Interview with Sakuya Morimoto, COO of Exvision Inc.

High-Tech Venture Begun in University Lab Enhances the Quality of Robots

By *Japan SPOTLIGHT*

Ishikawa-Watanabe Laboratory, at the Graduate School of Information Science and Technology at the University of Tokyo, has worked on a technology called High Speed Vision Technology for 25 years and delivered a large number of results. This technology has an expansive range of applications, including robots and driverless cars. Today we often see university labs as the source of new businesses, and this is a case in point, as the researchers in the lab have made this technology into a business. Exvision Inc., a company founded by a lab professor at the graduate school, Dr. Masatoshi Ishikawa, has been working on expanding applications of its university-originated technologies into various products and markets, and turning the venture into a global one. *Japan SPOTLIGHT* had an interview with Sakuya Morimoto, COO of Exvision Inc.

Company Introduction

JS: Exvision Inc. is a venture started in 2009 from the lab of the Graduate School of Information Science and Technology at the University of Tokyo. Could you briefly explain how the university lab started this new business?

Sakuya Morimoto, COO of Exvision Inc.

Morimoto: In image processing 20-25 years ago, all the researchers concentrated on work to produce a clear image by increasing pixel numbers or improving temporal resolution with instruments like the Change Coupled Device (CCD) image sensor or Complementary Metal Oxide Semiconductor (CMOS) image sensor. However, Prof. Masatoshi Ishikawa, the founder and CTO of Exvision who was working on computer vision at the University of Tokyo, tried to improve the temporal resolution instead of spatial resolution. He thought that computer vision will benefit more from increasing the number of images per second instead of pixel numbers, and so pursued the study of High Speed Vision Technology (HSVT). It had been theoretically self-evident that a computer could achieve this by increasing the number of frames per second, but it was so difficult to implement as a complete working system. No one has ever publicized as many working high-speed computer vision systems as Prof. Ishikawa has done.

Even now, 20 years after he started, there are very few companies specializing in this high frame rate image processing. The results of this technology are truly remarkable. We can now witness various eye-opening demonstrations of HSVT, including a bi-pedal running (not walking) robot, a camera to track a fast-moving ping pong ball, and an air-floating display that you can interact with using your fingers.

Prof. Ishikawa, while researching this technology, was in collaboration with some semiconductor-producing companies and Mr. Norio Tabuchi was the CEO. Prof. Ishikawa asked him if they could start up a venture for application of this new technology to the business. That was our starting point when Mr. Tabuchi accepted his offer.

How the Technology Works

JS: As we are non-experts on technology, could you explain what is distinctive about your technology? In particular, we are interested in your "gesture system" technology.

Morimoto: Several companies have been successful in producing camera technology that recognizes gestures and making it marketable. The most renowned is "Kinect", sold as an accessory for "Xbox 360", a Microsoft gaming machine. This was put on the market in 2010. It is probably the first product for consumers that recognizes gestures by 3D images. It has achieved remarkable success with record sales of more than 20 million units thanks to strong sales promotion efforts by Microsoft and its overwhelmingly cheap price in comparison to the products of preceding technologies.

Prof. Ishikawa, seeing this success, thought his technology could provide consumers with more accurate and stable recognition of gestures with a smaller time lag at a more affordable price than this one. Kinect is a camera technology embedding a 3D sensor and highly advanced algorism into a device and it has achieved success in recognizing gestures. However, there are some necessary conditions to make this possible. We need a powerful image-processing chip for analyzing the algorism and a pattern projector with a laser light source for projecting light. In addition, since we have to use infrared rays, this technology is sensitive to ambient light. Prof. Ishikawa was convinced that his HSVT could overcome these weaknesses with Kinect and provide a more practical gesture recognition.

