BU STUDENTSHIPS 2020
PROJECT DESCRIPTION

PROJECT DETAILS

PROJECT TITLE

Multimodal behaviour understanding in human-computer dialogues

Part of the Doctoral Training Centre - Centre for Digital Entertainment – Intelligent Virtual Personal Assistant (Intel-PA)

DOCTORAL TRAINING CENTRE SUMMARY

The Intelligent Virtual Personal Assistant consists of three landmark components that enable Intel-PA to perform human-like communications with real humans. These are: sensible conversations; appropriate body gestures and facial expressions; and ability to detect the motion and emotions of the human user. By performing such human-like functions, Intel-PA will be able to play a huge role in assisting the mankind in countless applications, such as healthcare, training, education, marketing and decision making, to name but a few. Taking healthcare as an example, we all are aware of the acute pressure the NHS is under and it usually takes a long wait even to see our GP. If some/most of the GP's tasks can be undertaken by Intel-PA, it will not only save the NHS a huge budget, but also significantly shorten the waiting time. In addition to the role of a GP assistant, in healthcare alone there are many other applications where Intel-PA can play an important assisting role.

Research in this area is still at a very early stage. The aim of this research is to investigate and create novel algorithms and techniques leading to the development of the three landmark components of Intel-PA. The initial six projects will define the research focus for the six PhD students. Working as a multidisciplinary team of researchers and supervisors, we will demonstrate the research by developing an Intelligent Neur-Assistant, an AI-powered conversational animated avatar that can have a plausible initial consultation with a patient with neurological conditions by integrating these six projects. All students will have to contribute to the overall objectives of the research with each of them concentrating on an individual research project.

PROJECT SUMMARY

In multimodal human computer dialogs, non-verbal channels, such as facial expressions, postures and gestures, combined with emotion in speech, are important in understanding the dialogue. This is because in human communications, the information is often transmitted from multiple channels, including facial expressions, emotional voices, postures and gestures. Understanding of people's intentions from their multichannel information plays a key role in natural interaction. This project aims at multichannel user behaviour detection and multimodal data fusion in multimodal human-computer interaction, which helps improve the believability in human computer dialogs.

AI technology has contributed to the improvements for single channel behaviour perceptions, including speech recognition [1], facial expression recognition [2], emotional understanding [3], gesture comprehension [4], posture analysis [5], gaze detection [6], etc. The computer has some ability to understand the user from single channel information. However, it is still a great challenge to understand human users' intention accurately from their multichannel behaviours. The main challenge is that we need to improve multimodal information fusion in theories, methodologies and practical systems.

Currently, many methods exist for multi-modal information fusion, including Bayesian decision model [7], Cross modal learning Neural Networks [8] and Graph model based information fusion [9]. However, accurately recognising the user's intension remains an unanswered research question. To improvement on these fronts, this research will focus on:

(1) A novel integration approach of multi-channel interaction information. Multi-channel user behaviour,

including voice, posture, emotional expressions, etc, in interaction is more likely to lead to uncertainty than its single channel counterpart. The richness and fuzziness of such expressions make it difficult to be accurately mapped to the traditional human-computer interface, which makes the system feedback inaccurate [10]. Our aim is to develop novel representation forms for such high dimensional multimodal data.

(2) Improvement of the accuracy of single channel information identification. In spite of the good performance of single channel behaviour computation, the change of environment will significantly influence behaviour recognition [6-9]. We will derive robust feature extraction methods/representations in channel filter processing.

(3) Novel multi-channel human-computer interaction model. An ideal interaction model should be able to learn, understand and integrate new knowledge into the existing knowledge bases. We will apply deep learning technology for model design.

References:

1) Chorowski J, Bahdanau D, Serdyuk D, Cho K, Bengio Y. Attention-based models for speech recognition. Neural Information Processing Systems (NIPS), 2015
2) L. Chen, R. Travis Rose, F. Parrill, X. Han, J. Tu, Z. Huang, M. Harper, F. Quek, D. McNeill, R. Tuttle, and T. Huang, "VACE multimodal meeting corpus", in Proc. Workshop on Machine Learning for Multimodal Interaction (MLMI), Edinburgh, Jul. 2005.
3) Wang S, Yan W, Li X, Zhao G, Zhou C, Fu X, Yang M, Tao J. Micro-expression recognition using color spaces. IEEE Transactions on Image Processing, 2015, 24(12): 6034–6047 DOI:10.1109/TIP.2015.2496314
4) Ge L, Liang H, Yuan J, Thalmann D. Robust 3D hand pose estimation from single depth images using multi-view CNNs. IEEE Transactions on Image Processing, 2018, 27(9): 4422–4436 DOI:10.1109/TIP.2018.2834824
5) Weiming H, Tieniu T, Liang W, Maybank S. A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2004, 34(3): 334–352 DOI:10.1109/TSMCC.2004.829274
6) Cheng S, Sun Z, Sun L, Yee K, Dey A K. Gaze-based annotations for reading comprehension. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Seoul, Republic of Korea: ACM, 2015: 1569–1572 DOI:10.1145/2702123.2702271
7) Town C. Multi-sensory and multi-modal fusion for sentient computing. International Journal of Computer Vision, 2007, 71(2): 235–253 DOI:10.1007/s11263-006-7834-8
8) Tzeng E, Hoffman J, Darrell T, Saenko K. Simultaneous deep transfer across domains and tasks. In: Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2015, 4068–4076 DOI:10.1109/ICCV.2015.463
9) Savran A, Cao H, Nenkova A, Verma R. Temporal Bayesian Fusion for Affect Sensing: Combining Video, Audio, and Lexical Modalities. IEEE Transactions on Cybernetics, 2015, 45(9): 1927–1941 DOI:10.1109/TCYB.2014.2362101
10) Cohen P R, McGee D R. Tangible multimodal interfaces for safety-critical applications. ACM, 2004, 47(1): 41–46 DOI:10.1145/962081.962103

