ABSTRACT

Title of Dissertation: OPTICS AND COMPUTER VISION FOR BIOMEDICAL APPLICATIONS

Bohan Wang, Doctor of Philosophy, 2018

Dissertation directed by: Professor Yu Chen, Fischell Department of Bioengineering

Bioengineering is at the cross sections of biology, clinical technology, electrical engineering, computer science and many other domains. The smooth translation of domain technologies to clinics is not just about accuracy and practicality of the technology. It also has to take into account the accessibility (cost and portability), the patients' comfort and the ease to adapt into the workflow of medical professionals. The dissertation will explore three projects, (1) portable and low-cost near infrared florescence imaging system on mobile phone platform, (2) computer aided diagnosis software for diagnosing chronical kidney disease based on optical coherence tomography (OCT) images and (3) the tracking and localization of hand-held medical imaging probe. These projects aim to translate and adapt modern computation hardware, data analysis models and computer vision technologies to solve and refine clinical diagnosis applications. The dissertation will discuss how the translation, tradeoffs and refinement of those technologies can bring a

positive impact on the accuracy, ease of conduct, accessibility and patients’ comfort to the clinical applications.

OPTICS AND COMPUTER VISION FOR BIOMEDICAL APPLICATIONS

by

Bohan Wang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Yu Chen, Chair
Professor Min Wu
Professor Behtash Babadi
Professor Yang Tao
Professor Julius Goldhar

Acknowledgement

I would like to appreciate my parents' and grandparents' support for both my undergraduate study and graduate study at University of Maryland, College Park. Without the solid support and foundation from my family, I could never be who I am today.

I would like to appreciate my graduate academic advisor Professor Yu Chen for giving me the academic freedom to explore the research topics that truly drive me. The academic research environment Professor Chen is able to provide me and the rest of the team is sincerely amazing. I enjoyed every single moment during my research here at University of Maryland and at Professor Chen's lab. I am constantly motivated to make progress every moment here at Maryland.

I would like to appreciate my undergraduate and graduate advisor Professor Julius Goldhar. Professor Goldhar has been my advisor since 2011 while I was an undergraduate student at University of Maryland. Under his advisory, I had the opportunity to work with many great researchers both at University of Maryland and at Laboratory of Physical Sciences. These past experiences have always been inspiring me to conduct better research and motivating me to innovate. When I think back, I could never feel more fortunate the moment I met Professor Goldhar at his circuit theory class, which led to all the opportunities afterward; without him, I would never be able to go so far.

Finally, I would like appreciate Dr. Anuja Sonalker for giving me the opportunity to work closely with her during the last year of graduate study at University of Maryland. Her abundant industry research experience and insights helps me grown so much in a short period of time. Her mentoring during my last year of graduate school had me prepared for the career after graduation. I could never see another person from the industry that would share their knowledge and wisdom so generously.

Last but not least, I would like to thank my lab members, Yi Liu and Qinggong Tang. Those helps and favors I got from you are small but they do add up. I would never forget the good time we worked together as a team under Professor Chen. I wish both of you the best of luck in your future works.

Table of Contents

Chapter 1 Introduction

Problems and Motivations

Modern medical diagnosis at the core consists of two stages, observation and deduction. Observation is to capture a set of measurable symptoms that are related to a certain illness. Whether a symptom is observable depends on the capability of the medical instruments, which convert and amplify the symptoms' physics into what human observers could perceive. Sometimes a symptom can be observed by more than one type of instrument. For example, in the case of measuring heart rate, stethoscope amplifies the sound while photoplethysmography (PPG) measures and digitizes the volumetric change in the microvascular bed of the tissue [1]. More capable medical instrument would enable medical professionals to detect various symptoms more accurate. More affordable medical instruments would enable more patients to have access to the medical services.

Deduction is to quantify the existence and severity of the illness by analyzing the observation data. Nowadays, deduction can be carried out (1) completely by a professional medical doctor, (2) completely by machine or (3) by the combination of the two. Till today, most of the medical diagnosis deductions are still carried out completely by medical doctors; this procedure has been done for centuries and is well accepted by the population. In the recent ten years, many consumer electronics start to have software that is capable to carried out certain relatively simple deductions. For example, wearable fitness trackers nowadays are able to measure and detect heart disease [2]. In addition, many on-going researches are aiming to diagnose complex illness from a big-data approach completely based on machines, such as the research of using IBM Watson supercomputer to perform cancer diagnosis [3]. The marriage between traditional physician-based diagnosis and the latest machine-based diagnosis would be the most practical and realistic solution. It is called computer-aided diagnosis (CAD). First, CAD would reduce time and labor that medical professionals have to spent on simple and redundant portion of the diagnosis procedure. Secondly, CAD provides a diagnosis from the big data and statistic perspective while the physicians can provide a diagnosis from the traditional clinical practice. Together, human and machine could not only boost the final diagnosis accuracy but also could act as a fail-safe for each other. Last but not least, with medical professionals in the loop, patients would still have the inter-person bond and trust with a human physician, which is a critical element in health care that would affect many patients' decisions [4]. Computer software that contains sophisticated logistics, statistics and visualizations would be an invaluable tool for medical professionals to acquire secondary diagnostic output and visual insight from large sets of observation data.

Accuracy, *ease of conduct*, *accessibility* and *patients' comfort* are among the most crucial elements of modern medical diagnosis. Bioengineering is at the cross sections of biology, clinical technology, electrical engineering, computer science and many other domains. The translation of domain technologies to clinics is not only about accuracy and practicality of the technology. It should also take into account the accessibility (cost and portability), the comfort and ease of use for both patients and medical professionals and smoothness of technology translation for the medical professionals. This dissertation explored three projects that aims to translate and refine modern computation hardware, modern data analysis models, modern computer vision technologies from electrical engineering and computer science to adapt and solve clinical diagnosis problems. This dissertation will discuss how the translation and various tradeoffs and refinement of the technologies could bring a positive impact on the accuracy, ease of conduct, accessibility and patients' comfort to those clinical applications. The three projects are (1) designing and characterizing a portable near-infrared fluorescence imaging device, (2) designing a computer aided diagnosis software for chronical kidney disease based on optical coherence tomography and (3) augmenting optical coherence tomography data by imaging probe localization.

Contributions

The dissertation author is the major contributor, creator and project leader of all projects and content discussed in the dissertation.

In designing and characterizing a portable near-infrared fluorescence imaging device,

(1) the author designed, constructed and characterized a conventional NIRF imaging system as a reference of the industry-grade NIRF devices that are currently being used and sold on the market.

(2) the author designed and fabricated a set of various fluorescent medical imaging phantoms in order to create a test bed of different NIRF imaging devices.

(3) the author designed, constructed and characterized a smart-phone-based portable NIRF imaging system and compared against the conventional NIRF imaging system.

(4) the author performed animal-based testing for both systems in order to quantify the real-world usage of the systems.

(5) the author’s work in this project was presented in Food and Drug Administration (FDA) ‘s student poster session, in IEEE-NIH conference on Healthcare Innovations and Point-of-Care Technologies and in SPIE BIOS conference [5].

In designing a computer aided diagnosis software for chronical kidney disease based on optical coherence tomography,

(1) the author designed and developed a C++ Windows PC application with user interface that can locate and measure kidney disease symptoms on optical coherence tomography (OCT) kidney images at high accuracy.

(2) the author used the aforementioned application to analyze ten-thousands high resolution OCT kidney images and performed statistics on the software’s analytic results. The statistics result indicates that software is able to detect and distinguish the severity level of the kidney disease.

(3) this work was published in Journal of Biomedical Optics [6].

In augmenting optical coherence tomography data by localization of imaging probe,

(1) the author designed an augmented conventional handheld OCT imaging probe with a stereo-camera mounted to the probe body.

(2) the author developed a C++ software application that localizes the handheld imaging probe using the stereo-camera and visualize the captured OCT 2D images in 3D virtual space.

(3) the publication of this work is in pending.

Thesis Organizations

Each project is assigned to one chapter. Chapter 2 discusses the project of designing and characterizing a portable near-infrared fluorescence imaging device. Chapter 3 discusses the project of designing a computer aided diagnosis software for chronical kidney disease based on optical coherence tomography. Chapter 4 discusses the project of augmenting optical coherence tomography data by localization of imaging probe. Chapter 5 concludes the dissertation as a whole and discusses the possible future work in the light of this dissertation.

Chapter 2 Portable Near Infrared Fluorescence Imaging System

Introduction

The revolution of smart phone platform in the recent ten years has changed many aspects of daily life. Many medical diagnosis applications, such as phone-based oximetry [7] and ophthalmoscope, have since been successfully ported to the mobile platform. Medical diagnosis applications are drawn to the smart phone platform for many reasons. Smart phones are gradually becoming a commodity as the average cost of a smart phone decreases every year. The increasing adoption rate of smart phones means medical diagnosis applications on this platform are widely accessible globally. With the continuous development of mobile phone processors, the computation capability of average smart phone now can rival that of desktop computers from 2 to 3 years ago *but* in a much smaller form factor. Such portability makes smart phones more suitable for point of care diagnosis in situations such as rural and global health, pandemic response, military and first-responder and medical countermeasures. Last but not least, smart phones are equipped with various long-range communication capabilities, including cellular network, Wi-Fi network, and various short-range communication capabilities, including Bluetooth and NFC (Near-field Communication). With various communication/networking capabilities, smart phones are capable to communicate the diagnosis result under various network connectivity environment.

Fluorescence imaging is a medical imaging technique that has many clinical applications. In treatment, fluorescence imaging surgery guidance present fluorescence-labeled structures to surgeons to perform accurate resection [8]. In diagnosis, retinal angiography and lymphatic imaging are both derivatives of fluorescence imaging [9], [10]. Near-infrared fluorescence (NIRF) imaging is a subdivision of fluorescence imaging that is specifically suitable tissue-based imaging. The 650 nm to 900 nm NIR biological window of tissue, allowing high penetration depth and minimal auto-fluorescence, make NIRF imaging suitable for application, such as ophthalmic imaging and intraoperative visualization of vessels/perfusion and lymph nodes (non-specific) as well as tumors (targeted molecular imaging). Minimal research has been found on near-infrared fluorescence (NIRF) imaging on mobile phone platform. Traditional NIRF imaging requires expensive and desktop-grade instrument. A smart phone based solution could lower the cost, increase the accessibility and portability of NIRF imaging. The purpose of this research is to *advance the mobile health technology for NIRF imaging*. In addition, there has been no standardized methods to quantify the imaging qualities of NIRF devices. As part of the research for the Food and Drug Administration (FDA), this study also aims to provide *a standardized performance quantification procedure of the NIRF devices*.

Design of NIRF Imaging Systems

Fluorescence Contrast Agent and Indocyanine Green (ICG)

Fluorescence dyes are usually used as contrast agent to create or enhance fluorescence image. To emit fluorescence light, a light source in a specific wavelength range is necessary in order to excite a dye. An excitation wavelength range and an emission wavelength range are associated with each fluorescence dye. Fluorescence occurs when an orbital electron got excited to a higher energy state then relaxes back to its original states (Figure 1). The emission photon energy is usually less than the excitation photon energy due the simultaneous generation of heat; the emission wavelength is in general slightly longer than the excitation wavelength.

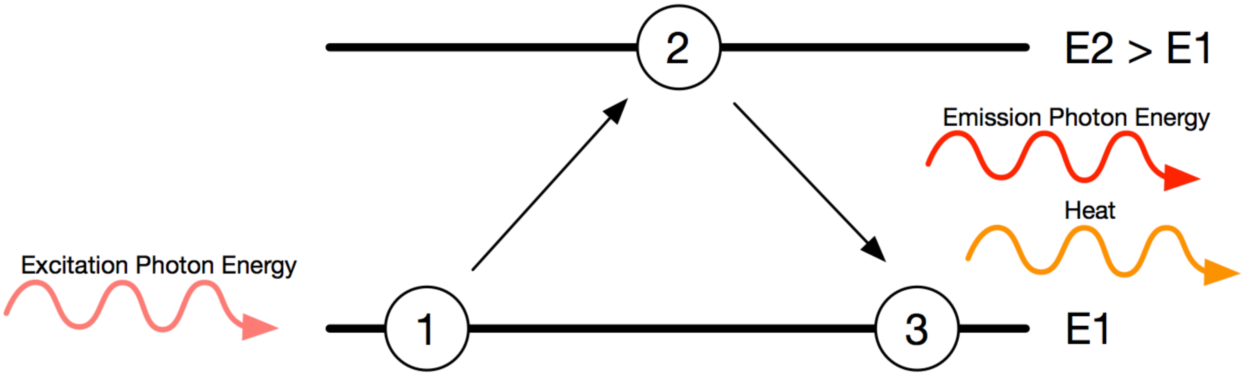


Figure 1. the excitation and emission process of fluorescence

For NIRF imaging for biological tissue, in order to achieve low loss of both excitation light and emission light, a suitable NIRF contrast agent should have both excitation and emission wavelengths within the 650 nm to 900 nm NIR biological window. In addition, it should be nontoxic and harmless for clinical usage. One of the very few fluorescence dyes that meets those requirement is indocyanine green (ICG). ICG is a well-established NIRF imaging contrast agent with wide clinical use, for example, in retinal angiography and intraoperative vascular imaging. The fluorescence properties of ICG in an aqueous solution with 2.5 µg/ml were characterized. Figure 2(a) shows the absorption spectrum of ICG measured with spectrometer. The figure shows that the peak absorption is at 780 nm. Figure 2(b) shows the excitation-emission matrix measured using a high sensitivity spectrofluorometer (QuantaMaster, Photonic Technology International, New Jersey, United

States) with dual-monochrometers for excitation and emission. An excitation-emission matrix (EEM) presents the emission intensity of a fluorophore at each excitation-emission wavelength pair. The result shows that the maximum emission intensity occurs at 780 nm excitation wavelength and 830 nm emission wavelength. The matrix also indicates that ICG has a broad emission spectrum ranging from 750 to 850 nm. Due to the wide clinical usage of ICG in NIRF imaging, we conducted the majority of the study specifically with ICG as the contrast agent.

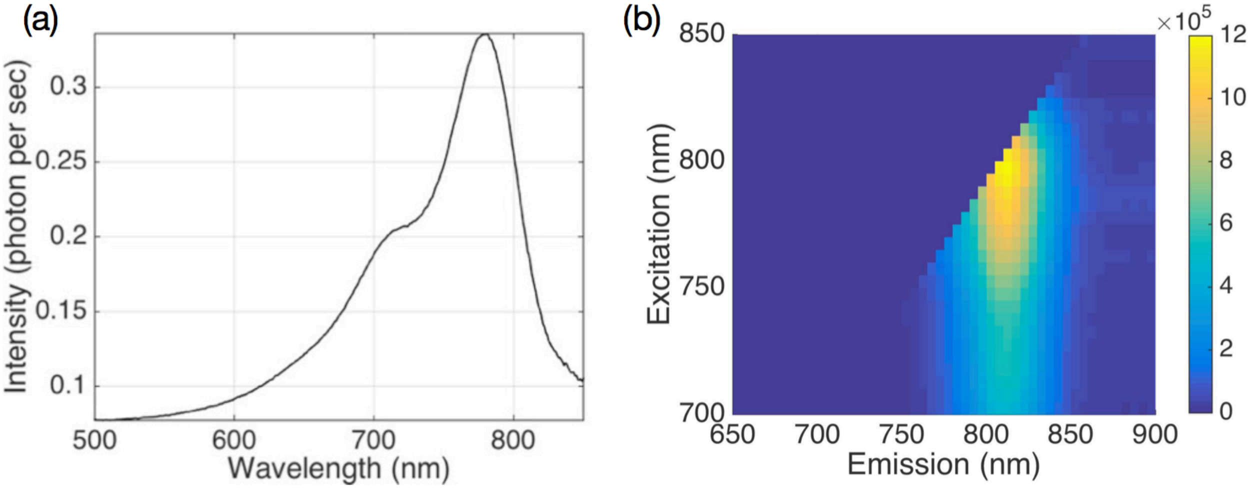


Figure 2. (a) ICG absorption spectrum (b) ICG excitation-emission matrix

Development of Desktop-Grade NIRF Imaging System

To create a reference point of the traditional industrial NIRF imaging system used in clinics, a professional desktop NIRF imaging system was built (Figure

3). In the system, a charge-coupled device (CCD) camera (Alta Series, Andor Technology, Belfast, Northern Ireland) is used for image capturing. This CCD camera has a resolution of 1600 x 1200 and 16-bit pixel depth. The lens of the camera and the distance of the camera to the imaging stage are adjusted (65 cm) so the the camera has a 10 cm by 10 cm field of view of the imaging stage. NIR LED excitation light source centered at 780 nm is used to match the absorption spectrum of ICG. Beam expander is installed onto the LED and the distance from the LED to the imaging stage is adjusted to achieve a beam spot size of 10 cm in diameter at the imaging stage. Maximum beam power of 2 mW/cm^2 is achieved at the imaging stage. 800 nm short-pass light is installed to the light source and 825 nm long-pass filter is installed to the CCD in order to prevent excitation light from mixing with the fluorescence signal when reaching the camera.

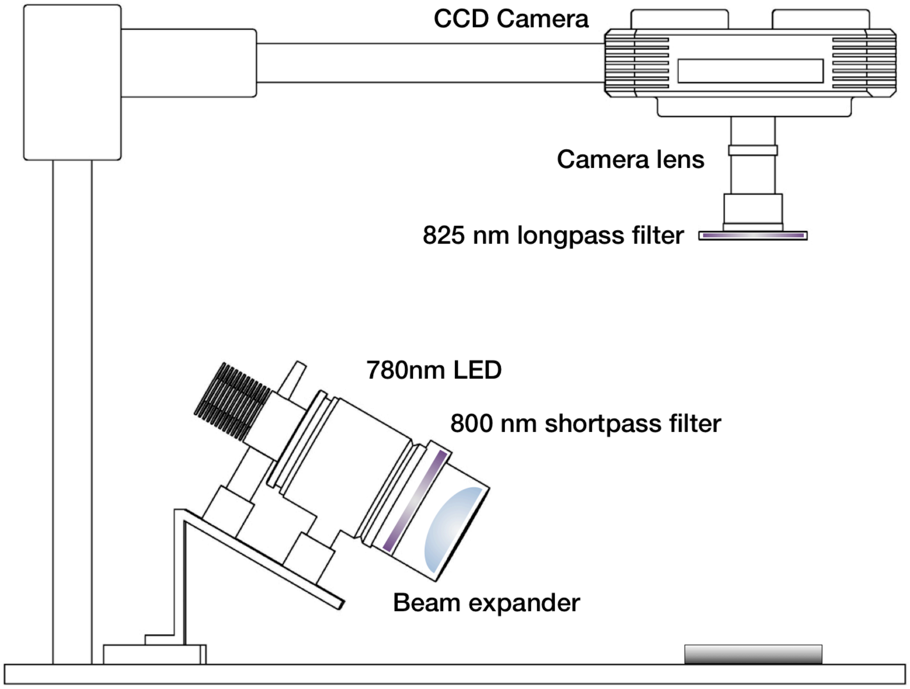


Figure 3. Desktop NIRF Imaging System

Development of Smart Phone Based NIRF Imaging System

A smart phone CMOS camera is used to replace the expensive CCD camera from the previous system (Figure 4). CMOS camera is intrinsically capable to sense NIR signal. In consumer CMOS camera, NIR filters are added to the CMOS sensor in order to simulate the photo sensitivity spectrum of human eyes. To enable the NIR capability of the smart phone camera, the NIR filter on the CMOS camera is removed. This CMOS camera has a resolution of 2448 x 3264. To isolate the fluorescence spectrum, the 825 nm long-pass filter is installed on the camera. The mobile phone is placed 12 cm away from the imaging stage to achieve the same field of view as the CCD camera.

Figure 4. NIR-Enabled Smart Phone

Biomedical Imaging Phantom Design and Fabrication

Imaging phantom is commonly used to evaluate certain performance of the biomedical imaging device. In order to characterize the two NIRF systems, multiple phantoms were designed and fabricated for different purposes. The objective is to create imaging phantoms that can become standardized tools to measure the performance of the NIRF imaging device. In addition, the imaging phantoms should be solid, reusable and stable for a long period of time in order to become part of the standardized performance quantification procedure.

Uniform fluorescence phantom

A uniform fluorescence phantom was fabricated to use with a non-fluorescent resolution target in order to evaluate the sharpness of the imaging system (Figure 5(a)). The phantom is made of an epoxy resin matrix with ethanol as a solvent, ICG as the fluorophore and titanium dioxide (TiO_2) as the scattering constituent. The scattering level was by design to be much higher than biological tissue in order to achieve a uniform fluorescence emission with minimal shading in bright regions of the acquired image. Epoxy resin as the solvent produces a solid and mechanically stable phantom. A glass resolution target chart with chrome bars and numbers (2” x 2” positive 1951 USAF, Edmunds Optics Inc., Barrington, NJ) was placed atop the phantom with the chrome surfaces adjacent to the Epoxy-ICG phantom and imaging was performed with both NIR methods (Figure 5(b)).

Photo-bleaching of the uniform fluorescence phantom is characterized to determine the reusability of the phantom. Photo-bleaching occurs when the contrast agent is under excitation light exposure for long period of time. If photo-bleaching effect is substantial over a short period of time, the target is not reusable or suitable for long-time experiment. To measure the photo-bleaching property of the phantom, aqueous ICG solution, as the reference, and ICG-resin phantom were exposed under 10 W/m^2 of NIR excitation light for 2 hours. Fluorescence emission intensity of both targets were measured repeatedly within the 2-hour experiment. The result showed that over the period of 2 hours, the emission power of ICG-resin mixture has reduced less than 5 percent of the original emission power while that of aqueous ICG solution has reduced more than 20 percent (Figure 5(c)).

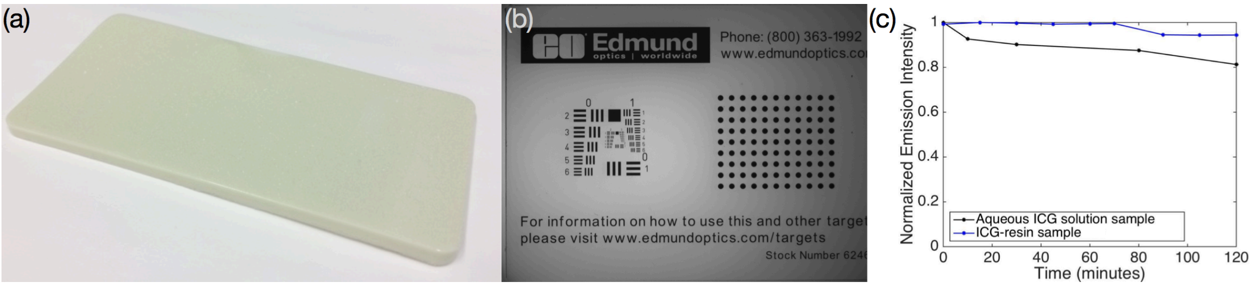


Figure 5. (a) uniform fluorescence phantom (b) NIRF image of glass resolution chart placed on top of the phantom (c) characterization of photo-bleaching effect

3D Printed Channel Phantom

Phantom with channels located at various depth was fabricated using photo-polymerization 3D printing (Figure 6(a)). By temporarily injecting aqueous ICG solution into the channels, phantoms with 3D channels could be used to measure the penetration depth of the NIRF imaging systems. Penetration depth for NIRF imaging system is affected by the combination of excitation light spectrum, fluorescence contrast agent emission spectrum and the sensitivity of the image sensor. The phantoms were fabricated with photo-polymerization 3D printer Objet 30 (Stratasys Ltd.), which is capable to produce prints with fine and accurate structures. Extrusion deposition and photo-polymerization are the two major 3D printing technologies that are widely used nowadays. Extrusion deposition creates layers by depositing small beads of material, such as acrylonitrile butadiene styrene (ABS), to the surface, where the material solidifies immediately. Photo-polymerization forms layers by drawing a laser beam on the surface of liquid photopolymer, which hardens on light exposure. Our group's previous study compared the printing quality between ABS-based extrusion deposition and photo-polymerization for phantom development [11]. In ABS phantoms printed with extrusion deposition process, internal leakage showed up after fluid had been injected into the channels. In comparison, the photo-polymerization technology can print water-tight channels that can hold fluid, such as aqueous fluorescence solution. In addition, HTTP resin, the 3D printing material used during the photo-polymerization process, is proven to have absorption spectrum and scattering spectrum that are similar to biological tissue [12].

