Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 67 (2015) 2 – 11
Procedia Computer Science

6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Infoexclusion (DSAI 2015)

Automatic description of SVG images for the visually impaired: a Gestaltic approach

Vítor Carvalho[a]*, Diamantino Freitas[a]

[a]*Faculdade de Engenharia da Universidade do Porto, Rua do Dr. Roberto Frias, 4200-465 Porto, Portugal*

Abstract

In this paper, a new approach to the automatic description of SVG images for the visually impaired based on Gestalt theory is presented, using levels with increased details. The description begins as a whole, considering relations of alignment, symmetry and group formation of the elements. Next, qualitative values for shape, locations, measures and colors are used to characterize each element. Finally, quantitative values are given for these properties. Tests carried out with users, compared automatic descriptions with human made. There were improvements of 9%, using automatic descriptions. Considering only the visually impaired, this figure rises to 18%.

Peer-review under responsibility of organizing committee of the 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2015)

Keywords: automatic image description; SVG; Gestalt theory; visually impaired;

1. Introduction

Visually impaired people represent 19% of world population. Of these, over 90% live in developing countries and over 18% have less than 50 years. Most of the visually impaired are not blind, suffering from various diseases such as glaucoma (which affects peripheral vision), age related macular degeneration (leading to loss of central vision), etc. [1].

* Corresponding author. Tel.: +351-220-413-061; fax: +351-225-081-537.
E-mail address: vitor@fe.up.pt

Peer-review under responsibility of organizing committee of the 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2015)
doi:10.1016/j.procs.2015.09.243

Assuming that education is a value and a right for everyone and that everyone should have access with the best conditions [2], it is of interest to create a method to automatically describe contents with a strong visual component (whose understanding is not directly transmitted through an alternative text), from the visual domain to other domains. One of these domains is hearing and, in this, one way of coding is verbal description. Another possibility is to perform the conversion to Braille and use a standard reading ruler to read. The conversion to text precedes both conversions feeding oral description and representation in Braille.

In areas such as engineering, those contents with a strong visual component may be technical drawings, graphs, charts and other complex documents, generally based on vector shapes produced by the respective editing programs. On the Web, the recommended vector format is SVG [3]. For this reason and at this stage, the development was carried out to this format but the paradigm can be applied to other kind of images, including raster.

Like everything that relates to usability and accessibility, this translation of the visual field to automatic text description, benefits all users and not just those who have visual impairment. A relevant example is the ability to search images or image elements based on their respective content and characteristics, obtained from the text, which forms the basis for the oral description.

Thus, the authors set out creating a new paradigm for the automatic description of simple SVG vector images to textual description in natural language. A custom application was developed to convey this automatic description to the users. Written in PHP, the custom application automatically converts the SVG image to textual description and uses a client side TTS (text-to-speech) to read the description. In order to control the cognitive load, the users can navigate through this same description using keyboard commands [4].

A review of the literature on this subject is presented in the next point followed, in chapter three, by the description of the proposed method. In chapter four the authors will present the preliminary assessment made on the automatic descriptions and, in chapter 5, a conclusion is stated.

2. Related Work

The literature is rich in examples of raster image analysis but there is no analysis for SVG images in the desired manner, i.e., there are algorithms to render SVG and not to do its automatic textual and natural spoken language description. We can advance some explanations for this:

- Raster images seem to be the preferred target of current investigation, perhaps because they are the most abundant form on the Web, being sought after by many people;
- The SVG language is a XML dialect, thought to be sufficiently descriptive of its base visual content. However, the SVG code is not the convenient way to convey the description of the generated image. Therein lies the interest of this work.
- The authors also think that there is still insufficient awareness on providing accessible Web content.

This lack of related work led to the adaptation of some of the methods employed in raster images in the description of SVG images.

Ordonez et al. [5] developed and demonstrated automatic methods for image description using a large collection of captioned photos. They developed a technique that automatically collects one million images from Flickr, with filtered noise until the results of associated subtitles were visually relevant. This collection allowed dealing with the extremely difficult problem of generating relatively simple description using non-parametric methods and produced surprisingly effective results.

Recently, Yao et al. [6] presented in their article, image analysis for textual description (I2T), a structure that generates text descriptions of image and video content based on image study. The proposed I2T structure follows three steps:

- Decompose the input images (or video frames) into their constituent visual patterns by an image analysis engine, in a similar spirit to analyzing sentences in natural language.

