Human-Machine Perception and Assistive Technology

Dr. Zhigang Zhu (朱志刚)
Herbert G. Kayser Chair Professor of Computer Science
City College Visual Computing Laboratory (CcvcL)
Grove School of Engineering
City College of New York and
CUNY Graduate Center
http://www-cs.engr.ccny.cuny.edu/~zhu/

City College Visual Computing Lab

• Assistive Technologies
  – Assistive Navigation and Smart Living
  – Sponsors: NSF, VentureWell, NYSID
• Deep Learning
  – Facial/Brain Computing and Crowd Analysis
  – Sponsors: NSF, DHS
• Multimodal Sensing
  – Surveillance and Security
  – Sponsors: AFRL, AFOSR, ARO, DHS, NSF
• Smart Transportation Hub
  – Transportation, Security and Services
  – Sponsors: DHS, NSF
Bio-Sketch 作者简介

Ten-years at Tsinghua shaped the **content**, Three-years at UMass provided the **context**. Personal faith stirs up my **compassion**, and City College is the place for my **contributions**. Computing vision is my major, which applies to transportation, security and navigation. Multimodal research helps people in need, which includes human & machine intelligence. Cutting edge research needs the best efforts, so students and professors work in harmony. My wish is that students are better prepared, and are more willing to contribute to society.

Abstract 摘 要

視覺甚精細，人腦更神奇；其一若有障，生活成問題。科研要聯合，輔助需科技；多方來資助，用戶亦參與。高校做研究，啟發新創意；培養有用才，工作有意義。

創新最前沿，機腦並生理；感知加計算，移動聯雲際。幻境與遊戲，皆能更有益；不拘舊觀念，感知互代替。交通在改進，智能又可及；導航到戶內，深度來學習。

雙目加智能，單目可測距；全向有優勢，三維更迅捷。設計全身目，嘗試群幫眼；肌電能辨聲，舌頭看圖案。機器讀人臉，盲者識表情；人腦之奧妙，或許可探知？
Abstract

Human vision is truly delicate, and the human brain is more amazing;
If either of them has problems, daily lives are very challenging.
Assistive living requires emerging technology, which needs collaborating;
Funding support is important, but more critical is user engaging.
The goals of research at higher education are not only to inspire novelty;
But also to train the next generation workforce, to do good to society.

Emerging Frontier in Research & Innovation (EFRI), includes M3 Control;
Multimodal perception is key, and mobile/cloud computing is the backbone.
Both virtual reality and gaming can be more positive and beneficial;
If not limited by conventional wisdom, alternative perception can be done!
Public transportation needs to be improved, to be both smart and accessible.
Location-based services go indoors, deep learning can also be here found.

Outline

1. Introduction: Assistive Technology
2. State of the Art: Human-Machine Vision
3. Machine Vision Techniques
4. Alternative Perception Techniques
5. Facial Computing and Deep Learning
6. Platforms and Applications
7. Education and Training
8. A Summary
1. Introduction: Assistive Technology

Among the 7 billion people worldwide, nearly 300 million are visually impaired. One in 70 is on ASD, & they’re facing challenges in both mobility and communication. For BVI, ASD and other disabled, assistive technologies can offer great helps: In navigation, in face recognition and in finding places, emerging technologies are in need.

City College Visual Computing Lab, is jointed by Media and Neuroscience Labs. Plus Mechanical and Biophysics at Georgia Tech, five disciplines are working on this EFRI. Within the topic Man Machine and Motor Control, we can only discuss a few projects. Even though there is a long way to go, we hope to provide some inspiration to peers.


The functions of human vision are so amazing, yet the human brain is really a mystery. Ten thousand researchers put their talents on them, fruitful yet still below expectation. Vision navigation is cross-disciplinary, including computing, neuroscience and physiology. See, think and action are in one, assistive vision needs research collaboration.

Funding support is important for research, and user engagement is even more critical. International forums are inspiring, with insights of experts from various fields. Alternative perception has multiple approaches, in physical, digital and medical. Ear, tongue, hand and body can all be used, let alone to say machine vision and implants.
3. Machine Vision Techniques

• 3.1. Vision-based indoor navigation
  – 室內導航視覺技術

• 3.2. Smart sampling for binocular vision
  – 雙目視覺智能採樣

• 3.3. Mobile vision with single images
  – 手機單目三維視覺

Outdoor navigation can readily use GPS, but indoor navigation still need new technologies. Wi-Fi, magnetic, etc. have been used, but vision techniques are still the most attractive. Panoramic vision can be used with a mobile phone, localization done w/ image retrieval. Mobile 3D vision is more effective, modeling & positioning are in one and in real-time. There’re always limitations in single techniques, so multisensor fusion shall do better. Planning and mapping can be undertaken before you go, by simply using a floor plan. Multimodal sensor techniques are the key, and mobile computing is the front-end. Cloud computing do the more sophisticated, and emerging technologies benefit all.
3.2. Smart sampling of binocular vision

