35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Smartphones as Image Processing Systems for Prosthetic Vision

Marc P. Zapf, *Student member, IEEE*, Paul B. Matteucci *Member, IEEE*,
Nigel H. Lovell, *Fellow, IEEE* and Gregg J. Suaning, *Senior Member, IEEE*

Abstract— **The feasibility of implants for prosthetic vision has been demonstrated by research and commercial organizations. In most devices, an essential forerunner to the internal stimulation circuit is an external electronics solution for capturing, processing and relaying image information as well as extracting useful features from the scene surrounding the patient. The capabilities and multitude of image processing algorithms that can be performed by the device in real-time plays a major part in the final quality of the prosthetic vision. It is therefore optimal to use powerful hardware yet to avoid bulky, straining solutions. Recent publications have reported of portable single-board computers fast enough for computationally intensive image processing. Following the rapid evolution of commercial, ultra-portable ARM (Advanced RISC machine) mobile devices, the authors investigated the feasibility of modern smartphones running complex face detection as external processing devices for vision implants. The role of dedicated graphics processors in speeding up computation was evaluated while performing a demanding noise reduction algorithm (image denoising). The time required for face detection was found to decrease by 95% from 2.5 year old to recent devices. In denoising, graphics acceleration played a major role, speeding up denoising by a factor of 18. These results demonstrate that the technology has matured sufficiently to be considered as a valid external electronics platform for visual prosthetic research.**

I. Introduction

In the recent decades, it has been demonstrated that visual perception can be elicited by electrical stimulation of the visual cortex [1], the optic nerve [2] and the eye [3]. While there are approaches to confine image processing to the site of electrical stimulation [4], most of the current prosthetic device prototypes feature an external image acquisition and computation framework to process the captured visual scene prior to feeding it into the implanted circuitry (Fig.1). Thus, as technology improves, the image processing part of the implant system can be steadily upgraded without the need for follow-up surgical intervention.

Considering the limited information that can be displayed with current low-resolution prosthetic vision due to physical constraints in stimulation, emphasis has to be put on implementing image processing strategies which extract useful features from the original image for final display to the implant recipient. For orientation and navigation, promising approaches are depth mapping and obstacle detection [5], [6]. Regarding social interaction, the ability to recognize facial features is of particular importance to implant recipients. To our knowledge there are no reports of applications of face detection in image processing for vision prosthetics. Yet, there have been promising algorithms made available in open-source code, namely the Viola-Jones method [7]; this algorithm could be used to detect faces in camera images and emphasize them for presentation. Thus magnified, facial features and expressions are likely to be more easily recognizable with low-resolution prosthetic vision.

Performing advanced image processing, the external electronics hardware and software have to be sufficiently powerful, yet remain portable enough to not interfere in everyday tasks. Previous research presented light-weight, affordable, wearable devices [8], [9], yet there were hardware design-specific limitations in camera resolution, data transfer speed between camera and processor as well as hardware upgrade potential. Recently, it has been demonstrated that a commercially available, portable device based on an ARM processor was powerful enough to perform image processing tasks that could drive a 98 electrode visual prosthesis with meaningful data in real-time, thereby producing interpretable visual output [10]. ARM chips are low-cost, heat- and power-efficient alternatives to the larger-scale processor architectures, making them prime candidates for use in light, battery-powered devices such as mobile phones. As a result, in 2011, there were 7.9 billion ARM-based chips shipped, a 30% increase over 2010, with 95% of smartphones driven by these processors [11]. Given their omnipresence and availability, the question arises whether smartphones with their increasingly powerful processors and integrated high-

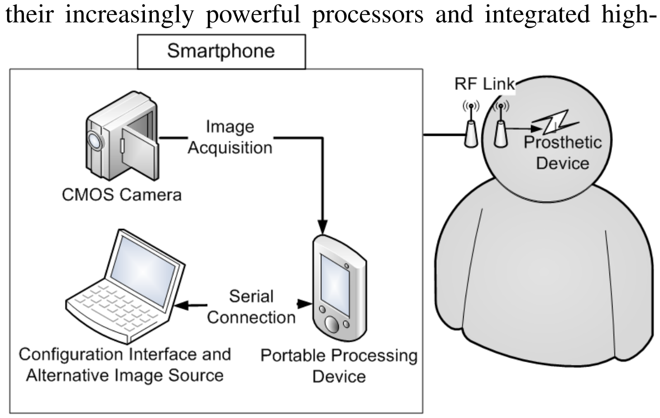

Fig. 1. Interaction between the image processing components. An external camera acquires images which are processed by the portable device and sent to the prosthetic device by means of the RF link. A monitor computer can be used to control sampling parameters. All external components are found in modern smartphones. From Matteucci, 2011.