Fluorescence property of the 3D printing material was measured to ensure the bulk material (HTTP resin) does not fluoresce in the same spectrum as the ICG contrast agent. EEM of the HTTP resin is measured with spectrofluorometer (Figure 6(b)). The absence of emission signal in the EEM indicates that the photopolymer material used in 3D printing does not have fluorescence excitation/emission spectrums that overlaps with that of the ICG. NIRF imaging phantoms made of HTTP resin photopolymer do not produce background fluorescence, which would otherwise decrease the contrast of the target image.

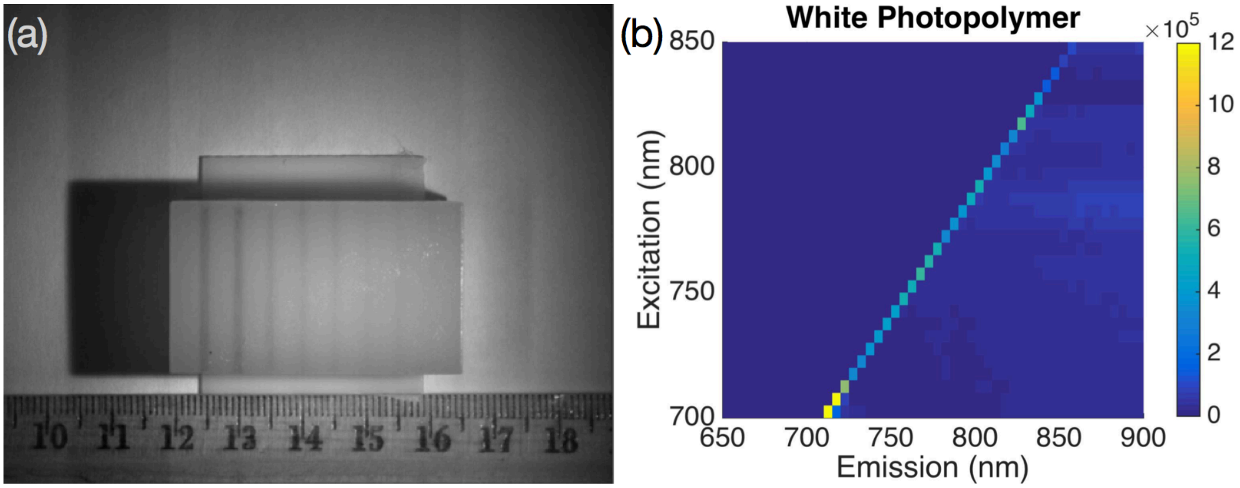


Figure 6. (a) 3D printed phantom with channels at various depth (b) excitation-emission matrix (EEM) of HTTP resin

3D printed bio-mimic phantom

The bio-mimic phantom (Figure 7) fabricated *from our group's previous study* was also included as part of the study [13]. This bio-mimic phantom was printed using photo-polymerization process based on a 3D model derived from human retina image. The phantom is 5cm x 5cm in dimension with 0.75 mm diameter channel at 0.75 mm depth.

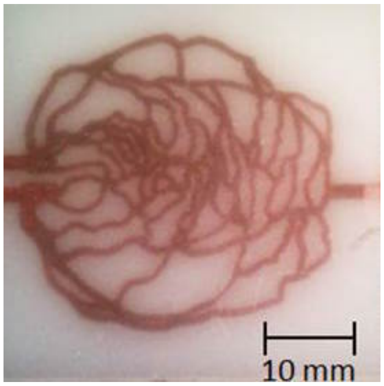


Figure 7. bio-mimic phantom fabricated from a previous study [13]

Characterizing Performance of NIRF Imaging Systems

The performances of the two NIRF imaging systems were characterized using the fabricated phantom. The performance measures include sharpness, signal-to-noise ratio (SNR) and penetration depth. In addition, the imaging result of the biomimetic phantom is compared to that of a real biological murine tissue to access the level of truthfulness of the biomimetic phantom to simulate or replace the animal testing in long term.

Sharpness

Sharpness is the ability of NIRF imaging system to resolve small details of the target object. Standard resolution test chart (USAF 1951, chrome on glass) was set on top of a flat homogenous fluorescence phantom was imaged with both systems. USAF test chart is composed of groups of three bars with spatial distance between the bars gradually decreases. Contrast transfer function, depicting the relation of contrast and spatial frequency, can be quantified by analyzing the image of USAF chart. Figure 8 shows the USAF images captured by both system and the CTF curves of the two systems comparing side by side. Michelson contrast is used as the contrast measurement.

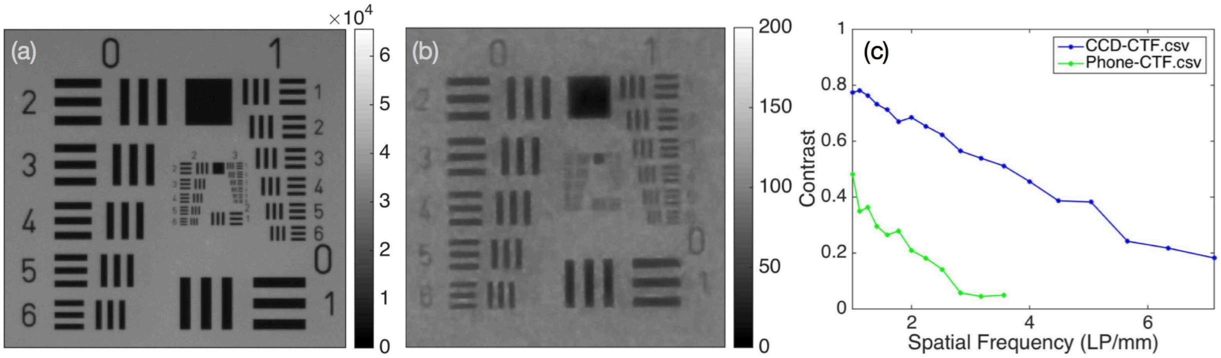


Figure 8. (a) USAF NIRF Image from CCD (b) USAF NIRF Image from Smart Phone CMOS (c) CTF curve of the two systems

Signal-to-Noise Ratio

SNR of the two systems are measured using the uniform fluorescence phantom without the USAF target. In imaging system, SNR is defined as the mean of the signal divided by the variance of the signal. The uncertainty of the pixel values arises from the extrinsic photon noise and from the intrinsic dark current noise of the sensor material and electronic noise from the circuit. The photon noise is independent of the sensing electronics and caused by the natural statistical variation of the photon arrival rate. The dark current noise is caused by the thermally generated electrons inside the sensor material. The electronic noise is caused by various electronics in the camera sensor, such as signal amplifiers and digitizers. However, SNR of the imaging system could also be adjusted by customizable settings of the system, such as exposure duration. To perform a fair comparison of the two systems, both systems are set to their default factory settings. Videos of the uniform fluorescence phantom were captured by both systems in a dark undisturbed environment with only excitation light turned on. The mean, variance and SNR of the fluorescence signal for each pixel over the time period of the video was determined. The SNR of the system is the average SNR computed from all the pixels. The SNR of the CCD system and the smart phone camera are 500 and 61 respectively. The cost of the CCD system justifies its higher SNR. The CCD systems has its dedicated cooling system that lowers the temperature of the sensor material that in turn lowers the dark current noise. In addition, the CCD system has a 16-bit digitizer comparing to the 8-bit digitizer in the smart phone camera system.

Penetration Depth

Aqueous ICG solution was injected into the phantom with channels located at various depth. The phantom was imaged with both systems. The CCD system produced a high contrast image with the contrast of the channel structures decreases as the channels reside deeper. In comparison, the CMOS camera on the phone produce lower contrast image. Channels at 3.2 mm and 3.7 mm depth is barely resolved. In addition, the auto focusing software of the phone camera constantly failed because of the low fluorescence light intensity the sensor can detect. The images and the quantification result are shown in Figure 9.

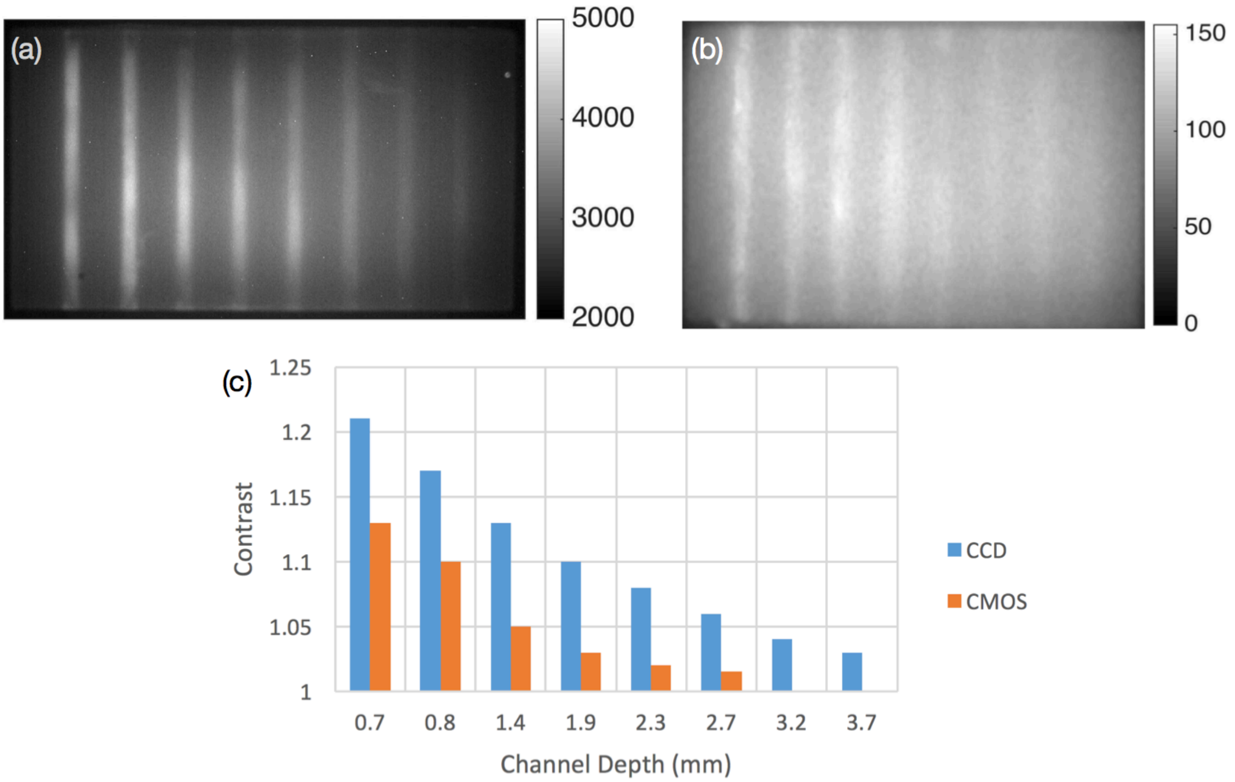


Figure 9. (a) CCD image of channel phantom (b) CMOS image of channel phantom (c) Contrast vs. channel depth for both systems

Biomimetic Phantom and Animal Experiment

While the NIRF imaging quality of the phone camera has been benchmarked being inferior than the CCD camera, the *feasibility* of the phone camera for biological NIRF imaging is still uncertain. In many cases of NIRF imaging for point of care diagnosis, such as cancer detection, achieving high resolution image of the structure is not necessary or even considered excess. If the phone-camera-based NIRF imaging at point of care could provide an accurate binary conclusion fast and at a low cost, it will still be a valuable solution. Two experiments, 3D biomimetic phantom NIRF imaging and an animal NIRF imaging, were conducted to evaluate the feasibility of phone camera based NIRF imaging of biological tissue.

3D printed human retina vascular channel phantom was injected with ICG aqueous solution and imaged with both systems (Figure 9). CCD camera produces a high-quality image with all the vessels well-resolved; the smart phone camera was able to capture the major vascular channels while the smaller vascular structures cannot be resolved from the smart phone NIRF image.

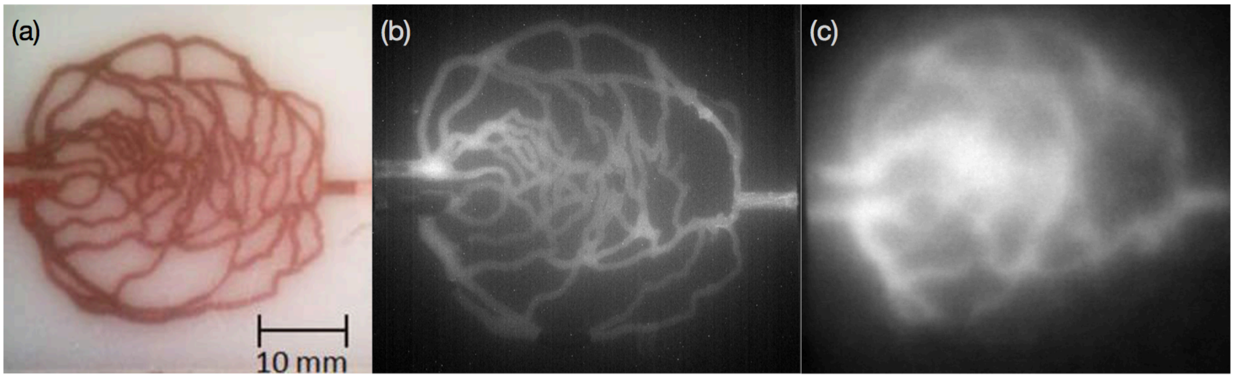


Figure 10. (a) human retina vessel phantom (b) CCD NIRF image (c) phone camera NIRF image

In the animal experiment, ex vivo imaging was performed on a mouse with ICG (2.5 μg/ml) injected in the tail vein and femoral vein region. ICG dispersal was limited to the region near the injection site due to the lack of perfusion. The animal target was imaged using both systems (Figure 11). The phone camera based imaging result can clearly show the locations where the ICG resided.

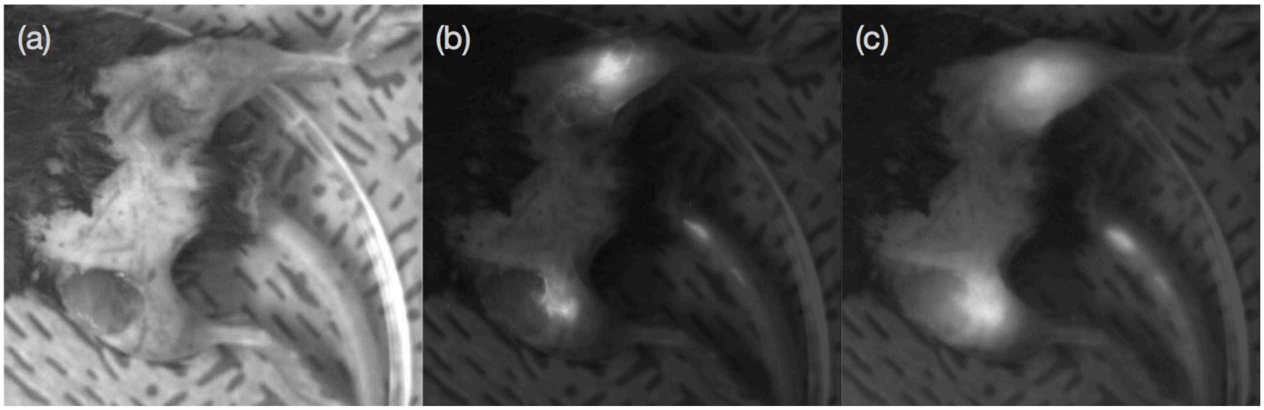


Figure 11. (a) RGB image of the animal (b) CCD imaging result overlaid on RGB image (c) phone camera imaging result overlaid on RGB image

Discussion

The NIRF imaging performance of a conventional CCD system and a mobile phone camera was compared. While in many technical aspects, the mobile phone camera has inferior NIRF imaging performance than the conventional CCD system. The cost of the mobile phone based system is one tenth of the cost of the CCD system. The cost difference could increase further as the smart phone market becomes more competitive.

Cost and portability are two important factors to deliver effective and affordable healthcare to developing regions and countries. Many medical diagnosis technologies are developed in response to the needs of the patients and medical professionals in the developed countries. These systems typically require highly maintained environments and operators who need to go through expensive trainings. However, the majority of the world’s population only have access to poorly resourced health care facilities with minimum clinical laboratory infrastructure [14]. The goal of point of care diagnosis is to bring diagnosis systems and professionals who are capable to operate the systems to the patients, especially those who do not have access to well-equipped clinical facilities. As more patients have access to the diagnosis systems, the more likely they would receive diagnosis at early stage, resulting more effective treatment and prevention of the disease [15].

Although mobile phone camera based NIRF system is unable to produce NIRF image at the same high quality of traditional CCD based system, its cost-effectiveness and portability justifies that mobile phone camera based NIRF system is well suited for point of care diagnosis of many NIRF imaging applications.

For example, NIRF imaging is applied in non-invasive imaging of the lymphatic system. Abnormal lymphatic system may be associated with a wide spectrum of diseases, including cancer metastasis, venous disease and complications of wound healing. Abnormal lymphatic system can be detected by intradermal administration of ICG and observing the appearance of the lymphatics in the medial arm[7]. The mobile phone based system would be sufficed for such application because of the relative large structure of the lymphatics.

Late-fluorescence mammography is another potential application of mobile phone based NIRF imaging. Due to the intravascular difference between normal breast tissue and carcinomas tissue, the fluorescence clearance capability of those tissues are different. After administrated ICG has been largely cleared from the blood by the liver, carcinomas tissue would produce a high contrast NIRF signal over the homogenous background[16]. To diagnose the potential existing carcinomas tissue at point of care, it is not necessary to capture the high definition NIRF mammography. Mobile phone camera based solution could potentially become a low-cost solution to increase the population to access mammography scanning.

Chapter Conclusion

In this study, NIRF systems and medical imaging phantoms were designed and built in order to compare various optical characteristics of a conventional CCD-based NIRF system and a noble mobile phone camera based NIRF system. The study has shown that mobile phone camera based NIRF system is able to provide enough sensitivity at a much lower cost to perform NIRF based diagnosis in point of care applications. The possibility of performing NIRF based diagnosis and preventive scanning on a mobile phone device would enable more people, especially those at developing regions, to gain access to NIRF based diagnoses, such as NIRF mammography and abnormal lymphatic systems detection.

Chapter 3 Computer Aided Diagnosis of Chronical Kidney Disease

Introduction

Chronical kidney disease (CKD) is a medical condition characterized by the progressive loss of kidney's function [17]. Nowadays, 26 million U.S. patients suffer from CKD condition while millions more are potentially at risk [18]. In estimate, more than half a million U.S. residents have end-stage renal disease (ESRD), which is associated with high mortality rates (163.8 deaths per 1,000 patients per year) and a huge economic burdens (>$30 billion per year) [18]. CKD is classified into five stages of increasing severity [19]. ESRD is the last stage of CKD when dialysis or transplant is needed to stay alive. The stages of CKD or the progressive loss of kidney function are mainly based on measured or estimated glomerular filtration rate (eGFR) [20]. Estimated GFR (eGFR) is obtained by blood test including serum creatinine level, together with age, sex, and sometimes other information [20]. Pathology can provide additional information about the microstructure of declined kidney function by viewing tubular atrophy, glomerulosclerosis, and interstitial fibrosis.

However, there are significant artifacts associated with excision biopsies and immersion fixation procedures.

Optical coherence tomography (OCT) [21]-[24] has the advantage to allow the analysis of the microscopic structure of kidney in a non-destructive manner. OCT has been used to image thermal tissue damage to the rat kidney resulting from laser ablation [25]. OCT's capability to resolve renal corpuscles and uriniferous tubules was first demonstrated on rat kidney *ex vivo* using high-resolution time-domain OCT[24]. With the advent of high-speed Fourier-domain OCT technology, 3D imaging of renal microanatomy *in vivo* was enabled [26]. The kidney microstructures prior to, during, and following exposure to renal ischemia can be observed in real time. In a recent study, Andrews *et al.* utilized OCT to visualize the characteristic histopathologic changes on aging rat kidneys *in vivo* [27]. With the onset of severe proteinuria at 10-12 months of age, OCT reveals tubular necrosis/atrophy, interstitial fibrosis, tubular dilation, and glomerulosclerosis [27]. With a further deterioration in kidney function at 16-18 months of age as indicated by rising creatinine levels, OCT reveals more extensive interstitial fibrosis and tubular atrophy, increased tubular dilation with cyst formation and more sclerotic glomeruli [27].

Wierwille *et al*. investigated the feasibility of Doppler OCT (DOCT) to image kidney microcirculation, specifically, glomerular blood flow [28]. Normal blood flow as well as the effects of acute mannitol and angiotensin II infusion has been observed and quantified. Using ultrahigh-sensitive optical microangiography (OMAG), Zhi *et al*. demonstrated highly sensitive imaging of renal microcirculation *in vivo* [29]. Normal peritubular capillary microcirculation as well as the changes in response to renal ischemia and reperfusion can be monitored and quantified using this method.

Since OCT has deeper penetration depth than confocal microscopy, it is able to penetrate the renal capsule surrounding human kidneys thereby enabling the characterization of renal tubules, glomeruli, and cortical blood vessels in human kidneys [30]-[32]. Using image-processing algorithms, key microstructural parameters such as tubular lumen diameter can be automatically quantified [32]. *In vivo* imaging of human kidney has been reported during kidney transplantation procedures [33], and clinical studies have indicated that the openness of tubular lumens observed by OCT has a strong correlation with the post-transplant recovery of renal function [34]. These results suggest that OCT may be a useful tool in intra-operative monitoring and evaluation of transplant kidneys for predicting post-graft function after ischemia-reperfusion injury. OCT has been applied to other clinical studies including assessing the morphological features of the endothelial and vascular injury induced by catheter-based renal nerve ablation (RNA) [35], and differentiating between normal renal parenchyma and renal cell carcinoma (RCC) [36], [37].

Computer aided diagnosis (CADx) systems can interpret medical images and provide an advisory diagnosis decision for medical doctors. Those systems aim to both decrease the diagnosis error rate and reduce decision making time [38]. A great amount of research in CADx systems has been conducted for different types of imaging modalities and imaging targets. CADx computation pipelines often include region proposing, feature extraction, and classification. Liu et al. [39] detected epidural masses on CT scans by region proposing with K-mean clustering, extracting a collection of texture features, and trained a support vector machine (SVM) classifier based on the texture features. Farag et al. [40] detected entire pancreas organ location in CT scans by region proposing with Simple Linear Iterative Clustering (SLIC) algorithm, extracted Dense Scale Invariant Feature Transform (dSIFT) features, and trained a random forest classifier based on dSIFT features. Jerebko et al. [41] detected colonic polyps in CT scans by region proposing with manual seed selection, extracted 12 hand-coded features, such as area, sphericity, surface pixel intensity mean, and trained a neural network classifier based on these 12 features. For OCT-related CADx system, Wan et al. [42] detected and classified breast tissues in OCT scans using a sliding window technique for region proposing, extracted local binary pattern (LBP) texture, and trained a neural network classifier based on the LBP texture feature vector. Qi et al. [43] detected and classified dysplasia in Barrett's esophagus in endoscopic OCT scans using a semi-automatic segmentation process based on global threshold, and extracted 6 texture features. They used principal component analysis (PCA) to obtain the top-2 significant feature spaces and found the classification threshold from the receiver operating characteristic (ROC) curve generated from 100 annotated training images. Qi et al. [44] detected colonic crypt morphology in *en face* image from OCT 3D scan using semi-automatic marker-based watershed segmentation, extracted 6 region features, such as area, density, eccentricity, and used PCA to obtain the top-2 significant feature spaces.

In recent years, deep convolutional neural network (ConvNet) has been increasingly adopted by CADx applications. With proper convolutional neural network architecture, the original computation pipeline of CADx, which consists of region proposing, feature extraction, and classification, can be condensed into region proposing and ConvNet classification or even simply a ConvNet that is trained to calculate both the class and bounding box of the object. Roth et al. [45] detected lymph node from CT scan by using a preliminary software to generate 3D lymph node candidate segments and a 5-layer ConvNet to classify the 3D segments based on 2D image generated from a random viewpoint of the 3D segments. Roth et al. [46] detected pancreas in CT scans by using SLIC algorithm for region proposing and deep convolutional neural network to classify the SLIC segmentation. Liu et al. [47] detected colitis in CT scans by using a "selective search" algorithm for region proposing, a pre-trained Alexnet on PASCAL 2007 dataset (natural images instead of medical images) for feature extraction, and an SVM for classification. Esteva et al. [48] classified 757 types of skin cancer captured on RGB camera using Google Inception Convolutional Neural Network architecture that was end-to-end trained on 1.4 million annotated skin-cancer images.

In this study, we used OCT to evaluate CKD in a murine model induced by intravenous Adriamycin(ADR) injection into Munich-Wistar rats [49]. We present a CADx system based on convolutional neural network to automatically detect and quantify tubular diameter and hypertrophic tubule population from OCT images at several post-Adriamycin induction time points.

