Recognizing Clothing Colors and Visual Textures Using a Finger-Mounted Camera: An Initial Investigation

Alexander J. Medeiros[1], Lee Stearns[1], Leah Findlater[2,5], Chuan Chen[3], Jon E. Froehlich[1,4]

[1]Department of Computer Science
[2]College of Information Studies
University of Maryland, College Park

[3]Poolesville High School
Poolesville, MD

[4]Computer Science and Engineering
[5]Human Centered Design and Engineering
University of Washington

ajmed@umd.edu, lstearns@umd.edu, leahkf@umd.edu, riverchen@gmail.com, jonf@cs.umd.edu

ABSTRACT
We investigate clothing color and visual texture recognition using images from a finger-mounted camera to support people with visual impairments. Our approach mitigates issues with distance and lighting that can impact the accuracy of existing color and texture recognizers and allows for easy touch-based interrogation to better understand clothing appearance. We classify image textures by combining two off-the-shelf techniques commonly used for object recognition achieving 99.4% accuracy on a dataset of 520 clothing images across 9 texture categories. We close with a discussion of potential applications, user evaluation plans, and open questions.

CCS Concepts
• **Human-centered computing → Accessibility technologies**

Keywords
Blind; visually impaired; wearables; texture recognition

1. INTRODUCTION
For people with visual impairments, color and texture information are frequently inaccessible. Standalone commercial devices such as *Color Teller* [2] or smartphone apps such as *Color Identifier* [5] allow users to identify colors and hear them read aloud. However, accuracy varies depending on ambient lighting and distance from the target surface. Also, these systems do not support recognition of visual textures or allow users to quickly query multiple locations—both of which are important to recognizing clothing [3]. Outside of commercial products, Yuan *et al.* [11, 13] investigated simple color and texture classification algorithms to identify clothing from static images. However, their approaches are again affected by distance and ambient lighting and do not allow for efficient interaction to identify more complex patterns.

In contrast, we introduce a wearable system using a small camera and co-located illumination source (an LED) worn on the finger. This system should enable access to color and texture information through touch, allowing users to move their finger across an article of clothing and combine knowledge of physical materials gained from their own sense of touch with automated audio feedback about visual appearance. This approach builds on a small but growing body of work, including our own [6, 10], that uses finger-mounted wearable cameras to support visually impaired users in a variety of activities. Our approach is most closely related to *Magic Finger* [12], which was not intended specifically for visually impaired users but which similarly uses a finger-mounted camera to classify touched surfaces including seven clothing textures. However, Magic Finger's classification approach was simplistic and would not scale well to a large database of textures. In contrast, our algorithmic approach is based on that by Cimpoi *et al.* [4], which combines two complementary features commonly used for object recognition to achieve state-of-the-art texture classification performance. To examine this approach with a finger-wearable camera, we conducted a classification experiment on a small custom dataset of 520 images across 9 visual texture categories and 29 articles of clothing, achieving 99.4% accuracy.

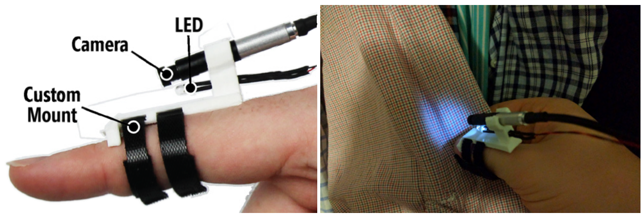

Figure 1. Prototype wearable camera and LED system: (a) Close-up view of system, (b) Identifying an article of clothing.

Our contributions include: (i) extending our finger mounted camera system [6, 10] to support robust color and visual texture recognition accessed through touch, (ii) a small but novel dataset of close-up clothing images captured using this system, and (iii) results of applying a state-of-the-art texture classification approach to our dataset. Because this work is at an early stage, we also cover open questions and challenges that will need to be addressed as we implement an interactive system for visually impaired users.

