

FPGA-Based Video Processing for a Vision Prosthesis

Benjamin Kwek¹, Freddie Sunarso¹, Melissa Teoh¹, Arrian van Zal¹, Philip Preston², Oliver Diessel¹

¹*School of Computer Science and Engineering, University of New South Wales, Australia*

²*Australian Vision Prosthesis Group, University of New South Wales, Australia*

¹ {zwbk038, freddies, mjst359, avza576, odiessel}@cse.unsw.edu.au, ² p.preston@unsw.edu.au

(Demonstration Paper)

Abstract—Restoring sight to the visually impaired poses significant challenges across multiple disciplines. In this demonstration, we present a prototype vision processing system to perform the external processing and to provide a technical user interface for a vision prosthesis. The system transforms an input video stream into intensity values suitable for transmission to a stimulation array implanted within the retina of an impaired eye. This FPGA prototype ensures the processing is fast, accurate, and has low power consumption. We demonstrate our progress to date and hope thereby to stimulate interest in this emerging and challenging application area.

I. INTRODUCTION

The Australian Vision Prosthesis Group at the University of New South Wales, as part of Bionic Vision Australia, is developing an artificial vision prosthesis to restore vision to patients suffering degenerative eye diseases such as macular degeneration and retinitis pigmentosa.

A. Vision prosthesis

The physical construction of their vision prosthesis consists of two parts – the external vision processing electronics and the internal implanted electrodes [1]. The external electronics consist of a camera, an image processing unit, an RF transceiver and the mechanical structures that support and connect these parts (see Fig. 1). Associated is a suite of software that provides image acquisition and processing with

transmission of processed information to the implant.

The internal electronics (see Fig. 2) consist of two interconnected modules. Unit 1 is a subcutaneous encapsulation containing an antenna, power supply and communication electronics which is to be implanted below the skin behind the ear. A pair of wires connects to Unit 2, a small module mounted externally to the sclera of the eye. A flexible electrode array passes from Unit 2 through the sclera to a position suprachoroidal within the retina. The implant simply follows commands from the external electronics, without further processing.

The purpose of the image processing unit is to maximize information transfer from the camera to the visual cortex in the brain. This involves transforming video data into currents emanating from 98 platinum electrodes that are stimulating retinal tissue. The electrodes in the stimulating array are arranged on a hexagonal grid (see Fig. 3). A nominal centre electrode is surrounded by 6 guard electrodes. This centre electrode is stimulated with charge balanced biphasic currents while the guard electrodes form a return path for current to focus stimulation. Each guard electrode plays the role of a

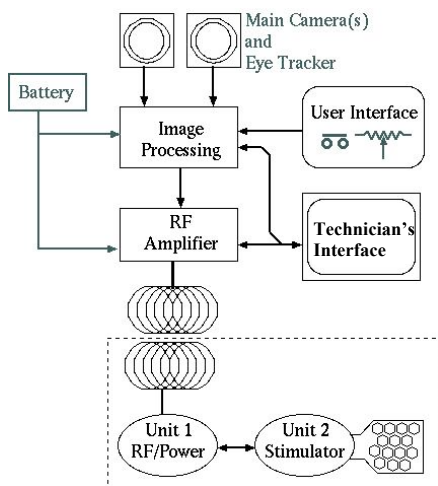


Fig. 1. Vision prosthesis

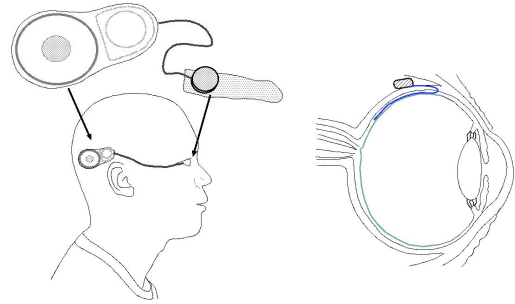


Fig. 2. Implanted electronics

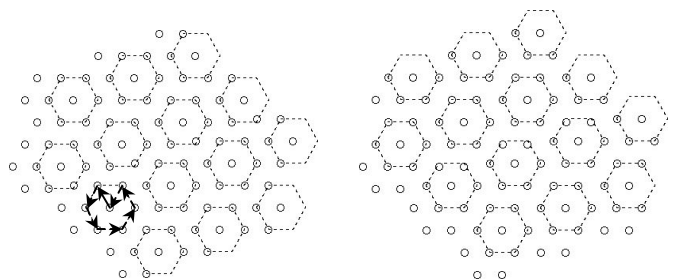


Fig. 3. Two successive stimulation patterns

centre electrode over 7 successive stimulation phases during which a video frame is transferred to the retina. Fig. 3 illustrates two such phases.