- Convert the image analysis results into semantic representation as Web Ontology Language (OWL), which permits integration with general knowledge bases.
- A text generation engine converts the results of previous steps in readable text reports, semantically meaningful and subject to consultation.

The case studies demonstrate two automatic systems I2T: maritime and urban video surveillance system and an automatic system of real-time understanding of driving.

Castillo-Ortega et al. [7] present, in their article, a preliminary proposal to linguistic description of images. The base of approach is a hierarchical fuzzy image segmentation, a set of linguistic features describing each area and the diffuse spatial locations and relationships. The process is independent from the origin of these elements and provides a description with the characteristics of a synthesis, i.e., a brief and accurate description of the entire image. They disclose that this can provide a description of disjoint regions containing phrases appearing in different levels of detail.

Zhang et al. [8] holds that digital images are increasing worldwide. Thus, there is a growing interest in finding images in large collections or remote databases. In order to find an image, they must show or describe certain characteristics. The shape is an important visual feature of an image. Finding images using features related to the shape has attracted much attention. There are many techniques of representation and description in the literature. Their article classifies and reviews these important techniques. It also examines the implementation procedures for each technique, discussing its advantages and disadvantages, presents the results of some recent research and identifies promising techniques.

Although describing images is not the goal, the contribution given by Ferreira and Freitas [9] regarding the automatic reading of mathematical formulas in MathML contributed substantially in the preparation of this work. Both situations assume a common point, a document-based format of XML and the way the "navigation" occurs in a mathematical equation can find some parallelism in the textual description of an SVG image.

3. Proposed method

The automatic description covers the perceptual organization of the constituent elements of the image. The approach given by Gestalt theory can fit this requirement concerning, for now, the occurrence of symmetries, alignments and group formation.

Besides these aspects, the description provides, for each element in the SVG image, information regarding shape, spatial location and color.

Finally, the methodology tries to synthesize all this information in a coherent automatic description. In order to prove this methodology, a custom Web application written in PHP was developed by the authors. This application automatically converts the SVG image to textual description. It uses a client side TTS (text-to-speech) to read that description. To control the cognitive load, users can navigate through this description using keyboard commands [4].

All these subjects will be covered in the next subchapters.

3.1. Perceptual Organization

Wagemans et al. [10] point that grouping is the most associated phenomenon in visual perceptual organization.

This effect results when some elements of the visual field appear to be more together. This also happens when the clues are weak and disparate. The human brain seems to have the ability to combine them synergistically in order to form strong evidence of grouping.

Some psychophysical and computational studies about groups, using carefully built stimuli [11], allowed quantifying some grouping principles used in this work, based on Gestalt theory:

- Proximity – grouping between two elements increases as these elements are close to each other; conversely, the strength of proximity grouping exponentially decays with increasing distance between the elements [11].
- Shape perception – symmetry and parallelism [10].

With only two elements in the image, there is the sought for horizontal and vertical relations of symmetry relative to the central vertical and central horizontal axis of the image. For this purpose, there is a function that receives the coordinates of the central point of each image element to determine if the x and y coordinates are almost the same.

If there are more than two elements in the image, a correlation analysis between the coordinates of the center points is performed to look for linearity by the linear correlation coefficient R (Equation 1) [12].

$$R = \frac{\sum x \times y}{\sqrt{(\sum x^2) \times (\sum y^2)}}; x = x_i - \bar{x}; y = y_i - \bar{y}; \qquad (1)$$

If the coefficient R is greater than 0.9 (positive linear correlation) or less than -0.9 (negative linear correlation), the elements are considered to be aligned. Another analysis of this level is the formation of groups.

For this, the authors developed an algorithm, based on a nearest neighbor analysis, given by R_n (Equation 2) [13].

$$R_n = \frac{\bar{D}(Obs)}{0.5 \times \sqrt{a/n}} \qquad (2)$$

In this formula, the numerator is the average distance of the observed nearest neighbor ($\bar{D}(Obs)$) and in the denominator, under the radicand, the study area (a) above the total number of points (n). This formula gives a value between 0 and 2.15. It takes the value 0 if the points are close together, 1.0 if randomly arranged and 2.15 if arranged regularly. Although the authors are considering to incorporate the formula, as is, in a future version of the application, at the present time, the decision was to implement an empirical formula to return groups of elements in the image (even if each group had few elements).

