Toward real-time virtual biopsy of oral lesions using confocal laser endomicroscopy interfaced with embedded computing

Patricia S. P. Thong
Stephanus S. Tandjung
Muhammad Moeen Movania
Wei-Ming Chiew
Malini Olivo
Ramaswamy Bhuvaneswari
Hock-Soon Seah
Feng Lin
Kemao Qian
Khee-Chee Soo
Toward real-time virtual biopsy of oral lesions using confocal laser endomicroscopy interfaced with embedded computing

Patricia S. P. Thong, a Stephanus S. Tandjung, b Muhammad Mobeen Movania, b Wei-Ming Chiew, b Malini Olivo, c Ramaswamy Bhuveswari, d Hock-Soon Seah, b Feng Lin, b Kemao Qian, b and Khee-Chee Soo b

1 Introduction

1.1 Oral Cancer Diagnosis

Oral cancer is becoming one of the most common types of cancer worldwide, particularly in developing countries where this form of malignancy ranks as the 7th and 9th most common cancers in males and females, respectively. 1, 2 Risk factors include smoking, drinking alcohol, using smokeless tobacco products and infection with the human papillomavirus. 3 Recent advances in cancer treatment techniques have not significantly improved the survival rate for oral cancer, which remains at about 50%, due to it often not being diagnosed until a relatively advanced stage. Early diagnosis is therefore crucial for a good treatment outcome. Currently, lesions of the oral cavity are diagnosed using white light endoscopy followed by histopathological examination of biopsy samples. Endoscopic examinations may also be extended to the larynx and esophagus to check for other possible lesions. 3

There are some challenges in the existing conventional techniques for oral cancer diagnosis. Firstly, early oral lesions are often flat, making it difficult to distinguish between benign and malignant lesions under white light illumination. Secondly, histopathology is time-consuming and requires specialized skills and experienced, trained personnel. Thirdly, while biopsies are generally safe, they carry a small risk of patient complications. Finally, it might be difficult to determine the margins of oral lesions, and multiple biopsies are often required to ensure a clear margin during surgical procedures. Therefore there is a need to develop minimally invasive “virtual” biopsy techniques that can provide accurate and real-time diagnosis of oral lesions in the clinic, which will help to target biopsy to abnormal regions, and thus reduce the number of biopsies needed to make a diagnosis. One emerging optical technique that has shown potential as a tool for virtual biopsy and guided biopsy procedures is confocal laser endomicroscopy.

1.2 Confocal Laser Endomicroscopy

Confocal laser endomicroscopy (CLE) is an endoscopic technique that complements conventional endoscopy by enabling in vivo imaging of tissue and cellular structures at microscopic resolution (about 1 μm laterally). 4–6 Through cellular, structural and molecular imaging, information can be extracted not only from the surface but also from deeper subsurface layers, thus offering a tool for optical or virtual biopsy in the clinic. 7–9 Currently there are two commercially available confocal laser endomicroscope systems. In the endoscope-based system (developed by Optiscan Pty, Ltd., Victoria, Australia), the focusing and scanning mechanisms are miniaturized into the distal tip of an endoscope. 10 In the probe-based system (developed by...
Mauna Kea Technologies, Paris, France), the scanning is achieved at the proximal end of a fiber optic probe.11

Recent studies demonstrate the potential of CLE as a clinical tool for surveillance and diagnosis of several cancerous and pre-cancerous conditions. These include clinical diagnostic applications in the airways,12–14 upper and lower gastrointestinal tracts,15–20 bladder neoplasia,21–23 cervical intraepithelial neoplasia24–27 and the oral and oropharynx.28–31 Results have been promising. For example, a recent study by Xie et al. reported that CLE could detect adenomas in colonic polyps at a sensitivity and specificity of 93.9% and 95.9%, respectively, when compared to histopathology results.23 In another recently completed multicenter randomized controlled trial for detection of high-grade dysplasia and early carcinoma in Barrett’s esophagus, the combined use of probe-based CLE (pCLE) and high-definition white-light endoscopy (HD-WLE) yielded a favourable sensitivity and specificity of 68.3% and 87.8% compared to 34.2% and 92.7%, respectively, for the use of HD-WLE alone.23

Confocal laser endomicroscopy has also been used as an aid for targeted biopsy procedures to improve effectiveness and reduce the number of biopsies needed. For example, Gunther et al. reported that targeted biopsies guided by either chromoendoscopy or CLE resulted in higher detection rates of intraepithelial neoplasia in the surveillance of inflammatory bowel disease.32 Confocal endomicroscopy also holds promise for image-guided surgery by aiding the assessment of lesion margins during or following surgical procedures.33