Background

Optical Coherence Tomography

Optical Coherence Tomography (OCT) is an established 3D tissue imaging technology that can acquire micrometer-resolution image for tissue structure located at as deep as 500 micrometers. The word "tomography" means visualizing the cross section of a solid object. OCT represents the cross section of an imaging target in terms of the reflectance of its composition materials at different depth location. In comparison to a pixel in traditional photography, an axial scan or in short A-scan is the fundamental imaging unit of OCT. The intensity of A-scan is proportional to the reflectance depth profile at the probed location. Multiple A-scans can be acquired by moving the imaging probe in a single direction with tiny step size. Combining these consecutive A-scans results a cross-sectional image called B-scan. Similarly, a 3D scan called C-scan can be acquired by combining consecutive B-scans.

The signal source of OCT commonly uses an optical source with a broad spectrum such as a broadband laser or super-luminescent diode. Infrared light sources centered at 1030nm are common for imaging tissue because tissue has low absorption (therefore better penetration) at infrared spectrum. Beam given off by the light source is directed into a Michelson interferometer where the it is split into two paths. On one path, the wave incidents vertically onto a mirror and reflects back into the interferometer. On the other path, the wave incidents on the imaging target. On target surfaces at each depth, a portion of the beam is reflected back into the interferometer. The power of each reflected beam depends on the reflectance at each depth. However, the reflection beams overlap into one beam and is indistinguishable from each other by itself. In OCT nomenclature, the optical path to the mirror is called the "reference arm" and the optical path to the imaging target is called the "sample arm." The output of the interferometer is the interference beam between the reflection on the reference arm and the reflection on the sample arm.

The core of OCT is to recover an A-scan $I(z)$ from the interference wave. In term of A-scan recover methods, OCT technology is further classified into time-domain OCT (TDOCT) and Fourier-domain OCT (FDOCT).

Time-domain OCT

In TDOCT, an A-scan $I(z)$ is recovered from the time-domain interference signal. An optical power meter is placed at the output of the interferometer in order to record the signal.

To understand how TDOCT works, the concept of “coherence” needs to be addressed. *Interference occurs if and only if two waves are highly coherent.* In general, the word “coherence” describes the likeness between two waves or between a wave and a delayed version of itself (self-coherence). Wave is a function of both time and space; coherence is further classified into temporal coherence and spatial coherence in terms of if the delay is caused by the difference in time or the difference in space. Two waves are “perfectly coherent” if they share the same frequency and a constant phase difference (proportional to delay). In practice, ideal coherent waves are impossible to generate separately; coherent waves are generated by sharing the same light source. In addition, even with the same optical source, waves can only be highly coherent within some maximum amount of delay. This maximum delay in time or space is called the “coherence length.”

Coherence cannot be observed directly because electromagnetic wave oscillation in optical spectrum is too rapid to be registered by any optical instruments. Only optical power or optical spectrum is perceivable. However, we can infer coherence through observing interference visibility because interference only occurs between coherent waves. For example, in Young's double slits experiment, when the two slits are close to each other, highly distinguishable interference fringes can be seen on the screen. As the two slits are moved apart, the interference fringes become dimmer and eventually disappear. This example demonstrates the effect of coherence length on interference visibility. If the spatial delay between coherent waves exceeds the coherence length, the waves are no more coherent so that the interference fringes disappear. Same phenomenon can be observed in time domain. The optical wave from a broadband light source has a short coherence length. The wave is self-coherent only within a small-time delay. In the Michelson interferometer, only if the time difference between traveling through the two optical paths are small, interference occurs at the interferometer output.

Interference visibility can be quantified with an optical power meter. When two non-interfering waves overlap, for example, two flashlights pointed to the same spot, the power of the overlapped beam is the summation of each beam power (Equation 1a). When two coherent waves overlap, the waves interfere either constructively or destructively. So, the power of the interference wave is the square of the added waves, where it has an extra term depending on the two waves' phase difference (Equation 1b). The exact OCT interference power function is only a slightly complicated version of the one showing here.

$$I_{no\ interference} \sim |E_1|^2 + |E_2|^2 \qquad \text{(a)}$$

$$I_{interference} \sim (E_1 + E_2)^2 = I_{no\ interference} + 2E_1E_2\cos\left(\Delta\phi\right) \qquad \text{(b)}$$

Equation 1. Interference Equations

In order to measure the reflectance at each depth, the power of each reflection needs to be known. However, the reflections from each depth of the imaging target are overlapped into one single sample arm beam. In order to decompose them, TDOCT applies the core idea that interference occurs only within the coherence length. TDOCT uses a broadband light source such that it has a short coherence length in time domain. With short coherent waves, only the reflection from the depth, where optical path difference between the reference arm and the sample arm is within the coherence length, would interfere with the reference arm beam. *In brief, TDOCT can select which constituent of the sample arm reflection interferes with the reference arm reflection by changing the length of the reference arm.* By first matching the reference arm length to the sample arm length referring from the sample surface, and then increasing the reference arm length by small step size, the power meter can record a series of value in terms of reference arm displacement. The reference arm displacement corresponds to the sample depth. Equation 2 shows the power meter reading as a function of reference arm displacement z_R.

$$i(z_R) \sim i_{DC} + \int_{z=0}^{\infty} \sqrt{R_R R(z)}\, e^{-(z_R - z)^2 \Delta k^2} \cos(\, 2k_0(z_R - z)\,).$$

Equation 2. Reference arm optical power as a function of arm length

In Equation 2, z is the sample depth referred from the surface of the imaging target. The term k_0 is the central wavenumber and Δk is the bandwidth of the light source in the unit of wavenumber. The term $e^{-(z_R - z)^2 \Delta k^2}$ represents the effect that interference visibility diminishes to zero if the delay between the reference arm beam and portion of the sample arm beam is larger than the coherence length. The term R_R is the reflectivity of the reference arm mirror which should closely equal to 1. The term $R(z)$ is the reflectivity of the imaging target at depth z. This function is the A-scan recovered by TDOCT.

Fourier Domain OCT

The moving mirror mechanism required in TDOCT bottlenecks the OCT image acquisition speed. FDOCT invented around the early 2000s recovers the A-scan through its Fourier domain signal, $I(k)$, called spectral interferogram. A-scan is computed by applying inverse Fourier transform to the spectral interferogram. FDOCT scans the frequency domain instead of the time domain. Therefore, it discards the needs of the moving mirror mechanism. As a result, the acquisition speed limit of OCT can increase dramatically. There are two methods to recover the Fourier domain signal that are different in term of frequency domain scanning mechanism. One is *spectral domain* OCT (SDOCT) that uses a spectrometer to capture the output of interferometer. One is *sweep source* OCT (SSOCT) that uses a swept source laser as the light source and an optical power meter at the interferometer output. Both setups can recover the same spectral interferogram shown in Equation 3.

$$\begin{aligned}
&I(k) \sim \frac{1}{8} I_{DC} \\
&+\frac{1}{4} S(k) \int_{z=0}^{\infty} \sqrt{R_R R(z)} \cos(\, 2k(z_R - z)\,) \text{ "}Cross - correlation\text{"} \\
&+\frac{1}{8} S(k) \iint_{z_1 \neq z_2 = 0}^{\infty} \sqrt{R(z_1) R(z_2)} \cos(\, 2k(z_1 - z_2)\,) \text{ “}Auto - correlation.\text{"}
\end{aligned}$$

Equation 3. Interferogram Equation

In the equation, $S(k)$ is the spectrum of the light source. Fourier domain considers the light source as a weighted combination of independent sinusoidal plane waves. In such perspective, each plane wave interferes not only with the reference arm beam (the cross-correlation term) but also with the reflection beam originated from other depths in the imaging target (the auto-correlation term.) The A-scan can be computed by conducting inverse Fourier transform on the interferogram (Equation 4).

$$\begin{aligned} i(z) \sim\ & \frac{1}{8} i_{DC} \\ & + \frac{1}{4}\int_{z=0}^{\infty} \sqrt{R_R R(z)}\, \left[\gamma[2(z_R - z)] + \gamma[-2(z_R - z)]\right] \text{"}Cross - correlation\text{"} \\ & + \frac{1}{8}\iint_{z_1 \neq z_2 = 0}^{\infty} \sqrt{R(z_1)R(z_2)}\left[\gamma[2(z_1 - z_2)] + \gamma[-2(z_1 - z_2)]\right] \text{“}Auto \\ & \qquad - correlation.\text{"} \end{aligned}$$

Equation 4. A-scan computed from interferogram

In the equation, function $\gamma(z)$, called coherence function, is the inverse Fourier transform of the light source spectrum $S(k)$. Comparing to TDOCT, A-scan recovered by FDOCT contains an extra auto-correlation term which brings artifacts into the image. There are a few techniques such as phase-shift interferometry that can minimize the effect from the artifacts.

Chronical Kidney Disease

Chronic kidney disease (CKD) is a general description for kidney structure and function disorder caused by kidney injuries, high blood pressure, diabetes or family CKD history [19]. With more than 200 thousand cases each year in United States, CKD has become a common disease that requires a systematic approach for screening, early detection and supervision [50]. Furthermore, the number of people treated for end- stage renal disease (ESRD) is increasing worldwide [51]. The two general CKD diagnosis methods are (1) measuring kidney waste filtration capability and (2) observing the visual presence of kidney damage[52].

Glomerular filtration rate (GFR) score is the most commonly used scoring system to describe Kidney filtration capability. GFR score is calculated in terms of blood serum creatinine level test result, age, weight and gender. Serum creatinine is a chemical waste produced by the body that shall be filtered out by the kidneys. CKD is classified into five stages by GFR score. The most severe stage, stage five, called end-stage renal disease (ESRD), requires patient to go through dialysis or transplant in order to survive. For early stages CKD, there are effective prevention practices and treatments, including life style modification and pharmacological treatments that can slow down the progression of the disease. It is the most cost-effective option for patients that CKD is detected and treated at early stages [51]. However, despite the widely use of GFR as the gold standard to estimate kidney function, study has suggested that it is not reliable to detect CKD at early stages. It has been shown that a significant change of serum creatinine concentration in blood does not occur until nearly 50% of the kidney is nonfunctional. The rest of the functional kidney can adapt and work harder in order to compensate the loss of filtration units [53].

The basic function unit of the kidney is a nephron. A nephron is a tube-shaped object that extracts wastes from the body's cardiovascular system. The start of a nephron is called the glomerulus. The glomerulus consists of many capillaries where small particles such as nutrients as well as wastes can filter through and enter the nephron's proximal tubule. Proximal tubule reabsorbs nutrients back into the cardiovascular system and sends the wastes into the urinal system.

Examine the morphology of the kidney structure mentioned above with kidney histology images is a more direct and accurate way to detect CKD. If a nephron loses its filtration capability due to injury or diseases, the proximal tubule of the nephron would shrink and cause nephron atrophy. The rest of the functional nephron would expand and cause nephron hypertrophy in order to compensate the loss of functional nephrons [53]. The correlation between proximal tubule morphology and CKD progression and nephron function has been discussed extensively and been proved with histology images in many literatures [49], [53]-[57]. Although observing kidney morphology is an effective way to detect CKD in early stages, the current procedure requires biopsy with a slice of kidney sample for histology. Kidney biopsy or invasive imaging procedures are usually only applied to patients in whom a definitive diagnosis would result in a change in either treatment or prognosis [52]. In addition, kidney biopsy process is a demanding and time-consuming task; it includes (1) delicate surgery procedure to acquire the sample, (2) careful sample preservation with correct storage temperature and solutions, (3) applying different types of stain onto the sample for microscopy imaging and (4) meticulous assessment of the sample image by two experienced histopathologists in order to reduce the error during the assessment {Williams:2002ws}.

Artificial Neural Network for Images Classification

In a linear regression problem, linear model $y = ax + b$ is fitted to a series of experimental data point (x_i, y_i) in order to find the unknown coefficients a and b. In other words, the linear model is trained with training examples

$(x_i \in R, y_i \ \in R)$. After the coefficients are found, we can use this linear model to estimate the output y' for a given input value x'. The estimation should be accurate given that the test input is generated under the same condition as the training examples.

In comparison to a linear regression problem whose function model is predetermined, artificial neural network (ANN) is a universal function approximator. An untrained ANN is essentially a function with numerous unknown coefficients. They are called weights in ANN terminology. The number of weights depends on the network architecture, such as layer number and layer size. In order to find the weight values, an ANN is fitted to a series of training example (x_i, y_i) until the it is able to compute y_i given x_i with very small error rate. This stage is called *learning*. After learning stage, an ANN is a function with known parameters. Given an input value x', the network can estimate the unknown output y'. This is called *inference*.

ANN Learning Model

Finding the right set of weights to minimize the output error rate is an optimization problem. An optimization problem is solved by either maximize or minimize a certain loss function, depending on the problem. The loss function for ANN describes the difference between the actual output and the inferred output, which we would like to minimize. There are many types loss function, for example square loss function (Equation 5).

$$C(w) = \frac{1}{2N}\sum_{n=1}^{N} | f(x_n|W = w) - y_n |^2.$$

Equation 5. Square loss function

In Equation 5, N is the population of training examples; $f(x_n|W = w)$ is the output of ANN giving input x_n and the set of weight $W = w$; y_n is the true value given by the training example. The cost function is a function of weights. It measures the mean squared error that the ANN, given the current set of weights, made on the training dataset. The best set of weights shall minimize this cost function. Analytic solution for a large number of weights is impossible to compute. Instead, it is solved by an iterative algorithm called the gradient descent.

The core concept of the gradient descent algorithm is first published in the 1960s for solving least squares estimation problem [58]. For a cost function with weight vector w of size M, at each iteration, the M-dimentional gradient vector of the cost function $\nabla C|_{W=w}$ is calculated. In the direction of the gradient vector, the cost function value varies the fastest. In order to reduce the cost function, the weights vector is updated by moving in the opposite direction of the gradient vector by a small step size η (Equation 6).

$$\underline{w} := \underline{w} - \eta * \nabla C|_{\underline{W}=\underline{w}} \cdot$$

Equation 6. Update the weights with gradient descent

The weight vector will be updated in the same manner each iteration until the difference of cost functions between the two consecutive iterations is less than a threshold. In this way, the weight vector, initialized randomly, is guaranteed moving to a M-dimensional location corresponding to a local minimum. The step size η, called the *learning rate*, controls the rate of gradient descent. A large learning rate allows the cost function to reach minimum with less iterations. A small learning rate allows a finer update so that the cost function is more likely to reach the actual local minimum.

However, calculate the gradient of the cost function for an ANN is slightly more complicated because its function model $f(x_n|W = w)$ is not predetermined. The gradient of cost function cannot be found by taking derivatives of the function with respect of each weight.

Calculate Gradient of Artificial Neural Network

The previous section describes ANN at an abstracted-level. It has been shown that ANN can learn using the method called gradient descent. This section discusses the *back-propagation algorithm* that can calculate the gradient of the ANN. A miniature ANN architecture will be used as an example.

First, we formulate the perceptron model. A perceptron has multiple inputs and a single output. For the jth perceptron located at lth network layer, each input is associated with a weight w_{kj}. The subscript k notates that this perceptron input is connected to the kth perceptron output in the previous $(l-1)$ th layer. The total weighted input to the jth perceptron at lth layer is $z_j^l = \sum_k w_{kj}^l * a_k^{l-1} + b_j^l$, where b_j^l is the *bias* associated with this perceptron. The output of a perceptron is called *activation* and notated by letter a. The activation is a function of the weighted input $a_j^l = \sigma(z_j^l)$, where the function σ is called the *activation function*. The following table summarizes the notations. An illustration of the perceptron model is shown in Figure 12.

Layer Index	l
Perceptron Index	j or k
Input Weight	w
Perceptron Total Input	z
Perceptron Activation (output)	a
Activation Function	$\sigma(\,)$

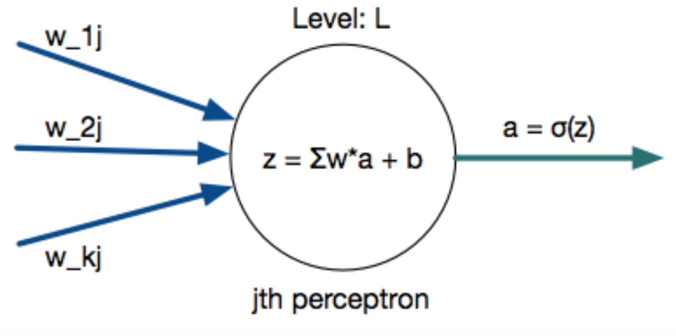


Figure 12 Perceptron model

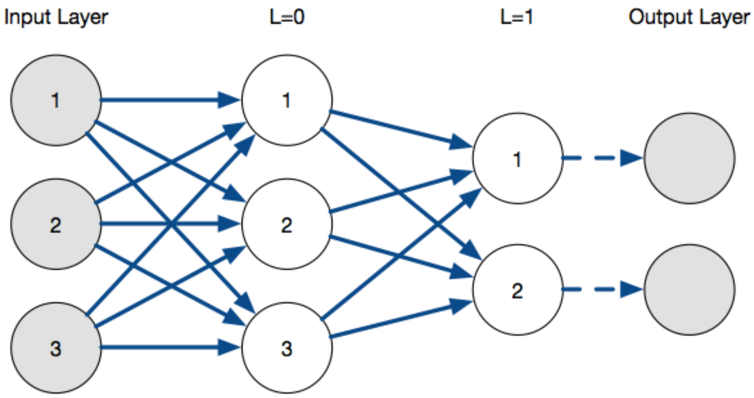


Figure 13 A miniature artificial neural network

With the perceptron model defined, a miniature ANN model is constructed (Figure 13). This specific type of ANN architecture is called *feedforward neural network* because the activation of each perceptron only affects those in the next layer. The layers used in this ANN is called *fully-connected layer* because each perceptron activation is connected to every perceptron input in the next layers. The number of weights and bias in this case is 15 weights (number of solid lines) and 5 biases (number of white circles). In total, there are already 20 parameters in this relatively small ANN.

The learning procedure starts with initializing the ANN with a set of random weights and biases. Then the ANN is presented with a pair of training example (x^1, y^1). The ANN would compute a result a^1, which is a vector contains activations of the last layer (layer 1). Since all the weights and biases are all random, a^1 would be different from y_1. The cost function of this single training example can be calculated (Equation 7). Notice both a^1 and y^1 are vectors of size 2 in this case.

$$C = \frac{1}{2} \sum_j \left(y_j^1 - a_j^1\right)^2$$

Equation 7. Square loss of the ANN given j training examples

Backpropagation algorithm computes the cost function gradient in the following steps [59]. First, an intermediate variable δ_j^l called "error" for each perceptron is computed (Equation 8).

$$\delta_j^l = \frac{\partial C}{\partial z_j^l}.$$

Equation 8. Perceptron error

Knowing the δ_j^l error of each perceptron, the gradient of the cost function $\frac{\partial C}{\partial w_{jk}^l}$ and $\frac{\partial C}{\partial b_j^l}$ can be represented by error and activation (Equation 9).

$$\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1}\delta_j^k \quad ; \quad \frac{\partial C}{\partial b_j^l} = \delta_j^k.$$

Equation 9. Cost function gradients in term of weights and biases

The process of calculating δ_j^l of each perceptron involves:

(1) computes the perceptron errors in the very last layer (Equation 10),

$$\begin{aligned}\delta_j^1 = \frac{\partial C}{\partial z_j^1} = \frac{\partial C}{\partial a_j^1}\frac{\partial a_j^1}{\partial z_j^1} = \frac{\partial C}{\partial a_j^1}\sigma'(z_j^1) \\ = (a_j^1 - y_j^1)\sigma'(z_j^1) \ \ \text{"}Errors\ of\ Last\ Layer\text{"}\end{aligned}$$

Equation 10. Perceptron error in the last layer of the neural network

(2) back-propagates the errors to each of the previous layer (Equation 11).

$$z_k^{l+1} = \sum_k w_{kj}^{l+1} * a_k^l + b_j^{l+1} = \sum_k w_{kj}^{l+1} * \sigma(z_j^l) + b_j^{l+1}$$

$$\frac{\partial z_k^{l+1}}{\partial z_j^l} = w_{kj}^{l+1} * \sigma'(z_j^l)$$

$$\delta_j^l = \frac{\partial C}{\partial z_j^l} = \sum_k \frac{\partial C}{\partial z_k^{l+1}} \frac{\partial z_k^{l+1}}{\partial z_j^l} = \sum_k \delta_k^{l+1} \frac{\partial z_k^{l+1}}{\partial z_j^l}$$

$$= \sigma'\left(z_j^l\right) * \sum_k \delta_k^{l+1} w_{kj}^{l+1}$$

Equation 11. Back-propagate the perceptron error to the preceding layers

The learning process of ANN with N layers can be generalized as the following steps:

1. Initialize weights and biases with a random number generator.

2. Given a new training example (x_i, y_i), input x into the ANN.

3. **Feed forward** the input. For each perceptron, calculate z_j^l, $\sigma\left(z_j^l\right)$ and $\sigma'\left(z_j^l\right)$.

4. Calculate output of the ANN and register the difference between the ANN output and the training example output.

5. Calculate the error $\delta_{j=N}$ of the last layer N.

6. **Back propagate** the error. Calculate the error $\delta_{j=N-1}$ of the layer N-1. Repeat until the error δ_1 of the first layer is calculated.

7. Calculate the gradient of the cost function for training example (x_i, y_i).

8. **Repeat** step 2-7 for all training examples.

9. Calculate the average gradient of all training examples.

10. Update weights and biases in the direction of the average gradient with learning rate η. **Repeat** step 2-10.

Image Classification and Convolutional Neural Network

Image classification or image identification, a subject of computer vision, has always been a popular topic due to its large range of applications. Early image classification methods extract pre-defined features from the image and perform the classification by comparing those feature values to the threshold. Such features can be colors, textures, morphologies, etc. [60] The idea of feature extraction is to compress an image, which has hundreds or thousands of pixels, into a feature vector in a much smaller size. However, there are a few drawbacks with those feature extraction methods. (1) It could be difficult to formulate features that can effectively differentiate visually-similar classes, for example, a green apple and a tennis ball. (2) Many features are variant to scale, rotation and affine transform (distortion), which means the features can work in one viewpoint but fail in another. It causes the image classification system to be fragile and error prone. (3) It is not a generalized image classification method to use hard coded features. For a different computer vision problem, a new set of features have to be invented, programmed and tested. Furthermore, such process could repeat several times until the system meets the expected precision. Many feature models have since published that can overcome those issues. For example, “Eigen-faces” and “Fisher-faces” use principal component analysis to decompose a set of training face images into several linearly uncorrelated major component images, called “Eigen-face.” A face image can be represented as a linear composition of those Eigen-faces. The set of linear coefficients can be used as the feature vector of the image [61], [62]. Another example is the scale-invariant feature transform (SIFT) algorithm. SIFT algorithm intends to extract a set of local features called SIFT key points that is scale invariant, for example, corners in the image. Each key point is then assigned with a 128-dimensional descriptor that carries the information of the gradient orientation around the key points in each direction. Image classification can be performed by matching the SIFT key points between test image and training image [63]. These algorithms all strive for generating features that are universal, highly generalized and independent of vision problem. Similar to the aforementioned algorithms, artificial neural network automatically extracts the major class features from training images. It is a general image classification system that is independent of the classification problems. Furthermore, trained with images captured under different conditions, such as different viewpoints and illumination, a learning-based classification system is more robust to classify images captured under various different conditions.

Neural network was first applied to classify images by considering images as single-dimension numerical arrays. An image is expanded into 1D input vector either in rows or in columns. However, this process removes the spatial information embedded in the image data. In addition, image input requires a large fully connected neural network with many weights. For example, a 32x32x3 RGB image has more than 3000 inputs. A fully connected ANN would need more than 3000 weights in the first layer just to read the input. Convolutional neural network (ConvNet) is an adaptation of artificial neural network that specializes for image classification. Its architecture preserves the spatial information in 2D dataset. In addition, its architecture, which will be discussed later, requires less weights and biases for the image input. One of the early successful application of ConvNet is the optical character recognition (OCR) system that is trained to recognize handwritten digits[64].