Thus, the proposed algorithm, measures the distance of each element in the image to the other elements. If this distance is less than one quarter of the maximum image diagonal, it is considered that these elements form a group. The algorithm ensures that each pair of elements appears only once.

3.2. Shape

The elemental form is a basic visual feature to describe the content of an image [8] and is an important property for the perceptual recognition of objects or elements and classification of images [14]. However, the representation of shape and its description is a difficult task in raster images [8]. Given the descriptive and parametric nature of SVG, the task is easier for the application.

Falomir et al. [15] present descriptions of some shape attributes, such as comparing lengths in a qualitative way like shorter, half the size, a little bit shorter, same size, etc. In this paper, the authors chose to calculate the ratio (r_l) between the length of an element attribute (diagonal of a circle, side of a square, etc.) and the width of the entire image, giving that length a qualitative name (L_n) based on a quantitative approach (Equation 3). The interval values were obtained from the mean values between each exact value for a given expression, rounded to three decimals.

$$L_n = \begin{cases} "five-sixths" & if \quad r_l > 0.825 \\ "four-fifths" & if \quad 0.775 < r_l \le 0.825 \\ "three-quarters" & if \quad 0.708 < r_l \le 0.775 \\ "two-thirds" & if \quad 0.583 < r_l \le 0.708 \\ "half" & if \quad 0.417 < r_l \le 0.583 \\ "a\ third" & if \quad 0.292 < r_l \le 0.417 \\ "a\ quarter" & if \quad 0.225 < r_l \le 0.292 \\ "a\ fifth" & if \quad 0.185 < r_l \le 0.225 \\ "a\ sixth" & if \quad 0.155 < r_l \le 0.185 \\ "a\ seventh" & if \quad 0.131 < r_l \le 0.155 \\ "an\ eighth" & if \quad r_l \le 0.131 \end{cases} \qquad (3)$$

The width or thickness of the perimeter for each element is given in a qualitative way. To name the thickness of the perimeter (T_n), a ratio (r_t) between the mean of the width and height of the entire image and the width of the perimeter is established (Equation 4).

$$T_n = \begin{cases} \text{"thin"} & if \quad r_t > 150 \\ \text{"medium"} & if \quad 75 < r_t \leq 150 \\ \text{"thick"} & if \quad r_t \leq 75 \end{cases} \tag{4}$$

By the direct analysis of the tags that have correspondence to each element of the SVG, the automatic description can locate lines, rectangles, circles, ellipses, and polygons. Having no particular tag in SVG to specify other kinds of shapes, the automatic description also extracts other types of shapes by analyzing the mathematical properties of the attributes belonging to the former tags, in search of squares, triangles, lozenges and naming some polygons.

3.3. Spatial location

Falomir et al. [15] proposes, following other models, a description of absolute and relative orientation of objects with the observer based on the division of space in eight directions surrounding each object: left, front-left, front, front-right, right, back-right, back, left-back. This work considered another approach. The authors divided the analyzed image into nine equal parts (Fig. 1a): upper left, upper center, upper right, center left, center, center right, lower left, lower center and lower right to locate the central point of each element.

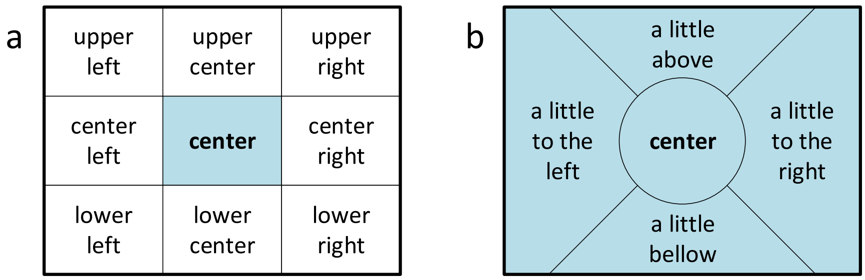

Fig. 1. (a) Division into Nine Equal Parts of the Analysed Image; (b) Central part of the analysed image: division into five areas.