With the use of suitable fluorescent dyes, confocal fluorescence imaging can be carried out. Fluorescein sodium is commonly used, as it is safe for human use34 and spectrally matched to the 488-nm excitation wavelengths of many confocal endomicroscope systems. However, fluorescein sodium is non-specifically absorbed by all cells and thus may result in false positives during diagnostic imaging. Other possible fluorescent dyes include hypericin, a photosensitizer extracted from the plant commonly known as St John’s wort, and the fluorescent pre-cursor; 5-aminolevulinic acid (5-ALA) that metabolises into the fluorescent compound protoporphyrin IX (PpIX). Hypericin and 5-ALA may be more selectively taken up by abnormal cells, and may thus enable fluorescence diagnostic imaging with higher specificity.35–39

1.3 Toward Real-Time Virtual Biopsy of Oral Lesions

We have previously described the use of endoscopy-based CLE for confocal fluorescence diagnostic imaging of the human and murine oral cavities.28–30 A prototype confocal endomicroscope with a rigid, hand-held probe was used with 5-ALA, fluorescein sodium and hypericin as contrast agents. Hypericin was used in mouse models while 5-ALA and fluorescein sodium were used in rat models.30,32 Fluorescence images of the normal rat tongue were compared to those from carcinogen-induced models of oral squamous cell carcinoma (SCC). Images of the normal rat tongue showed regularly arranged filiform papillae, while the architecture of the SCC rat tongue showed up as more irregular. In pilot clinical studies, 5-ALA was topically applied to the oral cavities of healthy volunteers and an oral SCC patient to compare ALA-induced PpIX fluorescence images from the normal and SCC human tongue.36–38 The results demonstrated the capability of the confocal endomicroscope to differentiate between the normal and SCC tongue by morphology and tissue architecture, leading to its potential to be used as a minimally invasive technique for oral cancer diagnosis. This was in agreement with Haxel et al. who reported that CLE could be used for the diagnosis of malignancy in the human oral and oropharyngeal cavity by means of the altered tissue architecture and irregularity in blood vessels.39

Conventional bench-top confocal microscopes are equipped with hardware and software to acquire and render 3-D confocal image stacks. However, these systems can only be used for ex vivo imaging of tissue sections. On the other hand, confocal endomicroscopes enable in vivo imaging but capture and display images from one single focal plane at a time. Real-time 3-D image registration, voxel-based processing and rendering software are unavailable, making it difficult to recognize 3-D structures. In order to bridge this gap and move toward a real-time “virtual” biopsy technique, we developed a 3-D fluorescence imaging system by interfacing a confocal laser endomicroscope to an embedded computing system.40–42 We used a high performance multimedia field-programmable gate array (FPGA) board as a reconfigurable platform.40 The FPGA board has the required interfaces, such as dual video support for Digital Video Interface (DVI), Thin-Film Transistor (TFT) flat panel display, Personal System/2 (PS/2) keyboard and mouse ports, a four-line-by-16-character Liquid Crystal Display (LCD) display, eight white user-programmable Light-emitting Diodes (LEDs), and general input/output (I/O) pins. Other peripherals, such as a keyboard, can also be used for user interface. In this study, we describe the development of the endomicroscope-embedded computing system for 3-D fluorescence imaging of the oral cavity.

2 Materials and Methods

2.1 Endomicroscope-Embedded Computing System

An endoscope-based confocal laser endomicroscope system (FIVE1, Optiscan, Australia) was fitted with a short, hand-held rigid probe (model RBK6315A) that is suitable for imaging the oral cavity. The excitation source is a 488-nm laser coupled into a single optical fiber acting as both a point source and as a point detection pinhole for confocal imaging.41 The lateral resolution is about 0.7 μm. The rigid probe houses the miniaturized components of the x-y scanning mechanism, allowing images to be captured with a field of view of 475 μm × 475 μm. The laser power can be adjusted between 0 and a maximum of 1000 μW at the distal tip of the probe in contact with the tissue sample. Fluorescence signals are collected via a 505- to 750-nm emission filter. Under the normal mode of operation, Z-depth sectioning can be achieved via a footswitch that controls the imaging depth from the surface down to deeper planes with a nominal step size of about 4 μm between consecutive slices. In biological samples, the maximum imaging depth is about 250 μm below the surface, depending on the tissue optical properties.