ACADEMIC IMPACT

Intel-PA is concerned with the research and interaction of three cross-disciplinary but complementary research subjects, which are computer animation and visualisation; computer vision; and conversational AI. They each have their own research challenges. Their inter-connection will further bring in new challenges and research questions. Our mission here in terms of academic research is to find solutions to these challenges.

The overarching research together with the individual PhD projects will lead to the formulation of many new algorithms. It will advance the state-of the-art of a number of research topics including the generation of plausible body and facial motions in real-time; tracking and detecting human body movements and facial expressions; natural language understanding and synthesis; expert systems and machine learning; conversational artificial intelligence.

The research outcomes will be published in peer-reviewed journals and conferences, which will inspire other researchers for further investigations of relevant or similar topics. The research achievements will be scalable and

applicable to other related research undertakings in computer animation, computer vision and conversational AI. The supervising academics can reuse the developed methods, computer programing codes and databases and can either continue with this work or apply them to relevant, but different research questions.

The research outputs will benefit REF2021 and future REFs.

SOCIETAL IMPACT

Intel-PA represents a new generation of virtual humans who can 'think' intelligently; 'understand' and participate in human conversations; 'see' the bodily actions of the human user; and enact appropriate body gestures and facial expressions. It will be a true Intelligent Virtual Human when it is fully developed. By initiating this hugely exciting research direction, Intel-PA can perform many human-like tasks having impactful effects on many aspects of the society.

In healthcare, it can assist the medical practitioners to collect data, have an initial consultation with the patient, generate preliminary diagnosis, suggest a range of treatment options. It can also perform many training tasks for different healthcare workers, such as nurses and clinic psychologists. Similarly, the other sectors will benefit from the functions of Intel-PA, such as in education, marketing, training and many more.

This new technology, once developed, will benefit the UN sustainable development goal of Good Health and Well-being (Goal 3). As Intel-PA will help significantly enhance the proficiency and productivity due to improved communication, training and decision making, it will benefit Goal 8: Decent Work and Economic Growth; and Goal 9: Industry, Innovation and Infrastructure.

DEVELOPMENT OPPORTUNITIES

One of the important aspects of our DTC is that all students will belong to one centre and have the same overall aim of research. We will arrange regular group supervision meetings where all students and supervisors are required to attend. All students will therefore benefit by understanding the global picture of the project; getting advice from all supervising academics; and seeing the progress of their fellow students.

For their individual projects, each student will be supported by their own supervisors on a weekly basis. They will be guided to research on the technical elements relevant to their investigation. In the same time, they will be exposed to the knowledge and practice of the applicable applications.

The training courses from the Doctoral College, such as research methodology, academic writing and presentation, research ethics, project management, career development, will help the students develop their research skills in a wide context.

Students will be provided some additional research training opportunities from our current Centre for Digital Entertainment will include: participating in CDE organised events (for technical and soft skills) and attending prestigious conferences and high-profile public engagement activities (e.g. the annual CDE research showcase at the British Film Institute, London https://youtu.be/WmEL2f6jK5s and the 2018 CDE showcase at the Digital Catapult London https://youtu.be/XOF0By9Wssw). CDE delivered the 2018 EPSRC funded Digital Economy Network Summer School (DEN SS) with 110 doctoral students from ten UK Universities attending http://www.digital-entertainment.org/research-impact/den-summer-school-2018-inspire-programme-develop/

SUPERVISORY TEAM	
First Supervisor	Dr Hongchuan Yu
Additional Supervisors	Prof Jian Jun Zhang