Both FPGA- and processor-based image processing solutions are being considered. FPGA devices offer the possibility of having a near single chip solution, with acquisition, processing and transmission to the RF amplifier all running simultaneously. While processor- or DSP-based solutions lack the raw speed that FPGAs offer, they are easier to design. It is taken for granted that the device will eventually operate from batteries—a necessity if it is to be portable—making power management a critical issue.

The prosthesis provides a “user interface” for use by the patient. This could allow the user to switch between or override processing modes selected for a given environment. For example, it is anticipated that quite different processing will be required to assist a user in an outdoor setting compared with an indoor seated setting, in which the user wishes to read.

A “technician's interface” provides a means for medical and systems specialists to evaluate and control the performance of the prosthesis. Crucially, a real-time display of the image information being sent to the user is required. The technician requires a tool with which to evaluate and control the system and a means to appreciate the patient's point of view. The technician will also be able to immediately assess the effects of adjusting system parameters—although obviously it is only the patient who really knows what is happening.

B. Design objective

Final year computer engineering students at UNSW are required to take a design project course involving FPGA design. In 2010, students were set the goal of developing an FPGA-based vision processing system, suited to the needs of a technician, which renders a visual representation of the stimulatory signals that would be transferred via radio link to the implanted (internal) part of the prosthesis. Their implementation was required to provide methods for adjusting the processing so as to match the output of the system to the specific needs of a recipient in various use case scenarios [2].

The project required students to interface a CCD camera and VGA monitor with a low-cost FPGA processing board in order to perform video processing. A method for displaying the array on the monitor in a visually intuitive and understandable manner had to be devised. The intensity at which each electrode would be stimulated was to be obtained by *averaging* the intensity of the pixels in the corresponding region of the input.

Suggested methods for allowing a technician to adjust the stimulation pattern included:

- rotating the mapping of the input stream to the electrode array to compensate for misalignment of the implanted array;
- providing the means to “zoom” in on the central portion of the input stream;
- providing the means to adjust the area (number) of pixels that are averaged per stimulation site at a fixed field of view i.e. while it is possible to use all pixels in the image to calculate the stimulation values, in the

minimum, just 98 pixels need to be sampled and all others are ignored;

- providing the means to “remap” which region of the image is associated with each stimulation site to compensate for transpositions in stimulated regions and perceived stimulations; and
- the ability to stimulate any two nominated sites at will to allow technical staff to calibrate the system's performance with a user's perception of the stimulation.

This demonstration is of the work completed by one of the 5 groups taking the course in 2010.

II. DESIGN

A. Equipment Used

The design was implemented using the following equipment.

- Digilent Spartan-3E 1600 prototyping board, comprising:
 - Xilinx XC3S1600E Spartan-3E FPGA
 - Micron MT46V32M16 512MBit DDR SDRAM
 - Rotary control, bush buttons, switches, LEDs
- Digilent Video Decoder (Analog Devices ADV7183B)
- Digilent VGA accessory board
- BULLET-C22HR CCTV PAL camera
- LCD VGA monitor
- Xilinx ISE 11.1

B. Approach

To meet the design requirements, the problem was decomposed in terms of the available physical components:

- Video Decoder Board: system input
- DDR SDRAM: de-interlaced video frame buffer
- hexagon datapath: processing input video
- VGA Board: system output

In addition, multiple smaller modules, such as a settings controller for making adjustments to video processing and a video mixer to be able to simultaneously display both the input video and the computed electrode intensities to a VGA monitor were developed. Fig. 4. depicts a block diagram of the implemented system.

The Input Decoder takes a PAL input from the video camera and produces the input stream that is used by the rest of the system. The signal at the electrodes can be expected to be so diffuse (affecting tens of thousands of photoreceptors) and strong that it will be experienced as intense light. The design therefore processes greyscale pixel values. Coordinate values are associated with the pixels for processing by the Hexagon Datapath.

