YOUNGKYOON JANG RESEARCH BRIEF

My research explores novel natural user interface technologies that aim to overcome challenges in interactions between humans and computers in a wearable AR/VR environment, specialising in **understanding human behaviours**, understanding scene and identifying users based on visual computing and mobile & wearable computing. Besides my research background in computer science, I draw on a diverse set of skills, including **machine learning** (particularly **Deep Learning (CNN)**, Random Forest), computer vision, video processing (including IR, colour, and depth images), **AR&VR**, and biometrics.

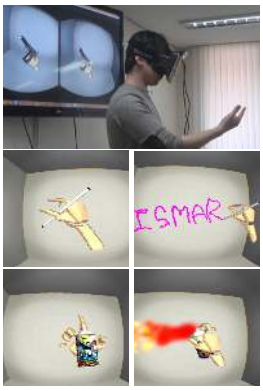
At the IEEE ICCV '17 in Venice, Italy

Interacting with AR/VR Objects in Natural Ways

The fundamental research questions of natural user interface (NUI) for supporting a wearable AR interface are apparent: how to overcome the challenges (e.g. self-occlusions leading to missing visual data) that occur when capturing an image sequence from the egocentric viewpoint. Because there is not yet an intuitive way to interact with AR/VR objects without simultaneously utilising a user's bare hands, NUI for a wearable AR/VR implies simplified hand gestures (e.g. hand shape classification, a skin-coloured region tracking, and fingertip detection) – all of which diminish usability and prevent us from understanding human behaviours. I ultimately selected this domain for my dissertation work. I have pursued three key strategies for addressing this challenge, and I highlight three exemplary projects below.

3D Finger CAPE [5] aims to infer an occluded fingertip position as well as clicking action from a single-depth sequence.

The most straightforward approach is to make maximal use of the preceding visual information which occurs in the frames prior to self-occlusions. For example, a clicking gesture has a contextual flow that could be used for interpreting a user's intention (clicking) and selection position. To explore this opportunity, I developed **3D Finger CAPE [5]** that estimates 3D finger-clicking action as well as clicked position simultaneously. 3D Finger CAPE offers sophisticated-direct-selection process in an arm reachable AR/VR space while self-occlusion is caused when a user interacts with VR objects in egocentric viewpoint. This strategy takes a simple context, which is a clicking action, to probabilistically estimate the occluded fingertip positions. While useful, if we want to expand the NUI so it supports complex scenarios, we need to consider multiple types of gestures and follow a more intuitive way that users have already learned from their daily lives.



MetaGesture [3, 4] estimates both (left) static and (right) dynamic hand gestures simultaneously.

One option is to imitate a user's behaviour of utilising familiar tools, including a stylus pen and a spray can. This metaphoric process for understanding a user's behaviour magnifies the interface usability of conventional approaches based on manually defined gestures. The technical challenge of making use of such methods is the combined processing of static and dynamic gesture estimation, which utilises missing visual information under self-occlusions. **MetaGesture [3, 4]** typifies this approach. By redesigning the conventional Random Forest structure, the system estimates static 3D hand postures for triggering a functional object on hand and estimates its action (i.e. function) status for manipulating the AR object. With MetaGesture, users intuitively summon a tool up to the hand and manipulate it as occasion demands without any additional device. This offered a dramatic expansion of interactive scenarios, including selection and manipulation processes, as well as retaining high accuracies of multiple gesture recognition while using incomplete visual data.

Smart Wristband [1] consists of a flexible touch screen panel and IMU sensor to capture multiple input sources outside of the camera's field of view.

Video-based understanding of the hand gesture alleviates the immediate problem of missing visual information in a given frame. However, we lose many available behavioural cues from other body parts due to the constrained interaction space of the camera's field of view (FOV). My third strategy has been to opportunistically appropriate wearable sensors such as a smart watch, armband, and ring. In this way, we can make the interaction space unconstrained for understanding a user's intention while retaining the benefits of direct interaction in the camera's FOV. **Smart Wristband [1]** shows the potential opportunities of the idea. Sensing touch points on a touch panel and arm rotation based on the inertial measurement unit sensor allow for the ability to specify and rotate a target object, respectively. Smart Wristband allows a user to utilise other parts of the body when the hand is outside of the viewpoint. Users can, for example, change a colour mode or brush type when drawing in the air by quickly rotating the arm, while their eyes remain focused on the target object.

Understanding Real Environments

As part of my investigations into appropriating everyday objects for interaction, I created a technical thread of research on real object recognition. Not only colour images, I'm also able to take advantage of other types of images captured from depth and IR cameras that are specially designed to characterise objects. In many respects, type specific feature representation / descriptions help an algorithm to understand the surrounding environment we are interacting in.

There are several significant challenges to achieving the vision of Understanding Real Environments. Appearance-invariant feature description is the foremost challenge. The object's appearance changes due to the occlusions, partial movements, rotation, translation, and surrounding clutter when using the hands to interact with real objects. The changes also occur when the extracted feature points begin disappearing along the time axis. Thus, to make features invariant to the challenges mentioned above, we developed a novel set-of-sets representation, which could be made by combining several patch tracks extracted from videos.

Video-based Object Recognition (**VbOR**) [12] was my definitive contribution to this topic, putting forward a novel set-of-sets feature representation to enable multiple 3D object recognition in a video. Through our proposed novel feature analysis and machine learning, we studied on a Unified Visual Perception Model (**UVPM**) [9] for context-aware wearable AR. In addition, to take better advantage of the depth image, Local Angle Pattern (**LAP**) [8] for describing shape information was proposed. The results of these deep explorations for utilising several types of images have revealed both limitations and opportunities, which point the way for developing novel types of user interfaces. My next direction for a novel hand gesture interaction would be to consider a situation where a user holds a real object. We can utilise both real object or scene understanding techniques and hand gesture recognition techniques together for making a novel NUI. To this end, we need to develop novel machine learning techniques defining visual features that adaptively analyse various types of object-specific patterns.

