Readable Image for the Visually Impaired

Sunil Kumar Kopparapu

TCS Innovation Labs Mumbai
Tata Consultancy Service,
Yantra Park, Thane (West), Maharastra 400601, INDIA
SunilKumar.Kopparapu@TCS.Com

Abstract. A picture is worth a thousand words is a well know adage and refers to the fact that any complex word description can be conveyed easily and quickly with a still image. The use of images, pictures and block diagrams to describe something is often used without a second thought. The use of images in any form becomes an accessibility issue for the visually impaired. With the concept of a document and the hyper text markup language (HTML) page blurring, W3C has come up with measures to make the images accessible on the web. The most recent being the use of `alt` attribute, which is designed to be an alternative text description for images on web pages and `longdesc` attribute which is a mechanism to give greater details of the image. In this paper, we propose an approach which enables accessibility of images. This paper has two parts, the first part describes a mechanism to build a *multi level description* of an image to enable accessibility or readability while the second part describes an user interface that enables navigation of the multi level description by hovering on the image.

1 Introduction

A picture is worth a thousand words is a well know adage and refers to the fact that any complex description of words can be conveyed, probably better, with a still image. The use of images to describe something is often used without a second thought and serves the purpose of compressing a large number of textual words. The use of images in any form becomes an accessibility issue for the visually impaired. Images can present a major obstacle to individuals that are blind or have low vision. With the proliferation of web usage the accessibility of a document has become synonymous with the accessibility of web document. In this paper we restrict ourselves to accessibility of images which are integral part of any document in general. There are several relatively simple techniques that can make an image accessible on the web [4]. "Section 508 Standard (a)" [15] of the US government addresses proper use of an image for accessibility. The world wide web consortium (W3C) [16] mandates the use of `alt` and `long-desc` attribute in hyper text markup language (HTML). While `alt` is designed to be an alternative text description for images the `long-desc` is a longish version of the alternative text description of the image. Most the discussion on accessibility

have been restricted to making known what the image represents at a very broad level, typically a a crypt caption of the image. Accessibility of graphics specifically in technical documentation is discussed at length in [10]. However discussion on making any image accessible or readable has not been addressed in literature specifically. In this paper we develop a method that allows making any image accessible so that it can be read by the visually impaired.

The rest of the paper is organized as follows, in Section 2 we describe the problem with an example. We look at an overview of related literature in Section 3 and introduce the contribution of this paper. We describe our work on image readability (accessibility) in Section 4 and conclude in Section 5.

Fig. 1. `converse.jpg`: "Two people conversing" (from [7])

2 The Problem

Consider the image for example Fig. 1. This as a standalone picture or in a document or in a web page [7] is not completely accessible to the visually impaired. The only description of the image is through the caption (*Description 1.*) of the image.

Description 1 *Two People conversing*

Notice that there is certainly a longer description that is possible which the normal person would *read* in the image, for example, *Description 2.*

Description 2 *Two people, one in a black coat and a red tie with a black eye wear, balding, fair, · · · sitting on the left · · · and the other in a white shirt and left leg over the right leg, dark complexioned, with black hair and beard · · · sitting on the left of the person in coat · · · sitting on a brownish wooden bench set against a dark brown checkered wall bearing a caption WINE SALES written in Roman all capital Font in light brown with bushes of flowers colored red and white to their right and red and violet to their left with all the flowers in front and some more flowers handing from the top colored yellow and violet with green leaves surrounding them, <more description> conversing.*

While *Description 2.* enables a person to *read* the image and make it more accessible, however it is not unique. Another possible description enabling accessibility of the image could be as shown in *Description 3.*

Description 3 *Two people, sitting on a brownish wooden bench set against a dark brown checkered wall bearing a caption WINE SALES written in Roman all capital Font in light brown with bushes of flowers colored red and white to their right and red and violet to their left with all the flowers in front and some more flowers handing from the top colored yellow and violet with green leaves, one person is in a black coat and a red tie with a black eye wear, balding, fair, · · · sitting on the left · · · while the other in a white shirt and left leg over the right leg, dark complexioned, with black hair and beard · · · sitting on the left of the person in coat · · · surrounding them, <more description> conversing.*

