Creating and climbing data mountains

Perspectives from optical imaging

Michael A. Choma, MD, PhD

Yale University
Diagnostic Radiology, Pediatrics,
Biomedical Engineering, and Applied Physics
michael.choma@yale.edu
Creating and climbing data mountains: perspectives from optical imaging

Michael A. Choma, MD, PhD
Yale University

Modern medical imaging is rooted in the ability to measure physical properties of tissues and organs as a function of space and time. When these spatially- and temporally-varying properties are displayed as images, features and lesions can be identified that are otherwise inaccessible to other testing modalities. Moreover, given that imaging is massively-parallelized measurement, there are rich opportunities for characterizing disease pathology in a truly quantitative manner. The daunting challenge ahead, however, is how to process large imaging datasets into actionable and reliable information. That is, after we generate a data mountain by performing an imaging study, what is the best way to climb down that data mountain?

Although they typically are not found in radiology suites, optical imaging modalities have an ever-growing importance in medical diagnostics. No longer limited to traditional microscopy, imaging modalities that use light can obtain information about previously inaccessible disease states. Moreover, imaging often is in vivo, non-destructive, and minimally- or non-invasive. In this seminar, I will discuss creating and climbing data mountains in optical imaging. I will focus on optical coherence tomography (OCT), which can be thought of as the optical analog of ultrasound imaging. In my 15 years in OCT research, technology development has matured from efforts to generate sufficient data to be clinically viable to generating data mountains of ever-increasing size. In conjunction with a changing healthcare landscape, the maturation of OCT provides an opportunity to frame important questions, including (1) how can imaging technology development be focused on information as well as data, (2) how can informatics guide image technology development, and (3) how can technology development be focused on generating shorter data mountains without sacrificing information content.
Acknowledgements

• Collaborators
  ‣ Hui Cao
  ‣ Brandon Redding
  ‣ Mustafa Khokha
  ‣ Rong Fan
  ‣ Vineet Bhandari

• Funding
  ‣ Yale Medical School
  ‣ Yale CHRC NIH 5K12-HD001401-12
  ‣ March of Dimes Foundation
  ‣ NIH R21EB016163
  ‣ NIH R01HL118419
  ‣ NIH R21HL125125
Outline

• Imaging, data, and information
  ‣ Data mountains
  ‣ Information as actionable knowledge

• The optical coherence tomography (OCT) story
  ‣ Data scarcity to data abundance
  ‣ Push for even more data

• Optical imaging and data mountains
  ‣ Planning climbs
  ‣ Have physics do climbing for you
Imaging, data and information

*Data mountains*

*Information as actionable knowledge*
My interests in informatics

• How can we use medical informatics to guide the development of pre-translational technologies that are based on fundamental, groundbreaking advances in the imaging and physical sciences?

• How can we use medical informatics to intelligently and rationally drive the translational process for new imaging technologies?
Imaging is not limited to radiology. Are there more general approaches to image informatics in a healthcare system?

- Traditional: radiology, pathology
- New: whole-body photography in dermatology, cameras and sensors everywhere

Climbing data mountains in the imaging sciences.
Bidirectional information flow

Technology

Bits

1 1 0 0 1 0 0
1 1 0 1 1 0 1
0 0 1 0 0 0 1
1 1 1 1 0 1 0
0 0 1 0 0 0 1

Atoms

Healthcare

Patients

Populations

Yale Biophotonics Laboratory
Imaging, data, and information

- Imaging
  - Massively parallel measurement

- Data
  - Documented features and quantities

- Information
  - Actionable knowledge
Thoughts about process and flow

Image Acquisition
- Single image series
- Bundled image series

Data Collection
- Pixel values
- Qualitative interpretation
- Quantitative interpretation

Information Synthesis
- Perspective
  - Prospective questions
  - Retrospective interpretation
  - Historical analysis
- Scope
  - Multi-silo
  - Within-silo
    - Single image series
Imaging

• Imaging is massively parallel measurement
  ‣ Spatiotemporal distribution of physical properties

• Qualitative feature/lesion identification
  ‣ Is there a tumor?
  ‣ Is there aortic stenosis?

