CS231M Mobile Computer Vision Project Proposal

Low Resolution Scene Representation for Retinal Prostheses

Pallabi Ghosh Ayesha Khwaja John Doherty

Introduction

There is currently research being done to develop a retinal prosthetic to restore some amount of vision to blind patients. These patients formerly had vision and lost it due to diseases that destroy photoreceptors like macular degeneration. This blindness can be debilitating, making simple daily tasks significantly more difficult. Additionally, many of these patients are older, so losing vision can increase the risk of devastating falls and can make moving around the house dangerous. Many of these patients would be grateful to restore even a small amount of sight to assist in basic ambulation tasks.

The retinal prostheses work is groundbreaking, but still in its early stages. One promising technique is being developed at Stanford in the labs of Daniel Palanker and EJ Chichilnisky [2]. These labs are developing a retinal implant with special goggles to partially restore sight. The goggles have a camera and project a near-infrared version of the camera image onto the retina. The retinal implant senses the near-infrared light and generates a signal that can be interpreted by the cells in the retina. This process effectively transmits some amount of visual information to the patient's brain.

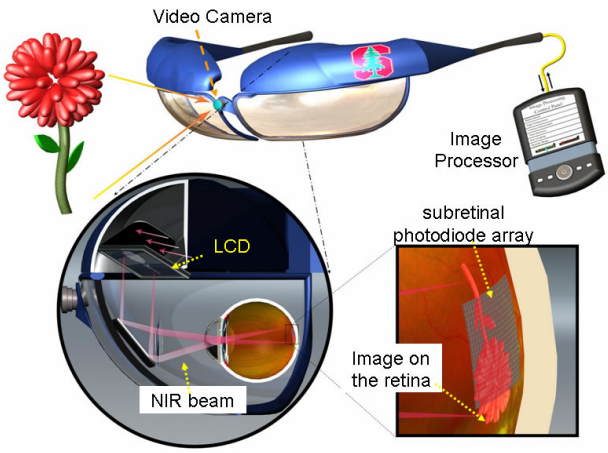

Figure 1: System Overview [2]

As state-of-the-art as this system is, the implant is still only able to transmit a very low resolution image to the brain. The team's most recent estimate was that the patient would be able to see about 64x64 pixels with 5 shades of gray. Additionally, since the implant only covers part of the retina the user would only get back about 15° field of view. At this resolution, simply outputting a down sampled version of the camera image will result in an image with no real information. Thus, we would like to generate a more information-rich representation of the scene that can be output to the implant.

There are a few areas that we want to focus our efforts to improve the scene representation. The most important tasks are indoor and outdoor ambulation. This involves generating a representation that will help patients navigate indoor and outdoor environments. Additionally, we would like to provide text detection and sharpening to allow patients to read signs in a scene.

While all of these tasks are doable with modern computer vision techniques, we have the key constraint that our system must run in real time. We plan on implementing algorithms on the Android tablets to simulate what the patient will see and improve on the baseline scene representation.

Technical details

Proposed Solution

We plan to focus primarily on scene navigation. We will use recent research on scene segmentation, both geometric and semantic, to understand the scene. There are numerous research work in this area, but one major constraint to be kept in mind is the real time execution speed necessary for our purpose. Usually this kind of image segmentation needs a couple of seconds per frame, which is not very useful for us. So we have looked into papers that have a rate of 1 frame per second [3][5]. Also we do not need multiple class recognition as the retinal prosthesis give a very low resolution view with maximum 5-6 gray levels. So we would try to detect major segments such as road or ground class, buildings, trees, cars, etc that will all be in the vertical class and sky. Also the device camera captures a high resolution image of the scene. We would try to intelligently label the scene into groups so that the blind person can have a maximum understanding of the scene with very low resolution display. We will use low level image processing techniques like edge detection, thresholding for this task.

