Acceptance speech

20 June 2024

Takeo Kanade, awardee in the Information and Communication Technologies category (16th edition)

I am most honored to receive the prestigious BBVA Foundation Frontiers of Knowledge Award in the Information and Communication Technologies category.

Computer Vision is the field that gives the abilities of "eyes" to computers and robots. I have had a chance to witness and participate in the tremendous advancement of the field from its 1970's infancy to today's indispensable and ubiquitous technologies.

Humans are superb vision machines. The early approaches to computer vision were heuristic and ad hoc, relying mostly on researchers' introspections. One might characterize them as "Let's program what I think I am doing" approaches. However, since visual data is the result of optical, geometrical and physical processes, I felt we should focus, instead, on underlying mathematical and scientific models. With that rigorous approach, my students and I were able to create several foundational computer-vision algorithms. One example is an algorithm commonly called the Lucas-Kanade Optical Flow, which turns out to be the basis for almost all video processing, such as motion tracking and video compression. Another is the Tomasi-Kanade Factorization algorithm, which became one of the first working algorithms for the long-defying problem of reliably reconstructing a 3D scene structure from a set of images.

Throughout my career, I have enjoyed working on a host of practical problems and creating solutions for them. The human face was one of my favorite topics. My PhD thesis work back in 1973 at Kyoto University, Japan, is recognized as the first computer program that automatically localized facial features, such as nose, eyes, and mouth, in digital face images to be used for face recognition. It could successfully process 800 images, modest by today's standards, but an uncharacteristically large number at that time. Computer face recognition is common today in everything from cellphone logins to passport control at airports. Then, in the mid 1990's, I demonstrated reliable face detection by using a neural-net machine-learning algorithm, a technology omnipresent in our cellphones today. I moved on to automated facial expression in the 2000's. Working with psychology researchers, our team developed methods for recognizing detailed micro facial expressions together with head and body

movements. They are now used for enriching human-robot interaction as well as for medical and psychiatric diagnosis.

Autonomous driving is another of my favorite topics. In the mid 1980's, I founded a project for it at Carnegie Mellon University. We developed a spectrum of capabilities for vision-based autonomous driving, including lane following, obstacle detection, car and pedestrian detection, and parallel parking. They were installed on a camera-equipped computer-controlled vehicle named NAVLAB, short for Navigation Laboratory. In 1995, its fifth generation NAVLAB-5 accomplished the "No Hands Across America" freeway drive from Pittsburgh to San Diego, approximately 5000 kilometers with 98.2% autonomous steering control. This was an important milestone in the history of driverless driving. During the drive, a Texas state policeman approached our vehicle, apparently suspicious of the "Nobody Onboard" sign attached to it. However, upon discovering that it was an experimental autonomous driving vehicle, he offered to accompany us to the state border for safety.

"Virtualized reality" is a concept that I coined in the mid 1990's. It creates a full space-and-time virtual representation of a real-world dynamic event by using a large number of cameras, and allows users to freely view it from any alternative viewpoints. For its study and development, I built a one-of-a-kind dome covered by many cameras, first with 50, then with as many as 400. In 2001, at the Super Bowl XXXV, the "EyeVision" system with 33 robotic-controlled cameras was demonstrated, broadcasting movie-"Matrix"-like spinning replays of a play to more than 50 million viewers worldwide. During the broadcast, I appeared briefly and explained how the system worked, which earned me the fun title ¨The only professor that has ever appeared on the Super Bowl¨. Today, the concept has become common and we observe similar media for all kinds of televised events.

Computer vision is now used everywhere; from our daily lives to space exploration to factory floor to medicine. Yet we have only scratched the surface of its potential. Far better capabilities and more applications are possible. We could even make the invisible visible. With advanced sensors, high-performance computers, and AI-learning algorithms, the computer vision field is in the middle of a "perfect storm," in a good sense, of course.

Finally, I must express my deepest gratitude to my numerous students, collaborators and colleagues. Without them, none of the computer vision systems I have described would have been possible. I thank my wife Yukiko, who is with me here today, and my two children Shunichi and Sayaka, both of whom unfortunately passed away young. They gave me their love and support, allowing me to pursue my research day and night.