If the central point of an element belongs to the central part of the image, usually with more information, the authors perform a second and thinner correspondence (Fig. 1b). If the distance of the element's central point to the central point of the image is lower than 6% of the image height and width average, the point belongs to the image center. Otherwise, the qualitative name of the position (P_n) follows Equation 5, where θ is the angle of the line that connects the central point of the image to the central point of the element.

$$P_n = \begin{cases} \text{"a little to the right"} & if \quad \theta > 315° \vee \theta \leq 45° \\ \text{"a little above"} & if \quad \theta > 45° \wedge \theta \leq 135° \\ \text{"a little to the left"} & if \quad \theta > 135° \wedge \theta \leq 225° \\ \text{"a little bellow"} & if \quad \theta > 225° \wedge \theta \leq 315° \end{cases} \tag{5}$$

3.4. Color

The color names are linguistic labels given by humans. They are used routinely and similarly with no effort to describe the world. Identifying the color by its name is a method of communication that all people understand [16]. The fields of visual psychology, anthropology and linguistics studied color naming.

The color information is critical in applications such as art, fashion, product design, advertising, film production and printing [17]. Mojsilovic [16] seems to have a similar opinion when she says that although the naming of colors is one of the most common visual tasks, it did not receive significant attention by engineers.

Today, with the emergence of visual technologies, sophisticated interfaces with the user and man-machine interactions, the ability to appoint individual colors, pointing out objects of a certain color and convey the impression of dithering, becomes an increasingly important task [16].

The advantages of naming colors have to do with image search, automatic image labelling, assisting individuals who are color blind and human-computer linguistic interaction [18]. Everyone can benefit from automated methods to describe and recognize color information [17].

Falomir et al. [15], in their qualitative description of color, begin to translate it from RGB (Red, Green, and Blue) system to HSL (Hue (H), Saturation(S) and Lightness (L)), claiming to be most suitable for color nomination by ranges of values. Guberman et al. [19] seem to have same opinion when they say that it is a more fruitful approach to color from the human point of view, i.e., based on the luminosity, gamma and saturation instead of the amount of additive primary colors like red, green or blue. Likewise, RGB is substituted by other color models similar to HSL. Authors like Arivazhagan et al. [20] use the HSI (hue, saturation, intensity) model to recognize fruit by color. Hatzidimos [21], uses the HSV model (hue, saturation, value) to recognize traffic signs.

Falomir et al. [15] present a model for the qualitative color description which separates only Lightness dependent colors (black, dark grey, grey, light grey and white) of the remaining specified colors (red, yellow, green, turquoise, blue, purple and pink). The latter still have name variations adding “pale”, “light” and “dark” adjectives.

This paper made a few adjustments to the Falomir et al. proposed model [15], adjusting the gray levels (G_l) (Equation 6) and introducing the brown, orange and yellow colors (C) (Equation 7). The authors translate variations of colors (C_v) dependent from other components besides Hue by the following adjectives: “pastel”, “light”, “whitish” and “dark” (Equation 8).

$$G_l = \begin{cases} \text{"}dark\ gray\text{"} & if \quad (0 \le S \le 20) \wedge (20 \le L < 50), \forall H \\ \text{"}gray\text{"} & if \quad (0 \le S \le 20) \wedge (50 \le L < 80), \forall H \\ \text{"}light\ gray\text{"} & if \quad (0 \le S \le 20) \wedge (80 \le L < 95), \forall H \\ \text{"}white\text{"} & if \quad (0 \le S \le 20) \wedge (95 \le L \le 100), \forall H \end{cases} \tag{6}$$

$$C = \begin{cases} \text{"}brown\text{"} & if \quad (\mathrm{S} < 80) \wedge (15 < \mathrm{H} \le 50) \wedge (\mathrm{L} > 20) \\ \text{"}orange\text{"} & if \quad (\mathrm{S} \ge 80) \wedge (15 < \mathrm{H} \le 50) \wedge (\mathrm{L} > 20) \\ \text{"}yellow\text{"} & if \quad (50 < \mathrm{H} \le 65) \wedge (\mathrm{L} > 20), \forall \mathrm{S} \end{cases} \tag{7}$$

$$C_v = \begin{cases} \text{"pastel"} & if \quad (20 < \mathrm{S} \le 50) \wedge (30 < \mathrm{L} \le 70), \forall \mathrm{H} \\ \text{"light"} & if \quad (50 < \mathrm{S} \le 100) \wedge (70 \le \mathrm{L} \le 95), \forall \mathrm{H} \\ \text{"whitish"} & if \quad (50 < \mathrm{S} \le 100) \wedge (\mathrm{L} > 95), \forall \mathrm{H} \\ \text{"dark"} & if \quad (50 < \mathrm{S} \le 100) \wedge (\mathrm{L} \le 30), \forall \mathrm{H} \end{cases} \tag{8}$$

3.5. Description and navigation synthesis

The image description and navigation through the description suits the notions considered as the basis of the Gestalt theory, presented by Wagemans et al. [11].