In the Video Datapath (blue), the input stream is stored in a frame buffer to synchronise between input (25 Hz) and output (60 Hz) frame rates. Moreover, the input frames arrive as two interlaced odd and even fields, which are de-interlaced as they are stored. The stored greyscale values are then retrieved display scanline by scanline and transformed into a VGA pixel stream.

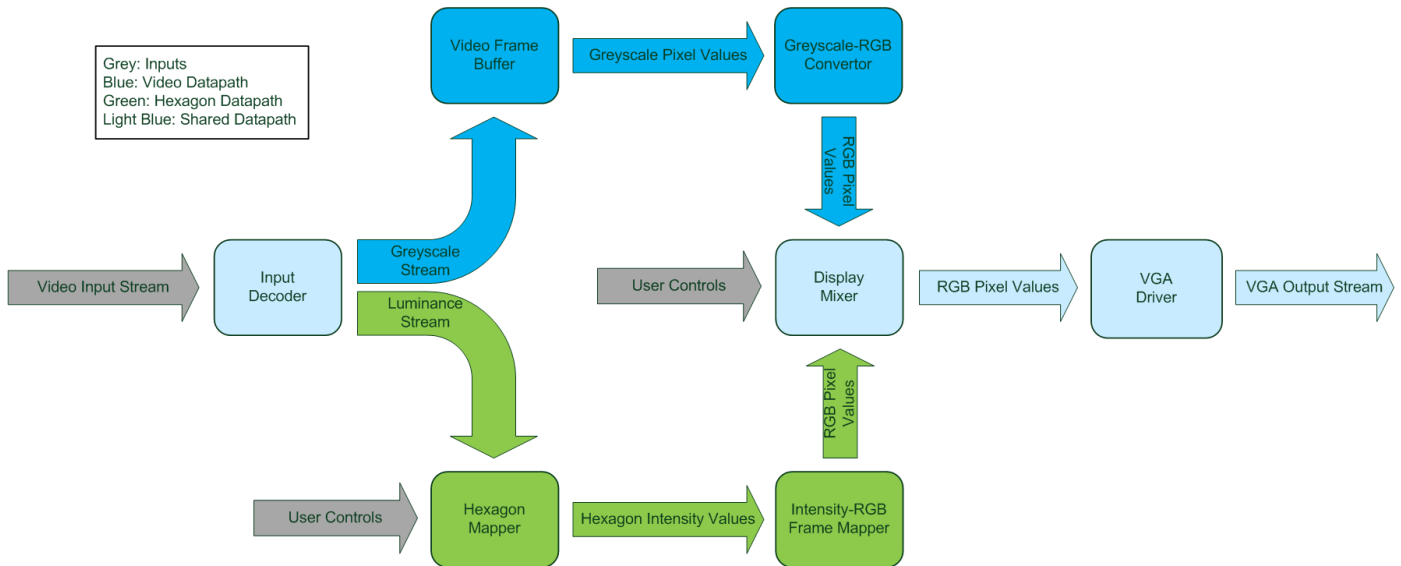


Fig. 4. Functional Block Diagram

The Hexagon Datapath processes a stream of luminance values from the Input Decoder and maps these to hexagon points. The average of the intensity values is output when requested. The Intensity-RGB Frame Mapper uses the current VGA display coordinate to determine which hexagon point (electrode) value to retrieve from block RAM for display on the monitor.

The Video Datapath and Hexagon Datapath are blended in the Display Mixer, which, based on user controls, combines their outputs for display. Finally, the VGA Driver produces the necessary VGA signals using the datastream received from the Display Mixer.

The independence of the Video and Hexagon Datapaths ensures the delay between video input and processing is minimized and allows power consumption to be reduced by disabling either datapath when it is not required. Power is further reduced when the system is used in the field since the Hexagon Datapath does not rely on a frame buffer or external memory.

III. DEMONSTRATION OVERVIEW

Photographs of the prototype and its output are provided in Fig. 5. Fig. 5a. depicts the setup with camera and display directly connected to the boards, whose arrangement is shown in Fig. 5b. The prototype is able to display the camera input signal in isolation (Fig. 5c.) and together with the intended electrode stimuli laid on top of the image regions used to calculate the intensity values (Fig. 5d.). The sample regions can be viewed in unmodified form on their own (Fig. 5e.) and as an array of derived average intensities (Fig. 5f.).