Figure 1 shows the schematic diagram of the confocal endomicroscope interfaced with an embedded computing system based on an RC340 FPGA (Mentor Graphics Corporation, USA). The embedded FPGA platform functions as the main board and controls the endomicroscope through a Z-depth control circuit called the daughter board.40 The fluorescence signal from the hand-held probe is converted by the endomicroscope into digital data that is displayed on the monitor and sent to the RC340 main board. The main board captures the images for real-time image processing and displays the processed data via the DVI interface. Under this mode of operation, depth control...
via the footswitch is disabled. The RC340 board automatically controls the endomicroscope system to capture confocal image stacks upon initiation by the operator via a keyboard user interface.

2.2 Volume-Rendering using GPU

For prototyping, the 3-D visualization of confocal image stacks was first developed on a PC equipped with a graphics processing unit (GPU). We utilized 3-D texture slicing to generate high quality volume renderings of the 3-D datasets. Our volume-rendering process starts with filtering the dataset to remove noise and other high frequency artifacts, after which a 3-D texture is generated to store the processed image stack. This stack is then classified using a transfer function. Finally, the volume is sampled using view-aligned proxy geometry to generate the 3-D texture slicing. These slices are then alpha-blended in back-to-front order using hardware-based alpha blending.

The overall volume-rendering process is illustrated in Fig. 2. The classification of datasets is critical to make sense of the volume-rendering. The process typically assigns a mapping \((R \to R^3)\) from a scalar value to a color value, implemented as a look-up table. We created a customized widget (Fig. 3) for assignment of the transfer function. This widget generates piecewise linear transfer functions and enables the user to create custom mappings for each individual color channel, alpha, red, green and blue (ARGB), by plotting control points. Each channel is rendered in its own colour to help distinguish it. The alpha channel curve is coloured black. The X-axis of the widget refers to the scalar value of the intensity, which ranges from 0 to 255 in our datasets. The Y-value stores the intensity (0-1) as an indication of how much of the value is being assigned. To assist the user in visualizing the output generated from the combined function, a colour preview strip is rendered under the curve while the background is rendered to portray the colour’s transparency. The higher the background is, the more transparent the colour. In addition, a specialized interpolation scheme is also developed to maintain smooth interpolation across acquired slices.

2.3 Volume Rendering using FPGA

In our embedded computing solution, we also programmed the RC340 FPGA board to support real-time 3-D visualization while the image stack is being acquired. The rendering of incrementally acquired slices will be produced and displayed “on-the-fly” through the video output on the FPGA board. The volume ray-casting technique is employed in our system to generate two-dimensional (2-D) projection output images viewed from arbitrary angles in 3-D. In this algorithm, imaginary rays are projected towards the dataset and sampling points along each ray are accumulated to produce the output, which is illustrated in Fig. 4. Each pixel on the image plane is initialized as a ray origin. Rays cast from each origin traverse through the confocal dataset volume. Consistent points on each ray are sampled within the dataset.
After a further incubation period of at least 45 min, fluorescence 3-D imaging was carried out.

3 Results

3.1 Automated Acquisition of Confocal Image Datasets

We have developed a prototype 3-D fluorescence imaging system comprising a confocal endomicroscope interfaced to an FPGA-based embedded computing system. In the normal mode of operation, an operator manually controls the imaging depth via a footswitch to collect individual confocal images at the desired depths. In our endomicroscope-embedded computing system, the image acquisition control is replaced by the daughter board and circuitry programmed on the FPGA board. In place of the footswitch, we have designed an interface with main components that act as relays to isolate the endomicroscope system from the FPGA board electronically. The controller circuitry replaces operator tasks from basic to higher level ones. With this controller, the user only needs to push one button for the FPGA board to capture a confocal image stack (dataset) automatically from the initial imaging depth and sequentially deeper focal planes until the desired imaging depth has been reached and the user initiates a stop signal. The current system is capable of acquiring one image per 1.4 sec, the fastest attainable rate limited by endomicroscope hardware. The automated acquisition of image stacks minimizes the data acquisition time and thus effectively minimizes the chances of movement in between consecutive images.

3.2 Rendering of Murine Datasets using GPU

Pilot testing of the prototype endomicroscope-embedded computing system was carried out using mouse models. Rendering of image stacks acquired from the murine tongue was carried out using a GPU program. Figure 5 shows a series of rendering results with each series having twice as many texture slices. This figure illustrates how adding more slices improves the rendering quality considerably.