VbOR [12] aims to stably recognise a 3D object based on a novel set-of-sets feature representation.

UVPM [9] aims to stably recognise objects based on a contextual understanding between surrounding object's relationships.

Contact [6] aims to detect touching points on a desk.

Beyond identifying Users

So far, I have focused on challenges inherent in a wearable computing environments / scenarios given with input (colour and depth) videos. However, there is a second, more subtle issue that is potentially significant: identifying a person for interacting with the user. For example, we can interact with not only AR/VR objects, but also users based on his identity and further analysed emotion of him. This direction for further research has been made based on my experiences of face-related researches, including a **registration-free smiling face detection model (SmileNet)** [2], a **person-reidentification (portable iris recognition) system** [10, 11] and a **smile training system** [7]. Based on the background knowledge of Biometrics (including iris, finger vein, and face recognition), I was able to notice the current challenges for each modality. Moreover, in terms of wearable AR environments, I have my great insight towards the modern challenges of NUI, which has to be redirected to the case of interacting with users in collaborative AR/VR environment.

SmilNet [2] aims to detect smiling faces in the wild w/o registration process.

Conclusion

In my research, I aim to expand and enrich the ways we interact with real & virtual objects in a mixed reality environment by **understanding human behaviours / analysing facial affect / interpreting scene**, or re-identifying a person. These advances make the best use of today's technologies, and also help to define and inspire the next generation of user interfaces of wearable AR/VR platforms with an egocentric vision sensor. Likewise, as underlying technologies improve, I hope to continue to lay the groundwork (e.g., machine learning, computer vision) for future human-behaviour-understanding-based interfaces. Although not discussed here, **my research threads on machine learning and video processing are central to these research objectives**. Overall, these efforts aim to unlock unrealised potential and advance the state of the art, allowing us to make the best use of **human behaviour / facial affect / scene analysis** and person re-identification as a natural way to interact.

Portable iris recognition system [11] aims to re-identify a person in any envrionment.

Contact

Youngkyoon Jang, Ph.D.

Multimedia and Vision (MMV) Research Group
School of Electronic Engineering and Computer Science (EECS)
Queen Mary University of London
Mile End Road, E1 4NS, London, UK

Mobile: +44 (0)7522 142643
Email: youngkyoon.jang[at]qmul.ac.uk
yj293[at]cl.cam.ac.uk
http://youngkyoonjang.bitbucket.io

References

[1] J. Ham, J. Hong, **Y. Jang**, S. H. Ko, and W. Woo. Smart wristband: Touch-and-motion-tracking wearable 3d input device for smart glasses. In *Distributed, Ambient, and Pervasive Interactions - Second International Conference, DAPI 2014, Held as Part of **HCI Interational** 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings*, pages 109–118, 2014.

[2] **Y. Jang**, H. Gunes, and I. Patras. SmileNet: Registration-Free Smiling Face Detection In The Wild. In *The IEEE International Conference on Computer Vision (**ICCV**) Workshops*, pages 1581–1589, October 2017.

[3] **Y. Jang**, I. Jeon, T.-K. Kim, and W. Woo. Multi-Layered Random Forest-based Metaphoric Hand Gesture Interface in VR. In *The IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**) Workshops*, July 2016. **(Best Poster Award)**.

[4] **Y. Jang**, I. Jeon, T.-K. Kim, and W. Woo. Metaphoric Hand Gestures for Orientation-aware VR Object Manipulation with an Egocentric Viewpoint. ***IEEE Transactions on Human-Machine Systems***, 47(1):113–127, February 2017.

[5] **Y. Jang**, S.-T. Noh, H. J. Chang, T.-K. Kim, and W. Woo. 3D Finger CAPE: Clicking action and position estimation under self-occlusions in egocentric viewpoint. ***IEEE Trans. on Vis. Comput. Graph.***, 21(4):501–510, April 2015.

[6] **Y. Jang**, S.-T. Noh, and W. Woo. RGB-D image-based touch points detection for hand-plane interaction. In ***HCI Korea** 2014, High1 Resort, S. Korea, February 12-14*, 2014.

[7] **Y. Jang** and W. Woo. Adaptive lip feature point detection algorithm for real-time computer vision-based smile training system. In ***Edutainment***, volume 5670 of *Lecture Notes in Computer Science*, pages 379–389. Springer, 2009.

[8] **Y. Jang** and W. Woo. Local feature descriptors for 3d object recognition in ubiquitous virtual reality. In *2012 International Symposium on Ubiquitous Virtual Reality (**ISUVR**), Daejeon, Korea (South), August 22-25, 2012*, pages 42–45, 2012.

[9] **Y. Jang** and W. Woo. Unified visual perception model for context-aware wearable AR. In *IEEE International Symposium on Mixed and Augmented Reality, **ISMAR** 2013, Adelaide, Australia, October 1-4, 2013*, pages 1–4, 2013.

[10] **Y. K. Jang**, B. J. Kang, and K. R. Park. A study on eyelid localization considering image focus for iris recognition. ***Pattern Recognition Letters***, 29(11):1698–1704, 2008.

[11] **Y. K. Jang**, B. J. Kang, and K. R. Park. A novel portable iris recognition system and usability evaluation. *International Journal of Control, Automation, and Systems (**IJCAS**)*, 8(1):91–98, 2010.

[12] Y. Liu*, **Y. Jang***, W. Woo, and T.-K. Kim. Video-based object recognition using novel set-of-sets representations. In *The IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**) Workshops*, June 2014. (* indicates equal contribution).