The image regions used to obtain the stimuli can be varied in size from 1x1 to 32x32 pixels in powers of 2. The spacing between the regions can also be varied in single pixel increments. In order to provide a sense of what the patient

would see, a mode for displaying the stimulus values at constant size and spacing (Fig. 5g.) has been provided.

Figures 5h-j. illustrate the use of the size and spacing controls to zoom in on text.

A. Conclusions

We have demonstrated a simple FPGA-based prototype for performing the video processing in a vision prosthesis. The results suggest FPGAs may well play a role in this interesting biomedical device. Such a view can only be reinforced by giving a moment's thought to the range of human activities the device should support, from going for a stroll and shopping at the local store to greeting friends and enjoying television. This variety of activities will demand concerted effort to develop suitable approaches that provide stimuli which are readily interpreted and understood by the user. Many will be compute intensive. Ideally, we will want the device to adapt to the present situation. Moreover, the time-spans over which adaptation will need to occur will probably allow for partial reconfiguration approaches.

ACKNOWLEDGMENTS

The authors gratefully appreciate Victor Lai's contribution to the development of the DDR memory interface and Liang Tang's assistance with the development of the video decoder's I²C controller. The donation of Digilent Spartan-3E 1600 Development Boards and ISE tools by Xilinx Inc. is also gratefully acknowledged.

REFERENCES

- [1] N. Lovell, et al., "External electronics for a vision prosthesis - Discussion and specification," Australian Vision Prosthesis Group 2009.
- [2] O. Diessel. (2010, 1 Aug 2010). *COMP4601 Project Description*. <http://www.cse.unsw.edu.au/~cs4601/10s1/project/4601-project.htm>

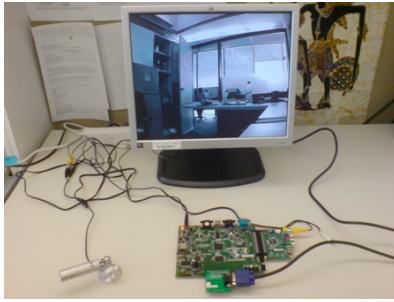


Fig. 5a. Equipment setup

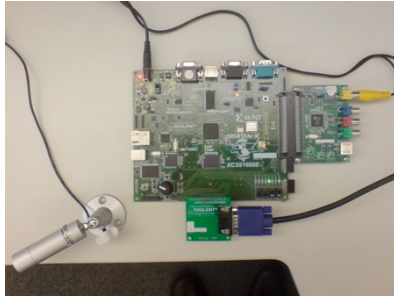


Fig 5b. Board arrangement



Fig. 5c. Office view



Fig. 5d. Stimulus values overlaid upon office view



Fig. 5e. Sampled region of office view

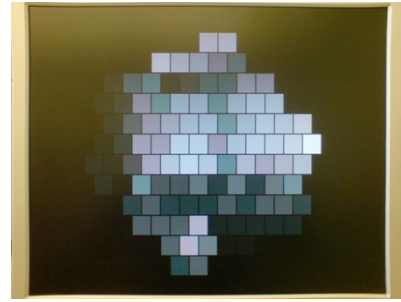


Fig. 5f. Averaged samples of office view

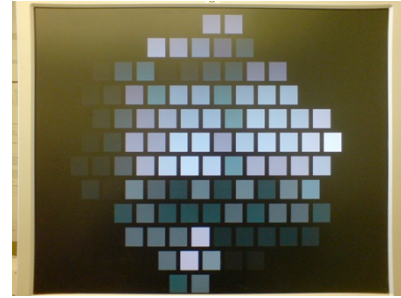


Fig 5g. Stimuli displayed at constant size and spacing

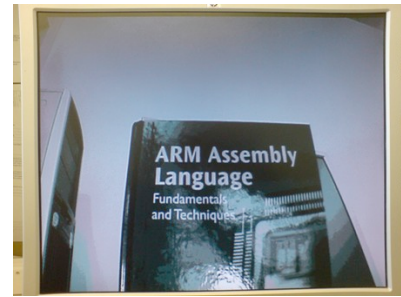


Fig. 5h. Text



Fig. 5i. Sampled region of text



Fig. 5j. Stimuli for sampled text displayed at constant size and spacing