While *Description 1.* is cryptic and conveys an *overall* description of the image (a typical caption associated with an image), the *Description 2.* is more *informative* and captures finer details embedded in the image. The description a user infers when looking at this image lies probably somewhere in between *Description 1.* and *Description 2.* Depending on the personal interest, background and the context in which the user is looking at the image the longish yet sequential description (*Description 2.*) can change. One way is to describe the surroundings (about the flowers etc) in the image first and then tell about the people in the image, say as in *Description 3.* Notice that there is (a) no one sequential way of describing an image and (b) there is no limit to the amount of description, say in words, that one can associate with the image. Several researchers, for example, [2, 5, 12, 9, 13, 11], have shown that the sequence in which the image is visualized by people depends on the cultural background of the person viewing the image. In this paper, given that the description of the same image could be very brief or very long and when description is long there is no one preferred sequence in which the image cab be described, we develop a method that allows an user to read the description of the image the way he desire, both in terms of the sequence and the density of description.

3 Related Literature

The issue of image accessibility has been discussed in literature sparsely and most of them are found in terms of patents. Accessibility of graphics in technical

documentation [10] discusses a mechanism to read out block diagrams and the likes of it which are usually accompanied in a technical document, in some sense it uses the descriptions associated with the generation of the block diagrams.

A recent published patent application [3] describes a method for rendering annotations associated with dense and huge map images, typically of size 5 GB with limited viewing capability, say 1 MB. They describe an interface to enable panning and zooming of the image which has an annotations with respect to a location on the image. The annotations content can be an audio loop, narrative audio, text labels, etc. The essential aspect is that the huge image has been annotated at a single level and the method of zooming and panning essentially brings into viewing focus a smaller part of a huge image and hence making only the description associated with that that part of the image active. In short, the image is annotated at one level only and only that annotation that corresponds to the part of the image in focus becomes active.

In [14] a method and system to process a Digital Image is describe. It describes a method to embed an audio data within an image to provide an embedded image wherein the audio data is freely recoverable from the embedded image. This process embeds an audio stream into an image and constructs a new image. The patent also speaks of storing more or less audio data on the image in terms of the the audio quality of the embedded audio.

In [6] a method to capture an image and encode the audio data using markings in the image that are substantially imperceptible to an unaided eye of a human viewer is discussed. It talks about making *a priori* markings in an image which would be mapped to a corresponding audio embedded into the image file. This patent also speaks of embedding audio file into the image and making that portion of the audio active which corresponds to the image. Along similar lines, in the patent [1] a method that enables visually impaired users to navigate websites and hear high quality audio of narration and description of each website is disused. The system involves creating an audible website corresponding to an original website by utilizing voice talent to read and describe web content and creating audio files for each section of the original website and then assign these audio files to the respective sections of the website. Text, images, and other rich media content on the website are represented by audio files. However the description of images is restricted to the caption associated with the image or the `alt` tag associated with the web page.

While the attempt is towards making image accessible they do not address the problem stated in Section 2 namely that of allowing the user to choose the amount of description that they can see and in the sequence in which they desire to see the description. The main contribution of this paper is in identifying a method which allows the user to read a description of the image in the sequence they desire (*Description 2.* or *Description 3.*) and at the level of description (*Description 1.* versus *Description 2.*) they desire. In brief, we describe a method that specifically describes an image at multiple levels and allows the user to jump the description sequence and density asynchronously to read the description of the image.