The basic building blocks of ConvNet include convolutional layer, pooling layer, rectified linear unit (RELU) layer and fully connected layer. A convolutional layer consists of N kernels. The input to the convolutional layer is a matrix of size $H \times W \times C$. Each kernel is a matrix of size $L \times L \times C$. *The kernel depth notated by C is by design always the same as the depth of the input.* The total number of weights in such a convolutional layer is $N \times L \times L \times C$. The convolutional layer convolves the input of the layer with its kernels. Each kernel produces one 2D matrix; the output of the convolutional layer is a matrix of size $H' \times W' \times N$. The terms H' and W' are always smaller than those of the input. The exact size of them depends on the pre-defined convolution stride size and whether zero paddings are applied. A pooling layer is usually connected right after a convolutional layer. The job of a pooling layer is to subsample the convolutional layer output so the rest of the network can use less weights to feedforward the input. Depending on the pooling size, which is usually 2 by 2, the pooling layer can either pick the maximum value (max pooling) or take the average of the values (average pooling) in the 2 by 2 block. There are no weights to train in the pooling layer.

In addition, it is optional to include pooling layer. RELU is another layer without weights. The function of an RELU layer is to set all the negative numbers in the convolutional layer output to zero. RELU layer brings non-linearity into the network and improves the training result[65]. Still, it is optional to include RELU layers. A fully connected layer is usually connected to the end of the ConvNet after several layers of convolutional layer. Each output of the last convolutional layer connected to input of each perceptron in the fully connected layer. From this point, the rest of the ConvNet is structured the same as an ANN. The fully connected layer performs the classification based on the filtered and down-sampled output of the convolutional layers. "AlexNet" is one of the state-of-the-art ConvNet architecture for image classification [66]. It has five convolutional layers. Each layer on average has more than 200 kernels. Each convolutional layer is followed with a max-pooling layer. Two fully connected layers are connected to the end of the network for classification. "AlexNet" has in total 11 layers and 60 million weights to learn. This type of ConvNet, due to its large structure, is called "Deep Learning Neural Network."

Training Database

A neural network learns from training examples. This section will discuss two topics associated with training a neural network. The first topic is about how to determine the minimum number of training examples; the relation between the training database size and the network size will be discussed. The second topic is about how to generate training datasets.

The minimum number of training examples required depends on the size of the ANN. *Overfitting* occurs when a function is given too few training examples. For example, a polynomial function with five coefficients will always fit to five data points perfectly. In such case, the function captures the random noise in those five instances that would not present in the actual data. As a result, overfitting leads to poor prediction on actual data. As a rule of thumb, an ANN needs about 7 to 10 times as many training examples as the number of weights to avoid overfitting. *Underfitting* occurs if a function with only few weights is fitted to a complicated dataset. For example, linear regression with two coefficients would not be able to fit well to a dataset modeled by a polynomial function of high degree. Under-fitting also leads to poor prediction. However, it is usually hard to quantify the complication of a problem. A heuristic approach is usually used to gradually increase the layer size and layer number as the performance of the ANN is monitored.

The most common way to create training examples is to collect, crop and label images for different classes. For example, MNIST handwritten digit database [64], CIFAR-10-class, CIFAR-100-class natural image database [67], imageNet [68], etc. are large well-labeled image dataset built manually in such way. However, creating a clean and accurate training set from scratch is a very demanding task[68]. An alternative method is to first manually create a small dataset and then artificially grow the dataset by randomly shifting, rotating, adding noise and distorting the image. Artificially boosting the size can expand the image data set by nearly 10-fold [66].

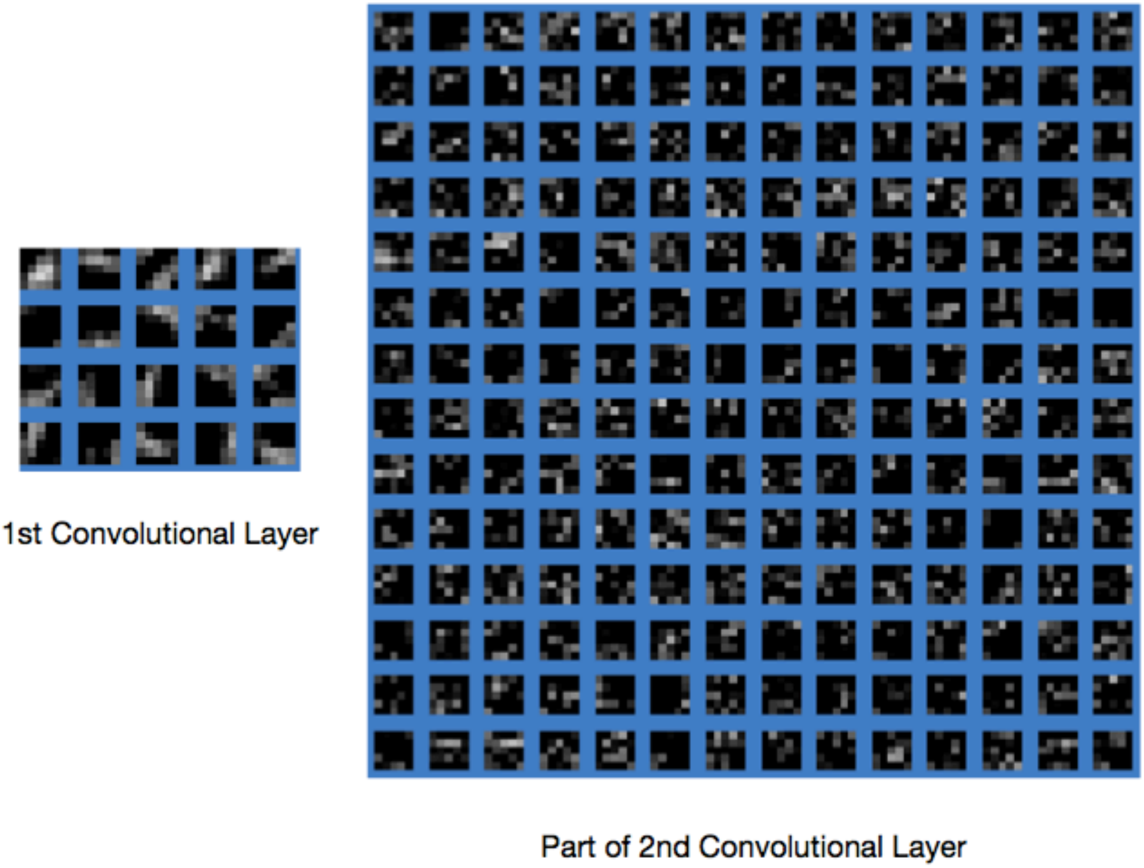
Transfer Learning

Transfer learning originates from the fact that human learners have the ability to share knowledges between different tasks {Yang:2013wx}. It has been observed that convolutional neural networks trained for recognition from scratch on natural images share a similar first-layer weights, which are similar to Gabor filters and color blobs [69]. This observation implies that ConvNet layers closest to the input are most likely to be shareable among different applications; the layers further away from the input layer are less likely to be shareable and are more specifically bounded by training examples. In transfer learning, instead of training ConvNets from ground up with randomly weights and biases, the weighs and biases can be initialized with those from a pre-trained network. The network can then be trained for the new classes by either (1) fine-tuning all layers or (2) retrain only the layers closer to the output while the rests are shared. Fine-tuning all layers improves the generalization capability of the ConvNet [69]. However, it still requires the same amount of training examples to avoid overfitting. Layer sharing works the best if the pre-trained network learns from a dataset that is similar in visual features, for example, use a pre-trained network trained to recognize cats to recognizing lions. However, layer sharing can still work among many applications because almost all the recognition problems seek for low-level features such as corners and edges. In general, imaging recognition problems that are visually similar can share more layers. There is a trade-off between the number of shared layer and the number of training examples. On average, the classification precision drops as more layers are shared [69]. However, as fewer layers need to be trained, smaller training set is required.

Visualize a Trained Convolutional Neural Network

Visualizing the ConvNet helps to verify if the network is actually trained to use the correct visual information in images for recognition. In addition, visualizing the weights and the response of the network can unveil architecture issues, such as overfit and underfit. This section discusses four methods to visually inspect a trained network. They are weight map, activation map, occlusion map and maximum-activation-image map.

Weight maps display the weights at each network layer as images. A well-trained ConvNet should have a smooth weight map with identifiable patterns. *A noisy weight map can be an indicator of overfitting or short training period.* As an example, the weight maps of the first convolution layer and part of the second convolution layer in “LeNet” for handwritten digit recognition are plotted below (Figure 14).


1st Convolutional Layer

Part of 2nd Convolutional Layer

Figure 14 Weight maps of the 1st and 2nd convolutional layer in LeNet

Activation maps show the activated neurons in the network for each given input image. Activation map helps to discover neurons that are not responsive to any input. *Unresponsive neurons indicate that learning rate is set too high during the training.* As an example, the activation maps of inputs handwritten “7” and handwritten “0” of “LeNet” is shown below (Figure 15). The image show that different neurons are activated (bright pixels) for different image input.

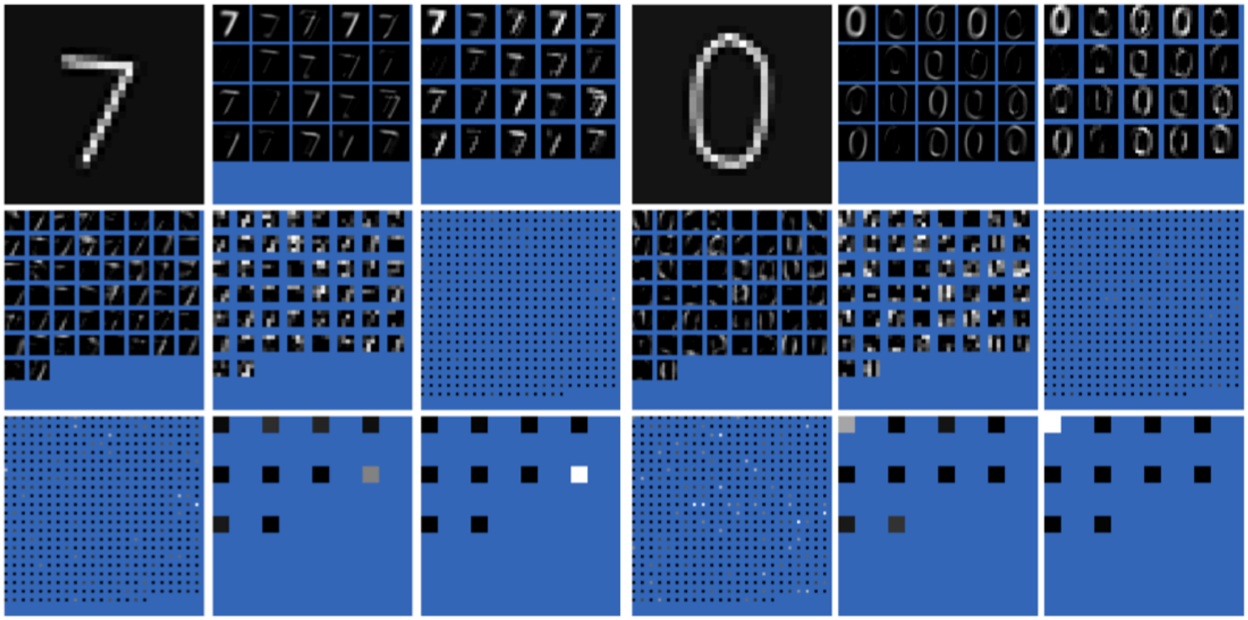


Figure 15 Activations of each LeNet layer given different input image

Occlusion map is created by recording the inference class score while a portion of the input image is occluded [70]. It is an intuitive method to *pinpoint the portion of spatial information in the image that "excites" the network to make the correct inference*. Image below shows an example of the occlusion map of handwritten digit "0" (Figure 16). The blue region are the occluded regions that could cause the network to output low class inference scores (wrong inference). Having low-score occlusion regions that are not covering the actual recognizable object indicates that the neural network has not been yet fully trained.

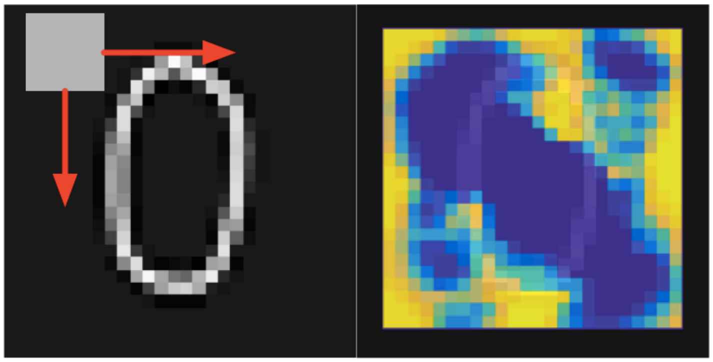
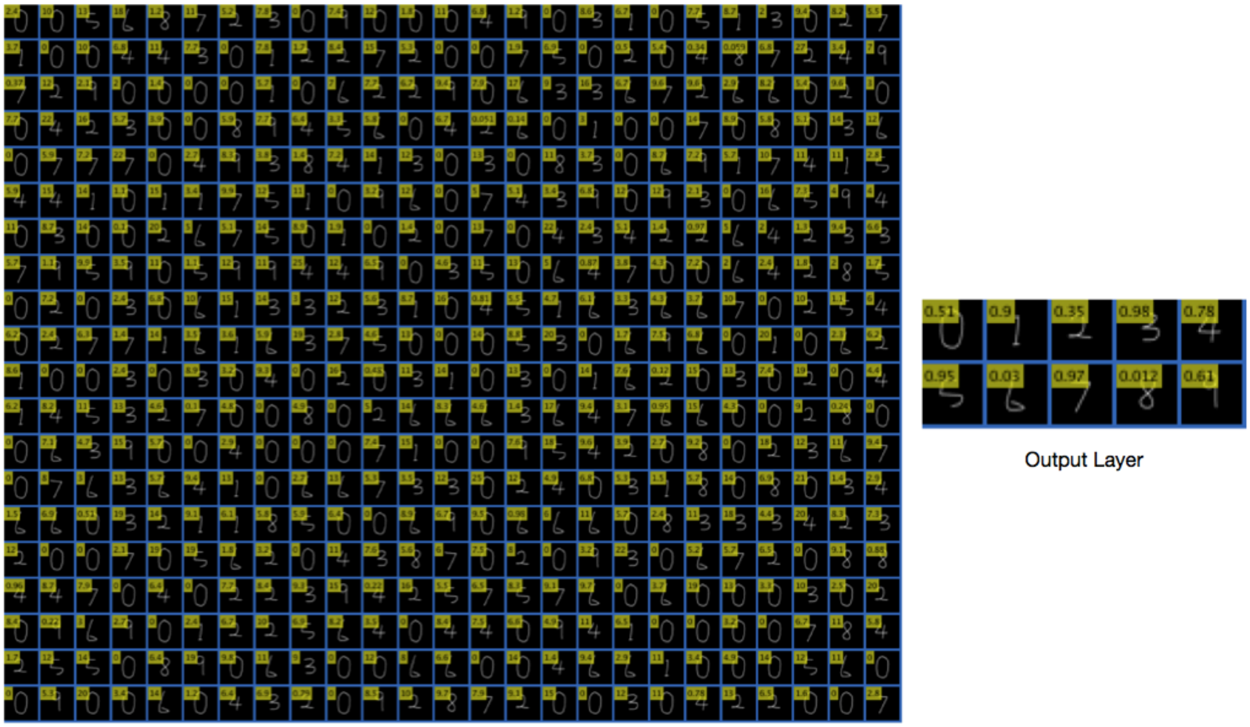


Figure 16 Occlusion map generated by blocking part of the input image with a gray patch each time

Maximum-neuron-activation-image map plots the image that mostly activated a certain neuron [70]. It helps to visualize what type of image a certain neuron is trained to recognize. As an example, the maximum-activation-image maps of the fully connected layer and the output layer of "LeNet" are plotted (Figure 17). The maps show that the work to recognize digits are distributed to different neurons in each layer. Each neuron in the output layer is optimized to max activate on or recognize exact one of the total ten digits.



Fully Connected Layer

Output Layer

Figure 17 Maximum-neuron-activation-image of two of the layers in LeNet

Machine Learning and Computer-aided Diagnosis

With the development of convolutional neural network for visual recognition, artificial neural network has been adopted increasingly in computer-aided diagnosis with medical images. However, there are still many CAD applications that have not yet adopted ANN because of the large number of different pathologies, different diagnosis approaches and different imaging modalities. In addition, there are many issues remained to solve for CAD with medical images. First, artificial neural networks have to learn each combination of different imaging modality and different pathological feature. For example, an ANN trained with x-ray images may not be able to recognize MRI images. Second, it is difficult to build or acquire large training database for emerging imaging modalities; fewer clinical images are captured with the new technologies because of the low device adoption rate at early stage. In addition, due to the nature of supervised learning, training images have to be annotated by professionals of the specific pathology in a coherent, explicit and digital manner. It could take years for recently-developed imaging modalities to build up large and publicly available training database like those for matured imaging modalities, such as x-ray, MRI, ultrasound and microscopy [71], [72]. To overcome those issues, in general, it is beneficial to publish pre-trained neural network for proceeding research to refer to or to use for transfer learning; it is beneficial to release professionally-labeled database; last but not least, it is critical to discover better computational models for machine to learn with fewer training examples. Solving those issues would push artificial neural network toward becoming a general medication tool that can be adopted by clinical professionals.

Method

Animal Models and Experimental Protocols

The animal protocol has been approved by the committee on animal care and use at the University of Maryland College Park. A murine model of CKD was induced by injection of Adriamycin (1.5mg/kg) into the tail vein of Munich-Wistar rats [49]. Once a week during the entire study protocol of 8 weeks, rats were weighed. 24-h urine volumes were collected in metabolic cages. Fresh urine samples were tested for albuminuria (Albustix), and blood samples taken from the tail vein were analyzed for serum creatinine and BUN values (Beckman Coulter Creatinine & BUN Analyzers). In each week for the 8-week period, two rats were anesthetized with isoflurane/O_2 (4% induction, 1.5% during operation, O_2 1 L/min). The abdominal cavity was opened through a midline incision, and the left kidney was exposed and imaged using OCT. Following *in vivo* OCT evaluation, the kidneys were fixed *in situ* by flushing with warm (i.e., 37 °C) oxygenated saline, followed immediately by phosphate buffered 2% paraformaldehyde and 0.1% glutaraldehyde. The fixed kidneys were excised and the rat euthanized by intracardiac injection of pentobarbital sodium to induce cardiac arrest. Blocks of fixed kidneys were embedded in paraffin, sectioned, and stained with hematoxylin and eosin (H&E).

Optical Coherence Tomography (OCT)

A custom-built Fourier-domain OCT system was used in this study (Figure 18) [73]-[76]. This system uses a swept-source laser operating at 1310 nm center wavelength with 100 nm bandwidth. The axial and lateral resolutions of the system are 12 µm and 6 µm, respectively. OCT image dimensions are 1024 pixels (X=1.4 mm) in lateral direction and 512 pixels (Z=2.0 mm) in axial direction. The sensitivity of the system is 90 dB. The A-scan acquisition speed is 16 kHz. 2D image acquisition speed is 16 frames per second. For each experiment, 3D OCT volumes were acquired from 5-10 kidney locations. Each volume consists of 475 consecutive 2D images.

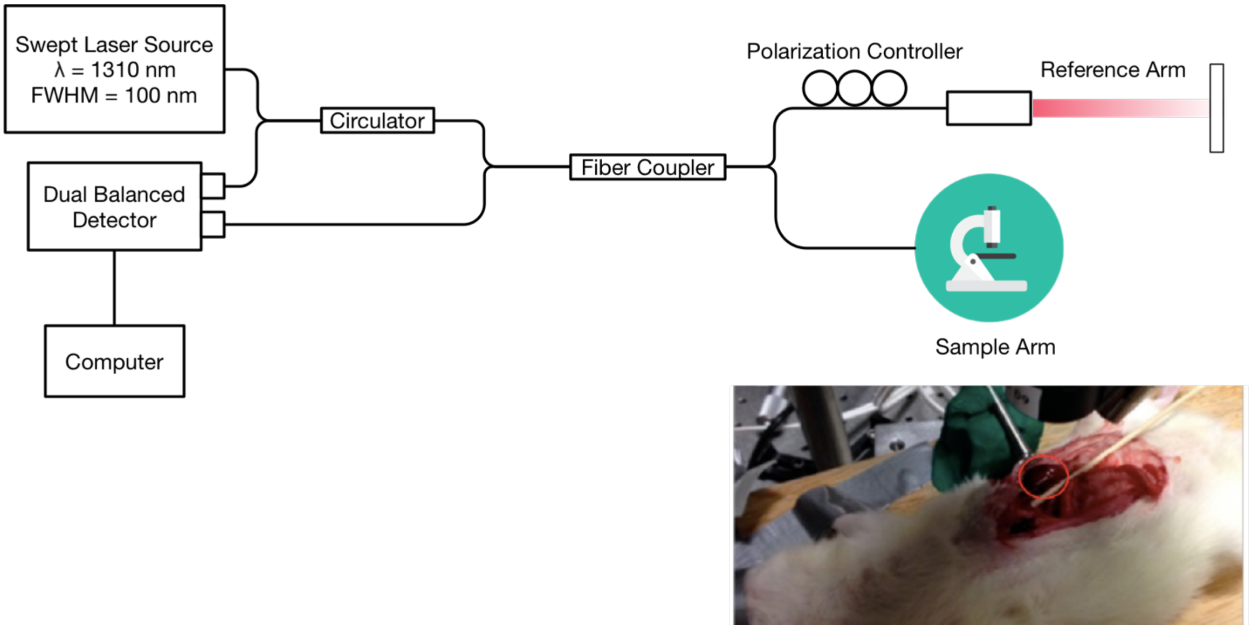


Figure 18. The schematic diagram of the custom-built OCT system used in this study. Inset image shows the abdominal cavity of a Munich-Wistar rat opened through a midline incision, and the exposed left kidney for OCT imaging.

Computer-Aided Diagnosis Software

About 3000 to 5000 OCT kidney images have been acquired each week for eight weeks. In order to determine CKD condition from the massive image dataset, an OCT image analysis software is developed to automatically detect and measure features that is related to CKD progression. In our application, possible features can be the proximal tubule morphology, such as tubular diameter and atrophy/hypertrophy [53]. The image analysis software is consisted of four sections: image preprocessing, feature region proposing, feature region classification, and feature region measurement.

Image Preprocessing

Gaussian blur is used to reduce the speckle noise of the OCT image. Gaussian blur is performed by convoluting the original OCT image with a 2D Gaussian kernel. As a low-pass filter, Gaussian kernel removes high frequency noise as well as useful high frequency information such as edges. A small Gaussian kernel (sigma = 2) is used to limit the blur effect and retain small and fast-varying features as much as possible.

Feature Region Proposing

To recognize multiple objects in the image, regions of interest (ROIs) have to be proposed to an image classifier. One of the easiest ways to propose ROIs is the sliding window algorithm. Specifying the range of window size and the stride size, the window slides from the edge of the image and proposes each window patch as an ROI. This method has been used widely and successfully in applications such as face recognition [77]. However, as a greedy search method, it proposes a large number of proposed regions and suffers from “curse of dimensionality” [78]. Sliding window method has to be combined with a very fast image classifier for speed-critical applications.

Another region proposing method is segmentation. Segmentation divides image into sub-images based on low-level image properties, such as pixel intensity and texture. For OCT images, image intensity-based segmentation with a global threshold cannot recall tubule lumens with high accuracy due to the non-uniform image brightness. A way to overcome this issue is to remove the effect of illumination in preprocessing, such as by the Retinex theory [79]. However, in practice, those enhancement algorithms are very slow. In addition, the illumination effect may not be removed perfectly. Another way is to use dynamic local intensity thresholding. “Bradley” adaptive thresholding technique uses computationally-efficient integral image algorithm to determine the local threshold value for each sub-image window [80]. In comparison to other dynamic thresholding algorithm, “Bradley” adaptive thresholding is robust and less computationally expensive [81]. A representative result of OCT image segmented by “Bradley” adaptive thresholding is shown in Figure 19a.

Superpixel is another region proposing method that groups similar pixels into a "superpixel." Those superpixels can then be proposed as ROIs. The state-of-the-art superpixel methods include graph-based algorithm [82], gradient-ascent-based algorithm [83] and simple linear iterative clustering (SLIC) algorithm [84]. A representative result of OCT image processed by SLIC superpixel method is shown in Figure 19b. The principle of SLIC superpixel algorithm is the following: (1) First N equally spaced pixels are initialized as cluster centers; (2) Each pixel is then associate itself with the most similar neighboring cluster centers; (3) The cluster centers are updated to be the average pixel location of all its associated pixels; (4) Repeat step 1 to step 3 until the new cluster center and the old cluster center converges. The number of cluster centers N controls the granularity of superpixels or the total number of proposed regions.