Manuscript received February 4, 2012. This research was supported by the Australian Research Council (ARC) through its Special Research Initiative (SRI) in Bionic Vision Science and Technology grant to Bionic Vision Australia (BVA).

M.P. Zapf, P. Matteucci, N.H. Lovell and G.J. Suaning are with the Faculty of Engineering, Graduate School of Biomedical Engineering, The University of New South Wales, Kensington, NSW, 2052, Australia. M.zapf@unsw.edu.au

resolution cameras can be used in lieu of specialised external electronics for vision prosthetics and perform complex image processing in real-time (Fig. 1).

In the present paper the authors investigated the performance of recent ARM-powered mobile phones running the Android operating system (Google, USA) in computationally demanding real-time face detection on still and dynamic image material to evaluate the evolution of mobile hardware capacity from 2010 to 2012.

II. METHODS

A. Hardware

Five different models of mobile phones were tested, representing subsequent generations of flagship portable devices released at regular intervals. A 2010 Galaxy S (Samsung Korea) smartphone was used to evaluate whether Android devices are capable of running image capture and face detection at useful speeds. They incorporate a 1000 MHz single-core ARM Cortex-A8 processor and 512 Megabytes (MB) RAM. The performance of the Galaxy S was compared to a 2011 Galaxy S2 smartphone, featuring a 1200 MHz dual-core ARM Cortex-A9 chip and 1 Gigabyte (GB) RAM, and a 2012 Galaxy S3 with a 1400 MHz quad-core ARM-Cortex-A9 chip and 1 GB RAM, released 2012.

Two models of HTC One mobile phones (HTC, Taiwan) were included, both running at 1500 MHz and with 1 GB RAM, differing only in CPU and Graphics Processing Unit (GPU) type. A detailed summary of relevant hardware characteristics can be found in Table 1.

B. Software

The operating system version running on the devices was Android 4.1, a stable and widespread update of the Android 4 major release. For the Galaxy S2, version 4.2 was used due to 4.1 being unavailable for this phone. On the Galaxy S, a custom image (Cyanogenmod) of 4.1 was installed to allow for modulating the chip speed for CPU performance tests. Wireless signal transmission functions as well as background applications were shut down to free resources and minimise confounding factors for the benchmark tests.

TABLE I

HARDWARE SPECIFICATIONS OF THE DEVICES USED IN THE STUDY

Device / Component	Galaxy S	Galaxy S2	Galaxy S3	HTC One XL	HTC One X
Release Date	06/2010	04/2011	05/2012	06/2012	04/2012
Processor Type	ARM Cortex-A8	ARM Cortex-A9	ARM Cortex-A9	Qualcomm Snapdragon S4	ARM Cortex-A9
CPU Speed	1000 MHz	1200 MHz	1400 MHz	1500 MHz	1500 MHz
No. of Cores	1	2	4	2	4
Memory (RAM)	512 MB	1 GB	1 GB	1 GB	1 GB
GPU	PowerVR SGX 540 200 MHz	ARM Mali-400 MP 400 MHz	ARM Mali-400 MP 400 MHz	Adreno 225 200 MHz	Nvidia ULP GeForce 520 MHz
Camera Res. (Megapixel)	5	8	8	8	8
Max. Video Capture Res. (Pixel)	1280x720	1920x1080	1920x1080	1920x1080	1920x1080
Weight (g)	119	116	133	129	130

The Eclipse Java IDE Juno (Eclipse Foundation, Canada) software development environment, on a development desktop computer, was used to install face detection software and run benchmark tests on the devices connected via USB. The Android software development kit (SDK) was integrated into Eclipse to permit development for Android devices.

The OpenCV 2.3.4.1 (Intel, Willow Garage, USA) image processing library provided the face-detection algorithm used for testing. The library is open-source, cross-platform, and can easily be adapted to the specific needs in prosthetic image processing. The face detector uses the Viola-Jones method, which permits detection in real-time by combining rapid, multiple feature detection with machine-learning methods. Due to its complexity, it is computationally intensive and thus a good indicator of device capability.

An OpenCV control program and the face detection application was installed on all devices and started from Eclipse. The application retrieved a continuous video stream from the internal rear camera, performed detection and displayed the feed with a rectangle overlay on detected faces. The time required in milliseconds (ms) for the algorithm to run face detection on captured frames was saved to a log file.