2. CAMERA AND ALGORITHMS
Our prototype system uses a small, endoscopic camera (*Awaiba NanEye GS Idule*) mounted on top of the finger to capture images of the touched surface (Figure 1). The camera provides 640×640px images with a 30º field of view and can focus from 15mm.

To identify textures, we use the classification approach described in [4], which combines two complementary features to improve classification performance. First, we detect deep convolutional activation features (DeCAF) using a pre-trained deep convolutional network. As in [4], we repurpose the *AlexNet* [7] image classifier for identifying textures by removing the last two softmax and fully connected layers. Second, we use scale-invariant feature transform (SIFT [8]) descriptors extracted densely at multiple scales. The SIFT features are combined into a single descriptor using the Improved Fisher Vector (IFV [9]) formulation. The result is two feature vectors of length 4,096 and 40,960 for DeCAF and IFV respectively, which are used as inputs (separately or concatenated together) to a support vector machine for classification.

We focus primarily on visual *texture* classification since few researchers have attempted to make this information accessible for visually impaired users. However, our *color* recognition approach

ASSETS '17, October 29-November 1, 2017, Baltimore, MD, USA

ACM ISBN 978-1-4503-4926-0/17/10.
https://doi.org/10.1145/3132525.3134805

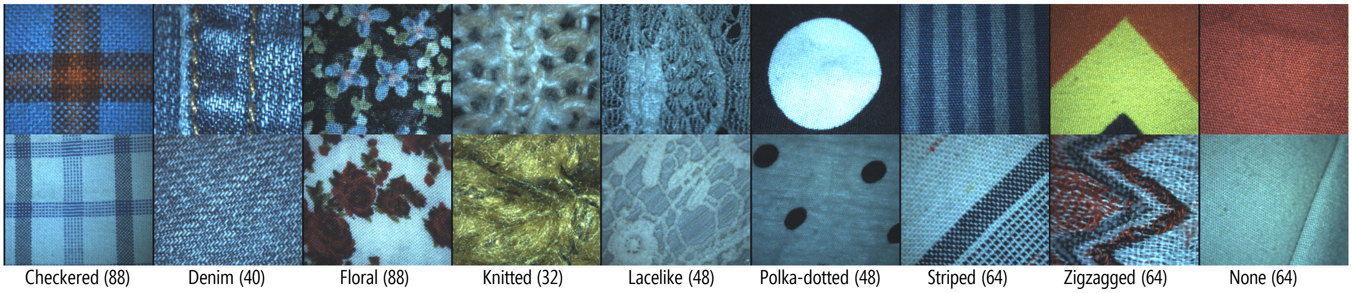

Figure 2. Examples of the 9 clothing textures included in our dataset. The numbers in parentheses indicate the quantity captured for each class.

also has some advantages over existing solutions. Our wearable system uses touch-based interactions to constrain the camera's distance from the target surface and includes a bright LED to overpower the effects of ambient light (allowing for more consistent color recognition). Furthermore, we use a color detection approach based on superpixel segmentation [1] that allows multiple colors to be detected simultaneously (one per superpixel) helping users to better understand how an article of clothing appears.

3. DATASET AND EXPERIMENTS

The results reported in [4] were promising but did not focus on clothing textures and used images from online sources that differed greatly from our target domain. It was unclear how well the approach would extend to close-up images captured by a finger-mounted camera. To answer this question, we collected a dataset of 520 images across 29 articles of clothing, which spanned 9 clothing texture categories (Figure 2). These categories are a subset of the 47 included in [4]; we eliminated categories that rarely describe clothing (*e.g.*, "bubbly"), combined those that are visually similar (*e.g.*, "striped", "banded), and added two new categories: "denim" and "none". We controlled for and varied the distance (5cm *vs.* 12cm), rotation (0° *vs.* 45°), and perspective of the camera (90° *vs.* 45°), as well as the tension of the fabric (taut *vs.* hanging naturally). The dataset was collected by one person over three months.

