Colors Identification for Blind People using Cell Phone

A.L. Dominguez and J.P. Graffigna
Gabinete de Tecnología Médica. Facultad de Ingeniería.
Universidad Nacional de San Juan, San Juan, Argentina.
email: anadominguez_22@hotmail.com, jgraffig@gateme.unsj.edu.ar

Abstract. Assistive Technology (AT) is an interdisciplinary research area that allows finding solutions to the individual with disability [1] by easing or improving the functions or the skills for accomplishing daily activities. A technology can be considered “assistive” if it is fit for the needs, skills and capabilities of the person, taking into account mainly the intended activity and the limitations of the context and environs where the person performs such activity.

The current work intends to solve the problems of vision impaired persons to recognize colors. To this aim, a Java application for cell phones has been made which lets complement the mobiles’ technology with that of image processing. The means to obtain the colors from a view are based on analysing the different color models join to a mechanism to reduce the collected data. This paper describes preliminary experiences, methodology and results considering the user perception.

1. Introduction
AT has as a main objective to ease or allow for accomplishing a certain task within a given setting. The professionals that work with AT are in charge of priorly pondering the various factors involved so as to recommend the proper technical assistance that better fits to the needs and capabilities of the user in his/her specific environment [2]. Then, it is possible to state that AT is not aimed at solving the impairing condition; instead, it tries to attain acceptable functional results that the user seeks for his activities.

In the design of AT systems, devices are engineered to make the user functionally independent. In the case of sensorial impairment, ATs may offer the necessary assisting means to let the user acquire such sensorial information. To decide what kind of AT has to be developed for a given case calls for a prior analysis on the impairment state itself, its residual function and level. With this information, it is then possible to respond to whether the sensorial enhancement devices can be used or not. For example, if the residual sensorial function is insufficient, alternative sensorial ways have to be used [3]. Therefore, for persons whose vision cannot be used as the interface with the intended technology, other senses, like hearing and tactile functions may be studied as alternate ways.

For blinds, specifically, ATs offer help in various areas, such as obtaining visual information from the context, manipulating and controlling objects, bringing mobility and orientation of AT users, and other applications.

At present, AT for blinds looks for integrating the blind individual with his daily context by rendering the information means for visual recognition. In such a situation, a problem a blind has to overcome is to recognize colors. For example, it is important when deciding what to wear and how to combine clothing, or how to discern from different medicines, and similar situations. As is the case for
anyone, it is very important for the blind to know they are dressed well, with a proper color combination when attending any social activity or event, so as to prevent scornful attitudes from others.

At present, blinds that need to know the color of some object, or to identify the object by just knowing its color, are able to do so only by asking it to another person, when this one is nearby. Therefore, the application proposed here allows the blind accomplish such a task independently.

This work has been motivated by observing the evolution and development experienced in recent years on the use of embedded programming of mobile telephone devices (MTD), along with the need to have concrete solutions for the visually impaired persons.

MTD technology offers a wide range of tools, along with ample advantages and salient features like a low cost of mobiles; a large amount of information offered by the manufacturers as regards the various functions and technical specifications that are relevant to develop the mobile’s software for this application; the availability of numerous free public licence development environments and programming languages for the various phone models and operative systems available in the market [4]; the possibility of acquiring images; the ever-decreasing size of these electronic devices; its scarce need for maintenance, altogether with the possibility of getting communicated.

In Cook [5], a model is proposed to define an AT system by analysing the solution as a whole; that is, considering the activity to be done, the functional characteristics of the person and the context of use (see Fig. 1). For ATs devised for sensorial impairment and interface is required that is able to receive information from the context, plus a processor that obtains the useful information to be transmitted to the person through another interface. Specifically for this current work, the proposed design resorts to using the cellphone camera, colors processing software and the cellphone speaker used as the communicating interface with the person.

![Figure 1. Interaction of the AT with the other components of the HAAT model.](image)

Image processing requires that the various color models be priorly studied. The color information for each pixel can be specified in standard ways, and accepted through these color models. Each model is specified by a coordinate system, generally tridimensional, where each color is represented by a single point. The linear combination of coordinates generates the entire color space. These coordinates do not indicate which color it is from the viewpoint of the user’s perception [6], but it rather
objectively shows its location within a given color space. The more general spaces of color pursue to capture the largest number of colors visible to the human eye.

The chromatic models oriented to image management are: RGB, HSI, HSV, among other approaches [7]. The first one outstands from the rest by being the closest one to the hardware of artificial vision. The remaining two use perceptive attributes on luminance or intensity, saturation and color hue. They are defined as psychological, intuitive spaces, or user-oriented spaces, because they are the optimal ones for human interaction.