Research in binocular vision is the most popular all times, and it can be used both in & out. It also serves as a window to understanding human perception, thus having hi-sci values. Unfortunately many indoor scenes are textureless, which troubles stereo correspondences. Researchers are always wondering how humans do the job: maybe we are using contents? If an image can be segmented into natural patches, then matching can be made easier. Even better is to provide 3D measures everywhere, thus providing basis for smart sampling. Priority can be given to the closest, and animated obstacles can also be very helpful. We believe smart data selection is the key, for an effective low-res retinal implants.

3.3. Mobile vision with single images

RGB-D camera uses IR for a short distance, & binocular vision has challenges in matching. A human being can estimate distances with a single eye, why not a machine too? Cameras are a default in smartphones, whose computing power is also stunning. It works in both indoors and outdoors, without the worries of distance and illumination. The best scenarios are corridors and streets, where VPs can be found for parallel lines. Both orientation and large distances can be obtained, & real-time computing is also for sure. Perceptual organization is the key principle, and HT and J-Linkage have been tried. Both corridors and doorways can be detected, which are very useful for indoor navigation.
4. Alternative Perception Techniques

• 4.1. Whole-body haptic eyes
  - 全身目的嘗試
• 4.2. Crowd-sourcing human vision sensors
  - 群幫眼導航初試
• 4.3. Tongue-eye and Electromyography (EMG)-ear
  - 舌頭看和肌電圖聽

Whole-body eyes are recorded in Revelation, and skin reading is also described in sci-fi. In the real-world researchers indeed tried haptic reading, with chest, back and even tongue. Reading brightness is rather confusing, why not trying depths and ranging? If users can be emerged into a field of distances, with haptic strengths to give the feeling! Simulated in virtual environments, and over thousand haptic cells can be on a body; But it would be a problem with a real person since vibrations will be interfering. Thus we designed modular haptic eyes with BLE connections, and put them on body to test. Both numbers and locations can be optimized, which is enjoyable for students and users.
4.2. Crowd-sourcing human sensors

Rapid advances’ve been made in machine vision, for navigation, obstacles and landmarks. Driverless cars rely on sensors but they are still not ready to be worn on a human body.

Around the globe millions of web surfers are online 24/7, both old and young; We witnessed the emerging crowdsourcing applications: Why not for assisted navigation?

Users only need to broadcast their cam video online, so crowd eyes can help; But data filtering and aggregation are a must, and real-time directions are essential.

We envision that users can cross streets safely, and crowd can tour the world for free; Further multimedia Big Data can be collected, for research in man-machine intelligence.

4.3. Tongue eye and EMG ear

Products for retinal implants are already there, and tongue-reading is also in trials; EMG can be used for speaker identification, and EEG might read human mind.

Impairment in vision and hearing are challenging, in both mobility and communication. Alternative perception is one of the approaches, which can also help understand brain.

EMG has been tried to identify speakers, using a deep belief network. Speaker ID can be obtained with a single word, comparable to using acoustic signals.

We also tested figure-reading with tongue, and found that this was very confusing; However it can readily tell directions, which’d be used for hand- & ear-free navigation.
5. Facial Computing and Deep Learning

Human emotions are written on faces, therefore facial recognition is vital for social lives. Unfortunately BVIs cannot see, and ASD friends also have challenges. We studied facial computing via deep learning, & a dataset from Web is the 1st step. CNNs are designed as the facial engine, and real-world data are collected via gaming. Multimodal ST features are integrated, to achieve the top 5 performance among 60. Both learned and hand-crafted features are used, and multi-kernel SVM is most effective. Action Units are the core, and learning transfer improves learning efficiency. Attention-coding is very effective, and LSTM net further improves the performance.

6. Platforms and Applications

• 6.1. GIVE-ME VISION: a platform
  – Gamification In Virtual Environment for Multimodal Evaluation
  – 虛擬技術助人遊戲技術

• 6.2. SAT-Hub: an Application
  – Smart and Accessible Transportation Hub
  – 智能可及交通樞紐
6.1. GIVE-ME VISION: a platform

Gaming can be good or harmful, and virtual reality (VR) are more attractive and lustful. Both can be better used for assistive vision, GIVE-ME VISION can be the first good step. Various sensors can be simulated, facilitating both rapid prototyping and low costing. Navigation training in VR can both be entertaining and provide peace of mind. The key is VR designs shall be close to reality, better directly mapped to real experience. Some initial tests have been carried out, showing GIVE-ME platform is effective. Tongue-seeing is simulated first, crowd-sourcing comes the second; Followed by the test of whole-body eyes, and the PhD graduate enjoyed it day and night!