Recent publications by supervisors relevant to this project	1. Deng, S., Chang, J., Kirkby, J.A. and Zhang, J., 2016. Gaze–mouse coordinated movements and dependency with coordination demands in tracing. Behaviour and Information Technology, 35 (8), 665-679. 2. Deng, S., Jiang, N., Chang, J., Guo, S. and Zhang, J., 2017. Understanding the impact of multimodal interaction using gaze informed mid-air gesture control in 3D virtual objects manipulation. International Journal of Human Computer Studies, 105, 68-80. 3. Feng. Y., Ji, M., Xiao, J., Yang, X., Zhang, J., Zhuang, Y., Li, X. (2015). Mining Spatial-Temporal Patterns and Structural Sparsity for Human Motion Data Denoising, Cybernetics, IEEE Transactions on. 4. Guo, S., Wang, M., Notman, G., Chang, J., Zhang, J. and Liao, M., 2017. Simulating collective transport of virtual ants. Computer Animation and Virtual Worlds, 28 (3-4). 5. Hu, W., Wang, Z., Liu, S., Yang, X., Yu, G. and Zhang, J., 2017, Motion Capture Data Completion via Truncated Nuclear Norm Regularization, in IEEE Signal Processing Letters, vol. PP, no. 99, pp. 1-1. 6. Liu, F., Southern, R., Guo, S., Yang, X., and Zhang, J., 2013. Motion adaptation with motor invariant theory. Cybernetics, IEEE Transactions on, 43(3):1131–1145. 7. Liu, S., Wang, Z., Yang, X., Zhang, J., 2017, Realtime Dynamic 3D Facial Reconstruction for Monocular Video In-The-Wild, The IEEE International Conference on Computer Vision (ICCV), 2017, pp. 777-785. 8. Nie, Y., Chang, J., Chaudhry, E., Guo, S., Smart, A. and Zhang, J., 2018. Semantic modeling of indoor scenes with support inference from a single photograph. Computer Animation and Virtual Worlds, 29 (3-4). 9. Nie, Y., Tang, Z., Liu, F., Chang, J. and Zhang, J., 2018. A data-driven dynamics simulation framework for railway vehicles. Vehicle System Dynamics, 56 (3), 406-427. 10. Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J., Campbell, C., and Levine, M., 2013. Bystander responses to a violent incident in an immersive virtual environment. PLoS ONE, 8(1): e52766, 01 2013. 11. Southern, R., and Zhang, J., 2011. Motion-sensitive anchor identification of least-squares meshes from examples. Visualization and Computer Graphics, IEEE Transactions on, 17(6):850–856. 12. Wang, M., Guo, S., Liao, M., He, D., Chang, J. and Zhang, J., 2019. Action snapshot with single pose and viewpoint. Visual Computer, 35 (4), 507-520. 13. Wang, Z., Feng, Y., Qi, T., Yang, X. and Zhang, J., 2016. Adaptive multi-view feature selection for human motion retrieval. Signal Processing, 120, 691-701. 14. Wang, J., Tian, F., Yu, H., Wang, X., Liu, C.H., Zhan, K., 2018. Diverse multi-view Nonnegative Matrix Factorization for Multiview Data Representation, IEEE Trans. on Cybernetics, Vol.48, No.9. 15. Wang, J., Wang, X., Tian, F., Liu, C.H., and Yu, H., 2017. Constrained Low-Rank Representation for Robust Subspace Clustering, IEEE Trans. on Cybernetics, Vol.47, No.12. 16. Yang, X., Chang, J., Southern, R., and Zhang, J., 2013. Automatic cage construction for 17. retargeted muscle fitting. The Visual Computer, 29(5):369–380. Zhang, Y., Liu, S., Yang, X., Shi, D., Zhang, J., 2016, Sign-Correlation Partition Based on Global Supervised Descent Method for Face Alignment, Asian Conference on Computer Vision (ACCV), 281-295. 18. Zheng, J., Qiu, H., Sheng, W., Yang, X., Yu, H., 2018. Kernel group sparse representation classifier via structural and non-convex constraints, Neurocomputing, Vol.296, pp.1-11. 19. Zheng, J., Lu, C., Yu, H., Wang, W., and Chen, S., 2018. Iterative Reconstrained Low-Rank Representation via Weighted Nonconvex Regularizer, IEEE Access, Vol.6, pp.51693-51707. 20. Zheng, J., Qin, M., Zhou, X., Mao, J., and Yu, H., 2019. Efficient Implementation of Truncated Reweighting Low-rank Matrix Approximation, IEEE Trans. on Industrial Informatics.

INFORMAL ENQUIRIES

Please contact the lead supervisor on the following email for informal enquiries: hyu@bournemouth.ac.uk

ELIGIBILITY CRITERIA

The BU PhD Studentships are open to UK, EU and International students.

Candidates for a PhD Studentship should demonstrate outstanding qualities and be motivated to complete a PhD in 4 years and must demonstrate:

- outstanding academic potential as measured normally by either a 1st class honours degree (or equivalent Grade Point Average (GPA) or a Master's degree with distinction or equivalent
- an IELTS (Academic) score of 6.5 minimum (with a minimum 6.0 in each component, or equivalent) for candidates for whom English is not their first language and this must be evidenced at point of application.

ADDITIONAL ELIGIBILITY CRITERIA

An ideal candidate should have a background in computer science, mathematics, engineering or a similar subject. Strong mathematical and computer programming skill is essential.

HOW TO APPLY

Please complete the online application form by **Monday 20 July.**

Further information on the application process can be found at: www.bournemouth.ac.uk/studentships