4 Readable Images

There are essentially two parts in making images accessible. The first part is to build and organize the description associated with the image with the express view to enable readability or accessibility of the image. We describe a mechanism to annotate an image at several levels. At one end of the annotation spectrum, namely the finest level is the description of each and every pixel corresponding to the image while at the other end of the spectrum is a gross description of the complete image, namely all pixels. In between there two levels of description are several levels of descriptions. The second part is a user interface that enables visually impaired person navigate the description of the image by moving the pointer device (a mouse) on the image. Specifically the position of the mouse on the image will make all the descriptions that have been annotated corresponding to that (x, y) location of the mouse active and the up and down scroll of the scroll button on the mouse will activate a description at the finer level or the coarser description of the image respectively. A text to speech engine (TTS) then reads out the description enabling the visually impaired person navigate and read the image he desires. The image (see Fig. 1, `converse.jpg`) could be a jpg, pgm, ppm, gif, bmp or any of the multitude of known image formats and the multi level description of this image is captured as a text file (see Fig. 2, `converse.des`). In this paper, we show (a) a mechanism (manual or semi supervised) to create the description file (Fig. 2), (b) representation of the description at different levels and (c) a method to access the multi level description of the image there by making the image accessible.

Consider an image I to be made up of $M \times N$ pixels, and to represent a pixel in the $(k, l)^{th}$ position we use the notation $I(k, l)$. Clearly k can take values from 1 to M while l can take values from 1 to N. Let there be K levels of descriptions, level K being the coarsest which could be the caption of the image (*Description 1.* , namely, "Two people conversing") and level 1 be the finest description of the image (every pixel described). Level n has a coarser descriptions than level $n + 1$.

4.1 Building Multi level Description

The method of creating multi level description of an image is semi supervised. The finest level details corresponding to each pixel can be captured automatically by identifying the color of the pixel so at `level 1` there is a description for each pixel, meaning there are MN descriptions associated with the image at `level 1`. The next level, `level 2` description can be achieved by image segmentation [8] which allows grouping of pixels that have similar properties, say pixels having the same texture. In Fig. 1 it could be the region associated with the brick wall behind the two people having a conversation. So a pixel in this area would have a `level 1` description as brown color while the `level 2` description would be brick wall. For example Fig. 3 has clearly very fewer details. For example, there are flower bushes, the brick wall, two people sitting next to each other, a bench. The annotation at this level would have only these broad description. Note that the

```
<description>
        <image>
                <name> converse.jpg </name>
                <size> M x N </size>
        </image>

        <pixel, x, y>
                <level 1>
                        <des> Black </des>
                </level 1>
                <level 2>
                        <des> Hair </des>
                </level 2>
                <level 3>
                        <des> Head </des>
                </level 3>
                <level 4>
                        <des> A person </des>
                        <des> Sitting to the right </des>
                </level 4>
                ...
                ...
                ...
                <level K>
                        <des> Two people having conversation </des>
                </level K>
        </pixel, x, y>
</description>
```

Fig. 2. A Typical Multi level description (`converse.des`)

same pixel will have multiple descriptions based on the level at which it has been annotated. For example a pixel on the head of the person sitting to the right (in white shirt) would have a `level` 1 description of black a `level` 2 description of hair, a `level` 3 description of head, a `level` 4 description of person to the right, sitting on a bench, in front of the wall, a `level` 5 description of conversing with a person on the left and so on until `level` K description Two people conversing or the caption of the Fig. 1. A typical `converse.des` would have a structure, as shown in Fig. 2. The method does not embed the description into the image file as an audio stream as discussed in a couple of patents in Section 3 but keeps the image file (`converse.jpg`) and the description file (`converse.des`) separate, further the description is stored as text information and not as an audio file.

