An Interactive Wearable Assistive Device for Individuals Who Are Blind for Color Perception

Troy L. McDaniel, Kanav Kahol, and Sethuraman Panchanathan

Center for Cognitive Ubiquitous Computing
Arizona State University, Tempe, Arizona, 85281
{troy.mcdaniel,kanav,panch}@asu.edu

Abstract. Color is inaccessible for individuals who are blind or visually impaired, as it is a purely visual feature. Given that many everyday tasks rely on color including coordinating clothing, social interactions, etc., the inaccessibility of color has an adverse effect on daily life. We propose an interactive, wearable assistive device that can recognize and convey colors of scenes or objects. As computer vision is challenging in real world environments due to, e.g., illumination or pose changes, computer vision algorithms can be augmented with sub-systems that can provide information on working environments of a recognition algorithm, and how it affects the recognition accuracy. In this paper, we introduce a framework that incorporates such measures, herein called *confidence measures*, in wearable assistive devices. By communicating to the user a quantitative measure that signifies the difference between optimal working conditions and the real environment working conditions, we can convey the reliability of system-made decisions, which enables the user to take action to improve confidence. Given that color recognition is challenging in real world settings, our system is built within our proposed framework for confidence measures. Finally, we present user recognition accuracies, both with and without confidence measures.

1 Introduction

Color information is inaccessible to individuals who are blind or visually impaired as it is a purely visible feature, unlike multimodal features such as shape or texture. Many everyday tasks rely on color such as coordinating clothing, sorting laundry, shopping, social interactions, etc., and hence, the inaccessibility of color has an adverse effect on daily life. Several commercial color recognizers are available including ColorTest Memo, Color Teller and Speechmaster Colour Detector [9]. However, these devices have several limitations: (1) they are only reasonably accurate; (2) they cannot detect the colors of scenes or distal objects; and (3) most cannot detect complex color profiles. In this paper, we propose an interactive, wearable assistive device that can recognize and convey colors of scenes and multicolored objects, both proximal and distal.

The challenge with most vision-based wearable assistive devices is obtaining satisfactory performance in real world conditions; illumination changes, pose changes,

C. Stephanidis (Ed.): Universal Access in HCI, Part I, HCII 2007, LNCS 4554, pp. 751–760, 2007.

scale changes, image blur and noise can be problematic for computer vision algorithms. Illumination conditions, motion blur and noise all adversely affect color recognizers. Some robustness to lighting variations can be obtained through the use of an illumination invariant color space, such as Normalized RGB (NRGB) [5]. Unfortunately, color spaces invariant to intensity changes, such as NRGB, cannot distinguish between colors that vary only with intensity such as white and black, or red and dark red. The color space selected for the work presented here is RGB, which is sensitive to lighting conditions, but can be used to recognize a wide gamut of colors. To ensure robustness against environmental variations (our focus here is lighting variations and motion blur), we propose a collaborative framework wherein the user and system work together to solve a task, that alone, the system can not.

The majority of wearable assistive devices, e.g., systems for assisting individuals who are blind in perceiving or navigating environments, do not seek assistance or feedback from users to help accomplish challenging tasks such as object recognition, face recognition, etc. Repetitive delivery of incorrect results to users causes a decline in system usability and reliability. A strategy to avoid leaving the user out of the "loop" is by communicating the confidence of a system-made decision to the user, which conveys the reliability of the decision, and allows the user to take action to improve confidence. Humans often associate a probabilistic measure with judgments and can state a confidence level involved in their judgment. For example, we might be 100% sure that the object on the table in front of us is a cup, but if lighting in the room is poor, we may not be as sure since lighting is not optimal. Pattern recognition algorithms such as Naïve Bayes, Bayesian networks, etc., are designed to provide a judgment on the recognized category and convey a measure of how reliable or accurate the recognized class is. Such systems are detrimental to wearable systems because humans are sensitive to wrong judgments.

In this paper, we propose a collaborative framework that can augment recognition algorithms by quantitatively analyzing the working conditions of the system, and comparing it to its optimal working conditions. A measure of the difference between the current working condition and the optimal working condition is herein referred to as a *Confidence Measure*. A framework for *confidence measures* exploits the human capacity by involving the user during problem solving to improve system usability and reliability. The novelty of the proposed approach lies in (1) development of an intuitive measure of the difference between real working conditions and optimal working conditions of an algorithm and (2) development of algorithms to derive confidence measures.