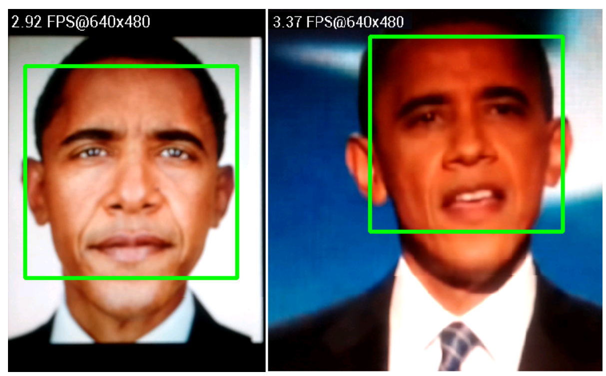

Fig. 2. Image Material Used in Benchmark Testing. A still frontal portrait (left) and a video with dynamic head movements and turning (right), causing the face detector to steadily lose and regain focus, were displayed at 100% screen height. A rectangle is displayed while a face is detected. Photo by Martin Schoeller, 2004; video from www.cbsnews.com.

C. Performance Testing

In this study, the effect of varying CPU rate with otherwise unchanged hardware, the development of performance in subsequent generations of smartphones and finally the effect of hardware acceleration on performance were observed.

For benchmarking, the phones were fixed parallel to a computer screen on which either an image or a video was displayed (Fig. 2). The images measured 100% of the image capture height. 1100 frames were processed and the first 100 frames discarded to allow for stabilization of data flow. The average ms $\pm$ standard deviation per frame and over multiple sessions (n=3) was determined.

CPU clock - dependent benchmarks were run on the Galaxy S2 phone at 640x480 pixel capture resolution with fixed clock speeds of 500, 800, 1000 and 1200 MHz.

The performance of one of the Galaxy S1, the S2, S3, and the HTC One X and XL was compared by running face detection at 640x480 pixels and native CPU clock speed.

Some newer devices feature ARM chips with integrated dedicated graphics processors partially optimized for OpenCV algorithms such as the Tegra 3 with a Geforce GPU (Nvidia, USA), which might speed up algorithms previously considered too demanding for vision prosthetics to usable levels. Since face detection had not been Tegra-optimized, we ran a computationally intensive Tegra-supported image denoising algorithm (median blur) on the HTC One X at 960x720 pixel resolution. Processing time with disabled versus enabled graphics acceleration was measured.

III. Results

A. Effect of CPU Clock Rate on Performance

The time needed for face detection decreased with increasing clock rate for both still and dynamic face recognition (Fig. 3). A considerable fluctuation of processing times was found at the lowest clock speed, stabilizing from 800 MHz upwards. For still images, slightly more time was needed, from on average 78.0±31.3 ms at the lowest tested clock rate of 500 MHz to 24.8±5.5 ms at 1200 MHz. Processing of dynamic faces ranged from 64.5±24.2 ms at 500 MHz to 20.1±4.9 ms at 1200 MHz.

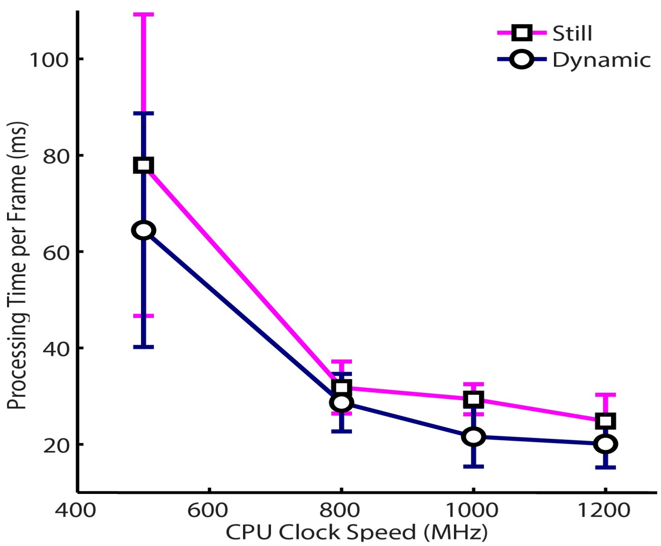

Fig. 3. 2011 (Galaxy S2) image processing speed throughout sessions (n=3) as a function of CPU clock speed. Squares: still source image; circles: dynamic source image. Standard deviations are shown as vertical bars.

