ROBUST GRAYSCALE CONVERSION FOR VISION-SUBSTITUTION SYSTEMS

Codruta Orniana Ancuti, Cosmin Ancuti and Philippe Bekaert
Hasselt University - tUL -IBBT, Expertise Center for Digital Media, Belgium

Keywords: Color-to-Gray, Visual saliency, Auditory substitution systems.

Abstract: Substitution systems have proved an important potential in mobility assistance for visually disable persons. Particularly, proficient users of auditory-vision substitution are able to identify and reconstruct visual targets. The content of non-visual image is simplified with the purpose to minimize the cognitive process for recognition and also to reduce the duration of the sound patterns. Motivated by these facts, many of the existing substitution systems discard the color information by dealing with grayscale images. This paper presents a robust and effective method of color-to-gray transformation, that preserves the original color contrast of the initial images but also the original saliency. The study is focused taken into consideration the hypothesis that visual salient areas are tightly connected with visual attention. We show that an appropriate translation allows a more accurate rendering of the important image regions but that creates a better mental representation of the environment.

1 INTRODUCTION

Humans orientation and mobility ability is highly correlated with the capacity of mental mapping the spaces and the possible navigation paths in the environment. Much of this information is gathered through the visual channel. Visually disabled persons lack this crucial information and as a consequence face great difficulties to orient in novel environments since they are not capable of creating mental maps of spaces. Recent technological advances have improved the development of portable non-invasive substitution systems (Meijer, 1992; Capelle C., 1998; Pun et al., 2007) designed for visually disabled persons. These systems aim to provide assistance at the perceptual level, by compensating the deficiency in visual sense. The main task of these systems is to translate the acquired image and to made it available to other senses such as audio, haptic or smell. For haptic systems the *white can* provides the low-resolution information about the nearby surroundings, the feet estimate the characteristics of the navigation surface while the palms and fingers provide the high-resolution information allowing the fine recognition of objects textures and forms (Pun et al., 2007). Complementary, the auditory channel usually supplies information about events (e.g. person presence), scene distances and rough interpretation over the environment (Hill, 1993).

The work presented in this paper is focused to image-to-sound substitution systems in order to generate more appropriate map representation of the scene. The aim of such systems is to induce representations or mental images for visually disabled users (in general proficient users) due to imaginary process. Investigations in neural rehabilitation field are explaining these phenomena through cross-modal brain plasticity, where large areas in brain cortex (of the visually disabled persons) are recruited to process non-visual tasks (Auvray M., 2005; Capelle C., 1998; Cronly-Dillon and Persaud, 1999). Most of the standard systems employ a faster scheme transformation by only selecting the intensity channel information and modulating the amplitude of signal proportional with the pixel intensity value. However, this straightforward technique may fail to interpret perceptually properly the scene appearance due to the fact that the color information is not considered. Separately the color does not provide enough information about objects forms or scene geometry. Nevertheless, by modeling a gray translation with color information we can implicitly identify color cues (mostly corresponding

Ancuti C., Ancuti C. and Bekaert P. (2010).
ROBUST GRAYSCALE CONVERSION FOR VISION-SUBSTITUTION SYSTEMS.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 130-135
DOI: 10.5220/0002846701300135

Figure 1: *Salient regions detection.* For the images from the left side part are displayed the salient regions detected by the algorithm of Itti et al. (Itti and E., 1998) when both color and intensity are considered while for the images from the right side part are shown the detected salient regions when only intensity is taken as a discrimination feature.

to texture or particular classes such as sky, grass or flowers) beside those cues that are provided by intensity variations. Additionally, a good interpretation of the most salient regions overcomes deprived information about the most attractive areas in the scene and leads to focus the attention to the most interesting regions. By this approach the chances of object and persons identification into the scene are substantially increased (see figure 1).