Two specific working conditions that can influence the recognition accuracy of computer vision systems include (1) illumination conditions and (2) motion blur. Consider the following example for clarity: a vision-based wearable assistive device, for an individual who is blind, reports that an object's color is red. Without the use of confidence measures, the system conveys the recognized class, and the user must trust that the system is correct. With confidence measures, the system conveys this same decision with a confidence rating of, e.g., 70%, and informs the user that a large amount of motion blur was detected in the image. Understanding the cause of low confidence, the user may act to increase the system's confidence (in this example, he or she may stand still while the system takes another image for analysis). Hence, this is a collaborative framework in which the user and system work together through a human-in-the-loop (HIL) strategy.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents the conceptual framework. Section 4 covers the experimental methodology and results. And finally, Section 5 presents possible directions for future work.

2 Background and Related Work

A number of handheld color recognizers are available commercially [9]. These devices recognize a color by analyzing the amount of reflected light when the sensor is held firmly against a surface. Hence, accuracy depends on illumination conditions, surface texture and density [9]. ColorTest Memo, from the ColorTest 2000 series developed by Caretec, can recognize over 1,000 color categories, from common colors (red, blue, etc.) to more complex colors (bright red, light yellow, etc.) [9]. ColorTest Memo can also convey a color's percentages of red, green and blue, and the amount of brightness and saturation. Moreover, the system is capable of conveying the colors of multicolored patterns. Similarly, Brytech's Color Teller offers color identification of common and complex colors, but lacks the additional analysis that ColorTest Memo offers and is much less accurate [9]. Finally, Cobolt System's Speechmaster Colour Detector can detect common colors and intensity variations (light, dark, etc.) with reasonably accuracy, but the operation of the device is challenging and requires a calibration step before each use [9].

In the way of vision-based color recognizers, Hub et al. developed a portable device for object identification [8]. An object's color can be estimated from a distance, but the device is limited to recognizing and conveying simple color profiles. Next we review the relevant literature on HIL strategies to improve system performance.

The human-in-the-loop strategy has been utilized mostly in content-based retrieval [2] and collaborative virtual environments [4,6]. Relevance feedback in content-based retrieval [2] improves query results by enabling the user to inform the system about the relevant retrieved items. Based on the user's feedback, the system adapts its search to find more relevant items. Collaborative virtual environments often lack realism due to problems such as a lack of realistic haptic feedback, network delay, etc. Rather than attempt to reduce or ignore network delay, Gutwin et al. [6] proposed to reveal it using visual ornaments called decorators to help users develop coping strategies. Experiments revealed that decorators for telepointers help users adapt to delay in the form of jitter and latency using a fading cursor effect and halo technique, respectively. These decorators are an example of computational aids that help individuals adapt to working conditions and work in symbiosis with systems.

Fraser et al. [4] investigated the limitations of virtual environments, and techniques to reveal these limitations to users to enhance communication and collaboration. Issues identified include limited field-of-view, lack of haptic feedback and network delays. A limited field-of-view causes confusion during interaction as it is not clear to users what is in another user's view. To alleviate this limitation, the authors suggested displaying the extent of a user's field-of-view using lighting. As haptic feedback devices are still in their infancy, the authors recommended conveying haptic feedback through other media such as audio or vision. And lastly, the authors suggested conveying network delay using visual indicators such as slider widgets.

In summary, confidence measures have been utilized in many fields such as content based image retrieval and online virtual environments. In this paper, we present a system to aid in recognition and analysis tasks. Lighting variations are often an impediment in visual recognition of stimuli. Another problem in wearable systems is motion blur, which can significantly affect recognition. We present a system for providing intuitive information on lighting conditions and motion blur to a user.