4.2 Accessing Multi level Description

Accessing the multi level description associated with an image is shown as a flow chart in Fig. 4. A user interface has the capability of rendering the image and

Fig. 3. Different level image annotation

also determining the spatial and the scroll position of the pointer on the rendered image. As shown in Figure 4, the accessibility of the image based on the location of the pointer (typically a mouse pointer) on the image. When the pointer is on the image, the location of the pointer is determined (say (x, y)) and the level identified (say `level` L). Then the string between the `<des>` and `</des>` tags corresponding to the pointer location (x, y) and `level` L is extracted from the description file (`converse.des`). This string is then read out using a text to speech engine or a screen reader. The accessibility of the image at different levels is possible through level sensing, which is enabled through the scroll button the mouse. Fig. 5 shows the control which allows the selection of the level of description by moving the scroll button of the mouse up or down. The level control makes sure that the `level` remains between 1 and maximum number of levels (say, K).

5 Conclusions

Use of images in documents or otherwise is very common in technical literature. Any kind of image in a document fuels accessibility to the visually impaired. While attempts have been made to make images accessible in technical documents by providing longish descriptions there is however no attempt to the best of the authors knowledge to make an image in general accessible. In this paper we described a method which allows images to be read (and thus accessible) by the

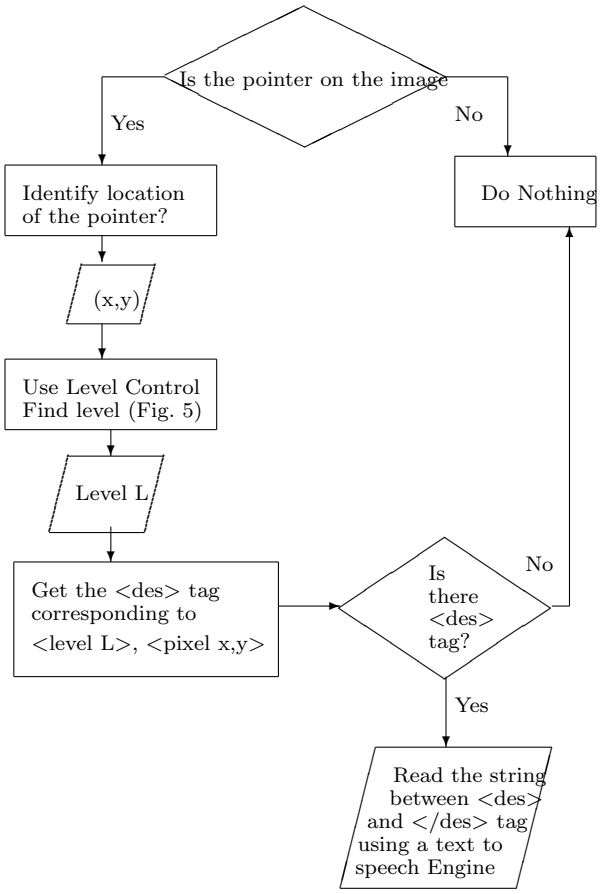

Fig. 4. Accessing multi level description of an image, given the image and the multi level description file.

visually impaired. We described a method that can be used to capture description of the image at multiple levels and showed how this multi level description associated with an image can be accessed by the visually impaired using a simple to use user interface. The proposed method not only enables a method to access image but also allows access information about the image to the desired depth and in the order in which the user is interested to read the image. This aspect removes the restriction that the user can read the image in only the sequence in which the image description has been captured by the creator of the description. The key contribution of this paper are (a) a method to annotate images at different levels, (b) a method to access information about the image at different levels, (c) a method to capture annotation in a description file at different levels and (d) a method that makes accessible an image in an non-sequential fashion.