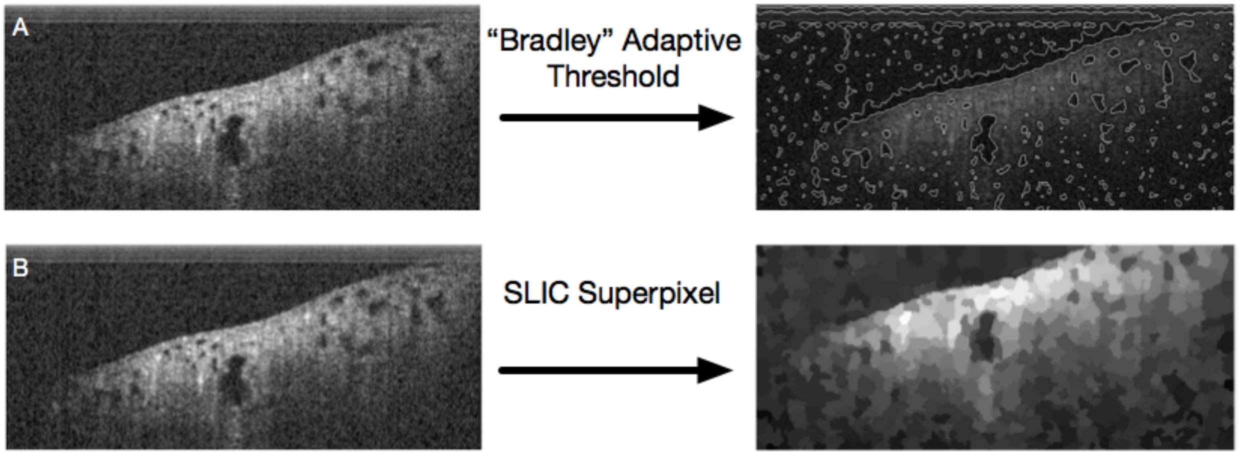


Figure 19. Same OCT image processed by (a) dynamic local intensity thresholding and (b) SLIC superpixel.

Ideally, the region proposing algorithm should have a 100% recall rate. Recall rate is defined as the percentage of total tubule lumens being proposed. Another key parameter is the total number of proposed regions. For example, sliding window algorithm with a small stride size and a large range of scale would have a near 100% recall rate because it can theoretically capture all tubule lumens in different scales and at different locations. However, its total number of proposed regions is massive, which is impractical for many applications.

We compared the ROI recall rate and the total number of proposed regions between “Bradley” segmentation and SLIC superpixel. To determine the recall rate, tubule cross-sectional regions in test images are manually labeled

(Figure 20a). A tubule region is correctly recalled if more than 60% of the recall region overlaps with the manually labeled region (Figure 20b). The number of proposed regions can be altered by changing the threshold level and window size for “Bradley” adaptive threshold algorithm or the initial number of cluster centers for SLIC superpixel algorithm. The recall rate versus the number of proposed regions for each algorithm was measured (Figure 20c). From the results, “Bradley” adaptive thresholding with a window size of 30 pixels is able to recall more than 80% of the tubule regions with under 500 proposed regions. SLIC superpixel algorithm underperforms “Bradley” adaptive thresholding for this application. Therefore, we choose “Bradley” adaptive thresholding with a window size of 30 pixels in this study.

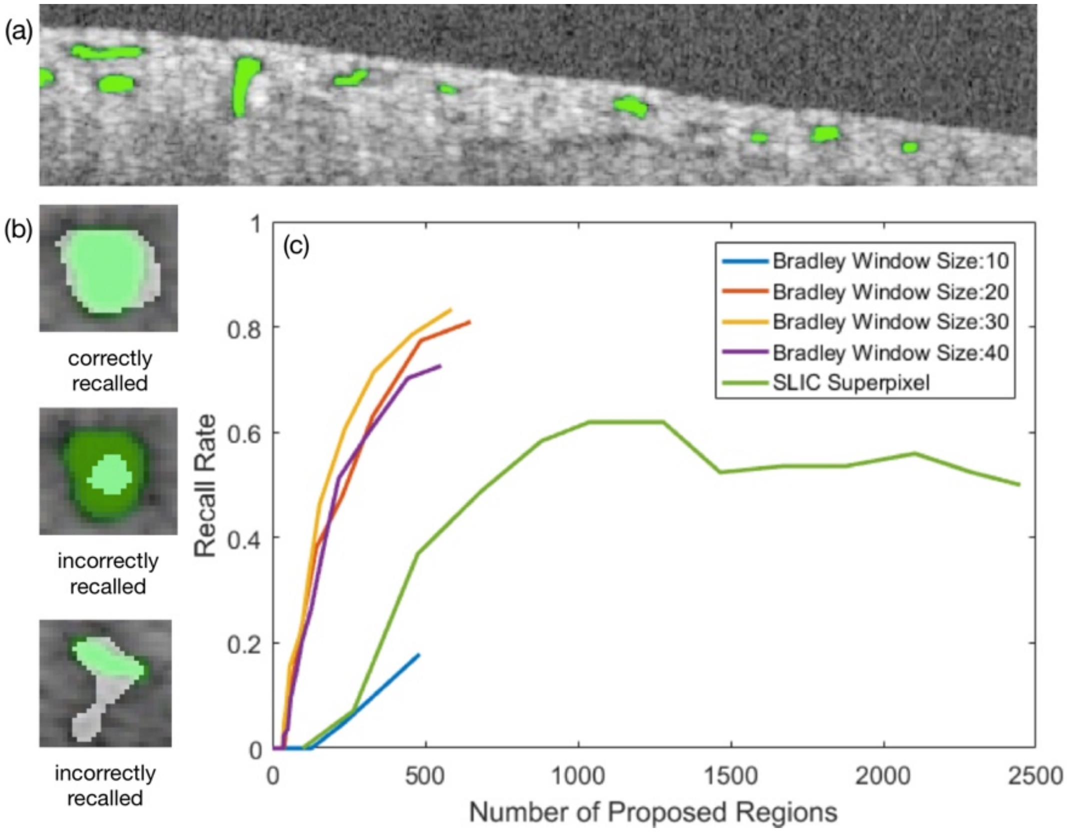


Figure 20. (a) Manually labeled tubule lumens (green regions). (b) Overlaps between proposed regions (white) and manually-labeled true regions (green). (c) Recall rate vs. number of proposed regions for different algorithms.

After candidate regions for proximal convolutional tubule lumens are chosen, rectangular image patches containing the candidate regions are cropped for further image classification. It is unavoidable that the set of proposed images would include many false positive regions, such as defects in the OCT image. From the perspective of adaptive threshold, they are similar to tubule cross-sections, which has a relative darker center than the surrounding regions.

Examples of the region-proposed images are shown in Figure 21. To differentiate them, an image classifier is necessary to distinguish the true tubule cross-sections (Figure 21 top row) from the falsely proposed regions (Figure 21 bottom row).

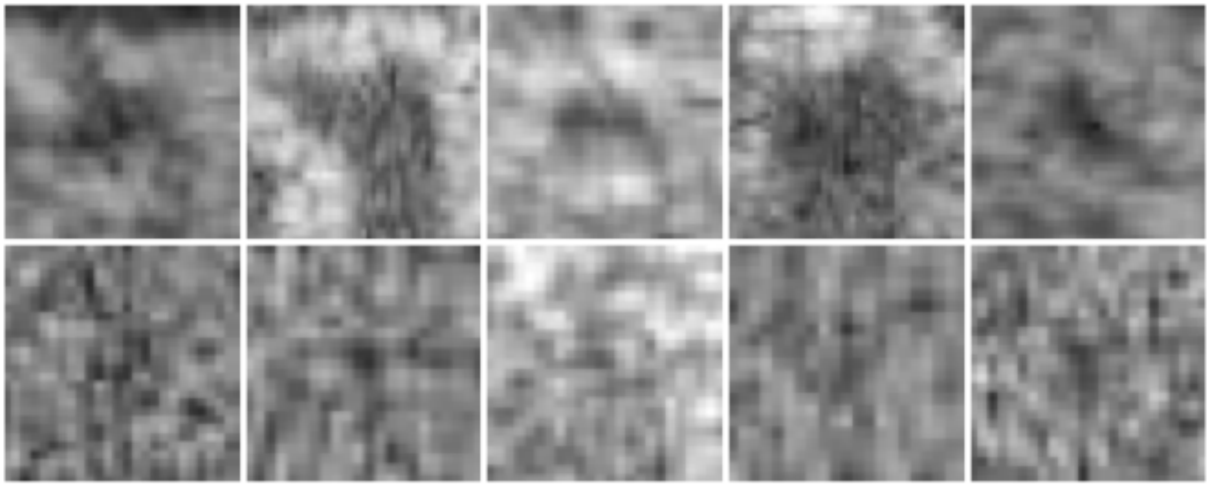


Figure 21. Examples of proposed region images. Images at the top row are tubule lumens. Images at the bottom row are OCT image artifacts and speckle noises. The brightness of some images has been increased for the ease of viewing. Sizes of the proposed region images range from 5 pixels to 60 pixels in width.

Feature Region Classification: Convolutional Neural Network Image Classifier

There are many image classifier models. Some of the well-known classifiers include hard-coded classifier, Eigen features classifier [85], and artificial neural network [86], etc. A classifier generated through learning tends to be more accurate due to the amount of prior information gained through the training process. In our study, OCT images contain large variations in tubule morphology, kidney condition, image quality, etc. Convolutional neural networks (ConvNet) [66] is eventually chosen as the image classifier due to its larger learning capacity.

ConvNet training requires thousands to millions of labeled images per detection class depending on the complexity of the network structure. A large and complex network, with more layers and more weights per layer, is capable to learn classifying more complicated visual features. However, a large network also requires a larger training sample sets. Overfitting happens if a large network is trained with a small sample set. Over-fitted neural network would quickly minimize the cost function of the training set, where the weights would no-longer be further fine-tuned. However, over-fitted neural network would have a very low classification accuracy in validation. There is a trade-off between learning capacity (neural network size) and training database size. In summary, to train a neural network to solve a more complicated problem would requires more annotated training samples of the problems.

In the case of training neural network to classify natural scenery and objects, for example, trees, cars, animals, etc., the impact of this trade-off is not significant because these classification problems are hard for neural network but easy for human; providing a large amount of annotated training samples for a large neural network can be done through crowd-sourcing [87]. In contrast, in the case of training neural network to classify medical images, the impact of this trade-off is much more significant because these classification problems are difficult both for neural network and for human. Crowdsourcing is only possible within a relative small crowd because only individuals that have the right medical background or individuals that go through training are qualified to annotate the medical images in order to create the training sample set. In practice, despite that sufficient image data is usually available, the labor cost for annotating the thousands to millions of medical images is a prohibitive task. While there are many advanced deep convolutional neural network, such as Alexnet [66], GoogLenet [88] that can classify visually complicated images, and region-based convolutional neural network (R-CNN) [89] that can perform both region purposing and classification, few medical imaging applications can apply those architectures because of the limitation on training dataset size. For those reasons, in our application, we used a hand-coded region proposed algorithm and trained the smaller LeNet [86] to perform the classification.

For our application, we built a training database with two classes, tubule class and non-tubule class. To crop and collect training images, we first created a less-accurate image classifier based on hard-coded features. Bootstrapping from this classifier, we were able to collect sufficient images per class quickly but with significant false classification rate. The incorrectly classified images in both classes were manually removed from the database by human inspection. To keep the manual inspection time minimal, we deliberately collected only 2000 training images per each class. However, 2000 images per class are far from enough to train a network that has thousands of weights and biases without overfitting. In order to increase the database size and variety, each of the original images was duplicated N times. Then the duplicated images went through random translation, rotation, scaling, and noise addition (Figure 22). The database size was artificially increased by N times. In our case, N equals to 10 and the final database size is 20,000 images per class.

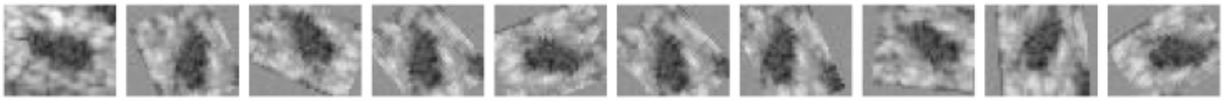


Figure 22. Multiple training images generated by manipulating the original tubule cross-section image (first on the left). All images are resized to 32 x 32 pixels for ConvNets training.

Two different ConvNets were trained with the database. The first network has a small structure size (less than the overfitting limit of the database) as shown in Figure 23a. The second network is the LeNet [86] (Figure 23b). LeNet is chosen because it is originally designed for recognizing single channel grey-scale handwritten-digit images (MNIST dataset), which is in the similar color space as OCT images. Many latest ConvNet architectures, such as AlexNet, are originally designed for RGB image classification. The smaller-sized network was trained end to end from the first layer to the last layer with each layer initialized randomly. LeNet was trained with transfer learning; instead of end-to-end training, it started from a pre-trained LeNet for handwritten digit recognition. Only the fully connected layers were fine-tuned using our training database. Transfer learning enables a large network to be trained on a smaller dataset without overfitting. The performance of the two networks were benchmarked with a separate testing database; the testing database contains OCT images randomly chosen from all CKD stages. Locations of the tubules in the testing database were manually labeled. The bench test result showed a better tubule recognition rate (90%) from LeNET with transfer learning. It is worth noting that whether transfer learning is a good option can be speculated based on the degree of visual differences between the original application of the pre-trained network and the current application [69]. The ConvNet training was performed using MatConvNet [90] on MATLAB using an Nvidia GTX660 GPU. Once the network was trained, the network weights were saved. Image classification based on the weights could be performed on machines without Nvidia GPU support. In this study, the image classification based on the trained ConvNet was programmed in C++, with multi-threading parallelism using OpenMP, and ran on an Intel 8-core Xeon Processor. The processing speed is about 10 frames per second.

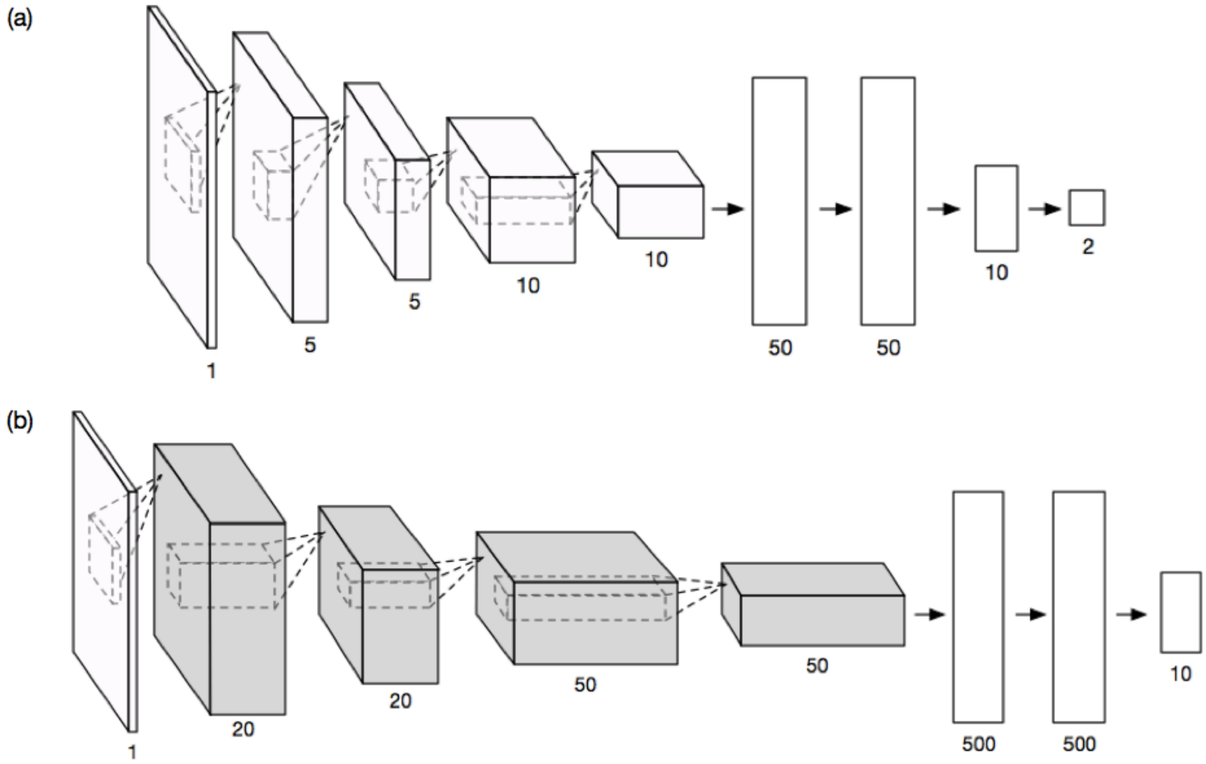


Figure 23 (a) An end-to-end trained small convolutional neural network. (b) LeNet with only the fully connected layer being trained while the rest (gray) layers inherited from the pre-trained network trained on MNIST database.

Feature Region Measurement

CKD is defined as the progressive loss of kidney filtration function in time. Nephrons are the basic filtration unit in kidney. Histology study of a rat CKD model has shown that when nephrons lose their filtration capability, the proximal convolutional tubules of those nephrons shrink and close [56]. The rest of the functional nephrons in kidney would compensate the dysfunctional ones. The proximal convolutional tubules of the functional nephrons enlarge and become hypertrophic. Histology study has shown that the average convolutional tubular diameter increases as CKD condition progresses [49].

For each tubule's cross-sectional region, the circularity of the region was first measured. Since a 2D cross-sectional OCT image could cut the proximal tubule in any random plane, the circularity of the region indicates whether the tubule cross-section is a perpendicular cut. Only the diameters of tubule regions with high circularity (greater than 0.8) are chosen for diameter analysis. The procedure to measure the circularity and diameter of a lumen is the following. (1) Locate the contour pixels of the region. The circumference of the region (L) is the total number of contour pixels. (2) Locate all the pixels belong to the region. The area of the region (A) is the total number of pixels in the region. (3) Circularity of the region is calculated by dividing the circumference square of the region by 4π times of region area, i.e., $L^2/(4\pi A)$. A perfect circle would have a circularity equal to 1. (4) Locate the region center by averaging the coordinates of each pixel inside the region. (5) Distances from the region center to each contour pixel are measured. The diameter of the region is 2 times the mean of the distances. Figure 24 depicts the process of determine the circularity and diameter of tubule cross-sectional region.


Tubule Cross Section
Region Contour
Region Area
Region Center
Center-to-contour Distance

Figure 24. Illustration of definitions used in region measurements.

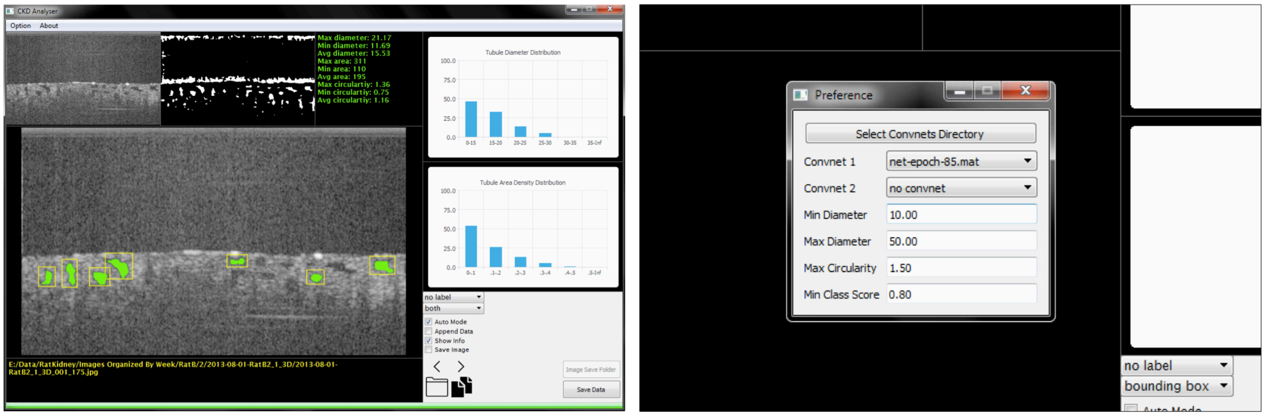
Results

Software Overview

The CKD CAD software is implemented on Windows PC with Visual C++ for backend logic and QT for frontend user interface. At the time of development, third-party libraries for convolutional neural network forward feeding, such as TensorFlow, is not well supported on Windows PC. We developed our own program in Visual C++ to process the pre-trained ConvNet and compute the output of the ConvNet given an image. Since both the architecture and weights of the ConvNet will be evolve overtime given more training datasets become available in the future, the software provides an interface to swap in different ConvNet file to perform the classification

(Figure 25b). As the user, he or she can customize which pre-trained ConvNet to use to process the images.

In the user interface for medical professionals, the software provides different ways to load in batches of images, displays the detections and measurements in real time and compute and displays the statistics in real time (Figure 25a). Through the interface, medical professionals can choose whether to analyze large bundle of images at once or review individual images one at a time by clicking the arrow keys. We believe having a professional and friendly user interface is a critical step to push the technology to clinics because it would help the medical professionals to see the potentials to include the technology into their day-to-day workflow.



```
Max diameter: 21.17
Min diameter: 11.69
Avg diameter: 15.53
Max area: 311
Min area: 110
Avg area: 195
Max circulartiy: 1.36
Min circulartiy: 0.75
Avg circulartiy: 1.16
```

E:/Data/RatKidney/Images Organized By Week/RatB/2/2013-08-01-RatB2_1_3D/2013-08-01-RatB2_1_3D_001_175.jpg

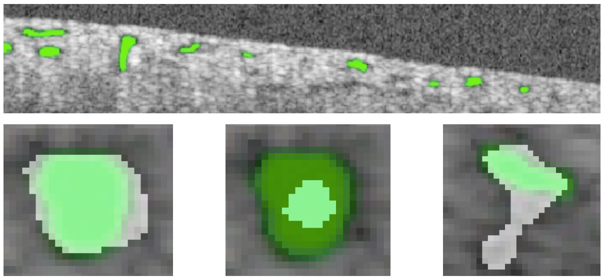
Convnet 1: net-epoch-85.mat
Convnet 2: no convnet
Min Diameter: 10.00
Max Diameter: 50.00
Max Circularity: 1.50
Min Class Score: 0.80

Figure 25. (a) CKD CAD software interface with real time detection, statistics displays (b) software supports ConvNet swaps and multiple ConvNet filtering

Comparing Different Tubule Classifiers

Three different tubule image classifiers were developed during the software development period. They are (1) a handcraft classifier based on segmented tubule morphology including contour curvature, circularity and eccentricity, (2) principal component based (PCA) linear discriminant analysis (LDA) classifier and (3) convolutional neural network model based on LeNet architecture.

Ground truth tubule segmentation of the OCT images is generated by manually labeled a set of images that are not used during the training or tuning of those classifiers. The same set of images are also gone through each classifier. The human labeled images and machine labeled images are compared. If the overlapped area between the two segmentations is greater than 85% of the total segmentation area, it counted as a successful detection; otherwise it is a false detection (Figure 26).



correctly recalled

incorrectly recalled

incorrectly recalled

Figure 26. Comparing human labeled detection with machine labeled detection

The receiver operating characteristic (ROC) curve of those classifiers have been computed (Figure 27). From the result, we could see that while all three classifiers have the capability to identify tubule images above 80% accuracy. The convolutional neural network based method has major advantage of rejecting false positive tubule images comparing to other classifiers. While the morphologies of the tubules are not much different from one to another, it is the image artifacts and various longitudinal cut of the tubules that are widely different from locations to locations. With much higher learning capacity, convolutional neural network is much better at learning different appearance of the false positive images and rejects them comparing to other techniques.