3 Conceptual Framework

An important element of wearable systems is to design human-in-the-loop methodologies wherein the system and user work collaboratively towards achieving certain tasks. We propose a framework for color recognition that utilizes *confidence measures* to enable the user and system to interact and achieve accurate color perception. In our system, confidence measures have been designed to assess poor lighting conditions and motion blur caused by excessive movement. The system generates a probabilistic measure of color of objects (distal or proximal), and evaluates if motion blur or poor lighting are encountered. The recognized color(s) and information on motion blur and lighting is conveyed to the user in an intuitive manner. This enables a collaborative interaction between the user and system to make judgments about color. In Section 3.1, we discuss the framework for color recognition, and in Section 3.2, we propose a framework for confidence measures.

3.1 Color Recognition and Segmentation

Given an image, each pixel is first classified independently of its neighboring pixels using Bayesian classification. A pixel is classified as the color category that maximizes the posterior probably conditioned on the pixel value:

$$P(C_i \mid x) = \frac{p(x \mid C_i)P(C_i)}{\sum_{j=1}^{n} p(x \mid C_j)P(C_j)} \tag{1}$$

where C_i is the i^{th} color category, x is the pixel value and n is the number of color categories. As shown in (1), the posterior probability is equal to the likelihood of C_i times the prior probability of C_i divided by a normalization factor, which can be ignored for the task of classification. The prior probability is the number of occurrences of a certain color category divided by the total number of pixels in the training set. The likelihood of C_i can be estimated using Maximum Likelihood Estimation (MLE). Assuming the densities are Gaussian, MLE is achieved by computing the mean and covariance matrix of each color category.

Given that vision-based wearable systems are (1) usually equipped with low-cost off-the-shelf video equipment, and (2) must operate in real world conditions with extreme environmental variations, point-based color classification often misclassifies pixels, resulting in noisy segmentation results. Instead, we can take into account a pixel's neighborhood to improve segmentation. In our framework, we use the methodology of [1], which uses the Iterated Conditional Modes (ICM) algorithm [3] to maximize a pixel's conditional probability based on its neighborhood. As in [1], we assume that the classes of neighbors are known, and each color category is treated as an independent process, modeled by the first order Gibbs-Markov random field:

$$P(C_i \mid N) = \frac{1}{Z} e^{-\lambda \frac{N_i}{N}} \tag{2}$$

where N is the neighborhood, N_i is the number of pixels in the neighborhood that fall into color category C_i, Z is the normalization factor, which can be ignored since it is constant across posterior probabilities, and λ is the clique potential, which determines the dependence of a pixel on its neighborhood. As $|\lambda|$ increases, a pixel's dependence on its neighborhood strengthens. In the ICM algorithm, (2) is applied to the image multiple times until a stopping criterion is met.

Once an image has been segmented into *n* or less color categories, colors are conveyed to the user as proportions. For example, if an object's color profile is half green and half red, the colors of the object would be conveyed to the user as approximately 50% red and 50% green. As motion blur, lighting variations and noise introduce segmentation errors, colors that make up a negligible portion of the image, determined by a certain threshold, e.g., 2%, are ignored.

The wearable system operates under two modes: distal and proximal. In the distal mode, the user may perceive all the colors present in a scene, and in the proximal mode, the user may hold an object in his or her hand, and move it in front of the wearable camera to assess its color profile. Before the colors of the proximal object are recognized, the foreground is segmented from the background, which can be accomplished through automatic cropping or more advanced techniques. Finally, once an image is captured and color segmented, color proportions are conveyed to the user.

3.2 Confidence Measure Framework

The proposed framework for confidence measures for wearable systems is shown in Figure 1. Given input from the environment, e.g., an image or video, the system analyzes the input and reports its decisions to the user on estimated motion blur and lighting conditions. Recognition algorithms are either deterministic or stochastic in nature. Deterministic algorithms assign a recognition class label. On the other hand, stochastic algorithms assign a probability value to a recognized class. Often, the probability value assigned to the recognized class can be employed to present users with some level of information on how well the algorithm performed. However, this measure is often not associated with the cause for performance degradation leaving users with the difficult task of guessing why the system is not performing with high reliability. Here, stochastic algorithms are grouped under algorithm-dependent factors, and may help provide an initial confidence measurement.