We propose an effective color-to-gray method that is able to preserve the high contrast appearance of salient regions. We aim to develop a suitable decolorization method that enhances the contrast of the grayscale image to better visually reflect the chromatic contrast of the initial color image. Additionally, we have been concerned to reduce the loss of visual information from converted image. The utility of this approach has been tested based on the well-known sound substitution system vOICe (Meijer, 1992; Meijer, 1998). The experiments prove that our decolorized images better preserve the saliency of the original color scene compared with the standard grayscale and other specialized techniques.

2 RELATED WORK

While the history of vision-substitution systems stretches back to the 1970s (y Rita, 1967; Fish, 1976), more recent approaches have been introduced. The common stages that are performed by auditory-vision substitution systems are: image acquiring, image processing algorithms (e.g. grayscale conversion, details visibility enhancement, gamma correction, automatic labels/objects recognition, etc.) and finally image translation into sound frequencies.

Several well known approaches including vOICe (Meijer, 1992; Meijer, 1998) and PSVA (Prosthesis for Substitution of Vision by Audition) (Capelle C., 1998) deal only with grayscale images. The rendering operation implies scanning the image from left to right, and computing per-pixel the audio amplitudes/frequencies (by various schemes) that are finally rendered to the user. In vOICe (Meijer, 1998) approach the pitch elevation is given by the position in the visual pattern, and the loudness is proportional with the luminance intensity, therefore in this approach white is played loudly and black silently. The PSVA (Capelle C., 1998) is based on

a raw model of the primary visual system with two resolution levels, one that corresponds to artificial central retina and one that corresponds to simulated peripheral retina. The Vibe approach (Auvray M., 2005) splits the image into configurable distributed *receptive fields* that interprets the mean value of the gray levels in their allocated areas. The basic components of the sound are sinusoidal, being produced by virtual placed sources.

Image simplification is employed in Cronly et al. (Cronly-Dillon and Persaud, 1999) approach. This system reduces the image information by selectively permitting (of user choice) the separately extraction of horizontal/oblique lines. The main idea is that feature extraction can segment the image before translation and contributes to recognize patterns (squares, circles, polygons). After this step, image-to-sound rendering follows the scheme where pixels in a column define a chord and the horizontal lines are played sequentially as a melody.

Recently, several color-to-gray algorithms have been introduced in literature in order to overcome the problems of the standard graysacale conversion that employs only the luminance channel. Although the results are quite promising, the computational complexity of the most proposed techniques is still an important bottleneck. The general goal of the transformation is to generate an image that preserve the image appearance rather than simply record light intensities (like in standard approach). Color plays a significant role in the scene interpretation in terms of visual perception. In general the distribution of the color contrast is obtained by evaluating the color differences of the image pixels. Gooch et al. (Gooch et al., 2005) proposed a technique that attempts to preserve the sensitivity of the human visual system by comparing each color pixel value with the average of its neighbor region. The algorithm is highly computational expensive and performs poorly for high resolution images. Rasche et al. (Rasche et al., 2005) introduced a method that computes the distribution of all the image colors previously quantized in a number of landmark points. Due to the color quantization, some image details can be lost. Grundland and Dodgson (Grundland and Dodgson, 2007) introduced a faster technique, that as will presented in the following inspired our approach, to decolorize images based on the chrominance and luminance fusion. Neumann et al. (Neumann et al., 2007) computes the gradient field with two different formulas, one that takes advantage of the Coloroid (Nemcsis, 1987) color space and the second that presents a generalized technique based on CIELab.