• Quantitative imaging
  ‣ How big is the tumor?
  ‣ What is the pressure drop across the stenosis?
Data

• Documented features and quantities

• Measurement values and derived summary statistics

• Obtaining digital images
  ‣ Climbing up the data mountain

• Interpreting digital images
  ‣ Climbing down the data mountain
Data

• Getting down from Everest is just as important as climbing up Everest
  ‣ Focus on operators: what do they need to climb down?

• Can new imaging technologies be built with the climb down in mind?
  ‣ Focus on the object, not the radiation (D. Brady, Duke).
Getting up the data mountain is physics, engineering, and statistics.

Getting down the mountain is human and computer interpretation.

Are data mountains higher than they need to be? Focus on knowledge generation and not data collection. Focus on the patient, not the radiation.
Shorter data mountains

• Has a bioresorbable stent dissolved?
  ‣ Focus on using established imaging modalities or develop a new, targeted technology with a lower data burden.

• What is the flow performance of a ciliated biopsy specimen?
  ‣ Have the physics of flow do some climbing for you.

• How do best use patient-specific organs-on-a-chip?
  ‣ New kinds of imaging, data, and information.

• Do you need an image?
  ‣ Will lower-dimensional datasets suffice?
Information

- Actionable knowledge

- What subset of data is translated into high-priority information?

- What new imaging technologies should be developed to address gaps in information?
  - Focus on the operator: what are new things that tech people can develop?
• What are the most important pieces of information that drive the adoption of new technologies?

• Focus on the operator: what concrete, reliable, actionable factors should technology developers focus on?

• Can we formalize a translational process that focuses on actionable knowledge as a unifying principle?
The optical coherence tomography (OCT) story

Data scarcity to data abundance

Push for even more data
THE HUNT FOR RED OCTOBER
Depth = Delay x Speed
What is light? What is optics?

- Light is electromagnetic radiation (energy) that we can see (400 nm to 700 nm)

- X-ray : UV : Vis : NIR : IR : Microwave : Radio

- Optics is the generation, manipulation, and detection of light
Human hair \(\sim 100 \, \mu m\)

Red blood cell \(\sim 10 \, \mu m\)

Respiratory cilia \(\sim 0.3 \times 10 \, \mu m\)

Alveoli
\(\sim 200 \, \mu m\) wide \(\times \sim 10 \, \mu m\) thick

Pulmonary artery
16th->23rd gen \(\sim 300->50 \, \mu m\)
Drexler, Nat Med, 2001
Optical coherence tomography (OCT)

Depth = Delay × Speed of light

<table>
<thead>
<tr>
<th>Depth</th>
<th>Delay (light)</th>
<th>Delay (sound)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 m</td>
<td>~10 ns</td>
<td>~1 ms</td>
</tr>
<tr>
<td>1 mm</td>
<td>~10 ps</td>
<td>~1 µs ✓</td>
</tr>
<tr>
<td>1 µm</td>
<td>~10 fs ✗</td>
<td>~1 ns</td>
</tr>
</tbody>
</table>
Optical interferometry for ranging in tissue

Optical coherence tomography

1991 to 2003 OCT progress

• New area of imaging research

• Routine *in vivo* use in retina

• Commercially available imaging systems

• ~1 frame per second for *in vivo* retina

• Inadequate speed for comprehensive *in vivo* coronary and GI imaging
Better technology, more data

Austria, 2003
Performance of fourier domain vs. time domain optical coherence tomography
R. Leitgeb, C. K. Hitzenberger, and A. F. Fercher
Department of Medical Physics, University of Vienna, Waehringerstr. 13, A-1090 Vienna, Austria
Rainer.Leitgeb@univie.ac.at
http://www.univie.ac.at/neph/lamba

Duke, 2003
Sensitivity advantage of swept source and Fourier domain optical coherence tomography
Michael A. Choma, Marinko V. Sarunic, Changhuei Yang, Joseph A. Izatt
Department of Biomedical Engineering, Duke University, Durham, NC 27708
izatt@duke.edu

Boston, 2003
Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography
Johannes F. de Boer, Barry Cense, B. Hylk Park, Mark C. Pierce, Guillermo J. Tearney, and Brett E. Bouma
Harvard Medical School and Wellman Center for Photomedicine, Massachusetts General Hospital, 30 E blown Street, BAR 724, Boston, Massachusetts 02114