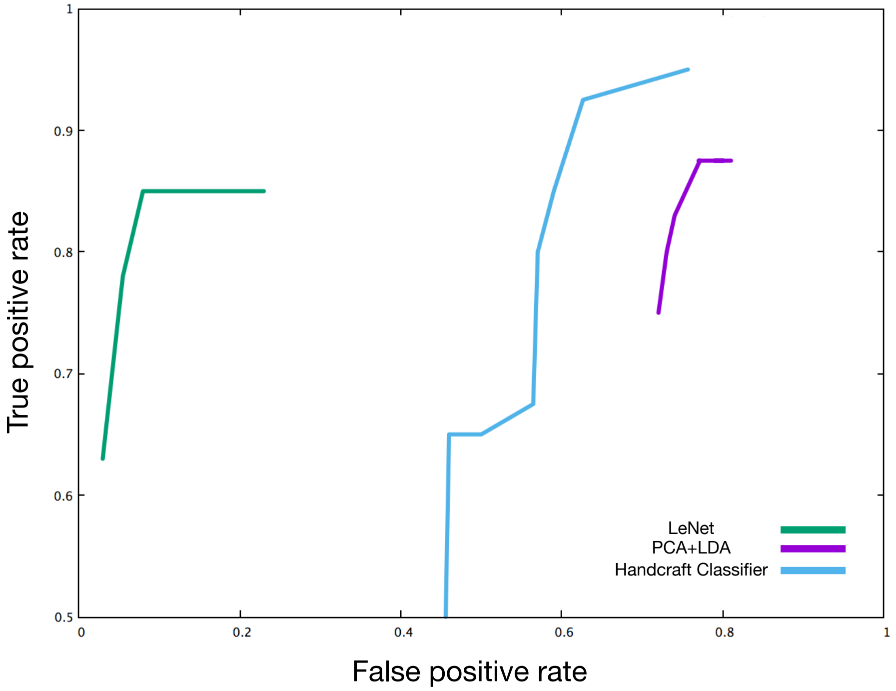


Figure 27. ROC curve of three different tubule image classifier

Robustness of the neural network model under different blurring effect

To test the robustness of the neural network model, positive image samples are blurred repeatedly with a growing Gaussian kernel. The output scores of the positive samples with various levels of blur effect are recorded and averaged (Figure 28). The confidence score gets below 50% when the Gaussian kernel sigma is larger than 9.5.

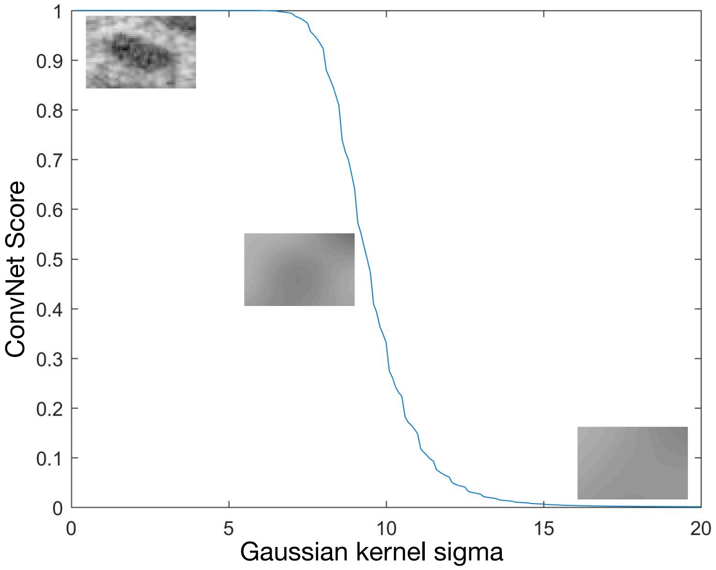


Figure 28. Prediction robustness of the neural network at different blur level

Robustness of the neural network with adversarial noise

The other way to test the robustness of the neural network is to perturb the image by adding adversarial noise to the image. Different noise level of white noise is add-on to a set of positive images that are not part of the training data. At each noise level, we generate the noise repeatedly and measure the likelihood that the neural network can make a correct prediction at this noise level (Figure 29). At Gaussian noise with variance of 2, the image is so noisy that it is difficult to recognize the tubule in the center of the image even for human.

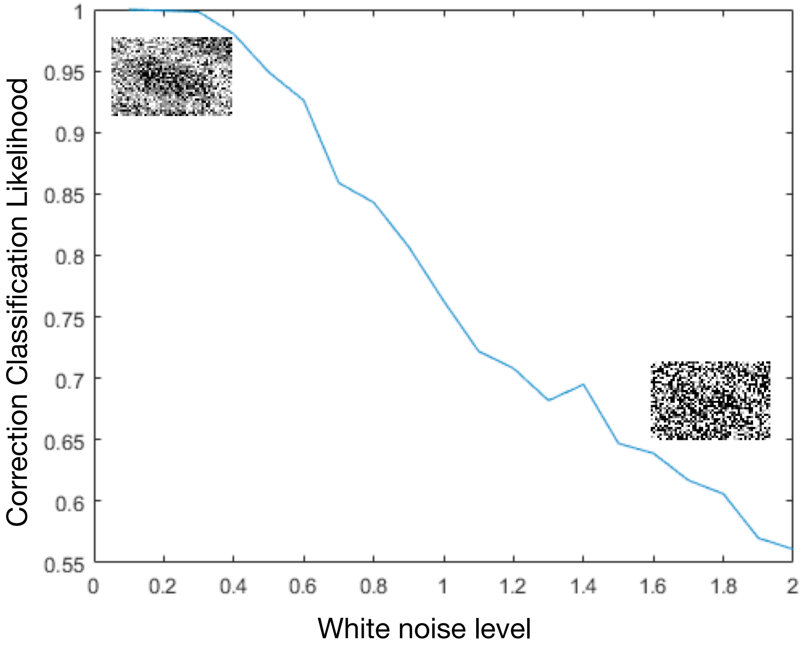


Figure 29. Prediction robustness of the neural network at different adversarial noise level

OCT Imaging of CKD Rats

Visual differences between normal kidneys and CKD kidneys can be perceived from OCT images of rat kidneys. Images in Figure 30 are collected from both healthy rats and rats with mid- to late stage CKD conditions. In healthy kidneys, the proximal tubule lumens are similar in size and distributed uniformly with a homogeneous distribution pattern (Figure 30A-C). In kidneys with late-stage ADR-induced CKD (Figure 30G-I), some proximal tubules either appear reduced in size or become indiscernible to OCT, i.e., atrophic. At the same time, the rest of the proximal tubules appear increased in size, i.e., hypertrophic. This is consistent with the published histology study in dysfunctional nephrons [49]. These are the two most extreme cases in CKD kidney morphology. In the mid-stage CKD (week 2-6), the observed kidney images (Figure 30D-F) can be a mixture of these two extreme cases, where the hypertrophic tubules are smaller than those at late-stage and atrophic tubules are more discernible than those at late-stage.

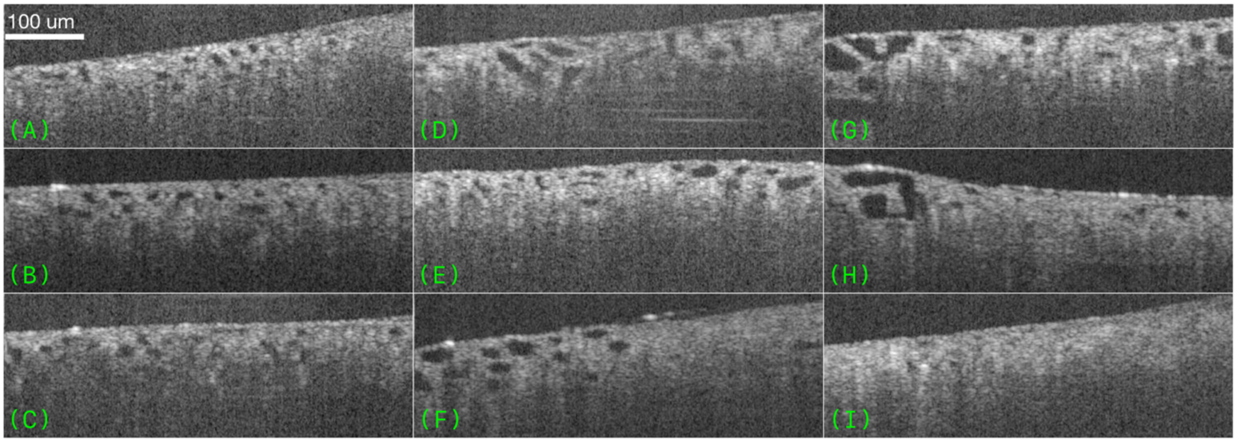
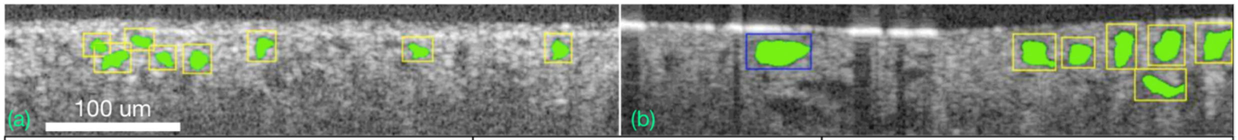


Figure 30. (A)-(C) are images captured from healthy kidney, where proximal tubules are uniform in size; (D)-(F) are images captured from mid-stage CKD kidney, where certain percentages of proximal tubules appear shrunk and certain percentages of proximal tubules appear enlarged; (G)-(I) are images captured from late-stage CKD kidney, where proximal tubules in some cross-sections appear further enlarged and in some other cross sections disappeared due to shrunk to extreme small sizes.

Individual Image Inspection

While it is possible to determine whether CKD exists, it is difficult to quantify the severity of the disease. To quantify the progression, we developed a CADx software to analyze the OCT images and measure the size of the proximal tubules. The CADx software was first tested on individual OCT images captured from different animals at different stages of CKD (Figure 31). The software can detect and measure the tubular morphology and present the statistics instantly as the user navigates through images. In addition, the software highlights the detection using color-coded bounding boxes, where yellow, blue, and green indicate normal-size tubules, hypertrophic tubules (diameter > 50 μm), and atrophic tubules (diameter < 15 μm) respectively. Figure 9 shows two images examined by the software. Figure 31(a) and (b) are from a healthy kidney and a late-stage CKD (8 weeks after ADR injection) kidney respectively. Visually, it is apparent that tubule lumens in Figure 31(b) are larger than those from Figure 31(a). The software measures the tubule lumen diameters and the results from the software show that tubules from healthy kidneys are mostly within the normal ranges, which is 20 - 30 μm. While the tubules from late-stage CKD kidneys are either in the hypertrophic range (diameter > 50 μm) or in the high end of normal ranges. The result is consistent with previous reports that the diameters of the hypertrophy tubules can be 50% larger in mice and nearly 100% larger in human than the diameters of the normal tubules [49], [53]. These results verify that the CADx software is able to detect and measure tubule cross-sections accurately, which can be used to differentiate healthy and non-healthy kidneys.



	Healthy	Late-Stage CKD
Max Diameter (um)	32.1	60.4
Min Diameter (um)	21.1	33.7
Avg Diameter (um)	25.4	44.2

Figure 31. Individual image analysis shows significant tubular morphological differences between (a) healthy kidneys and (b) late-stage CKD kidneys.

Automatic Batch Analysis

In addition to individual image inspection, the software is able to analyze the entire batch of image dataset automatically. The processing speed is about 10 images per second. It takes in total 10 minutes to process, measure, and generate statistics of a dataset containing 5000 images. The software records the information of each detected tubule, including diameter and circularity. The OCT image dataset can be quickly analyzed by the software without supervision of the user.

The complete OCT image dataset of rat kidneys was analyzed by the software. The dataset consists of 18 animals in total: 2 animals per week group for 8 post-ADR-injection weeks in total and 2 healthy animals for control. 3000-5000 2D OCT images were collected per rat. In total, the complete dataset contains more than 70,000 images. The CADx software measures and records tubule diameters and generates the statistics, including tubular diameter, percentage of hypertrophic tubules, tubule density, and tubule diameter range for each week.

Tubule diameters are measured directly by the software after the tubule being detected. Percentage of hypertrophic tubules is derived by recording the number of detected tubules whose diameters are larger than 50 μm. Tubule diameter range is the difference of the tubule diameter at 95 percentiles and that at 5 percentiles. Tubule density is computed by determining number of tubules per unit tissue area.

Figure 32 shows representative OCT images for the control and CKD-induced rat. For the healthy animal, most of the tubule diameters are within normal range (yellow). One week after injection, fewer tubules are visible and more atrophic tubules are present. During the mid-stages (week 2-6), number of both atrophic tubules and hypertrophic tubules increases. This is probably due to the fact that functional nephrons overwork in order to compensate the loss of atrophic tubules (green). At the late stages (week 7-8), more hypertrophic tubules (blue) are visible.

Figure 33 shows the statistical results from the analysis of the complete dataset. The statistics indicates a general trend of increase for average tubule diameter, diameter range, and percentage of hypertrophic tubules. The statistics also shows a fluctuation in these parameters during the mid-stages (week 2-6). We hypothesize it could be due to either variation of different rat's reaction to ADR injection, sampling bias, or the recovery of the organ before late-stage CKD (week 7-8). Previous studies have observed renal function recovery in drug-induced acute kidney injury [91]. Tubular density shows an initial increase (week 2-4) compared to normal, followed by a decrease in week 5-7. This could be explained by the closure of some atrophic tubules.

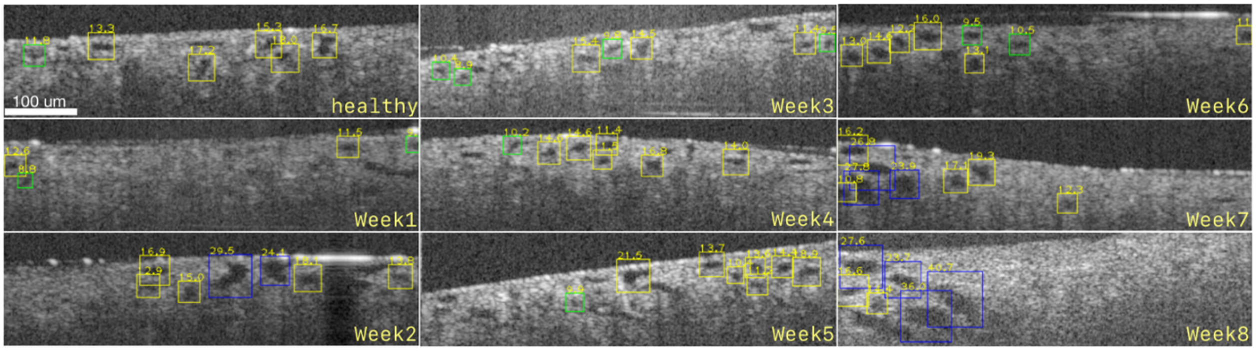


Figure 32. Sample kidney OCT Images from healthy control rats and from CKD rats at each week after ADR-injection. Bounding box color indicates the condition of the tubule; yellow color indicates normal tubule; green color indicates atrophy tubule; blue color indicates hypertrophy tubule.

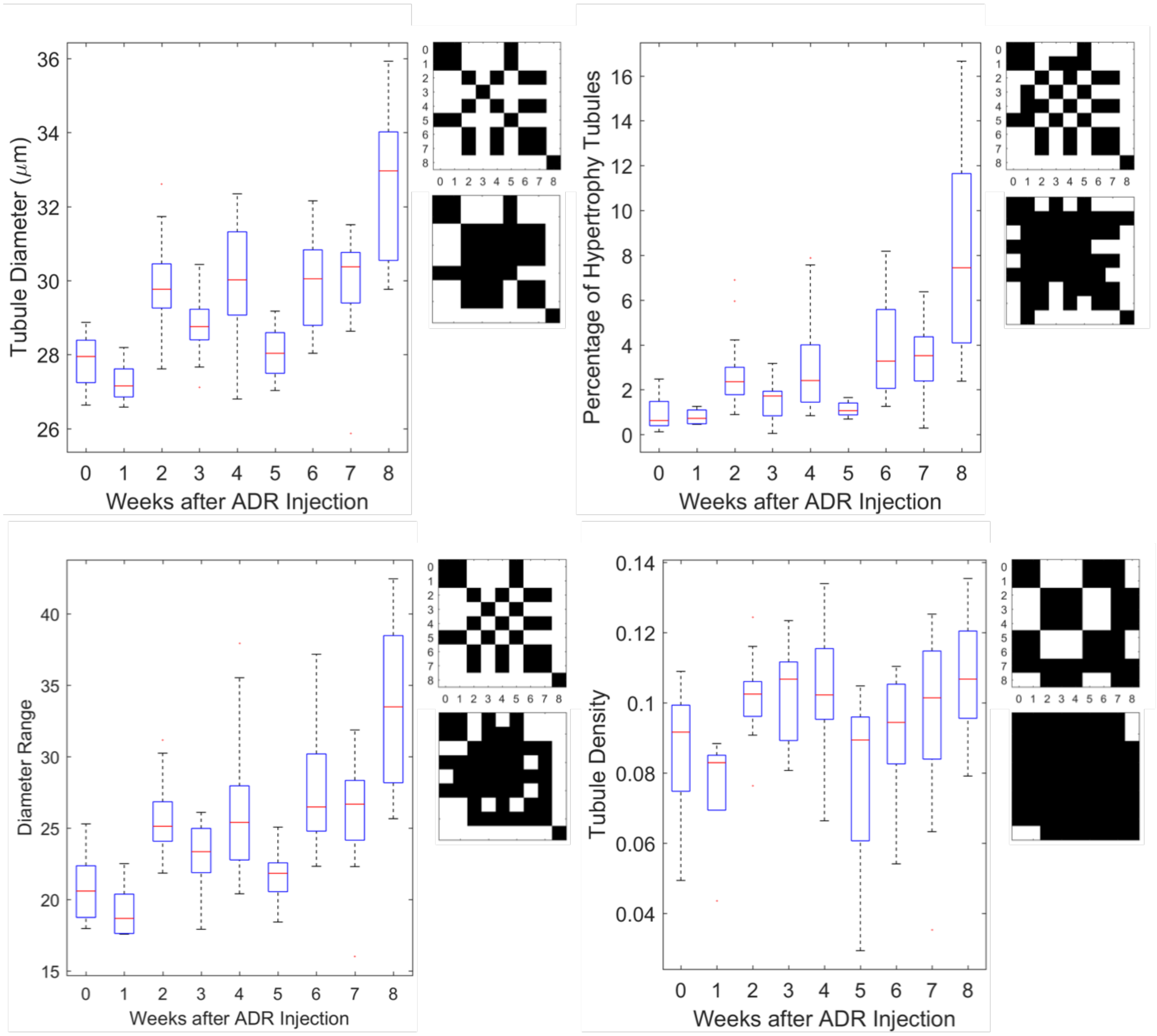


Figure 33. The trends of mean tubule diameter, percentage of hypertrophic tubules, tubule diameter range, and tubule density as CKD progression. Matrix of t-test between each week is attached to the upper right corner of each plot. The adjusted p test matrix using Bonferroni correction for multiple comparison for each plot is beneath each unscented p-test matrix. The dark regions indicate the t-tests between two weeks produce p value above 0.05 (insignificant); the bright regions indicate t-tests between two weeks produce p value below 0.05 (significant).

Discussion

It is crucial that CKD is detected and treated at early stages [51]. Currently, measuring glomerular filtration rate (GFR) score and observing the visual presence of kidney damage are the two commonly established CKD diagnosis methods in clinical nephrology [52].

GFR score is a scoring system that describes kidney filtration capability. GFR score is calculated in terms of blood serum creatinine level test in addition to patient’s age, weight and gender. Serum creatinine is a chemical waste produced by the body that is mainly filtered out by the kidneys. However, despite the wide use of GFR as the gold standard to estimate kidney function, studies have suggested that it is not sensitive to early-stage CKD. It has been shown that a significant change of serum creatinine concentration in blood does not occur until nearly 50% of the kidney is nonfunctional. It is because the rest of the functional nephrons can adapt and compensate the loss of filtration units [53].

Examining the morphology of the kidney structures, including glomerulus and proximal convoluted tubules, using histology images is a more direct and accurate way to detect CKD. The correlation among nephron tubule morphology, nephron function, and CKD progression has been discussed extensively and been corroborated with histology images in literature [49], [53]-[57]. If a nephron loses its filtration capability due to injury or diseases, its tubules would appear reduced in size. Meanwhile, remaining functional nephrons become enlarged in order to compensate the loss of filtration capability [53]. Although observing kidney morphology is an effective way to detect CKD in early stages, the current procedure requires biopsy in order to acquire a slice of renal tissue for histology. Such procedure has many disadvantages. (1) It requires the removal of certain amount of kidney tissues. Due to its invasiveness, only a small percentage of CKD patients receive diagnosis based on morphology images. (2) The histology sample size is restricted because of the limit amount of kidney tissue that can be sampled. The morphology observed with such technique is highly localized. Therefore, the observed condition may not accurately represent the condition of the entire organ. (3) It is a demanding and time-consuming task to acquire and prepare kidney biopsy samples. Such process includes delicate surgery procedure to acquire the sample, careful sample preservation with correct storage temperature and solutions and sample staining for microscopy imaging. (4) Similar to many imaging-based diagnosis procedures, histology image assessment is a meticulous and time-consuming task that requires two experienced histopathologists in order to reduce the diagnosis errors.

OCT imaging can observe nephron tubular morphology without the needs of acquiring tissue sample. This imaging method can examine the kidney structure in a similar fashion as histology while reduces the invasiveness of such procedure. In addition, since it does not require tissues to be removed from the organ for observation, OCT imaging is able to image multiple locations in the kidney or even a full surface scan without altering the condition of the organ. Being able to image multiple locations in the kidney can provide a much more comprehensive view of the overall health condition of the organ compared to biopsy/histology, which uses a very small percentage of the organ tissue to assess the condition of the entire organ.

A detailed multi-location imaging of the kidney drastically increases number of images to inspect per patient for a more accurate diagnosis. Being able to process such large quantity of information on time has a significant impact on the diagnosis workflow. The presented computer-aided diagnosis software showcases how to detect, measure, and statistically quantify kidney function based on OCT images. The software is not developed with the goal to replace the role of medical doctors; instead, it is developed as a useful tool for medical doctors to quickly scan through the large image database and compare their results with the statistics report generated by the software. The ultimate goal for such software is to decrease diagnosis time span, decrease diagnosis error rate, and improve the day-to-day workflow efficiency of medical doctors.

The current methods have limitations and room for improvement in several areas. (1) The animal experiment design to observe the progression of CKD can be improved. The current method can observe a single animal only once because it is very difficult to perform surgery on the same rat repeatedly for consecutive weeks. Therefore, the statistical results derived from this experiment design cannot avoid the biological variation between animals. Continuous monitoring of the kidney morphology of the same animal would definitely present a better picture of the morphology development with the CKD progression, especially at week 2-6. This will require an abdominal window [92]. (2) The selection of imaging locations on kidney can be biased by the operator. The bias induced by imaging location selection could also have an impact on the accuracy of the statistical results. The OCT instrument in use is a stationary microscope that performed scanning in a predefined rectangular region. Despite that multiple locations (rectangular regions) have been scanned per each kidney, a scan covering the entire organ surface will provide the most unbiased description of the kidney condition. (3) The OCT probe needs to be held closely to the kidney surface in order to image the subsurface structures of the organ. To reduce the invasiveness of such procedure, laparoscope- or needle-based OCT probe would produce much smaller incision size and shorten the recovery period post-surgery [93].

Chapter Conclusion

We demonstrated that OCT is able to evaluate living kidney microstructure and function in chronic kidney disease (CKD) models of Munich-Wistar rats. The preliminary results show that OCT can provide more information beyond proteinurea and serum creatinine. In addition, we presented a computer-aided diagnosis software that can automatically detect, measure, and quantify features in the kidney images that are related to CKD progression. The software is necessary to process a large number of OCT images in a short period of time. Our ability of discern pathological changes in rat models of CKD is important in view of our recent studies showing that OCT can be used to evaluate the status of human kidneys in the operating room and its potential use in conjunction with renal biopsies to evaluate kidney histopathology in patients.

Chapter 4 OCT Hand-held Probe Indoor Localization

Introduction

OCT imaging instrument is currently undergoing the transition from traditional desktop scanners to portable hand-held scanners. The portable hand-held scanners are easier for medical professionals to handle due to its light weight; it can be adopted to more applications due to its 6 degrees of freedom comparing to the 2 degrees of freedom of the traditional desktop scanners. Hand-held OCT scanners would very likely become the major OCT instruments for clinical applications in the next few years.

In the previous chapter, the kidney images of murine model and human model are collected using different types of handheld OCT scanners. During analyzing the medical image statistic result, we found that the lack of scanning location information of the organ could potentially bias the statistic result. During a clinical OCT imaging sessions, medical professionals holds up the scanners and traverse it around the entire exposed regions of the organ for several times while the 2D cross section images are captured and stored continuously. Medical professionals tend to hold the scanners above regions on the organ that displays interesting biological features for a long period of time and briefly scans through the rest of the organs that are normal or do not have distinctive features. The result is that the collected data would consist of a large number of images from organ regions that have abnormal features and only a small number of images from the organ regions that are healthy. While such human-in-the-loop process is completely natural and normal for medical professionals, the collected data is biased and is not a true reflection of the health status of the real organ. For example, we have seen medical professionals captured an image database that consists mostly images of regions with hypertrophy tubule while hypertrophy tubules actually only present in a small section of the entire organ. Statistical analysis of such database would produce a diagnosis that indicates the kidney is at late-stage chronic disease even most parts of the organ are healthy.