3 SUBSTITUTION SYSTEMS OVERVIEW

We have chosen to built our approach on the well-known vOICe[1] system (Meijer, 1998). Due to the fact that pixels intensity values are reflected in the amplitude of the frequency (perceived as sound loudness), we have investigated modalities to exploit the bandwidth to the optimal way in order to enhance important visual cues of the scene. The difficulties in scene understanding appears when visual salient regions are not accurately represented and this resumes in misinterpretations of local and global content of the scene. A general overview of the system is presented

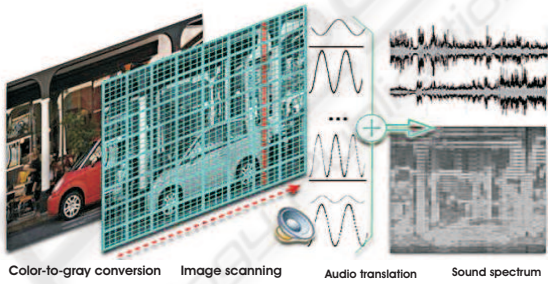

Figure 2: Overview of the *vOICe* auditory-substitution system.

in the following. The system translates the acquired frontal images into a time-multiplexed auditory representation. Each image is rendered with a resolution of 64×64 pixels in an approximate conversion time of $T = 1.05$ seconds. The translation operation is a per-pixel operation by encoding the vertical position into frequency and the horizontal position into time. The pixel intensity gives the oscillation amplitude, therefore white is mapped into *loudness* and black is mapped into *silence* of its associated oscillator.

Firstly the image matrix elements are associated with one of the G gray tones:

$$\mathcal{P}^k = (p_{ij}^k) \quad , p_{ij}^k \in \{g_i, ..., g_G\} \quad i, j = 1 \ldots N, N = 64 \tag{1}$$

where i and j represent the columns and lines indexes that are limited to the maximum values $N = 64$ (the input image has a resolution of 64×64 pixels).

Each of the N column that corresponds to the signal $s(t)$ is played in T/N seconds. As already presented, the amplitudes of sinusoidal components of the $s(t)$ signal are proportional with the intensity levels. Considering that $\omega_i = 2\pi f_i$ the sound pattern transformation is mathematical expressed as following:

[1] www.seeingwithsound.com

$$\begin{aligned} &s(t) = \sum_{i=1}^{N} p_{ij}^{k} \cdot sin(\omega_i t + \theta_i^k) \\ &t \in \left\{ t_k + (j-1) \cdot \frac{T}{N}, t_k + j \cdot \frac{T}{N} \right\} \\ &j = 1 \ldots N, k = 1, 2, \ldots \end{aligned} \quad (2)$$

The algorithm computes frequency distribution equidistant as expressed in equation 3. In addition to linear frequency distribution the approach allows also exponential distribution of frequency to render the patterns (see equation 4):

$$f_i = f_l + \frac{i-1}{N-1} \cdot (f_h - f_l), \quad i = 1 \ldots N \quad (3)$$

$$f_i = \left(\frac{f_h}{f_l} \right)^{\frac{i-1}{N-1}} \cdot f_l, \quad i = 1 \ldots N \quad (4)$$

where f_l (default $f_l = 500$Hz) and f_h (default $f_h =$ 5KHz) are the lowest and respectively the highest frequency.

Finally, after each image, as a distinct end-of-frame mark is inserted a synchronization click sound that indicates the end of the played image, respectively the beginning of a new input.

4 SALIENCY PRESERVING DECOLORIZATION

Decolorization or color-to-gray can be seen as an information compression operation since it maps three dimension information onto only one dimension. Standard transformation that employs only the luminance channel neglects the color information and as consequence in many cases visually important features are lost. This is due to the fact that different isoluminant colors are mapped on the same intensity level. Recently introduced decolorization methods aim for a better conservation of the original scene content after compression. Since the majority of existing methods are computationally expensive, in this work we have chosen to adapt Grundland's approach (Grundland and Dodgson, 2007) mainly due to the fact that this technique can preserve effectively the original image chromatic contrast but with low computational cost. However, our experiments disclosed several important limitations of these technique. Because the technique considers only a single dominant color axis it may fail to preserve a consistent appearance of images that are characterized by uniform hue distribution, since a single hue is highly advantaged (see figure 3).