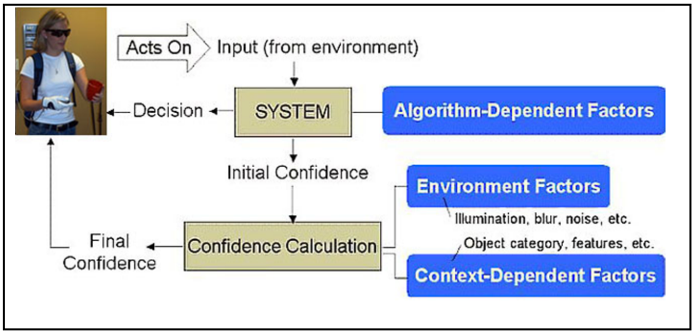

Fig. 1. Confidence measures framework

Environment factors such as illumination, motion blur, and noise can be taken into account for confidence calculation in our framework. Based on these factors, penalties or rewards are issued for the current estimate of the recognized class. In this paper, we propose algorithms for confidence measurement based on illumination and motion blur, which are presented in the next two subsections. At the final stage, context-dependent factors are considered, which may include object category or features, context, etc., depending on the application. As an example of context-dependent factors, consider an object that is classified as a bowl with 95% confidence. The system may further investigate other features of this object such as its material or size. If the material of the object is classified as cloth with high confidence, then the object is most likely not a bowl, and the original 95% confidence is penalized. On the other hand, if the material is classified as ceramic, then the object is likely a bowl, so confidence is increased. (Confidence cannot exceed 100%.)

Illumination Classification. Illumination classification attempts to match the illuminant of the scene to one of several illuminant models. One algorithm for illumination classification is an object-based approach developed by Hel-Or and Wandell [7]. For each illuminant, given the set of all surface reflectances of an object and the set of camera sensors, a cluster of points in 3D sensor output space is generated. When an image is captured, object segmentation is performed, and each set of pixels belonging to the object in question forms a cluster in 3D sensor space. Each image cluster is classified as the closest illuminant using Mahanalobis distance. The disadvantage to this approach is that object segmentation is required, which, in general, is still an unsolved problem. We propose an algorithm for real-time illumination classification that does not require scene segmentation. Our illuminant classifier categorizes the current illuminant into one of several coarse illuminant classes: poor (too dark), good (a little dark), great, good (a little bright) or poor (too bright).

To obtain an estimate of the illuminant of a scene, its luminance is estimated by computing the mean grayscale value of its image. This value is then classified using Bayesian classification to categorize the illuminant as one of the five proposed classes. The priors reflect how often each of the five illuminants occurs. Class-conditional densities are estimated using MLE. Assuming the probability density functions are Gaussian, the unknown parameters mean and variance are estimated by the sample mean and sample variance of the mean grayscale values of the training data, where sample mean and variance calculations are done separately for each class.

Motion Blur Classification. A variety of approaches exist for estimating motion blur in an image [5]. These algorithms estimate motion blur parameters, utilizing a degradation function H, to restore a blurred image. In this paper, we instead wish to classify the amount of motion blur using a measure that can be categorized as one of several motion blur levels. One approach is a no-reference, perceptual blur metric developed by Marziliano et al. [10]. This algorithm measures edge spread through edge detection followed by computing the average edge width, where edge width is defined as the local extrema locations closest to the edge. In the context of confidence measures, it is important to know why confidence has been penalized as this communicates to the user what must be done to increase confidence, and hence, improve reliability. Unfortunately, the approach of [10] is sensitive to any type of blur, and thus is not useful in this context.

We propose a real-time algorithm for classifying the amount of motion blur contained in an image, which requires no reference image and is sensitive to only motion blur. Our algorithm aims to classify the overall motion blur present in an image, whether this blur is caused by user and/or object movement. Further, our algorithm takes an indirect approach in that it does not directly classify motion blur, but overall movement, which is a good indicator of the amount of motion blur in an image. Our proposed algorithm for motion blur classification is described next.

The algorithm takes as input the current frame of the video stream as well as the previous frame. A difference image is computed by subtracting the current and previous frames. If a pixel value remains the same between the previous and current frame, it will have a value of zero in the difference image; otherwise, it will have a value greater than zero. A binary threshold is then performed on the difference image, and the image is scanned both horizontally and vertically to detect vertical and horizontal lines, respectively. The average width of vertical and horizontal lines is computed. A line is vertical if its height is more than its width, and a line is horizontal if its width is more than its height. The larger of these two averages is then classified as one of four motion blur classes: *no motion blur*, *small motion blur*, *large motion blur* or *extreme motion blur*. The range of average line widths is divided into four sub-ranges representing the four classes. These sub-ranges are determined though experimentation such that motion classifications predict the corresponding motion blur levels.