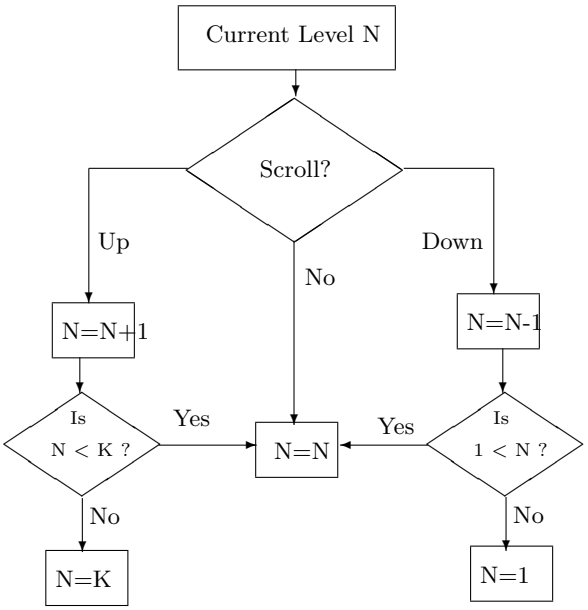

Fig. 5. Level Control to find `level` and K is the maximum number of levels at which an image is described.

Acknowledgments

The author would like to thank several members of the TCS Innovation Labs - Mumbai who have been instrumental in refining the idea of making an image readable.

References

1. Bradley, N.T.: Method and apparatus for website navigation by the visually impaired. US Patent Publication Publication No. US 7653544 B2 (2010), `http://ip.com/patent/US7653544`
2. Chua, H.F., Boland, J.E., Nisbett, R.E.: Cultural variation in eye movements during scene perception. Proceedings of the National Academy of Sciences of the United States of America 102(35), 12629–12633 (2005), `http://www.pnas.org/content/102/35/12629.full.pdf+html`
3. Cohen, M.: Rendering annotations for images. US Patent Application No. US 2010/0085383 A1 published on 08-Apr-2010 (2010), `http://ip.com/patapp/US20100085383`
4. Creating accessible images. `http://www.doit.wisc.edu/accessibility/online-course/standards/images.htm`
5. Evans, K., Rotello, C.M., Li, X., Rayner, K.: Scene perception and memory revealed by eye movements and receiver-operating characteristic analyses:

Does a cultural difference truly exist? The Quarterly Journal of Experimental Psychology 62(2), 276– 285 (2009), http://www.informaworld.com/10.1080/17470210802373720
6. Inness, G.: Picturs with embedded data. US Patent Publication No. US 2005/0068589 A1 published on 31-Mar-2005 (2005), http://ip.com/patapp/US20050068589
7. IVCNZ98: (1998), http://www.citr.auckland.ac.nz/~ivcnz98/pictures/wineHouse_8.jpg
8. Kang, W.X., Yang, Q.Q., Liang, R.P.: The comparative research on image segmentation algorithms. Education Technology and Computer Science, International Workshop on 2, 703–707 (2009)
9. Miellet, S., Zhou, X., He, L., Rodger, H., Caldara, R.: Investigating cultural diversity for extrafoveal information use in visual scenes. Journal of Vision 10(6) (2010), http://www.journalofvision.org/content/10/6/21.abstract
10. Murphy, S.: Accessibility of graphics in technical documentation for the cognitive and visually impaired. In: Proceedings of the 23rd annual international conference on Design of communication: documenting & designing for pervasive information. pp. 12–17. SIGDOC '05, ACM, New York, NY, USA (2005), http://doi.acm.org/10.1145/1085313.1085320
11. Press, A.: In Asia, the eyes have it. http://www.wired.com/culture/lifestyle/news/2005/08/68626
12. Rayner, K., Castelhano, M.S., Yang, J.: Eye movements when looking at unusual/weird scenes: Are there cultural differences? Journal of Experimental Psychology: Learning, Memory, and Cognition 35(1), 254–259 (2009)
13. Roach, J.: Chinese, Americans truly see differently, study says. http://news.nationalgeographic.com/news/2005/08/0822_050822_chinese.html
14. SIM, W.H., HII, Toh Onn, D.: Method and system to process a digital image. World Intellectual Property Organization Publication No. WO/2005/05983 (2005), http://www.wipo.int/pctdb/en/wo.jsp?WO=2005059830
15. USGovernment: Resources for understanding and implementing Section 508. http://www.section508.gov/
16. World wide web consortium: Accessibility. http://www.w3.org/standards/webdesign/accessibility