Taken into consideration these aspects we develop a new technique addressing several issues: chromatic contrast adaptation based on hue distribution identification, image intensity manipulation and final constrains that provide a consistent output even when the parameters (e.g. chromatic contrast λ) are stretched on high values. We introduced several additional parameter constrains that generates more pleasant results but also a better control in comparison with the original technique.

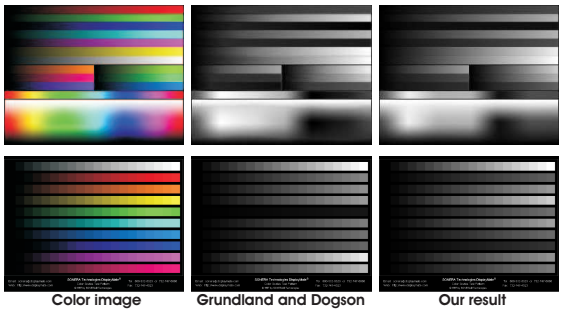

Figure 3: *Spectrum representation.* As can be seen the Decolorize (Grundland and Dodgson, 2007) approach fails to preserve a consistent appearance since a single hue is highly advantaged. Notice the appearance of the red flowers.

In the following are described the main steps of the algorithm. The presentation is focused mainly onto the modifications that have been added to fit properly this strategy into our system.

The transformation of the color image is performed in YPQ linear color space. The channel $Y \in [0,1]$ is the achromatic luminance channel and the pair channels $P \in [-1,1]$ and $Q \in [-1,1]$ represent the opponent-colors channels: yellow-blue and red-green.

$$\begin{bmatrix} Y \\ P \\ Q \end{bmatrix} = \begin{bmatrix} 0.2989 & 0.587 & 0.114 \\ 0.5 & 0.5 & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (5)$$

Beside luminance channel Y the computation of the chromatic channels (Hue-H and Saturation-S) is performed in a straightforward way as follows:

$$H = \frac{1}{\Pi} tan^{-1}\left(\frac{Q}{P}\right) \quad (6)$$

$$S = \sqrt{Q^2 + P^2} \quad (7)$$

The first step after image is converted into YPQ color space is to analyze the distribution of the image feature chromatic contrast. This is performed by computing the color difference between pairs randomly chosen and sampled by Gaussian pairing. The main idea of this approach is that nearby pixels may represent similar color since they might be part of the same feature, while more distant pixels have increased chances of having different colors. To identify the main principal chromatic contrast axis the method uses the predominant component analysis.

This represents a derivation of the well-known dimensionality reduction technique - principal component analysis (PCA) (Dunteman, 1989). The method optimizes the differences between observations by projection onto the two principal chromatic contrast axes. The purpose of these chromatic axes is to recover within a single direction the color contrast magnitude that is not contained by the luminance channel. This search maximizes the covariance between chromatic contrast and the weighted polarity of luminance contrast. However, this approach has the drawback that a single chromatic axis is not able to depict differently color contrasts that are perpendicular to it. Despite of this disadvantage, in general the image contrast is relatively pleasantly enhanced. The main advantage of this approach is the processing time that is linear with the image resolution. Gaussian pairing sampling technique reduces the time spent to compute color differences in comparison with similar techniques (Gooch et al., 2005; Rasche et al., 2005). Following the predominant component analysis step that decides the representative color contrast axis, the next step is to fuse the chromatic information with the luminance channel Y. Predominant chromatic channel contributes to the luminance with a λ degree of contrast enhancement (default value is 0.5). Our extensive experiments of approximately +200 images have shown that this parameter should be also correlated with the chromatic distribution (see figure 4). Hue histogram analysis controls the parameter η. The parameter is equal with 1 if the image hue distribution does not contain both red-green/yellow-blue opponent pairs and is equal with 3 if the hue distribution covers the entire range.

$$U_i = (\eta Y_i + \lambda C_i) / \eta \quad (8)$$

In order to maintain the luminance polarity the chromatic axis orientation needs to generate similar chromatic contrast with the luminance contrast.