Kinect is a depth image sensor that measures the distance between the device and the object per each pixel by using light projected from the light source and traces the original shape or movement of the object. Our gesture system in Exvision recognizes a gesture without a depth image sensor but with only an ordinary image sensor such as the one used in a mobile phone. Kinect applies a theory similar to stereo vision to recognize a distance, but our gesture system recognizes a gesture by a high-speed analysis of a picture taken by an image sensor. While image processing in general deals with 30-60 images per second, our gesture system processes 100 or 120 images per second obtained by a CMOS sensor. In the laboratory, we deal with even 500-1,000 images. Of course, we know that as the number of images increases, we will get a more precise analysis. But in that case, the data to be analyzed also increases. Our algorism is characterized as constraining the possible increase in the data to be analysed as much as possible in spite of the increase in the number of images and thus shortening the time lag between images significantly. With this technology, we can recognize gestures in a responsive, accurate, and stable manner that is not possible with other technologies. With neither stereo vision, nor a structured light method, nor a 3D depth sensor, but with just a CMOS sensor available anywhere, you can achieve the highest level of gesture recognition. This has never been done anywhere before.

JS: Your technology seems very inexpensive and highly competitive. But in general, the greater the added value of a technology, the more expensive it is, isn't it?

Morimoto: No, not necessarily. What determines the price of hardware is whether customized components are needed for the technology or not, how many parts and components are needed, and how many units need to be produced. We do not use any special component but use only a CMOS sensor and processor available on the market. Since we do not need any component to be specially devised for our technology, we do not have any additional cost for developing a new hardware.

In addition, in comparison with the number of components used for Kinect, we need far fewer components in our gesture system. Since Microsoft shipped tens of millions of pieces all over the world through large-scale promotion efforts, they could minimize the material costs of Kinect. However, with the large difference in the number of hardware components between Kinect and our gesture system, our price could go much lower than Kinect.

JS: The algorism used for your gesture system can enable you to do what Kinect cannot do. Is that right?

Morimoto: I would not say that. I would rather say that if you want to play a casual game with gestures, you would not need a 3D sensor. A 2D image sensor should be good enough. Thus far, gestures have been thought to be recognized only by a 3D sensor, but with high-speed image processing 2D image processing technology would create a more precise and stable result with a smaller time lag and at a cheaper cost.

JS: Can your technology be categorized as AI, since you apply a certain algorism in your system?

Morimoto: No. You can interpret our technology as providing information for AI, but it is not AI itself.

Applications of the Technology

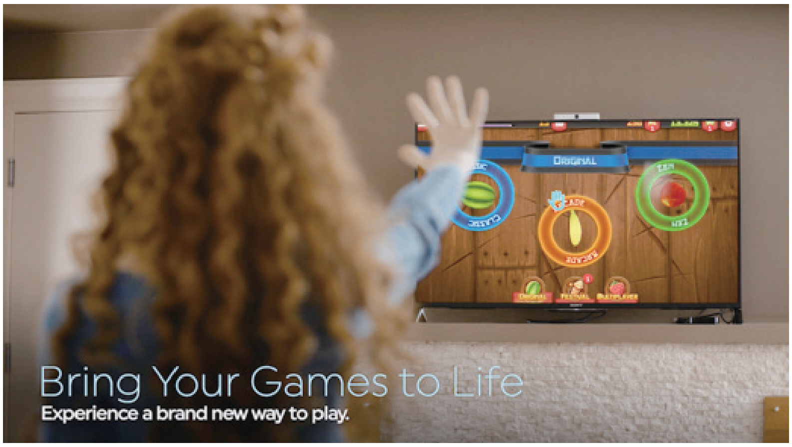
JS: You said that this technology was applied to a game machine. What do you think are other applications of your technology?

Morimoto: First, you can operate a large TV set in your living room by a gesture system. For example, you can browse websites on a TV screen or watch a news site by scrolling. You can turn a page by hand movement or click without a remote control. You can also play back a video and in particular you can play back the music by choosing a video site like YouTube by clicking with a hand.

You can also type texts directly on the TV screen using a software keyboard application. In this case, you can send a message on SNS like Twitter by using a TV screen. As such, TV will become an Internet device.