B. Development of Performance 2010-2012

On the common Android 4.1.0 processing a 640x480 pixel image, the oldest 2010 model Galaxy S performed worst, requiring on average 197.6±13.9 ms for still and 170.8±15.8 ms for dynamic images (Fig. 4), showing higher absolute fluctuations in processing time despite running at 1000 MHz clock rate. A significant performance increase and stabilization was found in more recent models, to 24.8±5.5 ms for still and 20.1±4.9 ms for dynamic sources was recorded testing the 2011 model Galaxy S2. The Galaxy S3, released 2012, reduced the time needed to 9.7±4.6 and 8.4±4.4 ms, respectively.

The HTC One X and XL phones performed similarly and consistently well throughout all conditions. The HTC X outperformed the XL slightly with a mean of 7.6±1.74 as opposed to 8.8±1.9 ms.

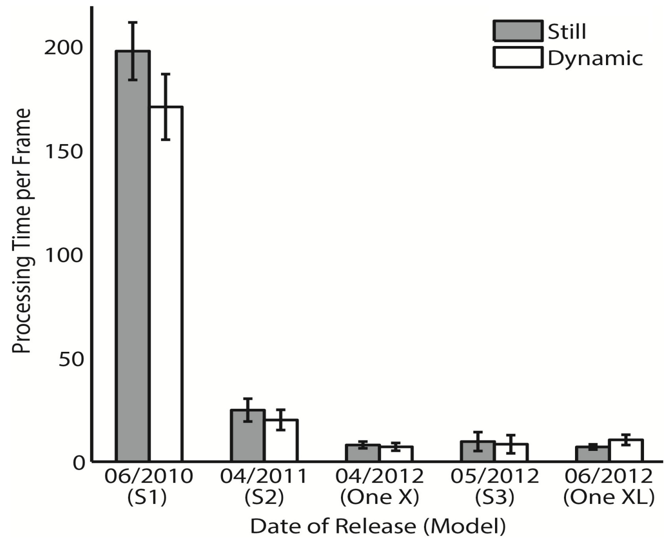

Fig. 4. Development of image processing speed throughout sessions (n=3) between mid-2010 (Galaxy S1) and mid-2012 (Galaxy S3, HTC One XL and X) running still and dynamic face tracking. Dark bars: still image; white bars: dynamic image. Standard deviations are shown as vertical bars.

C. Effect of Hardware Acceleration on Performance

Performing the median blur algorithm with disabled GPU acceleration, 275.4±7.0 ms were needed for still and 286.4±8.1 ms for dynamic scenes (Table 2). Turning on the Tegra features, these times dropped to 15.3±1.0 ms and 16.6±1.2 ms, respectively.

TABLE II
HTC One X image denoising speed in ms with enabled and disabled Tegra acceleration (n=2)

Tegra Acceleration / Source Image	Enabled	Disabled
Still Image	275.4±7.0 ms	15.3±1.0 ms
Dynamic Image	286.4±8.1 ms	16.6±1.2 ms

IV. Discussion and Conclusions

In the present paper we have shown the marked effect the evolution of ARM-based technologies has had on performance in affordable, portable devices. Higher CPU clock rates alone were found to speed up image processing considerably (Fig. 3). Furthermore, the integration of advanced multi-core CPUs in conjunction with other hardware upgrades like the doubling of RAM has led to a more significant rise in processing speed than may be predicted by pure clock rate increase alone (Fig. 4). On average, within two years, speed increased by a factor of 22 from 184.2 ms to only 8.4 ms needed for face detection.

In order to determine whether smartphones have become fast enough to perform image processing between capture and display, the frame rate and thus the temporal resolution of stimulation required for flicker-free image presentation (flicker-fusion frequency) in prosthetic vision has to be reviewed. While in optic nerve stimulation, a 10 Hz stimulation rate has been reported to eliminate flicker [12], 40-50 Hz were required for retinal prostheses [13]. Setting 40-50 frames per second as an aim, i.e. displaying an image every 20 ms, completing all processing faster than 20 ms would eliminate lag. This value has already been undercut by 2012 devices. It is probable that higher speeds will be reached by next-generation devices with more RAM and higher CPU clock rates. The final processing speeds of a certain device will be influenced by the computational demand and extent of image processing done in an actual prosthesis. The dependence of frame rate on image capture resolution will have to be further investigated.

Previous studies demonstrated image processing using portable, customized ARM-based development kit computers as an alternative to professional consumer hardware, however frame rates dropped considerably [8] or drastic image downsampling was required [9]. It can be expected that the performance of these devices has improved in the meantime; however, principal limitations such as the low data transfer speed from camera to the processor persist. It is due to this that the authors suggest smartphones as ideal portable image processors and stressed the importance of dedicated graphics processors alongside the ARM chip.