While it is possible to instruct the medical professionals to perform the scan in a certain way that images from the entire organ are evenly distributed, it has few difficulties. First, a human-in-the-loop process would always have a fair amount of errors; the integrity of the captured image database cannot be guaranteed. Second, considering the amount of stress and many tasks that keep the attention of the medical professionals during the clinical sessions, adding one more task to the list is hard to justify and would give the impression that the instrument is too complicated to adopt. We propose the solution that the OCT handheld scanner should be able to independently associate a location data to each 2D cross section images it captured. This solution requires to perform indoor localization of the OCT handheld probe.

Background

There are many ways to perform indoor localization. It includes inertial based solution, radio based solution and camera based solution. This section would provide a background overview of various indoor localization techniques. It would also discuss the advantages and disadvantages to apply those techniques in our application.

Inertial based Indoor Localization

Inertial based indoor localization has its long history and was first appeared in digital devices such as the pedometer. Inertial measurement is mostly carried out by an inertial measurement unit (IMU). Modern IMU includes three microelectro-mechanic systems (MEMS) devices, an accelerometer, a gyroscope and a magnetometer. The accelerometer measures the change of velocity in three dimensions; the gyroscope measures the angular velocity in pitch, roll and yaw; the magnetometer measures the heading of the unit with reference to the earth magnetic north. The advantage of inertial based indoor localization is that the solution is a complete independent solution by itself that can be mounted to the tracking object. The disadvantage of inertial based indoor localization is its lack of correction mechanisms.

It is currently impossible to use IMU alone to acquire high precision localization for more than a few seconds [94]. The small acceleration measurement error accumulates in the accelerometer would accumulate cubically in terms of time in positioning because distance is the double integration of acceleration. The angular positioning error accumulates in a similar way as well. In addition, the heading measurement from the magnetometer is very likely to be disturbed by the magnetic field from the surrounding environment, especially in the case of a clinical environment where many medical instruments are within close range. A prior knowledge of the kinetic model of the object could provide a correction mechanism for IMU to zero the accumulated error and improve the tracking accuracy. NavShoe is a pedestrian tracking product that performs pedestrian localization using a single IMU mounted inside the shoes. It detects the stationary stance walking phase and applies zero-velocity updates to Extended Kalman Filter (EKF) that filtering the output of the IMU [95]. Yuan el al. showed pedestrian tracking using 3 IMUs that mounted at waist, upper leg and lower leg of the tracking person. With the knowledge of human lower body movement kinetic model during walking, the output of the IMU is fused with the velocity estimation from the human body kinematics through a EKF in order to achieve a more precise estimation of the pedestrian location [96].

Radio based Indoor Localization

The most well-known radio based localization method is the global positioning system (GPS). Since radio based localization method is based on time-of-flight measurements, which requires the line of sight between the emitter and receiver, GPS based localization methods cannot be used indoor. Radio based indoor localization methods are essentially replacing GPS satellites with indoor radio emitters and it shares many fundamentals with the GPS technology.

A radio-based indoor localization solution consists of radio base stations and a radio receiver that mounted to the tracking object. A radio base station constantly broadcasts its current clock and its positions. For the radio receiver to compute its 3D location (3 unknown variables), it has to receive from at least three base stations at line of sight. Upon receiving the clocks and positions of the base stations, the receiver compares the clock difference with its own clock, computes the distances from each base station and uses triangulation to fix its location. Not receiving the base station broadcast at line of sight (multi-path) would cause the receiver to compute a distance longer than the actual physical distance, which in turn introduces errors in the localization.

Wi-Fi based indoor localization solution is one of the most popular and researched solution in recent years most because of the wide adoption of Wi-Fi based station both in private and public areas. Wi-Fi base stations broadcasts its signal strength and MAC address to the environment; a database storing the location of each MAC address can be accessed by the tracker in order to compute its location based on the received MAC addresses. The distance between the receiver and the base station could be inferred from the signal strength decaying (instead of clock difference). The localization error of Wi-Fi based solution is about 6-8 meters because of the existence of distinct locations with similar Wi-Fi signatures (similar to multipath problem in the case of GPS) [97]. Recent study shows leveraging the relative locations of other receiver units nearby could reduce the localization error of Wi-Fi based solution to 1-2 meters [97]. However, it would further increase the level of dependency of the infrastructure.

Ultra-wideband (UWB) radio based indoor localization solution can reduces the localization error of radio based solution down to 5-10 cm. UWB radio signal comparing to Wi-Fi radio signal has a wider spectrum and a narrower pulse width in time domain. The narrower pulse in time domain increases the precision of the time of flight measurement of the pulse, which is a better source for distance estimation than the signal strength decay [98]. However, since most Wi-Fi base station does not use UWB radio due to its lower data rate. Most UWB indoor localization solution relies on special hardware that needs to be installed to the environment, which are very expensive till today.

The advantage of radio based indoor localization solution in general is that it is a much more accurate solution than the IMU based solutions. The disadvantage of the radio based solution is that it relies on the existence of infrastructure hardware built into the environment. For a complicated environment with many occlusions, multiple radio base stations are necessary to ensure the line of sight coverage. The more precise of the system, the costlier is the environment setup. Radio based indoor localization solution is much practical for applications that tracks multiple objects indoor so that the tracking cost per device can go down tremendously. For OCT handheld probe tracking, such scenario does not apply most of the time. The cost for using radio based probe tracking would be too high for this technology to be widely adopted.

Camera based Indoor Localization

Visual odometry and visual simultaneous mapping and localization (visual SLAM) are two closely related areas that aims to localize the 6 DOF camera pose only based on the video captured by itself. In recent years, with the trend of augmented reality and virtual reality entertainment, a large amount of research has been conducted in visual odometry and visual SLAM to localize the mixed reality helmet indoor in real time with low cost hardware that is affordable by the consumer market.

Visual odometry is the process of estimating the motion of the camera through a sequence of monocular or stereo camera images. Visual odometry system matches the visual features that are observed by the camera from various locations and compute the motion of the camera from the perspective changes of those visual features. Visual odometry could be carried out either by monocular cameras, RGBD cameras or stereo cameras. The difference is that monocular camera based visual odometry can only produce odometry measurement in an arbitrary scale; RGBD camera and stereo camera could produce odometry measurement in metric scale because these cameras have range measurement for each pixel of the image. The first visual odometry system is purposed by Nister et al. where the location of the ground vehicle is estimated using the images captured by the monocular and the stereo cameras mounted on the vehicle [99]. The advantage of visual odometry system is that it is a compact, low-cost, independent and real-time system that can be used both for indoor and outdoor applications. The disadvantage of such system is that similar to many camera based solution, the environment requires to have constantly good lighting and the scene captured by the camera should have enough unique visual features that are long-term static. For example, visual odometry system would fail to track the camera if the full camera scene is a white wall. In addition, visual odometry system works the best when the movement of the camera is smooth. Sudden camera motion could cause the visual odemetry system to lose tracking because of the motion blur and the minimal visual features that overlap between the frames during the sudden camera pose change. Visual SLAM not only estimates the location of the camera using various type of visual odometry, it also creates a map of the environment by estimating the location of the environment features relative to the camera location. SLAM in general can be implemented using various types of range measurement sensors, such as infrared sensors[100], sonars[101] or LASERs[102]. Visual based SLAM is more popular in general because CMOS cameras can provide a large amount of information at a relative low cost.

Discussion

In consideration of the ease of use, accessibility and accuracy, the optimal solution for an imaging probe tracking device should produce a tracking accuracy within 1-10 cm; the cost of the system should be relatively low for a high adoption rate; the tracking device should be relatively compact that it would not impact the ease of use of the imaging probe; the tracking system should be as independent of the environment and infrastructure as possible; no modification to the environment should be required for the add-on tracking capability. Last but not least, minimal training should be required for the medical professionals to use such device and the add-on of such device should not change the workflow of medical professionals.

In consideration of all the requirement for the optimal solution, tracking the motion of the probe using visual odometry stands out among other solutions. Since the indoor clinical environment is constantly under good lighting, minimal modification of the environment is required to adopt this solution. In our application, stereo camera based visual odometry is chosen over monocular camera based visual odometry. Monocular camera based visual odometry usually requires an initialization process because the first image only of the sequence could not provide depth information [103], [104]; the initialization process can be either pointing the camera to a structure with known depth or moving camera from side to side pointing to a planar surface. The initialization process is impractical for a system for clinical usage because surgical time is valuable. Furthermore, fail to initialize the system would cause the rest of the visual odoemtry data become erroneous. In addition, stereo camera based visual odometry can provide odometry measurement with accurate scale in comparison to the arbitrary scale from the monocular camera based visual odometry.

OCT Probe Localization using Visual Odometry

Hardware Add-on

A commercial stereo camera (ZED camera, Stereolabs Inc. San Francisco, CA) is mounted to the OCT handheld probe (Figure 34). This specific camera has a depth range from 0.5 - 20 meters and a stereo baseline (distance between the two cameras) of 1.2 cm. The viewing angle of the camera is set to point 90 degrees upwards from the direction of the OCT scanner. In real world usage, the viewing angle of the camera would not have an effect on the quality of the visual odometry as long as the viewing angle offset is fixed and known and there are enough visual features within the depth range of the camera. The camera is set to recording VGA footage of the environment during the OCT scanning session. The visual odometry data can be computed either in real-time during the scanning session or in post process. However, since there is no incentives and benefits to provide visual odometry data to the medical professionals during the scanning session. Only VGA footages from the two cameras are recorded together with the OCT images during the session. The probe locations for each OCT images are computed in post processing. It is worth to mention that the commercial stereo camera does come with a software development kit that providing visual odometry capability. However, the visual odometry capability in the software development kit is rudimentary and dependent on NVidia GPUs, which is not always available in a clinical setup. As a result, only the RGB images from the left and right cameras on the commercial stereo camera are used in this project. The functionalities provided in the software development kit are not used in this project.

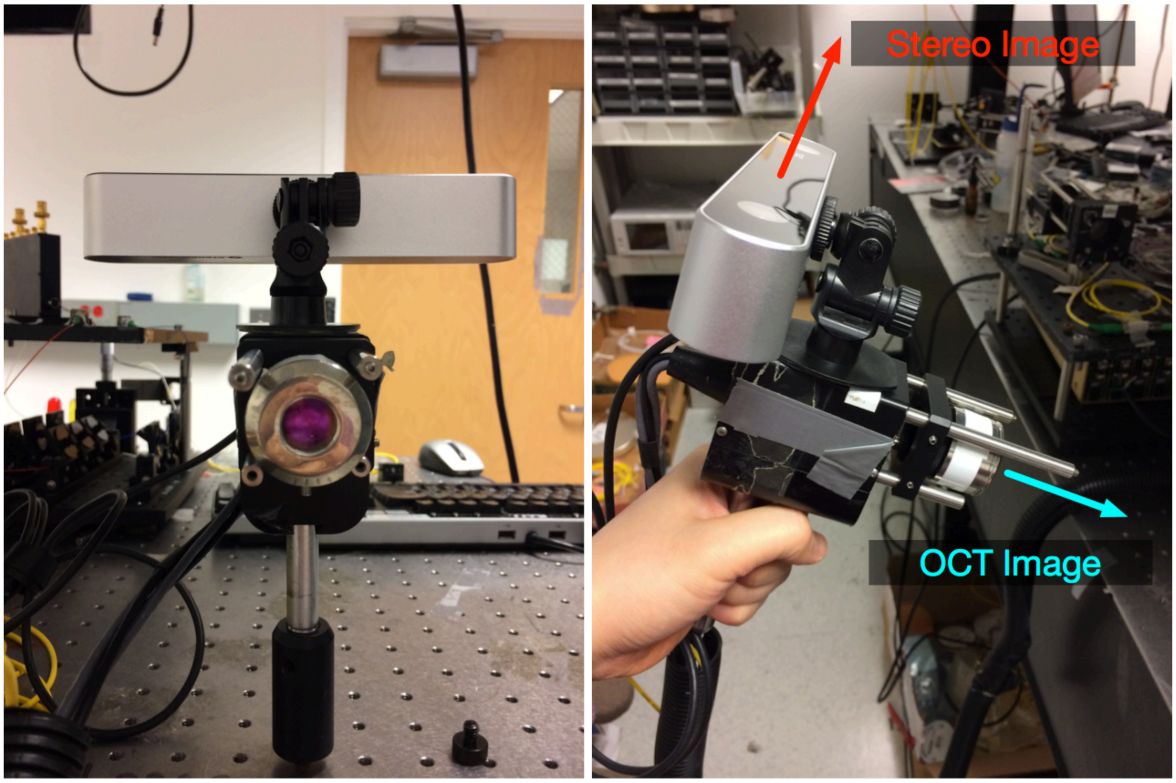


Figure 34. Augmented OCT handheld scanner with stereo camera

Visual Odometry Software

The visual odometry post processing software is based on ORB SLAM algorithm [105] in C++. ORB SLAM is a visual SLAM system that consists of visual odometry tracking, mapping and loop closing.

The fundamental visual feature descriptor of the ORB SLAM system is the oriented FAST and rotated BREIF (ORB) feature descriptor [106]. The ORB feature is both scaled and orientation invariant that it provides high likelihood of feature points matching between images captured from different view port

(Figure 35c). With stereo camera, the depth information of each pixel can be computed from the disparity map of the stereo pair (Figure 35b). With depth information, each ORB feature point is a 3D point referring to the current camera pose. As the camera pose is computed through visual odometry in real time, a 3D map consisting of ORB feature points can be generated in real time (Figure 35d). Visual odometry, similar to many other odometry algorithm, is computed based on the prior sequence of measurements; small estimation errors in the previous measurements accumulate and cause the visual odometry measurement drift from the ground truth especially for long term sessions. Loop closing uses bag of words with ORB features to detect places that the camera has captured before, comparing the visual odometry measurement difference between the two frames and perform global bundle adjustment to correct the visual odometry measurements in the previous frames [107], [108]. However, visual odometry with loop closing is no longer real time because the corrected visual odometry measurement in the current frame is dependent on the measurement in the future frame. However, in our application, visual odometry with loop closing further improves the localization of the imaging probe in post video processing.

The raw stereo images are pre-processed before input to the SLAM system (Figure 35c). The image is first converted from RGB space to Lab color space [109]. Then contrast limited adaptive histogram equalization (CLAHE) is performed on the L channel to reduce the light differences and increase the contrast of the under-exposure and over-exposure image regions [110]. This process is to make sure more static visual features are picked out even in regions in the images that are not well represented due to lighting. Finally, a Gaussian blur with a small 5x5 kernel is applied to the image in order to reduce tiny artifacts caused by the CLAHE algorithm.

To visualize the quality of the mapping, 3D ORB feature points are plotted, which depicts the room shape of the environment (Figure 35d). In Figure 36, the trace of the OCT handheld scanner is plotted together with the environment mapping to show the relative scale of the movement of the handheld probe comparing to the environment. The localization accuracy of the software is characterized by comparing the algorithm output with the actual physical displacement of the imaging probe measured with linear translation stage, for multiple times. While the uncertainty of the localization measurement is gradually increasing with the displacement, the mean of the localization measurement grows linearly with the physical displacement of the imaging probe (Figure 35e).

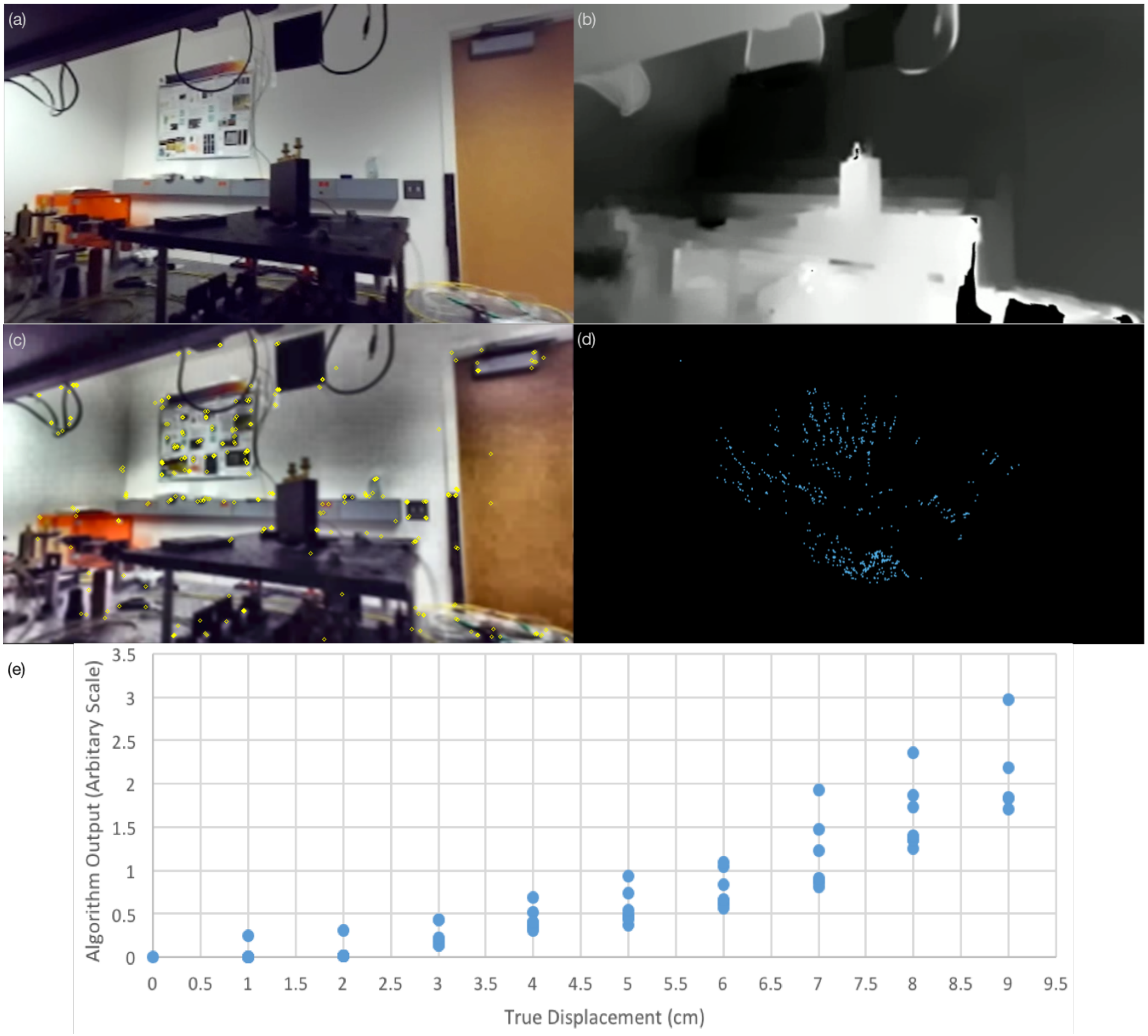


Figure 35. (a) RGB image from the left camera (b) depth image computed from the stereo image pair (c) ORB feature points overlaid on pre-processed RGB left image (d) 3D Map of the environment generated from the 3D ORB points (e) Characterization of the linearity of the localization algorithm

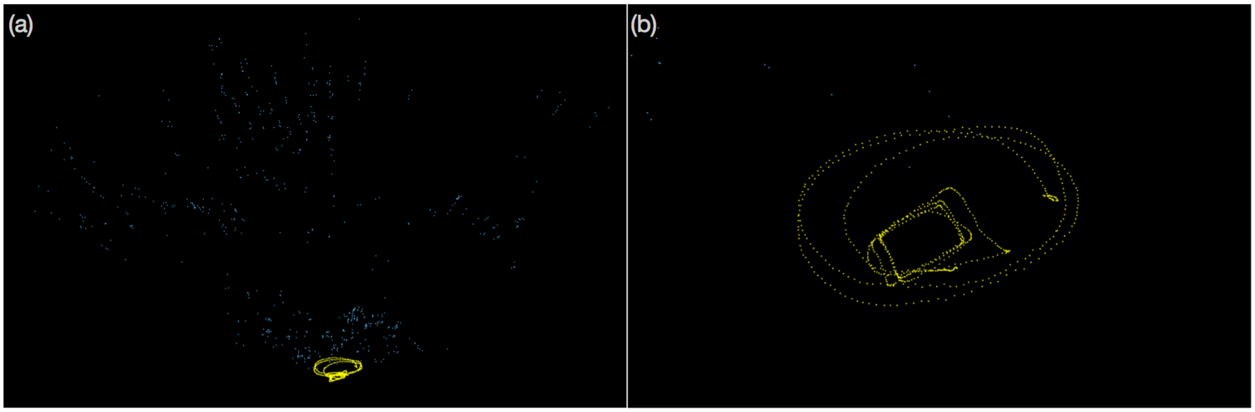


Figure 36. (a) Trace of the OCT scanner plotted with the environment map (b) trace of OCT handheld scanner in close

Visualize OCT Images in 3D

Since RGB images of the stereo camera and the OCT images are captured simultaneously, the imaging location of each OCT image can be computed from the RGB images captured at the same time stamp. In addition to save the imaging location of each OCT image, we added an extra layer of software to visualize the 2D OCT images in 3D virtual space (Figure 37). This software enables users, such as medical professionals to query and visualize the location of each frame (Figure 37a-b). We believe this visualization tool would be helpful for medical professionals to review OCT scan record. The software is implemented in C++ using OpenGL, OpenCV and Point Cloud Library.

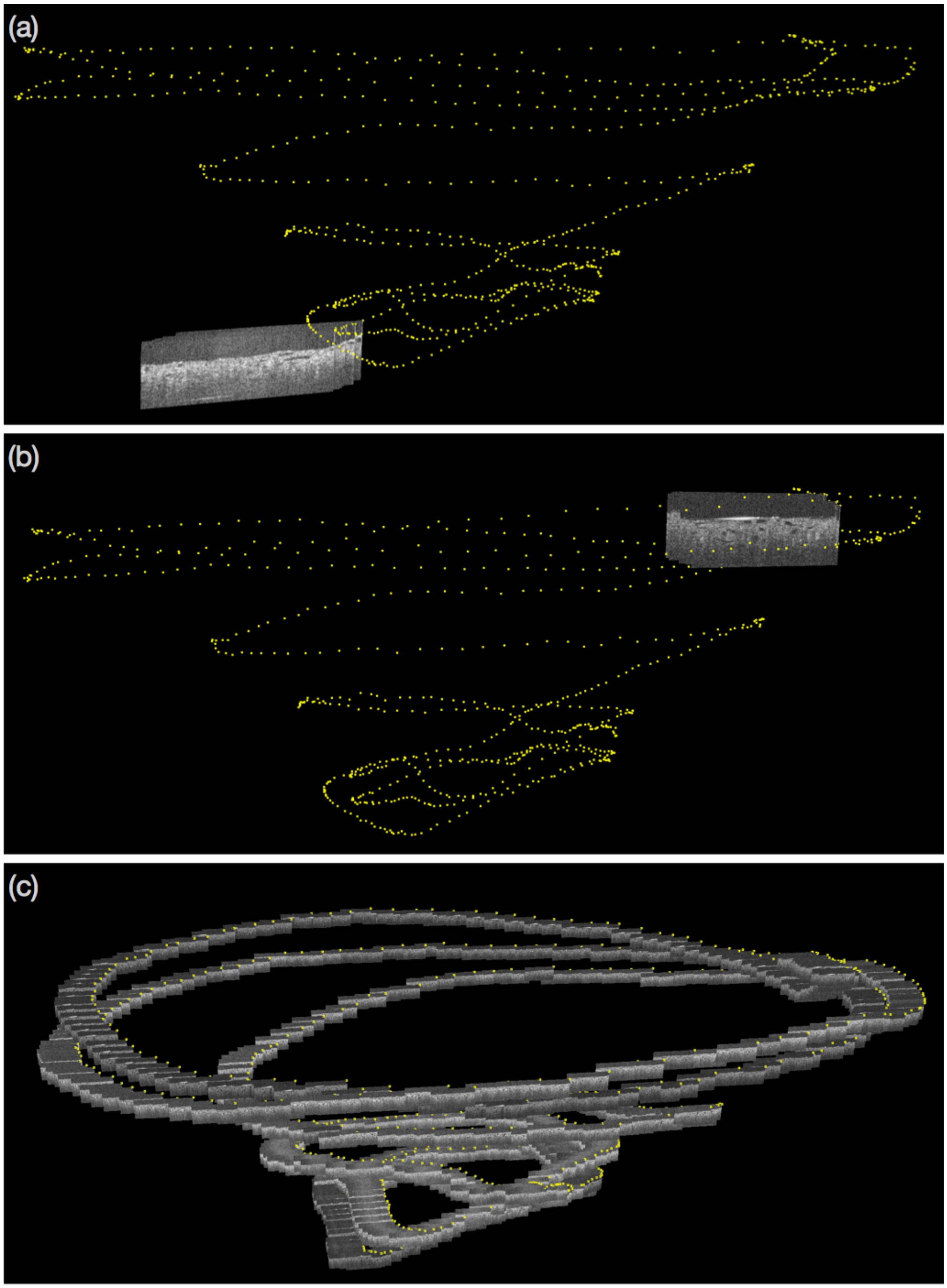


Figure 37. (a-b) the visualization tool enables user to query and visual the relative location of each OCT image (c) the whole scanning session and the location of each frame plotted in one image

Chapter Conclusion

In this chapter, we implemented a software that could compute and visualize the location of the OCT imaging probe in post processing. The design and implementation of such system is practical in terms of the cost, ease of use and accuracy. We have demonstrated that our hardware and software implementation is able to track OCT imaging probe location at high accuracy and provide a simple to use visualization tool to review OCT 2D scan images in 3D. We believe that the extra probing location information of each OCT 2D scan provided by the system would give the medical professionals a better understanding and visuals of the organ health status.