JS: Could it be used for robots?

Morimoto: Well, yes. But it is another implementation of HSVT, which has nothing to do with gesture recognition. For example, we can think about increasing the throughput of production lines using industrial robots. The industrial robots currently used for assembling cannot recognize the parts and components unless they are displayed in order. However, if we embed a high-speed camera into a robot's arm, Photo: Exvision Corp.



ZKOO offers a very responsive, accurate, and robust gesture experience.

the robot will be able to recognize the direction of the production line and do its work even though the parts are not displayed in order, and thus we can eliminate the effort required to align the components in order. The throughput of the production line would rise. Another possibility is that a robot could avoid a collision with any barrier or obstacle at such places as an airport or station platform. Even now we can avoid it to some extent, but with our HSVT even a robot moving around at high speed can avoid collisions. Oh, I forgot to mention that gesture recognition techniques can also be used for interaction with robots.

JS: Would it be possible to apply it to a driverless car as well?

Morimoto: Yes. In the long run, it would definitely be possible. A driverless car running at high speed could avoid collisions using this technology just as in the case of industrial robots.

Interaction with Silicon Valley

JS: Your company has two offices, one in Japan and the other in Silicon Valley. Could you explain why you have an office in Silicon Valley?

Morimoto: Yes, of course. The main reason was that I was working in Silicon Valley even before joining Exvision Inc. It worked in favor of our company to consolidate our business relations with business partners in Silicon Valley, such as processor or chip producers and game developers. In addition, it worked well for us to win new customers, since in Silicon Valley there are many potential customers.

JS: In Silicon Valley now we see a significant increase in the number of robot and AI companies, so is that why you see your potential customers increasing there?

Morimoto: Yes. Robots and AI are in general categorized as IT and Silicon Valley must be the best venue for information transmission as well as information collection with respect to IT.

JS: Besides that, as you have Stanford University with its large number of innovative scientists in Silicon Valley and can take advantage of well-established business-academic partnerships, would it be a comfortable place for a start-up company to merge into the working environment?

Morimoto: Yes, definitely. Stanford University provides its researchers with limitless laboratory space and other research facilities, such as incubators or accelerators, on condition that it will acquire a certain percentage of the stock of the start-up companies initiated by these researchers.

JS: How do you think Japan could accommodate such a business environment that works to the benefit of entrepreneurs?

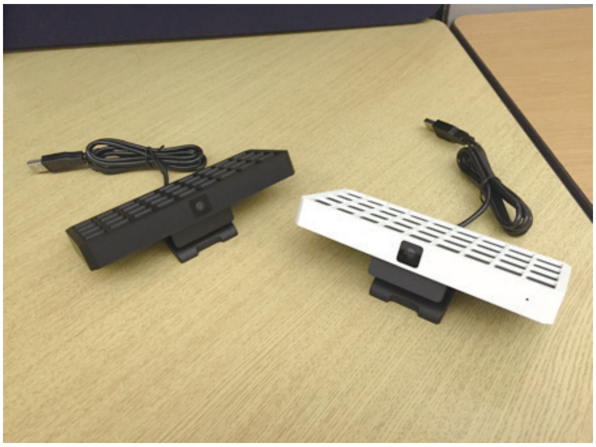
Morimoto: We would need to define "a friendly business environment for entrepreneurs" more clearly. Silicon Valley certainly provides a business environment friendly for start-up companies. First of all, the venture capital market has a long history in Silicon Valley and is well developed there and those start-up companies can take advantage of it in starting and expanding their businesses. However, it is not the case in Japan yet. This makes a big difference. Second, Japanese ventures would find it difficult to be acquired by or merge with large companies even if they have earned sufficient profits and growth.

As a result, most of the ventures active in Japan are ones related to Internet business which could reap profits in the short term. Unless such start-ups can earn profits in the short term, they would not be able to be financed or acquired by another company in the Japanese business environment. Internet-related start-ups could find some venture capital ready to finance them rather easily and also they could merge or be acquired by some large companies in the same Internet business. But ventures whose hardware or technology would need long-term investment would find it very difficult to tap into financial resources or M&As under Japanese business circumstances today.