- **Holism** – the perceptual experiences are intrinsically holistic and organized by rejecting atom-ism and associationism, as well as any summative approach. Whatever parts (properties, elements) are perceived holistically and not in a separate or independent manner. However, per-sons perceive shape and color separately.
- **Emergency** – emergent properties and superiority of configuration. Emergent properties belong to the whole and not the individual parts (e.g., the density of a forest applies to the whole forest and not to a single tree).
- **Configuration superiority** – persons perceive the parts after the whole (using the same example, an observer perceives the existence of a forest before focusing on the trees that compose it).
- **Global precedence** – processing happens from the global structures to the analysis of local properties. The persons process first the overall properties of a visual object, followed by the analysis of the local properties.
- **Primacy of the whole** – the properties of the whole cannot derive from the properties of its constituents. They are born of interparty relations: symmetry, regularity, closing, etc.

Thus, the authors picked four levels of detail, so that the description can achieve the accuracy required by the growing complexity and level of detail while providing precise and cognitively correct brief descriptions.

The first level focuses on the aspect of image, background color and number of elements.

The second level breaks down the elements regarding the shape, indicating how many elements of each type exist. It also concerns some aspects of the Gestalt theory, regarding the picture as a whole, evaluating the existence of alignments, symmetries and group formation.

The third level is concerned with elements description, indicating its type, approximate dimensions in relation to the image width, approximate location relative to the image, the name of the fill color and, for the perimeter, the name of its color and type of thickness.

The fourth level is similar to the third but introduces a deeper technical description, in which all values are numeric, in order to disambiguate any approach or names given in the third level.

4. Evaluation

4.1. Test images

To test the automatic descriptions, the authors chose eight images (Fig. 2) created with Adobe ® Illustrator ® CS6 software, thought to represent some of the problems of analysis and interpretation, including Gestalt theory, spatial visualization, semantic and cognitive load.

Each image has an associated set of attributes that users need to memorize:

- Image background color;
- For each of its constituent elements: location, size, shape, fill color, perimeter thickness and perimeter color.

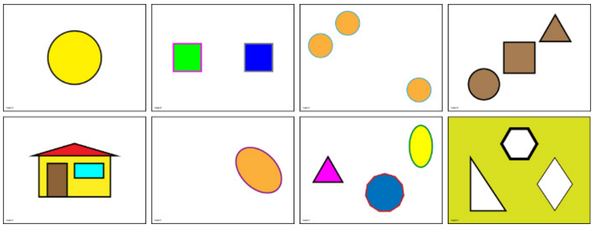

Fig. 2. Test images

4.2. Example of an automatic description

The custom application automatically describes the first image in the top left corner of Fig. 2 as:

Level 1: The image is a horizontal rectangle, white and with one element.
Level 2: The element is a circle.
Level 3: The circle 1 has a diameter of one-third the length of the image, filled in yellow, with a medium thickness black perimeter and is located in the center of the image.

4.3. Test images description by expert users

First, three expert users from University of Porto and working in areas such as fine arts, computer graphics and technical drawings, described the images.

On average, each image took the experts four minutes to describe.

4.4. Test subjects without visual impairment

The authors recruited eight people, college graduates at University of Porto in areas where SVG images may be relevant such as geography, informatics, mathematics, physics and computer engineering.

In the first part of the test the participants were asked to read the description of four images on paper (made by one of the expert users), taking the desired time to understand the image depicted. As soon as they completed the reading of an image description, the description was removed and the participants were asked to draw on paper the image corresponding to the description using felt-tip pens of different colors (Fig. 3).

In the second part of the test, after a period of explanation on the operation and navigating through the description using the custom application, participants were asked to hear the description of four images, taking the desired time to understand the images depicted. As soon as they completed the image description hearing, the application was withdrawn and the participants were asked to draw on paper the image corresponding to the description using felt-tip pens of various colors (Fig. 3).

The authors had the care to distribute randomly the expert users' descriptions, always having one male and one female as a reader.

