results, we conducted a Friedman test [17], which is a non-parametric alternative to the one-way ANOVA with repeated measures. This is useful to test for differences among groups when the dependent variable being measured is ordinal. It is suitable in our case since the intervals in the four Likert type rating scales used may not be equal. To examine where the differences occur, this study ran a separate Wilcoxon signed-rank test [17] on the related groups: no cue to random cue, no cue to similar cue, and random cue to similar cue. The table in Figure 2 shows the test results. The results show a significant effect of sample example cues on the quality of image description written by the users. Effect of random cues over no cue is small, whereas effect of similar cues over no cue and random cues is moderate.

V. CONCLUSIONS

This study investigated the effectiveness of providing real time sample example cues as an alternative to a set of guidelines for the users who have no or minimal knowledge about how to write an image description and the one who do not have enough time or do not want to read long guidelines before writing an image description.

The results demonstrate that similar example cue provides significant help than no example cue and random example cues in writing image descriptions in compliance with the NCAM guidelines to make them accessible.

As a future work, the study could be extended further with more images. Also, the effect of sample cues in different contexts and usability of the method by the real users with accessibility issues can be investigated.

REFERENCES

[1] S. L. Henry, "Introduction to Web Accessibility, 2005. [Online] available from: https://www.w3.org/WAI/fundamentals/accessibility-intro/, Retrieved: Jan. 2019.

[2] E. Eggert, and S. Abou-Zahra, "Web Accessibility Tutorials", 2014. [Online] available from; https://www.w3.org/WAI/tutorials/images/, 2014, Retrieved: Jan. 2019.

[3] R. Bavani, A. Jaafar, and N. F. M. Yatim, "A study on web experience among visually impaired users in Malaysia", Proceedings of the International Conference on User Science and Engineering (i-USEr), pp. 11-15, 2010.

[4] H. Francis, D. Al-Jumeily, and T. O. Lund, "A framework to support e-commerce development for people with visual impairment", Proceedings of the 6th International Conference on Developments in eSystems Engineering, pp. 335-341, 2013.

[5] R. Goncalves, J. Martins, and F. Branco, "A review on the Portuguese enterprises web accessibility levels-a website accessibility high level improvement proposal", Procedia Computer Science, vol. 27, pp. 176-185, 2014.

[6] J. P. Bigham et al., "VizWiz: Nearly Real-time Answers to Visual Questions", ACM User Interface Software and Technology Symposium (UIST), 2010.

[7] TapTapSee, "Assistive Technology for Blind and Visually Impaired", 2014. [Online] available from: http://taptapseeapp.com/, Retrieved: Jan. 2019.

[8] B. Splendiani, M. Ribera, "How to textually describe images in medical academic publications", Proceeding of the XV International Conference on Human Computer Interaction, vol. 67, pp. 1-3, 2014.

[9] V. S. Morash, Y. Siu, J. A. Miele, L. Hasty, and S. Landau, "Guiding novice web workers in making image descriptions using templates", ACM Transactions on Accessible Computing (TACCESS), vol. 7, pp. 4-12, 2015.

[10] NCAM, "Guidelines for Describing STEM Images", 2009. [Online] available from: http://ncam.wgbh.org/experience_learn/educational_media/stemdx/guidelines, Retrieved: Jan. 2019.

[11] S. Wu, J. Wieland, O. Farivar, and J. Schiller, "Automatic alt-text: Computer-generated image descriptions for blind users on a social network service", CSCW, pp. 1180-1192, 2017.

 ISBN: 978-1-61208-692-7

[12] C. Cundiff, "Alt text bot", 2015. [Online] available from: https://connectability.devpost.com/submissions/37785-alt-text-bot, Retrieved: Jan. 2019.

[13] K. Ramnath, S. Baker, L. Vanderwende, "AutoCaption: automatic caption generation for personal photos", Proceedings of the IEEE Winter Conference on Applications of Computer Vision, pp. 1050-1057, 2014.

[14] M. R. Morris, A. Zolyomi, C. Yao, S.Bahram, J. P. Bigham, and S. K. Kane, "With most of it being pictures now, I rarely use it: Understanding Twitter's evolving accessibility to blind users", Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5506-5516, 2016.

[15] E. I. Allen and C.A. Seaman, "Likert scales and data analyses", Quality progress, vol. 7, pp. 40-64, 2007.

[16] H. N. Boone, and D. A. Boone, "Analyzing Likert data", Journal of Extension vol. 50(2), pp. 1-5, 2012.

[17] Lund-Research, "Friedman test in SPSS statistics". [Online] available from: https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php, 2013, Retrieved: Jan. 2019.

 ISBN: 978-1-61208-692-7