An important desired feature in many cases is to control the contrast effects. The goal of image decolorization is to obtain a perceptual preservation of the original saliency (see figure 4). During our extensive tests we have noticed that for higher values of λ (λ=1) the saliency regions are better preserved when applying the well-known Itti algorithm (Itti and E., 1998) that identifies the most salient regions. We assume that the saliency is preserved only if the detected salient regions in the color image are maintained almost in the same regions as in the decolorized image.

In addition, we observed that when increasing chromatic contrast parameter λ, several undesired artifacts are introduced into the output image. Since increasing the chromatic contrast has a similar impact with changing the illumination color (blue areas become darken when opponent yellow areas becomes lighter), there is a requirement to limit the impact to a certain boundaries that may assure a decent visibility of details but also to maintain the original image appearance. In the biological system proposed in (Land, 1971) the signal of simulated neural path travels until it finds an inhibitory signal that is larger or equal with the sequential product. The principle is that the signal is blocked rather than transformed into negative value. Correlated to our approach this can be seen as an effective slicing technique of the intensity level. Therefore, in order to reduce the level of undesired artifacts we enforce the decolorization results to remain in the range of $[l * Min(R,G,B), l * Max(R,G,B)]$ ($l = 1$ is default value).

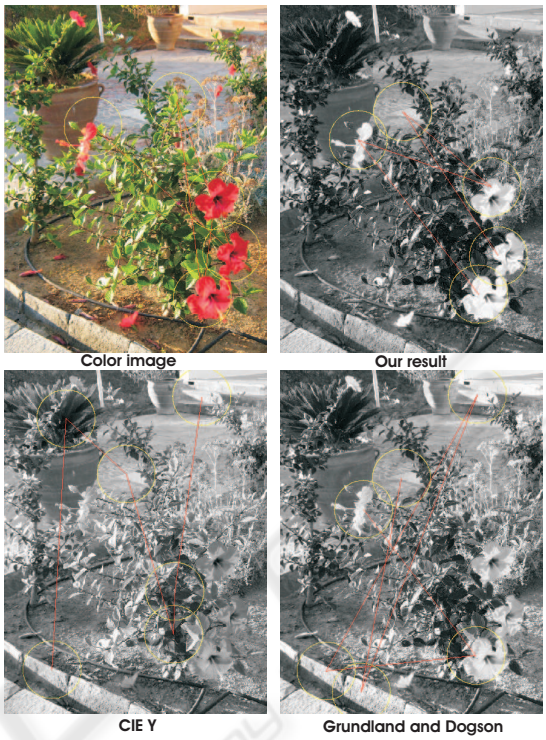

Figure 4: The Decolorize (Grundland and Dodgson, 2007) conversion fails to maintain the color salient regions while our result is capable to better preserve the original salient regions.

5 DISCUSSIONS AND CONCLUSIONS

This paper introduces a novel decolorization method in order to support auditory-vision substitution systems to translate efficiently grayscale images into sound patterns. A challenging problem in image grayscale conversion is how to interpret the scene

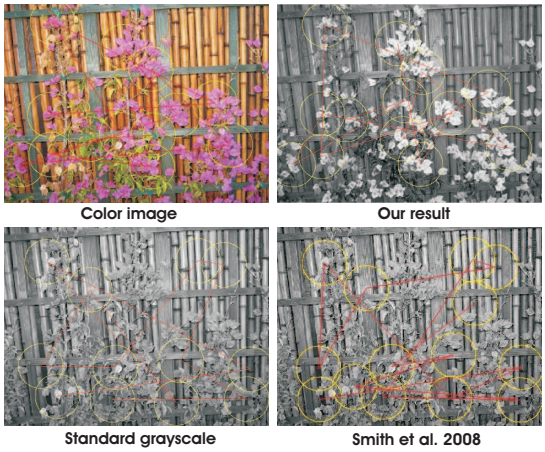

Figure 5: The contrast enhancement approache of (Smith et al., 2008) but also the standard grayscale risk to not preserve the original color salient regions.