6.2. SAT-Hub: an application

Publication transportation is vital in a big city, and IT really changes the landscape of ITS. However this creates more challenges for disabled, making digital divide even more obvious. Smart and Accessible Transportation Hub, is a collaboration between CUNY & Rutgers. Security, Transportation and Services, three are in one to serve all citizens. Digital 3D modeling is for the environment, crowd analysis is for the pedestrians. Multi-sensors are used for localization, customized guidance is provided both in & out. Partnerships are among government, industry & academia, service institutions are also in. User engagement is an indispensable part, and Smart Cities apply to anywhere in the world.
7. Education and Training

• 7.1. Capstone Project Design on Assistive Tech
  – 輔助技術本科實習

• 7.2. REM：Research Experience and Mentoring
  – 研究經歷指導

7.1. Capstone/Senior Design on AT
  輔助技術本科實習

The joint Cs/CpE/EE Senior Design course, has been running for six years. The theme is Assistive Technology (AT), over 120 students have graduated. It all started with assistive vision, and then smart living was also included. Further AT covers challenges in hearing, motion and memory, ASD and elderly are also on. Each project starts with users’ real need, & students are motivated via learning by doing. Collaborating with multiple institutions, for both education and social services. Students at City are truly creative, in technology, entrepreneurship and research. A number of grand prizes have been won, and UG students can also publish papers!
7.2. Research Experience and Mentoring

REM: 研究經歷指導

REM was a pilot program at NSF EFRI, and it focuses on underrepresented groups. The CCNY REM has graduated over 70 RPs, and achievements are visible in various levels. RPs included high schoolers, HS teachers, college students & community college professors. Formal interviews were carried out, and PhD students learned to be better mentors. Ten weeks were spent in a team environment, with daily interaction and discussions. Weekly presentations are required in all meetings, and most PRs also attended ERN. RPs and mentors were both advanced, and HS teachers started to offer new courses. CC students transferred to senior colleges, & HS students attended prestigious colleges.

8. A Summary 小結

Binocular vision is made smart, and monocular vision can better estimate distances; Omnidirectional vision sees everything, and 3D vision is more direct and effective. Whole-body eyes have been designed, and crowd-sourcing human sensors are tried; EMG can be used to recognize speakers, even tongues may be able to read images. Smart machines read facial expressions, so the blind can see emotion of others; The mystery of the human brain, may be further explored.
CcvC L Research Grants (Total >$5.5M)
Prof. Zhigang Zhu, CS/CCNY

• **AFRL, AFOSR, ARO and DARPA**
  – Exploitation and Fusion for DoD missions
  – Multimodal Sensing for ISR applications

• **DHS SRT and NSF S&CC**
  – DHS SRT Crowd Analysis (2015), and 3D+Crowd+Service (2016)
  – NSF Smart and Connected Community (2017)

• **NSF EFRI (Emerging Frontiers in Research and Innovations)**
  – Man, Machine and Motor Control (M3C) for Visually Impaired, $2M
  – NSF Research Experience and Mentoring (REM), $420K

• **NSF I-Corps (Innovation Corporation) and SBIR**
  – Transforming 2D Video into an Interactive 3D Viewing Experience
  – Vista Wearable, Inc. (startup by three students)

• **Industrial and Local Supports**
  – Smart Offices, Urban & Homeland Security
  – 3D Visualization and Modeling

• **VentureWell (formerly NCIIA)**
  – Novel Ideas for Vital Applications
  – Course and Program Development (Senior Design)

---

References


References


[17] Q. Chen, M. Khan, C. Tsangouri, C. Yang, B. Li, J. Xiao and Z. Zhu, CCNY Smart Cane, IEEE-CYBER 2017-The 7th Annual IEEE Int. Conf. on CYBER Technology in Automation, Control, and Intelligent Systems, July 31 – August 4, 2017, Hawaii, USA


[27] The Holy Bible, Revelation, Chapter 4 Verse 8.


[34] F. Abtahi, W. Li, Z. Zhu, T. Ro, Multimodal Speaker Recognition using Deep Belief Networks, CVPR Women in Computer Vision Workshop, 2015

References


References