Running considerably more demanding algorithms such as denoising, the impact of optimized algorithms for the Tegra chip has been shown. While face detection is already sufficiently fast to not require further acceleration, future prosthetics desirably feature simultaneous extraction of multiple image features, finer details or more sophisticated algorithms, posing a challenge for reaching the frame rates required for lag-less display. Optimizing each single algorithm to the built-in hardware, graphics acceleration could reduce the overall computation time below this limit.

The evolution of ultra-compact mobile phones may make image processing in vision prosthetics increasingly lightweight, affordable and, noteworthy, inconspicuous. The latter addresses the concerns of prospective implant recipients that modern prosthetic devices are more noticeable, letting them stand out against healthy individuals; integrating common consumer electronics serves to blur this boundary. It might suffice installing an application on the personal phones of the patients in order to establish a link with a newly implanted vision prosthesis. Google's announcement [14] to release a minimalist spectacle-integrated device with camera and face detection by 2014 feeds hope that in the near future external image processors will be so discreet that blindness will be even more masked.

In this preliminary phase of visual prosthetics, with the first routine use of retinal implants years ahead and several technical issues still unsolved, mass production of increasingly potent mobile hardware renders it unnecessary to assign resources for dedicated external hardware development. It is unlikely that the bionic vision niche market will be able to out-engineer professional consumer hardware developers. However, given the impact of the software components in directing communication of hardware components, it will remain advisable to evaluate the options for optimizing algorithms for performance. Besides the implant itself, combining available hardware and software resources into a powerful external image processing package might form one foundation for meaningful and intuitive prosthetic vision.

References

[1] W. H. Dobelle, "Artificial vision for the blind by connecting a television camera to the visual cortex," *ASAIO journal (American Society for Artificial Internal Organs: 1992)*, vol. 46, no. 1, pp. 3–9, Feb. 2000.

[2] C. Veraart, C. Raftopoulos, J. Mortimer, J. Delbeke, D. Pins, G. Michaux, A. Vanlierde, S. Parrini, and M.-C. Wanet-Defalque, "Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode," *Brain Research*, vol. 813, no. 1, pp. 181 – 186, 1998.

[3] E. Zrenner, "Will retinal implants restore vision?" *Science*, vol. 295, no. 5557, pp. 1022–1025, 2002.

[4] E. Zrenner, K. D. Miliczek, V. P. Gabel, H. G. Graf, E. Guenther, H. Haemmerle, B. Hoefflinger, K. Kohler, W. Nisch, M. Schubert, A. Stett, and S. Weiss, "The development of subretinal microphotodiodes for replacement of degenerated photoreceptors," *Ophthalmic research*, vol. 29, no. 5, pp. 269–280, 1997.

[5] P. Lieby, N. Barnes, C. McCarthy, N. Liu, H. Dennett, J. Walker, V. Botea, and A. Scott, "Substituting depth for intensity and real-time phosphene rendering: Visual navigation under low vision conditions," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 8017 –8020.

[6] A. Stacey, Y. Li, and N. Barnes, "A salient information processing system for bionic eye with application to obstacle avoidance," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 5116 –5119.

[7] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[8] D. Tsai, J. Morley, G. Suaning, and N. Lovell, "A wearable real-time image processor for a vision prosthesis," *Computer Methods and Programs in Biomedicine*, vol. 95, no. 3, pp. 258 – 269, 2009.

[9] W. Fink, C. X. You, and M. A. Tarbell, "Microcomputer-based artificial vision support system for real-time image processing for camera-driven visual prostheses," *Journal of Biomedical Optics*, vol. 15, no. 1, pp. 016 013–1 – 016 013–10, 2010.

[10] P. Matteucci, P. Byrnes-Preston, S. Chen, N. Lovell, and G. Suaning, "Arm-based visual processing system for prosthetic vision," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 3921 –3924.

[11] ARM, "Arm holdings plc annual report & accounts 2011," 2012. [Online]. Available: http://financialreports.arm.com

[12] C. Veraart, M.-C. Wanet-Defalque, B. Grard, A. Vanlierde, and J. Delbeke, "Pattern recognition with the optic nerve visual prosthesis," *Artificial Organs*, vol. 27, no. 11, pp. 996–1004, 2003.

[13] M. S. Humayun, E. de Juan Jr., J. D. Weiland, G. Dagnelie, S. Katona, R. Greenberg, and S. Suzuki, "Pattern electrical stimulation of the human retina," *Vision Research*, vol. 39, pp. 2569–2576, 1999.

[14] Google, "Project glass," April 2012. [Online]. Available: https://plus.google.com/+projectglass/about