Chapter 5 Conclusion and Future Work

This dissertation showcases three projects I have worked during my doctoral research. Portable near infrared fluorescence imaging project aims to provide a low cost and portable NIRF imaging device that is more accessible for patients in the under-developed regions. The result shows the that although imaging quality produced by the device is inferior than the desktop counterpart. Many clinical applications are still able to adopt this technology and become more widely available to the community. Computer-aided diagnosis of chronical kidney disease using OCT and machine learning explores the possibility of providing more advanced diagnosis tools for the medical doctors with software that integrated the latest image analysis technology. We hope the software, including the user interface at the front and the computational model behind the scene, would prompt the medical professional community to think more about integrating machine learning and big data powered tools to their day-to-day workflow. The indoor localization of OCT project showcases that by integrating latest technology from the AR/VR industry to medical instruments, more layers of data is able to be extracted from the same procedure without modify the workflow of the medical professional. The extra layers of data and the visualization it brings can provide more meaningful insights to the data for the medical doctors during the record review process, which would lead to more accurate diagnosis. These three projects are great demonstrations of successfully apply, refine and transfer the latest technologies from electrical and computer engineering to bio-medical clinical applications.

Bibliography

[1] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.

[2] D. W. Kaiser, R. A. Harrington, and M. P. Turakhia, "Wearable Fitness Trackers and Heart Disease," *JAMA cardiology*, vol. 1, no. 2, pp. 239–239, 2016.

[3] Y. Chen, J. E. Argentinis, and G. Weber, "IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research," *Clinical therapeutics*, vol. 38, no. 4, pp. 688–701, 2016.

[4] D. Roter, "The enduring and evolving nature of the patient–physician relationship," *Patient education and counseling*, vol. 39, no. 1, pp. 5–15, 2000.

[5] B. Wang, J. Wang, Q. Wang, Y. Chen, and T. J. Pfefer, "Performance Testing of a Mobile-Phone-Based Near-Infrared Fluorescence Imaging System," p. JTu2A. 75, 2017.

[6] B. Wang, H.-W. Wang, H. Guo, E. Anderson, Q. Tang, T. T. Wu, R. Falola, T. Smith, P. M. Andrews, and Y. Chen, "Optical coherence tomography and computer-aided diagnosis of a murine model of chronic kidney disease," *J. Biomed. Opt*, vol. 22, no. 12, p. 121706, 2017.

[7] E. M. Sevick-Muraca, "Translation of near-infrared fluorescence imaging technologies: emerging clinical applications," *Annual review of medicine*, vol. 63, pp. 217–231, 2012.

[8] W. Stummer, U. Pichlmeier, T. Meinel, O. D. Wiestler, F. Zanella, H.-J. Reulen, and A.-G. S. Group, "Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial," *The lancet oncology*, vol. 7, no. 5, pp. 392–401, 2006.

[9] J. C. Rasmussen, I.-C. Tan, M. V. Marshall, C. E. Fife, and E. M. Sevick-Muraca, "Lymphatic imaging in humans with near-infrared fluorescence," *Current opinion in biotechnology*, vol. 20, no. 1, pp. 74–82, 2009.

[10] N. J. DAVID, E. W. NORTON, J. D. GASS, and R. SEXTON, "Fluorescein retinal angiography in carotid occlusion," *Archives of*

neurology, vol. 14, no. 3, pp. 281–287, 1966.

[11] J. Wang, J. Coburn, C.-P. Liang, N. Woolsey, J. C. Ramella-Roman, Y. Chen, and T. J. Pfefer, “Three-dimensional printing of tissue phantoms for biophotonic imaging,” *Opt. Lett., OL*, vol. 39, no. 10, pp. 3010–3013, 2014.

[12] J. Wang, P. Ghassemi, A. Melchiorri, J. Ramella-Roman, S. A. Mathews, J. Coburn, B. Sorg, Y. Chen, and J. Pfefer, “3D printed biomimetic vascular phantoms for assessment of hyperspectral imaging systems,” pp. 932508–932508–9, 2015.

[13] P. Ghassemi, J. Wang, A. J. Melchiorri, J. C. Ramella-Roman, S. A. Mathews, J. C. Coburn, B. S. Sorg, Y. Chen, and T. J. Pfefer, “Rapid prototyping of biomimetic vascular phantoms for hyperspectral reflectance imaging,” *J. Biomed. Opt*, vol. 20, no. 12, pp. 121312–121312, 2015.

[14] P. Yager, G. J. Domingo, and J. Gerdes, “Point-of-care diagnostics for global health,” *Annu. Rev. Biomed. Eng.*, vol. 10, pp. 107–144, 2008.

[15] H. Zhu, S. O. Isikman, O. Mudanyali, A. Greenbaum, and A. Ozcan, “Optical imaging techniques for point-of-care diagnostics,” *Lab on a Chip*, vol. 13, no. 1, pp. 51–67, 2013.

[16] A. Hagen, D. Grosenick, R. Macdonald, H. Rinneberg, S. Burock, P. Warnick, A. Poellinger, and P. M. Schlag, “Late-fluorescence mammography assesses tumor capillary permeability and differentiates malignant from benign lesions,” *Opt. Express, OE*, vol. 17, no. 19, pp. 17016–17033, 2009.

[17] H. Abboud and W. L. Henrich, “Stage IV chronic kidney disease,” *New England Journal of Medicine*, vol. 362, no. 1, pp. 56–65, 2010.

[18] M. E. L. Nahas, “The global challenge of chronic kidney disease,” *Kidney International*, vol. 68, no. 6, pp. 2918–2929, Dec. 2005.

[19] A. S. Levey, J. Coresh, E. Balk, A. T. Kausz, A. Levin, M. W. Steffes, R. J. Hogg, R. D. Perrone, J. Lau, and G. Eknoyan, “National Kidney Foundation Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification,” *Ann Intern Med*, vol. 139, no. 2, pp. 137–147, Jul. 2003.

[20] L. A. Stevens, J. Coresh, T. Greene, and A. S. Levey, “Assessing Kidney Function — Measured and Estimated Glomerular Filtration Rate,” *New England Journal of Medicine*, vol. 354, no. 23, pp.

2473–2483, Jun. 2006.

[21] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, “Optical Coherence Tomography,” *Science (New York, N.Y.)*, vol. 254, no. 5035, p. 1178, Nov. 1991.

[22] J. M. Schmitt, “Optical coherence tomography (OCT): a review,” *IEEE J. Select. Topics Quantum Electron.*, vol. 5, no. 4, pp. 1205–1215, 1999.

[23] R. Reif and R. K. Wang, “Optical Microangiography Based on Optical Coherence Tomography,” in *Optical Coherence Tomography*, no. 45, Cham: Springer International Publishing, 2015, pp. 1373–1397.

[24] Y. Chen, P. M. Andrews, A. D. Aguirre, J. M. Schmitt, and J. G. Fujimoto, “High-resolution three-dimensional optical coherence tomography imaging of kidney microanatomy ex vivo,” *J. Biomed. Opt*, vol. 12, no. 3, pp. 034008–034008–7, May 2007.

[25] S. A. Boppart, J. Herrmann, C. Pitris, D. L. Stamper, M. E. Brezinski, and J. G. Fujimoto, “High-Resolution Optical Coherence Tomography-Guided Laser Ablation of Surgical Tissue,” *Journal of Surgical Research*, vol. 82, no. 2, pp. 275–284, Apr. 1999.

[26] P. M. Andrews, Y. Chen, M. L. Onozato, S.-W. Huang, D. C. Adler, R. A. Huber, J. Jiang, S. E. Barry, A. E. Cable, and J. G. Fujimoto, “High-resolution optical coherence tomography imaging of the living kidney,” *Laboratory Investigation*, vol. 88, no. 4, pp. 441–449, Apr. 2008.

[27] P. M. Andrews, H.-W. Wang, H. Guo, E. Anderson, R. Falola, and Y. Chen, “Optical Coherence Tomography of the Aging Kidney.,” *Experimental and clinical transplantation: official journal of the Middle East Society for Organ Transplantation*, vol. 14, no. 6, pp. 617–622, 2016.

[28] J. Wierwille, P. M. Andrews, M. L. Onozato, J. Jiang, A. Cable, and Y. Chen, “In vivo, label-free, three-dimensional quantitative imaging of kidney microcirculation using Doppler optical coherence tomography,” *Laboratory Investigation*, vol. 91, no. 11, pp. 1596–1604, Nov. 2011.

[29] Z. Zhi, Y. Jung, Y. Jia, L. An, and R. K. Wang, “Highly sensitive imaging of renal microcirculation in vivo using ultrahigh sensitive

optical microangiography," *Biomedical Optics Express*, vol. 2, no. 5, p. 1059, 2011.

[30] P. M. Andrews, H.-W. Wang, J. Wierwille, W. Gong, J. Verbesey, M. Cooper, and Y. Chen, "Optical coherence tomography of the living human kidney," *Journal of Innovative Optical Health Sciences*, vol. 7, no. 2, p. 1350064, 2014.

[31] M. L. Onozato, P. M. Andrews, Q. Li, J. Jiang, A. Cable, and Y. Chen, "Optical Coherence Tomography of Human Kidney," *The Journal of Urology*, vol. 183, no. 5, pp. 2090–2094, May 2010.

[32] A. Cable, A. Paek, C.-W. Chen, J. Jiang, M. L. Onozato, P. M. Andrews, Q. Li, R. Naphas, S. Yuan, and Y. Chen, "Automated quantification of microstructural dimensions of the human kidney using optical coherence tomography (OCT)," *Opt. Express, OE*, vol. 17, no. 18, pp. 16000–16016, Aug. 2009.

[33] P. M. Andrews and Y. Chen, "Using optical coherence tomography (OCT) to evaluate human donor kidneys prior to and following transplantation," *J Nephrol Ther*, vol. 4, no. 151, pp. 2161–0959.1000151, 2014.

[34] P. M. Andrews, M. Cooper, J. Verbesey, S. Ghasemian, D. Rogalsky, P. Moody, A. Chen, P. Alexandrov, H.-W. Wang, and Y. Chen, "Mannitol Infusion Within 15 Min of Cross-Clamp Improves Living Donor Kidney Preservation," *Transplantation*, vol. 98, no. 8, pp. 893–897, Oct. 2014.

[35] C. Templin, M. Jaguszewski, J. R. Ghadri, I. Sudano, R. Gaehwiler, J. P. Hellermann, R. Schoenenberger-Berzins, U. Landmesser, P. Erne, and G. Noll, "Vascular lesions induced by renal nerve ablation as assessed by optical coherence tomography: pre-and post-procedural comparison with the Simplicity® catheter system and the EnligHTN™ multi-electrode renal denervation catheter," *European heart journal*, vol. 34, no. 28, pp. 2141–2148, 2013.

[36] K. Barwari, D. M. de Bruin, E. C. C. Cauberg, D. J. Faber, T. G. van Leeuwen, H. Wijkstra, J. de la Rosette, and M. P. Laguna, "Advanced Diagnostics in Renal Mass Using Optical Coherence Tomography: A Preliminary Report," *Journal of Endourology*, vol. 25, no. 2, pp. 311–315, Feb. 2011.

[37] K. Barwari, D. M. de Bruin, D. J. Faber, T. G. van Leeuwen, J. J. de la Rosette, and M. P. Laguna, "Differentiation between normal renal

tissue and renal tumours using functional optical coherence tomography: a phase I in vivohuman study," *BJU International*, vol. 110, no. 8, pp. E415–E420, May 2012.

[38] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," *The British journal of radiology*, 2014.

[39] J. Liu, S. Pattanaik, J. Yao, E. Turkbey, W. Zhang, X. Zhang, and R. M. Summers, "Computer aided detection of epidural masses on computed tomography scans," *Computerized Medical Imaging and Graphics*, vol. 38, no. 7, pp. 606–612, Oct. 2014.

[40] A. Farag, Le Lu, E. Turkbey, J. Liu, and R. M. Summers, "A Bottom-Up Approach for Automatic Pancreas Segmentation in Abdominal CT Scans," in *Abdominal Imaging. Computational and Clinical Applications*, vol. 8676, no. 10, Cham: Springer, Cham, 2014, pp. 103–113.

[41] A. K. Jerebko, J. D. Malley, M. Franaszek, and R. M. Summers, "Multiple Neural Network Classification Scheme for Detection of Colonic Polyps in CT Colonography Data Sets," *Academic Radiology*, vol. 10, no. 2, pp. 154–160, Feb. 2003.

[42] S. Wan, H.-C. Lee, X. Huang, T. Xu, T. Xu, T. Xu, X. Zeng, Z. Zhang, Y. Sheikine, J. L. Connolly, J. G. Fujimoto, and C. Zhou, "Integrated local binary pattern texture features for classification of breast tissue imaged by optical coherence microscopy," *Medical Image Analysis*, vol. 38, pp. 104–116, May 2017.

[43] X. Qi, M. V. Sivak, G. Isenberg, J. E. Willis, and A. M. Rollins, "Computer-aided diagnosis of dysplasia in Barrett's esophagus using endoscopic optical coherence tomography," *J. Biomed. Opt*, vol. 11, no. 4, pp. 044010–044010–10, Jul. 2006.

[44] X. Qi, Y. Pan, Z. Hu, W. Kang, J. E. Willis, K. Olowe, M. V. Sivak, and A. M. Rollins, "Automated quantification of colonic crypt morphology using integrated microscopy and optical coherence tomography," *J. Biomed. Opt*, vol. 13, no. 5, pp. 054055–054055–11, Sep. 2008.

[45] H. R. Roth, Le Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations," in *Medical Image*

Computing and Computer-Assisted Intervention – MICCAI 2014, vol. 8673, no. 65, Cham: Springer, Cham, 2014, pp. 520–527.

[46] H. R. Roth, A. Farag, Le Lu, E. B. Turkbey, and R. M. Summers, " Deep convolutional networks for pancreas segmentation in CT imaging," *SPIE Medical Imaging*, vol. 9413, pp. 94131G–94131G–8, Mar. 2015.

[47] J. Liu, N. Lay, Z. Wei, Le Lu, L. Kim, E. Turkbey, and R. M. Summers, " Colitis detection on abdominal CT scans by rich feature hierarchies," *SPIE Medical Imaging*, vol. 9785, pp. 97851N–97851N–7, Mar. 2016.

[48] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[49] V. W. S. Lee, Y. Wang, X. Qin, G. Zheng, D. Mahajan, J. Coombes, G. Rangan, S. I. Alexander, and D. C. H. Harris, "Adriamycin nephropathy in severe combined immunodeficient (SCID) mice," vol. 21, no. 11, pp. 3293–3298, Sep. 2006.

[50] A. S. Levey, L. A. Stevens, and J. Coresh, "Conceptual Model of CKD: Applications and Implications," *American Journal of Kidney Diseases*, vol. 53, no. 3, pp. S4–S16, Mar. 2009.

[51] A. M. El Nahas and A. K. Bello, "Chronic kidney disease: the global challenge," *The Lancet*, vol. 365, no. 9456, pp. 331–340, Jan. 2005.

[52] A. S. Levey and J. Coresh, "Chronic kidney disease," *The Lancet*, vol. 379, no. 9811, pp. 165–180, Jan. 2012.

[53] R. L. Chevalier, "The proximal tubule is the primary target of injury and progression of kidney disease: role of the glomerulotubular junction," *American Journal of Physiology - Renal Physiology*, vol. 311, no. 1, pp. F145–F161, Jul. 2016.

[54] V. W. LEE and D. C. HARRIS, "Adriamycin nephropathy: A model of focal segmental glomerulosclerosis," *Nephrology*, vol. 16, no. 1, pp. 30–38, Jan. 2011.

[55] W. Metcalfe, "How does early chronic kidney disease progress?A Background Paper prepared for the UK Consensus Conference on Early Chronic Kidney Disease," *Nephrol Dial Transplant*, vol. 22, no. 9, pp. ix26–ix30, Sep. 2007.

[56] Y. Wang, Y. P. Wang, Y.-C. Tay, and D. C. H. Harris, “Progressive adriamycin nephropathy in mice: Sequence of histologic and immunohistochemical events,” vol. 58, no. 4, pp. 1797–1804, Oct. 2000.

[57] A. P. Evan, G. A. Tanner, W. T. T. A. O. P. Blomgren, and L. C. Knopp, “Proximal tubule morphology after single nephron obstruction in the rat kidney,” *Kidney International*, vol. 30, no. 6, pp. 818–827, Dec. 1986.

[58] D. W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” *Journal of the society for Industrial and Applied …*, 1963.

[59] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning Internal Representations by Error Propagation,” Sep. 1985.

[60] L. Shapiro, *Computer vision and image processing*. 1992.

[61] P. N. Belhumeur and J. P. Hespanha, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” *IEEE transactions on …*, 1997.

[62] M. A. Turk and A. P. Pentland, “Face recognition using eigenfaces,” *... Vision and Pattern Recognition*, 1991.

[63] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *Int J Comput Vis*, vol. 60, no. 2, pp. 91–110, 2004.

[64] Y. LeCun, L. Bottou, and Y. Bengio, “Gradient-based learning applied to document recognition,” presented at the Proceedings of the …, 1998.

[65] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” *Proc ICML*, 2013.

[66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” pp. 1097–1105, 2012.

[67] A. Krizhevsky and G. Hinton, “Learning multiple layers of features from tiny images,” 2009.

[68] J. Deng, W. Dong, R. Socher, L. J. Li, and K. Li, “Imagenet: A large-scale hierarchical image database,” *Computer Vision and …*, 2009.

[69] J. Yosinski, J. Clune, and Y. Bengio, “How transferable are features in deep neural networks?,” *Advances in neural*, pp. 3320–3328, 2014.

[70] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” in *Computer Vision – ECCV 2014*, vol. 8689, no. 53, Cham: Springer International Publishing, 2014, pp. 818–833.

[71] B. van Ginneken, M. B. Stegmann, and M. Loog, “Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database,” *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, Feb. 2006.

[72] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-I. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, “Development of a Digital Image Database for Chest Radiographs With and Without a Lung Nodule,” *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, Nov. 2012.

[73] Z. Li, Q. Tang, L. Jin, P. M. Andrews, and Y. Chen, “Monitoring Kidney Microanatomy Changes During Ischemia-Reperfusion Process Using Texture Analysis of OCT Images,” *IEEE Photonics Journal*, vol. 9, no. 2, pp. 1–10, 2017.

[74] Q. Tang, J. Wang, A. Frank, J. Lin, Z. Li, C.-W. Chen, L. Jin, T. Wu, B. D. Greenwald, H. Mashimo, and Y. Chen, “Depth-resolved imaging of colon tumor using optical coherence tomography and fluorescence laminar optical tomography,” *Biomedical Optics Express*, vol. 7, no. 12, pp. 5218–5232, 2016.

[75] Z. Ding, Q. Tang, C.-P. Liang, K. Wu, A. Sandlerc, H. Li, and Y. Chen, “Imaging Spinal Structures With Polarization-Sensitive Optical Coherence Tomography,” *IEEE Photonics Journal*, vol. 8, no. 5, pp. 1–8, 2016.

[76] Q. Tang, C.-P. Liang, K. Wu, A. Sandler, and Y. Chen, “Real-time Epidural Anesthesia Guidance Using Optical Coherence Tomography Needle Probe,” presented at the CLEO: Applications and Technology, Washington, D.C., 2014, p. AM2O.3.

[77] P. Viola and M. J. Jones, “Robust Real-Time Face Detection,” *Int J Comput Vis*, vol. 57, no. 2, pp. 137–154, 2004.

[78] J. H. Friedman, “On Bias, Variance, 0/1—Loss, and the Curse-of-Dimensionality,” *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 55–77, 1997.

[79] E. H. Land and J. J. McCann, “Lightness and Retinex Theory,” *J. Opt. Soc. Am., JOSA*, vol. 61, no. 1, pp. 1–11, Jan. 1971.

[80] F. C. Crow and F. C. Crow, “Summed-area tables for texture mapping,” *ACM SIGGRAPH Computer Graphics*, vol. 18, no. 3, pp. 207–212, Jan. 1984.

[81] D. Bradley and G. Roth, “Adaptive Thresholding using the Integral Image,” *Journal of Graphics, GPU, and Game Tools*, vol. 12, no. 2, pp. 13–21, Jan. 2011.

[82] P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient Graph-Based Image Segmentation,” *Int J Comput Vis*, vol. 59, no. 2, pp. 167–181, 2004.

[83] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, “Turbopixels: Fast superpixels using geometric flows,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 12, pp. 2290–2297, 2009.

[84] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[85] M. A. Turk and A. P. Pentland, “Face recognition using eigenfaces,” pp. 586–591, 1991.

[86] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[87] T. Finin, W. Murnane, A. Karandikar, N. Keller, J. Martineau, and M. Dredze, “Annotating named entities in Twitter data with crowdsourcing,” pp. 80–88, 2010.

[88] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” pp. 1–9, 2015.

[89] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” pp. 91–99, 2015.

[90] A. Vedaldi and K. Lenc, “MatConvNet - Convolutional Neural Networks for MATLAB,” *arXiv.org*, vol. cs.CV. 15-Dec-2014.

[91] N. P. Singh, A. Ganguli, and A. Prakash, “Drug-induced kidney diseases,” *JAPI*, 2003.

[92] L. Ritsma, E. J. A. Steller, S. I. J. Ellenbroek, O. Kranenburg, I. H. M. Borel Rinkes, and J. van Rheenen, “Surgical implantation of an

abdominal imaging window for intravital microscopy," *Nat Protoc*, vol. 8, no. 3, pp. 583–594, Feb. 2013.

[93] H.-J. Tan, J. S. Wolf Jr., Z. Ye, K. S. Hafez, and D. C. Miller, "Population Level Assessment of Hospital Based Outcomes Following Laparoscopic Versus Open Partial Nephrectomy During the Adoption of Minimally Invasive Surgery," *The Journal of Urology*, vol. 191, no. 5, pp. 1231–1237, May 2014.

[94] E. Foxlin, "Motion Tracking Technologies and Requirements," *Handbook of Virtual Environment Technologies*, pp. 163–210, 2002.

[95] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," *IEEE Computer graphics and applications*, vol. 25, no. 6, pp. 38–46, 2005.

[96] Q. Yuan and I.-M. Chen, "Localization and velocity tracking of human via 3 IMU sensors," *Sensors and Actuators A: Physical*, vol. 212, pp. 25–33, 2014.

[97] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Push the limit of WiFi based localization for smartphones," pp. 305–316, 2012.

[98] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE signal processing magazine*, vol. 22, no. 4, pp. 70–84, 2005.

[99] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," vol. 1, pp. I–I, 2004.

[100] F. Abrate, B. Bona, and M. Indri, "Experimental EKF-based SLAM for Mini-rovers with IR Sensors Only.," 2007.

[101] M. F. Fallon, J. Folkesson, H. McClelland, and J. J. Leonard, "Relocating underwater features autonomously using sonar-based SLAM," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 3, pp. 500–513, 2013.

[102] H. Kretzschmar, C. Stachniss, and G. Grisetti, "Efficient information-theoretic graph pruning for graph-based SLAM with laser range finders," pp. 865–871, 2011.

[103] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," pp. 834–849, 2014.

[104] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse,

"MonoSLAM: Real-time single camera SLAM," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[105] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[106] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," pp. 2564–2571, 2011.

[107] D. Gálvez-López and J. D. Tardós, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[108] R. Mur-Artal and J. D. Tardós, "Fast relocalisation and loop closing in keyframe-based SLAM," pp. 846–853, 2014.

[109] R. G. McGuire, "Reporting of objective color measurements," *HortScience*, vol. 27, no. 12, pp. 1254–1255, 1992.

[110] K. Zuiderveld, "Contrast limited adaptive histogram equalization," pp. 474–485, 1994.