So, I guess the shortcut to achieving many more active ventures in Japan is not by trying to create a venture-friendly business environment like in Silicon Valley, but rather to activate venture businesses either at universities or large corporations, and pave a path to their exits including M&A. Such ventures continuing to research without worrying about exits might have achieved the highest-level technologies in the world.

Exvision is, I believe, one such Japanese venture. Our technology

Photo: Exvision Corp.



Exvision ZKOO (Zi-Koo), the world's finest gesture tracking camera

developed by Prof. Ishikawa's lab is overwhelmingly the best in the domain of computer vision. I think technologies enjoying such absolute advantages should be well noted and considered to ensure Japanese technological competitiveness in the long term.

JS: Well, then, do you think there will be more ventures in Japan starting from universities?

Morimoto: Yes. I think it is possible to see many young Japanese interested in starting up businesses. However, they are not ready yet to turn their ventures into mature businesses. They need to accumulate knowledge in business development after successfully turning their technology into a product, and consider what business models to pursue or what financial plans to adopt.

JS: What do you think about the AI and robot industries in Japan compared with Silicon Valley? Are they competitive enough?

Morimoto: My answer is yes and no. It depends upon how you look at the sources of competitiveness. If you consider technological edge as a source of competitiveness, the Japanese and robot industries are competitive enough. However, in my view, Japanese robot producers tend to concentrate on technological competitiveness. They need to take note that a product sells well not because of its technological edge but because of the customers who want it. While Japanese robot-producing companies are paying more attention to how it moves and how it looks, US producers are more business oriented and are turning their technology into businesses, such as the "Jibo" personal assistant robot created by researchers from MIT, or the "Relay" produced by Savioke, a robot for delivery service in hotels.

These American robots do not look how robots are perceived to be in Japan. Japanese robot producers are so enthusiastic about having humanoid robots resemble human beings as much as possible. The American ones do not appear so attractive and are not what a young Japanese engineering student would dream of creating. But the "Relay" has now been adopted by many hotels and grown into a good business in spite of its simple structure. They think about the venue where a robot is working or creating added value from the viewpoint of business, and thus in the US there are many robots like this that concentrate on functional utility or value. Even those robots that could be considered technologically less attractive in Japan can work well in business in the United States.

Open Innovation & Exvision

JS: Open innovation is an idea promoted by the Japanese Ministry of Economy, Trade and Industry to encourage international R&D collaboration on modern technologies of a complex nature. Could this idea be applied to your HSVT?

Morimoto: Yes, of course. In the case of our technology, it is a component technology and as such we cannot make it a product on our own. It will need to be engaged in or adopted by other companies to provide a product.

Concerning our gesture system, its algorism has been developed by ourselves but the processors to achieve it are provided by collaborating companies. The platform to be used for activating it will be provided by Android or Windows. The games to which it is applied will be created by our game partners. Thus, we cannot imagine working without open innovation and collaboration with other firms.

Future Plans

JS: If possible, could you please tell us about your future plans?

Morimoto: In the short term, I would like to achieve a good business starting point for our "gesture recognition camera" which will be on the market this year. Over the long term, HSVT will be applied to many other usages than just gesture recognition. Taking advantage of being a spin-out start-up of a university, we can start applying the existing seeds of technologies already developed at our laboratory to various markets with high potential growth in an appropriate order. In this way, we can expand our business and company, and eventually make it a global technology venture starting from Japan.

JS: Does this mean you will develop your business into Asia or Europe, beyond the US and Japan?

Morimoto: Yes. If we are successful in the US and Japan, I believe we can reach other markets. JS

Written with the cooperation of Naoko Sakai who works for the NPO Yokohama Community Design Lab and is also a Hama-link Project leader and writer for the Yokohama Keizai Shimbun.