2. Methodology

The work proposed here consisted in obtaining the predominant color of a scene, and communicating it verbally to the user through a high-end cell phone (MTD). The methodology was divided into two parts involving, first, the color models study and, second, the implementation of an application into the device.

2.1. Color Models.

The models allow identifying distinctly the colors of an image. There are more than five different proposals according to their application. In this work, the models RGB (red, green and blue) and HSI (hue, saturation and intensity) have been studied. The first model is based on the characteristics of the video hardware, while the second one is based on the human interpretation of color.

For each studied model, it was established a data reduction approach that allows reducing the number of colors in the view, to a quantity that can be verbalized by the device. This implies devising a univocal function that, for each pixel color of the domain space there is, correspondingly, a value in the image space. The domain is composed by $2^m$ colors, whereas the image space is composed by $2^n$, with $n < m$. In the proposal for reduction, it is assumed that for each image space color $c_i$ there is, correspondingly, $2^{m-n}$ colors in the domain (see Fig. 2) that conform a subdomain $S_i$.

![Figure 2. “F” transformation. Venn Diagram representation](image)

The image obtained by the MTD is RGB 8:8:8, using 8 bits to represent each color. Therefore, the domain will have 224 possible colors ($m=24$). From this, four strategies of color number reduction are evaluated (see Table 1).

| Representation | $N$ | Number of colors: $2^n$ | Number of colors for each element of the image space: $2^{m-n}$ |
|----------------|-----|-------------------------|-------------------------------------------------|
| RGB 2:2:2      | 6   | 64                      | 262144                                          |
| HSI 2:2:2      | 6   | 64                      | 262144                                          |
| RGB 3:3:3      | 9   | 512                     | 32768                                           |
| HSI 5:2:2      | 9   | 512                     | 32768                                           |
Each strategy was evaluated according to the complexity to implement it and to the quality of reduction. The complexity is pondered according to the computational costs in working it by the MTD. In order to measure the image quality, for each value \( c_i \), a color palette (CP) was made, that contained the entire subdomain \( S_i \). Then, it was analyzed for each CP its capability to represent a given color. A palette of good quality is one that, for the observer, does not contain contrasting mixtures of colors.

After choosing the best strategy for data reduction, the color for each element \( c_i \) was determined. To this aim, the various \( S_i \) palettes were shown to different observers, and the color denominations were recorded for each palette. From this data, a definite color was assigned for each palette.

2.2. Application Development

An application development on a MTD necessarily implies to know the software programming language J2ME [8], which is widely used with this equipment type. Besides, the characteristics of the proposed development require a thorough study of the image acquisition hardware, data processing, the management of the audible information, and other aspects.

The requirements of the development are as follows:

- A simple algorithm
- A minimal time of execution
- A small memory for software execution due to MTD limitations
- A limited number of colors in order to have a simple verbalization
- Color characterization of the image as a function of the occurrence analysis of each color.

Basically, the system captures the image sent by the MTD camera and stores it in memory, to be later processed one pixel at a time. The analysis consists in taking the image’s center, with a pre-established size of window, where the information corresponding to the RGB color is extracted.

For each RGB pixel, the representation levels were reduced in order to simplify the number of colors and to assign them to ranges experimentally determined in the previous stage. The software computes the number of pixels belonging to each range and, by means of an audible output, it tells the user the colors of the view. To attain this, it is necessary to complete a number of steps of image processing. Figure 3 shows the block diagram of the application.

For the audible output, the maximum values attained for each color are processed. This way, when there is a dominant maximum, the system automatically verbalizes the color name. When there are two dominant colors, both names are verbalized and, finally, when there are no relevant maximum counts, a color range analysis is performed.

3. Results

The results allow evaluating the preliminary tests and the final development of the system. From the preliminary tests, the comparison of the four proposals for color representation is made, along with the interpretation of colors by the observers. In the phase of final experiences, tests are performed on the implementation of the system.

3.1. Analysis of the representation models

The representation models listed in Table 1 were evaluated using the complexity and quality analysis of the CP obtained from \( S_i \). The study results for two 6-bit strategies and two 9-bit strategies are presented.
3.1.1.  RGB 2:2:2

In Figure 4 some CP are shown (represented as a color square) obtained through Matlab®, where each one represents $2^{18}$ possible colors for each element of the image space of the transformation. To the left, it can be noted those palettes that present some difficulty to differentiate some colors, and to the right there is another group of well-defined CPs that can be easily discriminated one from another.

In this model, most images were obtained with very definite colors and easy to discern, a situation that led to a low complexity in determining the color names. Besides, the reduction is simple, thus allowing a shorter software execution time due to the small number of colors.