To assess performance, we conducted a classification experiment computing accuracy as the number of test samples classified correctly. We also explored the effect of training set size to determine whether a small user-gathered training set would be sufficient. Figure 3 shows classification accuracy using DeCAF and IFV features separately and together as the training set size increases averaged across 40 random samples.

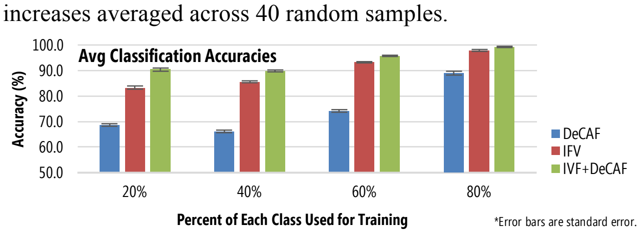

Figure 3. Accuracies using individual and combined features.

As with [4], in each case the combined result is best, demonstrating the advantage of using complementary texture features. However, unlike [4], our DeCAF results are significantly lower than IFV. This is likely because DeCAF requires a large amount of training data to perform well while IFV does well even with the small amount that we provided.

4. DISCUSSION AND OPEN QUESTIONS

Our results demonstrate the feasibility of recognizing clothing textures using close-up images from a finger-mounted camera. Even with a small amount of training data across a variety of variables, we achieve high classification accuracy. However, our dataset was captured under controlled conditions, so future work will need to investigate performance with a much larger, more varied dataset and conduct a user study with a real-time system.

We envision a few different scenarios in which visually impaired users could benefit from our proposed system. First, users could train the system to recognize customized articles of clothing in their own closet gathering training examples like those included in our dataset. Our results suggest reliable performance even with a small number of training examples. Second, we could draw upon larger texture datasets such as [4] to train a general-purpose classifier, which would allow users to access information about unfamiliar articles of clothing (*e.g.*, while shopping). Our approach should allow users to quickly explore a surface and combine their sense of touch with visual texture and color information to make informed decisions about what to wear or buy. How best to convey this information to visually impaired users is an open question.

5. ACKNOWLEDGMENTS

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

6. REFERENCES

[1] Achanta, R., *et al.*. 2012. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *TPAMI*. 34, 11: 2274–2282.

[2] Brytech Color Teller: *http://www.brytech.com/colorteller/*.

[3] Burton, M.A., *et al.*. 2012. Crowdsourcing Subjective Fashion Advice Using VizWiz: Challenges and Opportunities. *Proc. ASSETS 2012*, 135–142.

[4] Cimpoi, M., *et al.*. 2014. Describing Textures in the Wild. *Proc. CVPR 2014*, 3606–3613.

[5] GreenGar Studios. Color Identifier: *https://itunes.apple.com/us/app/color-identifier/id363346987?mt=8*.

[6] Findlater, L., *et al.*. 2015. Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch. *Proc. ASSETS* 2015, 383–384.

[7] Krizhevsky, A., *et al.*. 2012. ImageNet classification with deep convolutional neural networks. *Proc. NIPS 2012,* 1097–1105.

[8] Lowe, D.G. 2004. Distinctive image features from scale-invariant keypoints. *IJCV*. 60, 2 (2004), 91–110.

[9] Perronnin, F., Sánchez, J. and Mensink, T. 2010. Improving the Fisher Kernel for Large-Scale Image Classification. Springer, Berlin, Heidelberg. 143–156.

[10] Stearns, L., *et al.* 2016. Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. *TACCESS*. 9, 1: 1–38.

[11] Tian, Y. and Yuan, S. 2010. Clothes Matching for Blind and Color Blind People. *Computers Helping People with Special Needs SE - 48*. Springer Berlin Heidelberg. 324–331.

[12] Yang, X.-D., *et al.*. 2012. Magic finger: always-available input through finger instrumentation. *Proc. UIST 2012*, 147–156.

[13] Yuan, S., *et al.*. 2011. Clothing matching for visually impaired persons. *Technology and Disability*. 23, 2 (Jan. 2011), 75–85.