Figure 6(a)–6(f) show six consecutive images from a confocal image stack acquired from the dorsal surface of the mouse tongue following topical application of fluorescein sodium. The arrows in (a) indicate filiform papillae on the surface of the mouse tongue. The images show that while bright fluorescence is observed at the surface (a), there is a gradual reduction in fluorescence intensity as the imaging depth increases (b–f). Figure 6(g) shows the 3-D volume rendering result obtained using a GPU-based program and displayed in false colours. The conical shapes of the filiform papillae are well rendered.

2.4 Fluorescence 3-D Imaging of the Murine Oral Cavity

Fluorescein sodium and hypericin were used as fluorescent agents in 6- to 8-week-old Balb/c nude mouse models. Fluorescein sodium (Novartis Pharma AG, Switzerland) was freshly prepared as a 1% solution while hypericin (Molecular Probes, USA) was prepared as a 0.004% solution. Topical application to the murine oral cavity was carried out by the insertion of cotton buds soaked in the freshly prepared fluorescein or hypericin solutions for 5 to 10 min. Following an incubation period of 30 min, the mice were sacrificed and the tongues excised for imaging. Excised tissue of the mouse tongue was sectioned and processed for haematoxylin and eosin (H&E) staining.

2.5 Fluorescence 3-D Imaging of the Human Oral Cavity

A pilot clinical study to test the prototype system for fluorescence 3-D imaging was approved by the Centralised Institutional Review Board of the Singapore Health Services Pte. Ltd. Four healthy volunteers with no history of oral malignancies and two patients who were undergoing surgical procedures for lesions in the head and neck were recruited for the study following informed consent. Topically applied hypericin was used as the fluorescent agent. Additionally, fluorescein was also used in the volunteer group only to compare the results from topically applied hypericin and fluorescein.

Hypericin (Molecular Probes, USA) was freshly prepared in 1% serum albumin in phosphate buffered saline (PBS) and diluted in saline to give an 8 μM instillation solution. The solution was filtered and topically administered to both the volunteer and patient groups by oral rinsing using 100 ml of solution over 30 min. After a further incubation period of at least 45 min, fluorescence 3-D imaging was carried out.

Fluorescein (Novartis Pharma AG, Switzerland) was freshly diluted in PBS to obtain a 0.1% solution. The solution was filtered and topically administered to the volunteer group only by oral rinsing using 100 ml of solution over 30 min.
Figure 7 shows the H&E stained image of a cross-section of a mouse tongue, showing filiform papillae on the surface (arrows) similar to those seen in the confocal images. Compared to the individual confocal images captured from single focal planes and the H&E stained cross-sectional image, the 3-D image provides additional topographical information that is not easily visualized from the 2-D images.

Figure 8 shows the rendering results for a confocal image stack acquired from the dorsal surface of a mouse tongue following topical administration of hypericin. In addition to the normal composite rendering with classification (a), we also used an omni-directional “light” source to illuminate surface details that are otherwise less visible (b).

3.3 Rendering of Human Datasets using GPU

Confocal fluorescence image stacks of various sites in the human oral cavity were acquired in vivo from healthy volunteers following topical administration of fluorescein sodium or hypericin. The sites investigated include the dorsal surface of the tongue, base of tongue, floor of mouth, the buccal mucosa and the lip. Real-time volume-rendering of the acquired image stacks was achieved using a GPU-based program. In Fig. 9(a) and (b) show confocal images from the same dataset that was acquired from the dorsal surface of the human tongue following topical administration of fluorescein sodium. The image in Fig. 9(a) was captured at the surface and Fig. 9(b) image from a focal plane approximately 30 μm below the surface. Filiform papillae [solid arrows in (a)] and cellular structures [dotted arrows in (b)] can be seen in the individual images. The GPU volume-rendering result of the entire dataset, displayed in false colours in (c), show the depth relation of the filiform papillae with respect to the cellular structures below the surface. Such information on depth relation is not easily visualized from the individual 2-D confocal images.

In Fig. 10(a) and (b) show confocal images from the same dataset that was obtained from the buccal mucosa of a healthy volunteer following topical application of fluorescein sodium. Image (a) was captured from the surface and image (b) from a deeper layer approximately 30 μm below the surface. Cellular structures at the surface [solid arrow in (a)] and below the surface [dotted arrow (in (b)] are observable from the individual images. The GPU volume-rendering result of the entire dataset, displayed in false colours in (c), shows the depth relation between the cellular structures from different focal planes.
Figure 11 shows images acquired from the human buccal mucosa following topical administration of hypericin. Image (a) was captured at the surface and image (b) approximately 15 μm below the surface in the same dataset. Cellular structures at the surface (solid arrow) and below the surface (dotted arrow) can be observed from the individual images, while the GPU volume rendering result of the entire dataset, displayed in false colors (c), shows the depth relation of the filiform papillae to the cellular structures.