**Figure 4.** Color Pallet for RGB222 Image with and without problems of differentiation (Matlab)
3.1.2. HSI 2:2:2  
In this model, by analysing the CPs obtained through tests, it was noted a better differentiation of colors, but also a significant lack of color hues. Its complexity and software execution time were greater due to the need to make an intermediate transformation between the native color model and this one. To the left of Figure 5, it can be noted the CPs that present difficulty to differentiate some colors; on the right, there is another group with well-defined CPs, where each one represents 218 possible colors. In this reduction strategy, colors like yellow do not have any representation in the image set.

![Figure 5. Color Pallet for HSI222 Image with and without problems of differentiation (Matlab)](image)

3.1.3. RGB 3:3:3  
In Figure 6, it can be observed that most CPs for this representation are well defined, but operative problems arise in the manipulation of 512 colors in the transformation, therefore causing more complexity in assigning names and ranges, and a longer software execution time.

![Figure 6. Color Pallet for RGB333 Image](image)
3.1.4. HSI 5:2:2
This strategy was approached differently by interpreting the information that is stored for each model component. Since hue H contains all the information referred to the predominant color, S represents the saturation and I is referred to intensity, the function used for this model gives prevalence to the first parameter. This way, from the 9 bits used for the reduction, 5 bits are used for H, 2 bits for S and 2 bits for I. Even though the number of colors of the image space is equal to the number of RGB 3:3:3 strategy, the variation may be perceived in the representation of the color, because the modified parameters are different. To the left in Figure 7, there are CPs which show the difficulty for discerning colors, as opposed to the right-hand portion showing images with well-defined colors. In this model, a great percentage of non-defined images was obtained, which made cumbersome the task of defining names and ranges, in addition to a much longer software execution time on account of the large number of colors.

![With problems of differentiation](image1)

![Without problems of differentiation](image2)

Figure 7. Color Pallet for HSI522 Image with and without problems of differentiation.

3.1.5. Comparison among strategies
After performing the tests, the representation system chosen to carry out the application was the RGB 2:2:2 due to a number of better characteristics and advantages respecting the remaining strategies. A key feature that pointed to its selection is the fact that RGB is the native model of the hardware and of the J2ME language, in contrast with the HSI which needs a previous conversion step. This way, the RGB representation system allows to adequately discerning the colors as a function of in straight connection to the experience, while it also allows for an optimal processing within the system.

In addition, it was opted for reduction 2:2:2 due to its balanced characteristics which permit a proper identification of colors and a greater computational speed. This strategy makes it easier for the observers to group colors and defines the ranges.

3.2. Definition and interpretation of colors.
The determination of color names in order to allow for a proper verbalization was made by analysing each CP by 10 observers having normal vision. The perception of some of these observers was considered for defining the color naming, which resulted in 31 different names. The ten observers analyzed each CP and gave them a color name. Table 2 shows in part the result from such a task. For each CP, the chosen color is presented, along with the number of observers which opted for such a color name (name/ number of observers). When the CP has very different interpretations, it is referred as undefined. Only 1 CPs were classified as undefined.
### Table 2. Analysis for color determination (color name/number of observers)

| Color | 1st Dominant Color | 2nd Dominant Color | 3rd Dominant Color | 4th Dominant Color | 5th Dominant Color | 6th Dominant Color | Adopted Color |
|-------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|---------------|
| 0     | Black / 10        | -                 | -                 | -                 | -                 | -                 | Black         |
| 1     | Dark blue / 6     | Blue / 2          | Navy blue / 2     | -                 | -                 | -                 | Dark blue     |
| 2     | Blue / 6          | Navy blue / 3     | Dark blue / 1     | -                 | -                 | -                 | Blue          |
| 3     | French blue / 5   | Bright blue / 4   | Light blue / 1    | -                 | -                 | -                 | French Blue   |
| 4     | Army green / 5    | Green / 2         | Olive Green / 1   | Moss Green / 1    | Dark green / 1    | -                 | Military Green|
| 5     | Oil blue / 3      | Greenish blue / 2 | Green / 2         | Gray / 2          | Blue / 1          | Indefinite / 1    | Oil blue      |

For each of the 31 colors and its ranges, it were generated audio files with theirs names. This allows communicate color to a blind person using the audible output of the MTD, based on the processing of the dominant color.

### 3.3. System Test.

The experience tests were carried out into two stages. The first one was done by a normal-vision person, using as a tool two mobile devices (a Nokia 5800 and a N73), under different conditions of lighting (natural light, artificial light, etc.) to determine the optimal light conditions for work. The second stage was performed by three persons with visual impairment (blinds) who used a N73 mobile.