4 Experimental Methodology

We built a vision-based wearable assistive device for color perception based on our proposed framework. The system consists of a pair of sunglasses with an embedded video camera, headphones for audio output, USB number pad for input and a laptop. The proposed framework for color recognition and segmentation was implemented. Training data consisted of manually segmented color images taken from the COREL color image database. Our color categories, and respective pixel counts for training, included *white* (1600), *gray* (300), *black* (1072), *red* (1100), *light red* (400), *dark red*

(400), *green* (600), *light green* (100), *dark green* (300), *blue* (1100), *light blue* (400), *dark blue* (400), *orange* (500), *purple* (600), *yellow* (500) and *brown* (600). Both priors and class-conditional probabilities were estimated and used in (1). We assumed Gaussian densities, and used MLE to estimate the means and covariance matrices.

Through experimentation, we estimated parameters for (2). We found a neighborhood size of 3 and 3, and a clique potential of -5 and -0.1, to work well for proximal and distal mode, respectively. Our stopping criterion was when the number of classification updates is below a threshold, which is recommended by [1]; we found a threshold of 1000 to work well. As a preprocessing step, noise is reduced using median filters before point-based classification. See Figure 2 for segmentation results.

Distal mode uses the entire 320x240 captured image, whereas in proximal mode, to segment the foreground from the background, we crop the image from each side by 50 pixels. Hence, to use proximal mode successfully, the user must hold small objects close and large objects farther from the camera. This segmentation technique, although simple, was effective for our purposes.

Algorithms for illumination and motion blur classification were implemented and integrated into our system. Motion blur categories *no motion* and *small motion* did not penalize confidence, but *large motion* and *extreme motion* each generated a penalty of 30%. Both *poor* illumination categories generated a penalty of 20%, and both *good* illumination categories generated a penalty of 10%. Illumination category *great* did not penalize confidence. In our system, the initial confidence begins at 100%. The confidence measure based on illumination was trained using 1200 images of indoor and outdoor environments, captured from the wearable system. Each image was manually labeled as one of the five illuminant classes. Ten-fold cross-validation was used to evaluate the algorithm, which provided a classification accuracy of 96%. To train the confidence measure based on motion blur, we manually adjusted the decision boundaries until the four motion levels corresponded with the four motion blur classes. To test the algorithm, we collected 400 image pairs from video recorded from the wearable system. These image pairs were manually labeled, and then classified by our algorithm with an accuracy of 95%.

Two experiments were designed to test the validity of the system. Both were performed in an office setting, and involved three participants who are totally blind and one participant who is visually impaired but has some color perception (this participant was blind-folded during the experiment). The first experiment tested system usability (in the proximal mode) without confidence measures. During the training phase, participants learned how to use the system and were asked to perceive the color profiles of five, randomly presented objects. Before the experiment, participants were informed that motion blur and lighting can affect system accuracy. Participants repeated the training phase until all color profiles were estimated correctly. During the testing phase, ten novel objects are randomly presented, and participants were asked to estimate each color profile. Moreover, for five of these objects, which were randomly selected, the lights were turned off to simulate an environment with poor lighting. The test objects are the same for all participants; the set consists of a red bowl; black cup; yellow and white cup; green pail; white bowl; red and green block; red, blue and yellow block; blue and green block; block wrapped in wallpaper with a floral pattern; and a block covered in brown sandpaper. (The blocks were made from LEGO® blocks.)

During the training phase for experiment two, participants learned about confidence measures and how to use them. For example, if there is motion blur in the image, the participant may try again, but this time, reduce movement. Or, if lighting is not optimal, then the participant may, e.g., change their orientation or the orientation of the object with respect to the lights, or instruct the experimenters to adjust the lighting in the room (e.g., turn the lights up), and then try again. Participants first went through a training phase similar to before but with confidence measures, and then tested on the set of ten objects described before while using confidence measures. Similarly, objects were perceived in the proximal mode, and lights were turned off for five randomly selected test objects.