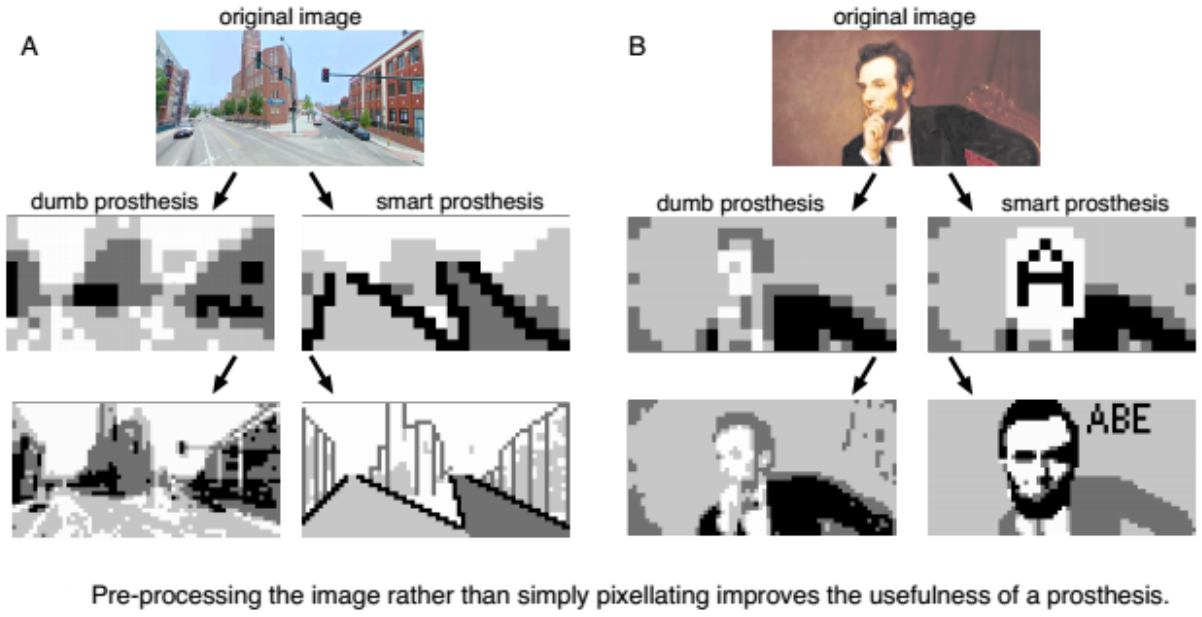

Figure 2: Naive prosthesis vs a smart one (First row: Original image, Second row: Very low resolution display of these images using naive and smart prosthesis, Third row: Slightly higher resolution doing the same tasks)[1]

In the figure above we can see the difference that a smart prosthesis can make to the information content of the image as compared to the naive prosthesis. Our goal is to develop such a smart prosthesis with emphasis mostly on navigation as studies have shown that navigation is the most important task that blind people want to be able to do. Facial recognition and other such tasks are far less important to them. So for navigation, not only segmentation, but road sign recognition is also important. We would try to develop a robust road sign recognition system. Not only recognition but proper low resolution display or emphasis of these road signs during navigation are also an important part of our goal.

Target mobile platform

Our target mobile platform is Android NVIDIA Tegra 4 supplied in class for this project. Later we plan to use Google Glass.

Milestones (dates and sub-goals)

- Things done till now:
    - Literature Review: Review all the quick image segmentation segmentation algorithms both for indoor and outdoor scenes ([3],[4],[5])
- 19th May - 23rd May:
    - Implement algorithm used in some of the papers to come up with an almost real time output with labeled segments on either one of scene categories: indoor or outdoor (on Matlab)
- 24th May - 31st May
    - Develop an app to represent the segmented output as a low resolution (64x64) image in 5 shades of gray
    - Use the results to simulate what the blind would see (on an app developed by the Palanker's Lab)
- 1st June - 4th June
    - Generate a dataset of all the road signs and detect these on the test images
    - Represent and highlight these road signs, again in some low resolution format, on the original low resolution output images
- 4th June - 8th June
    - Test the method for various image or video sequences

References

[1] E.J. Chichilnisky, "Glass as a Prototype and Inspiration for a Smart Retinal Prosthesis"
[2] Daniel Palanker, "Photovoltaic Retinal Prosthesis for Restoring Sight to the Blind" http://www.stanford.edu/~palanker/lab/retinalpros.html
[3] Joseph Tighe,Svetlana Lazebnik, "SuperParsing:Scalable Nonparametric Image Parsing with Superpixels," European Conference on Computer Vision, 2010.

[4] Varsha Hedau, Derek Hoiem, David Forsyth, “Recovering Free Space of Indoor Scenes from a Single Image”, Conference on Computer Vision and Pattern Recognition, 2012.
[5] Cadena and Košecka, “Semantic Parsing for Priming Object Detection in RGB-D Scenes,” 2013.