Pilot testing of the endomicroscope-embedded computing system in a clinical setting was carried out on two patients who were undergoing surgical procedures for lesions in the head and neck. Hypericin was topically applied prior to both in vivo and ex vivo imaging using the prototype imaging system. As the initial testing yielded weak fluorescence signals and noisy images, the results are not shown here. Further improvements are being made to the system to achieve better performance in a clinical setting.

3.4 Rendering of Image Stacks using Embedded FPGA

The FPGA board in our endomicroscope-embedded computing system was also programmed to support real-time 3-D visualization while the image stack is being acquired. The rendering of incrementally acquired slices will be produced and...
displayed “on-the-fly” through the video output on the FPGA monitor. The rendering pipeline developed on the embedded FPGA system is capable of rendering datasets while they are being acquired. The current rendering module deals with grayscale pixel values from the confocal endomicroscope. Shading effects are omitted in the current stage to simplify the rendering pipeline. Upon retrieval of an image slice, it is stored in the on-board synchronous Dynamic Random-Access Memory (SDRAM) module. The maximum capacity of the SDRAM is 256 MB, which is sufficient for storing up to 256 slices with a resolution of $1024 \times 1024$ pixels using direct address indexing. This capacity is sufficient for our typical CLE datasets. Subsequently, the output pixel values from rendering pipeline are stored in a frame buffer driven by the display module.

**Fig. 12** Real-time volume rendering of a mouse tongue dataset acquired following topical application of hypericin. The results were obtained under different thickness: (a) 2 slices, (b) 10 slices, (c) 15 slices, (d) 20 slices, (e) 25 slices, (f) 30 slices, (g) 35 slices and (h) 38 slices.
minimizes the chances of movement in between consecutive images, thus providing for more effective volume rendering in real-time. Even with this automation, image acquisition in the Z plane can be influenced by operator movement during image capture, for example if the pressure on the tissue changes and there is compression of the tissue. Selecting a different resolution changes the speed of image capture and thus may influence the effect of operator movement during the acquisition time. Currently, the best attainable rate limited by endomicroscope hardware is about 1.4 sec per frame. While this rate is slow by real-time standards, our results demonstrate the potential for 3-D endomicroscopic imaging of the oral cavity. Further hardware acceleration to increase the image capture rate will help to improve the performance of the system.

We tested the prototype endomicroscope-embedded system on murine models, healthy human volunteers and patients with head and neck lesions. Fluorescence 3-D imaging was carried out using fluorescein sodium or hypericin, both of which are safe for diagnostic use in humans. 

4 Discussion and Conclusion

Oral cancer is becoming one of the most common forms of cancer worldwide and early diagnosis is the key to a good prognosis. Current conventional techniques used for diagnosing oral cancer have their limitations and there is a need to develop newer and better techniques. We have previously shown the potential for confocal laser endomicroscopy (CLE), a minimally invasive endoscopic technique, to be used for fluorescence diagnostic imaging of oral cavity lesions. In this study, we present the further development of both hardware and software to optimize CLE for fluorescence 3-D imaging and to move toward a real-time virtual biopsy of oral lesions in a clinical setting.

We developed a prototype 3-D fluorescence imaging system comprising a field-programmable gated array (FPGA)-based embedded computer system interfaced to a confocal laser endomicroscope. The system is designed for automated acquisition of confocal image stacks and real-time volume rendering and display of 3-D tissue structures. In the normal operation of the endomicroscope, the manual acquisition of image stacks is accomplished by the operator using a footswitch. This process is slow and subject to movement by both the operator and the subject being imaged. With the prototype system, the image stack acquisition control is automated and needs only a start and stop input from the operator. Automation effectively
Acknowledgments
The authors would like to thank Cheong Lee Sing of the School of Computer Engineering, Nanyang Technological University, Singapore; Sasidharan Swarnalatha Lucky and Tee Chuan Sia of National Cancer Centre, Singapore; visiting student Agostino Guida of Seconda Università degli Studi, Naples, Italy; and Peter Delaney of Optiscan Pty., Ltd., Australia, for their assistance. This research project was supported by a grant from the Singapore Bioimaging Consortium (SBIC RP C010/2006).

References
42. T. J. Cullip and U. Neumann, Accelerating Volume Reconstruction with 3-D Texture Hardware, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA (1994).