Table 1. User/System accuracies for Experiment 1 and 2. Accuracies are documented as the number of objects correct out of the total number of objects.

	Experiment 1	Experiment 2
Participant 1	2/10	7/10
Participant 2	3/10	7/10
Participant 3	3/10	6/10
Participant 4	4/10	7/10

User/System accuracies are shown in Table 1. An object is classified correctly if all its colors are identified and proportions are within 20% of the actual proportions. Colors with proportions of 10% or less were considered noise and ignored. These results show that our system can allow accurate perception of complex color profiles with the use of confidence measures. Further, these results show a major improvement in user/system accuracy when confidence measures are utilized as they allow the user to interact with the system to achieve better results.

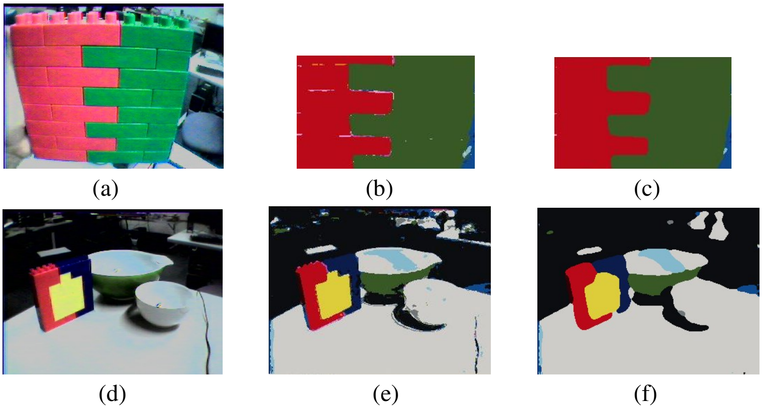
(a) (b) (c)
(d) (e) (f)

Fig. 2. Segmentation results for proximal (a-c) and distal (d-f) mode. (a) and (d) are the original images, (b) and (e) are point-based classified images (cropped if proximal mode) and (c) and (f) are the final, segmented images.

5 Conclusion and Future Work

An interactive, wearable assistive device for color perception was proposed. Further, a framework to integrate confidence measures with wearable assistive devices was presented, along with algorithms for classifying illumination and motion blur, which can be used for confidence measurement. Results demonstrate that confidence measures can help make systems more reliable and usable by involving both the user and system in problem solving. Future work will involve continued testing with more participants, and the use of additional environmental and context-dependent factors.

References

[1] Abdel-Hakim, A.E., Farag, A.A.: Color segmentation using an Eigen color representation. In: Proc. of 8th International Conference on Information Fusion, pp. 25–29 (2005)

[2] Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley, Reading, MA, USA (1999)

[3] Besag, J.: On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society 48(3), 259–302 (1986)

[4] Fraser, M., Glover, T., Vaghi, I., Benford, S., Greenhalgh, C., Hindmarsh, J., Heath, C.: Revealing the Realities of Collaborative Virtual Reality. In: Proc. of Collaborative Virtual Environments, pp. 29–37 (2000)

[5] Gonzales, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley, Reading, MA, USA (1992)

[6] Gutwin, C., Benford, S., Dyck, J., Fraser, M., Vaghi, I., Greenhalgh, C.: Revealing Delay in Collaborative Environments, In: Proc. of Human Factors in Computing Systems, pp. 503–510 (2004)

[7] Hel-Or, H.Z., Wandell, B.A.: Illumination classification based on image content, In: Proc. of Annual Symposium of the Optical Society of America (OSA) (1998)

[8] Hub, A., Diepstraten, J., Ertl, T.: Design and development of an indoor navigation and object identification system for the blind, In: Proc. of ACM SIGACCESS Accessibility and Computing, pp. 147–152 (2004)

[9] Kendrick, D.: What color is your pair of shoes? A review of two color identifies, AFB AccessWorld, 5(3), May 2004, (February 14, 2007) http://www.afb.org/afbpress/pub.asp?DocID=aw050308

[10] Marziliano, P., Dufaux, F., Winkler, S., Ebrahimi, T.: Perceptual blur and ringing metrics: application to JPEG2000. Signal Processing: Image Communication 19(2), 163–172 (2004)