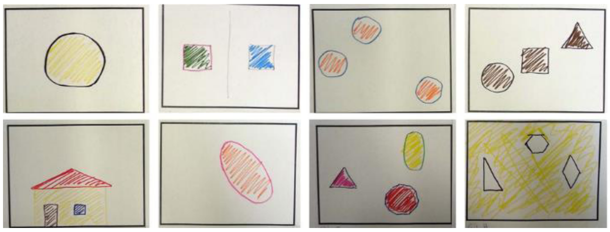

Fig. 3. Example of drawings from a user without visual impairment

The images were given in random order for each pair of the previously referred users. Four image descriptions to read on paper and four to hear in the custom application. However, if the first element of the pair read the description of the image, the second element of the pair heard the description of the image in the custom application and so on. The user U6 heard all the images in the application in a random sequence.

4.5. Test subjects with visual impairment

The authors recruited three people, college graduates at University of Porto in areas where SVG images may be relevant, such as information science, literature and history.

In the first part of the test, after a period of clarification on the operation and navigation in the description using the application, the users were asked to hear the description of four images, taking the needed time to understand the image depicted. Once they concluded hearing the description of the image, the application was withdrawn and they were asked to compose, using various geometric sponge figures, the image corresponding to the description (Fig. 4).

In the second part of the test the users were asked to read the expert users' description of four images (stored, each one of them, in a text file) using their usual screen reader and taking the desired time to understand the image depicted. As soon as they completed the reading of the image description, the computer was removed and they were asked to compose, using various geometric sponge figures, the image corresponding to the description (Fig. 4).

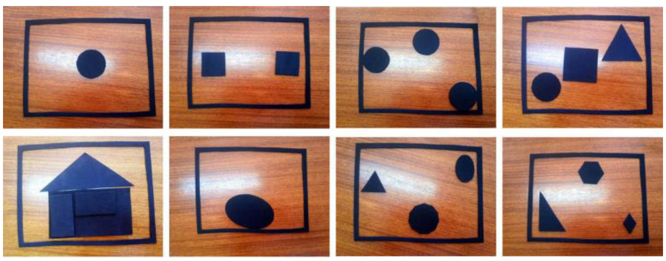

Fig. 4. Example of compositions from a user with visual impairment (blindness)

4.6. Score from assimilated descriptions

The drawings and compositions analysis was not made for its artistic quality, but rather for what users wanted to represent. Notes were taken during user testing and, jointly, an exempt person evaluated the drawings in a later stage of the data collection. The authors sought for the accordance of the aspects for each image in the qualitative characteristics described in paragraph 4.1. For a correct aspect, 1 point was assigned. Being incorrect, no points were assigned. At the end of each image and for each user, the average of all the attributes was found and multiplied by 100 to obtain a percentage termed "perfection".

4.7. Conclusion from the tests

The first conclusion to draw is that the description of images by expert users is a time consuming task, with an average of four minutes for each image. Furthermore, the result of the expert users' descriptions was not ideal, because in this evaluation work more than half of the end users pointed out that the descriptions were not homogeneous and used ambiguous terms to describe the images. In fact, the average "perfection" (see paragraph 4.6) of all users who read the expert users' descriptions was 70%. The average "perfection" for users who heard the descriptions in the application was 79% (9% more than reading on paper). Considering only the visual impaired users, the average stands at 87% using the application and the automatic description and 70% using the screen reader and the expert users' descriptions, a fall of 17%.

Taking into account all users, there was an improvement of 7% when they heard the automatic description on the application rather than reading the human made descriptions on paper or with a screen reader. Users who decreased their performance when using the automatic descriptions and the application, worsened on average 1%. Users who improved their results when using the automatic descriptions and the application, improved on average 15%.

Given the above, the authors consider that the application is an asset to describe SVG images, benefiting general users and even more the visually impaired.

5. Conclusion

The authors created a new paradigm for the automatic description of SVG images based on Gestalt theory.

The automatic descriptions were tested with a custom application and a representative set of users.

The best results were achieved using the automatic description read by the custom application. Taken all users, there were improvements of 9%. If the analysis focuses only on users with visual impairments, this figure rises to 18%. A larger population will further test statistical proof.

It is provided, for the first time, access to the full contents of SVG images to users with visual impairments in a user-friendly automatic description that addresses sensitive issues such as the cognitive loads.