It can be noted in Table 3 for the first stage the coincidence percentage between the scene and the captured image, as well as that of the captured image and the audible output. Both values affect directly to the quality of the audible output respecting the scene. The main problem arises in the correspondence between scene and image. Particularly, it was noted that with recurrent image takes, when the MTD camera makes a self-adjustment (thus permitting a stronger or weaker input of external light), a significant change of color of the captured image was observed. It was also noted that when having objects or textures of glossy surface, a problem arises on lack of coincidence between the scene and the captured image. Reflex or gloss seen at various inclination angles caused gray tones in the captured image.

### Table 3. Results obtained with Color Recognition Application

| Cell Phone | Light               | Scene/Image | Image/Audible Output | % COINCIDENCES |
|------------|---------------------|-------------|----------------------|----------------|
| N73        | Natural             | 36.66       | 73.33                |                |
| Nokia 5800 | Natural             | 50          | 86.67                |                |
| N73        | Natural Indoors     | 66.67       | 70                   |                |
| Nokia 5800 | Natural Indoors     | 66.67       | 70                   |                |
| N73        | Artificial          | 66.67       | 63.33                |                |
| Nokia 5800 | Artificial          | 50          | 50                   |                |
The tasks for the blinds in the second stage were performed with the help of a person who guided them to capture the image and acquire the ability to use the MTD. The goal is to acquire capability to do the focusing and centring of the object by encouraging the autonomous usage of the application embedded in the MTD. The tests were carried out with lended MTDs, because there was no user owning high-end mobiles.

The results from these tests were highly satisfactory as regards the usefulness of the application to the MTDs. More practice is needed to increase the ability to focus and use the MTD, as well as to handle the regulation of the room’s light. The three test cases presented a high coincidence ratio between the captured image and the audible output, but there were some flaws between the scene and the captured image.

4. Conclusions
In this work, it was posible to design and implement an application on a MTD which allows determine and communicate audibly the predominant colors of an image taken by the cell-phone’s camera. To this purpose, it was studied the problems associated to the activity and the technological aiding system that fits more properly.

Priorly to the development, the color models were studied, as well as some approaches for reducing the number of colors. Representation RGB 2:2:2 resulted to be the most convenient one when analysing the complexity and quality of results. From the perception of ten observers, it was possible to associate each color with a set of colors easy to verbalize.

The application thus developed allows capturing, determining and emitting an audible rendering of color names of an image taken by a MTD. The view that is wished to be captured gets modified by a number of factors external to the processing, which affect teh image to be analysed, and to the perception of the image by the human eye. Such factors can be noted in artifacts produced by the camera and artifacts stemming from external sources, such as shades, reflexes of natural light, the objects’s own glow, the artificial lighting, and the like. These inconveniences are then translated into audible reproduction errors.

The implementation can be made on any high-end MTD, acondition set by the basic needs for dynamic memory and standardization. The application has been satisfactorily implemented on two mobile devices, Nokia models 5800 and N73.

In order to solve the inconvenience presented by the hardware, an analysis is under way to acquire the image in a controlled environment. For example, in the room of the user define distance to the object and fixed lighting conditions.

References
[1] OMS, ORGANIZACIÓN MUNDIAL DE LA SALUD: Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud (CIF) 2001.
[2] GRAFFIGNA, Juan Pablo: “Sustitución Parcial de la Visión mediante Realimentación Táctil”. 2009. Universidad Nacional de San Juan. ISBN: 987-05-7233-6.
[3] ZELEK, Jhon: The E. (Ben) & Mary Hochhausen Fund for Research in Adaptative Technology For Blind and Vissually Impaired Persons. 2002.
[4] GEORGE APPLING, Giulio Pappalardo: “The Rise of Mobile Applications Stores Gateways to the World of Apps”. Booz and Company Consulting Firm. 2010.
[5] COOK A. HUSSEY .S.: Assistive Technologies, Principles and Practice. 2da Ed.Mosby. 2002
[6] MOJSILOVIC, Aleksandra. A Computational Model for Color Naming and Describing Color Composition of Images, JEEE, 2005
[7] GONZALEZ R, Y WOODS Richard: “Digital Image Processing (3rd Edition - 2007). Prentice Hall; ISBN-10: 9780131687288.
[8] GÁLVEZ ROJAS Sergio, ORTEGA DÍAZ Lucas “Java a Tope: J2ME (JAVA 2 MICRO EDITION)”, Dpto. de Lenguajes y Ciencias de la Computación E.T.S. de Ingeniería Informática Universidad de Málaga. ISBN: 84-688-4704-6