- 100 to 1000x speed increase
- Clinical utility in retina
  ‣ Anti-VEGF treatments
- Clinical viability of intracoronary and GI
  ‣ Not far behind: pulmonary
- ~$50k lab instrument
- Sustained technology development
  ‣ Lasers, Doppler, probes
Time Domain OCT
~1 million pixels per second

1000x

Fourier Domain OCT
~1 billion pixels per second

>10x
Parallel imaging

Image on a 2D digital camera

Object

Channels

Imaging as parallel communication

Yale Biophotonics Laboratory
Limitations of current light sources

Many dim communications channels
low spatial coherence

One bright communication channel
high spatial coherence
A new regime of power per mode
A new regime of power per mode

- **High Power per spatial mode**
  - **Few spatial modes**
    - Single Mode Laser
  - **Many spatial modes**
    - Chaotic Laser

- **Low Power per spatial mode**
  - **Few spatial modes**
    - SLD
  - **Many spatial modes**
    - Random Laser, Thermal Source, LED

---

Yale Biophotonics Laboratory
Speckle-free laser imaging using random laser illumination

Brandon Redding\textsuperscript{1*}, Michael A. Choma\textsuperscript{2,3†*} and Hui Cao\textsuperscript{1,4†*}
Low spatial coherence electrically pumped semiconductor laser for speckle-free full-field imaging

Brandon Redding, Alexander Cerjan, Xue Huang, Minjoo Larry Lee, A. Douglas Stone, Michael A. Choma, and Hui Cao

Departments of Applied Physics, Electrical Engineering, Biomedical Engineering, Yale University, New Haven, CT 06520; and Departments of Diagnostic Radiology and Pediatrics, Yale School of Medicine, New Haven, CT 06520

PNAS

Edited by David A. Weitz, Harvard University, Cambridge, MA, and approved January 1, 2015 (received for review October 13, 2014)
Optical imaging and data mountains

Planning climbs
Have physics do climbing for you
Planning climbs in ciliary physiology

How can imaging technology development be focused on information as well as data?

**Ciliary motion**
- 10 Hz beat frequency
- $10^3$ pixels per volume
- 10-100 volumes per beat
- 0.1 to 1 GHz pixel rate
- ~1 μm resolution
- Max pixel rates for serial confocal ~10 MHz
Planning climbs in ciliary physiology

How can imaging technology development be focused on information as well as data?
How can informatics guide image technology development?

- New pathophysiology
- New diagnostics
- Drop in the ocean?
- What can be actionable?
- Focus on the operators, that is, those who develop technology.
- Value proposition of new information
  - Outcome ÷ cost
HEALTH CARE REFORM AND THE FUTURE OF AMERICAN MEDICINE

Ezekiel J. Emanuel, M.D., Ph.D.

Vice Provost for Global Initiatives
Chair, Department of Medical Ethics and Health Policy
University of Pennsylvania
The Vicious Cycle of Biomedical Research

- The cost of health care is the largest driver of long term federal debt
  - Health care consumes an ever-increasing portion of the federal budget
- Why?
- Increasingly expensive, low-value technologies—which stem from research funded by the NIH
Bidirectional information flow

Technology

Bits

Atoms

Healthcare

Patients

Populations

Yale Biophotonics Laboratory
Planning climbs in ciliary physiology

How can technology development be focused on generating shorter data mountains without sacrificing information content?

\[ \mathbf{v} = v_x \mathbf{i} + v_y \mathbf{j} + v_z \mathbf{k} \]

*Three dimensions, three components*
Planning climbs in ciliary physiology

How can technology development be focused on generating shorter data mountains without sacrificing information content?

300 μm/s

0.22 μm/s
Mixing: have physics do climbing for you

Integrated readout of flow performance

1x viscosity

3x viscosity

Viscosity increased using HMW dextran
n=21 (1x), n=18 (3x)


Mixing speed ($t_{25-75}$)

- 1x relative viscosity: p<0.01
- 3x relative viscosity:

Yale Biophotonics Laboratory
Conclusion

• Fundamental challenges in imaging sciences
  ‣ Images $\rightarrow$ Data $\rightarrow$ Information

• Fundamental challenges in lung physiology
  ‣ Data/Information $\rightarrow$ New disease understanding

• Fundamental challenges in healthcare
  ‣ Bits/Atoms $\rightarrow$ Patients/Populations

May you live in interesting times