References

[1] Pun T, Roth P, Bologna G, Moustakas K, Tzovaras D. Image and video processing for visually handicapped people. *Journal on Image and Video Processing* 2007: 12.

[2] SAEDUP. Serviço de Apoio ao Estudante com Deficiência da Universidade do Porto [online]. 2013. Available from: URL: http://sdi.letras.up.pt/default.aspx?pg=saedup02.ascx&m=11

[3] W3C. Scalable Vector Graphics (SVG) 1.1 (Second Edition) [online]. 2011. Available from: URL: http://www.w3.org/TR/2011/REC-SVG11-20110816/

[4] Carvalho V, Freitas D. Converting SVG Images to Text and Speech. In: *Proceedings of the International Conference on Enabling Access for Persons with Visual Impairment*: 2015 Feb 12-14; Athens, Greece. Athens: National and Kapodistrian University of Athens; 2015.

[5] Ordonez V, Kulkarni G, Berg TL. Im2Text: Describing Images Using 1 Million Captioned Photographs. *Advances in Neural Information Processing Systems* 2011: 1143-1151.

[6] Yao BZ, Yang X, Lin L, Lee MW, Zhu SC. I2T: Image Parsing to Text Description. *Proceedings of the IEEE* 2010; 98(8): 1485-1508.

[7] Castillo-Ortega R, Chamorro-Martínez J, Marín N. Describing Images Via Linguistic Features and Hierarchical Segmentation. In: *IEEE International Conference on Fuzzy Systems (FUZZ)*: 2010 Jul 18-23; Barcelona, Spain. IEEE; 2010.

[8] Zhang D, Lu G. Review of shape representation and description techniques. *Pattern Recognition* 2004; 37: 1-19.

[9] Ferreira H. Leitura Automática de Fórmulas Matemáticas [M. Sc. thesis]. Porto: Faculdade de Engenharia da Universidade do Porto; 2005.

[10] Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson MA, Singh M, Heydt R. A Century of Gestal Psychology in Visual Perception: I. Perceptual Grouping and Figure-Ground Organization. *Psychological Bulletin 2012*; 138(6): 1172-1217.

[11] Wagemans J, Feldman J, Gepshtein S, Kimchi R, Pomerantz JR, Helm PA, Leeuwen C. A Century of Gestalt Psychology in Visual Perception: II. Conceptual and Theoretical Foundations. *Psychological Bulletin* 2012; 138(6): 1218-1252.

[12] Stat Trek. Linear Correlation Coefficient [online]. 2012. Available from: URL: http://stattrek.com/statistics/correlation.aspx.

[13] Barcelona Field Studies Centre. Nearest Neighbour Analysis [online]. 2013. Available from: URL: http://geographyfieldwork.com/nearest_neighbour_analysis.htm.

[14] Prasad BG, Biswas KK, Gupta SK. Region-based image retrieval using integrated color, shape, and location index. *Computer Vision and Image Understanding* 2004; 94: 193-233.

[15] Falomir Z, Jiménez-Ruiz E, Escrig MT, Museros L. Describing Images Using Qualitative Models and Description Logics. *Spatial Cognition & Computation: An Interdisciplinary Journal* 2011; 11(1): 45-74.

[16] Mojsilovic A. A Computational Model for Color Naming and Describing Color Composition of Images. *IEEE Transactions on Image Processing* 2005; 14(5): 690-699.

[17] Syeda-Mahmood T, Petkovic D. On describing color and shape information in images. *Signal Processing: Image Communication* 2000; 16: 15-31.

[18] Weijer J, Schmid C, Verbeek J. Learning Color Names from Real-World Images. In: *Conference on Computer Vision and Pattern Recognition*: 2007 Jun 17-22; Minneapolis, USA. IEEE; 2007.

[19] Guberman S, Maximov VV, Pashintsev A. Gestalt and Image Understanding. *Gestalt Theory* 2012; 34(2): 143-166.

[20] Arivazhagan S, Shebian R, Nidhyanandhan S, Ganesan L. Fruit recognition using color and texture features. *Journal of Emerging Trends in Computing and Information Sciences* 2010; 1(2): 90-94.

[21] Hatzidimos J. Automatic traffic sign recognition in digital images. In: *Proceedings of the International Conference on Theory and Applications of Mathematics and Informatics - ICTAMI*: 2004 Sep 16-18; Thessaloniki, Greece. 2004.