cues elements that permit to represent more accurately the scene content. The selection of a good information reduction method is fundamental for the effectiveness of image understanding or attention focus guidance. In comparison with other approaches (Cronly-Dillon and Persaud, 1999) our model takes advantage of the color contrast. Regardless of scene complexity if the target object is not distinctively rendered the participants risk to inaccurately locate it. Comparing with existing approaches, our translation model is able to improve the user perception over the chromatic contrast image content. In low illuminated scenes many decolorization methods fail to convert accurately images while increasing the contrast. Our improved decolorization method has shown promising results against standard and recent algorithms. For images with isoluminant areas the system is able to translate with a higher recognition rate the visual cues. Even if for the moment the visual substitution systems are far from being comparable with the visual feedback, due to the limitation imposed by the input sensory, these systems can be designed suitable for basic specific tasks. For the moment all the available systems require costly training period in order to obtain reliable interpreted results. For future work we aim to perform extensive tests for more complex tasks such as object localization and mobility assistance.

REFERENCES

Auvray M., Hanneton S., L. C. (2005). There is something out there: distal attribution in sensory substitution, twenty years later. *Journal of Integrative Neuroscience*, 4:505–521.

Capelle C., Trullemans C., A. C. V. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Transactions on Biomedical Engineering*, (45):1279–1293.

Cronly-Dillon, J. and Persaud, G. R. (1999). The perception of visual images encoded in musical form: a study in cross-modality information. *Biological Sciences*, pages 2427–2433.

Dunteman, G. (1989). rincipal components analysis. *Sage, Thousand Oaks, CA*.

Fish, R. (1976). An audio display for the blind. *IEEE Transactions on Biomedical Engineering*, 23(2):144–154.

Gooch, A. A., Olsen, S. C., Tumblin, J., and Gooch, B. (2005). Color2gray: salience-preserving color removal. *ACM Trans. Graph.*, 24(3):634–639.

Grundland, M. and Dodgson, N. A. (2007). Decolorize: Fast, contrast enhancing, color to grayscale conversion. *Pattern Recogn.*, 40(11):2891–2896.

Hill, E., R. J. H. M. H. M. H. J. H. R. (1993). How persons with visual impairments explore novel spaces: Strategies of good and poor performers. *Journal of Visual Impairment and Blindness*, pages 295–301.

Itti, L., K. C. and E., N. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Patt. Anal. and Machine Intell. (PAMI)*.

Land, Edwin H., M. J. J. (1971). Journal of the optical society of america. 61(1).

Meijer, P. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2):112–121.

Meijer, P. (1998). Cross-modal sensory streams. *In Conference Abstracts and Applications, ACM SIGGRAPH*.

Nemcsis, A. (1987). Color space of the coloroid color system. In *In Color Research and Applications*.

Neumann, L., Cadik, M., and Nemcsics, A. (2007). An efficient perception-based adaptive color to gray transformation. In *In Computational Aesthetics*.

Pun, T., Roth, P., Bologna, G., Moustakas, K., and Tzovaras, D. (2007). Image and video processing for visually handicapped people. *Journal Image Video Process.*, 5:1–12.

Rasche, K., Geist, R., and Westall, J. (2005). Re-coloring images for gamuts of lower dimension. *Comput. Graph. Forum*, 24(3):423–432.

Smith, K., Landes, P.-E., Thollot, J., and Myszkowski, K. (2008). Apparent greyscale: A simple and fast conversion to perceptually accurate images and video. *In EUROGRAPHICS*.

y Rita, B. (1967). Sensory plasticity. applications to a vision substitution system. *Acta Neurological Scandinavica*, 43(4):417–426.