Online Human Action Recognition using Deep Learning for Human-Robot Collaboration.

Keywords: artificial intelligence (IA); Deep Learning (DL); Human Action Recognition (HAR); Industry 4.0; Human-Computer Interaction (HCI); Human-robot collaboration (HRC).

Context:

In the context of the industry 4.0, technological developments such as collaborative robotics (Pilat et al., 2017), Internet of Things (machines, products, tools, etc.), Artificial Intelligence (AI), processing and decisions' algorithms (gesture analysis, environment understanding, etc.) and the new human-machine interfaces (augmented reality, gestural interfaces, etc.) open up many possibilities to make production systems more flexible.

Since humans and robots are working more closely together nowadays as technology improves, this is increasing the productivity of companies and the quality of products, leading to efficiency and growth. Researchers and companies are improving the safety of robot systems so humans can work close beside robots that become co-workers, more than mere tools.

Human-Robot Collaboration can free the workers from difficult tasks and make production systems more flexible. In order to convey visual messages to a receiver, a human expresses "actions" that are variable but distinct and have an associated meaning therefore there is a need for appropriate modes of communication and interaction between humans and robots in production workshops which will require structured forms of communication are required. For example, the human must be able to interrupt the robot in its action to reconfigure it and the robot must be able to communicate on its current task and its intention. Communication and interaction between the operators and their environment (machines, robots, etc.) therefore represent a major performance and safety issue and require upstream training of the users.

In this context, using the Industry of the Future platform of the LINEACT laboratory and its Digital Twin, a dataset called InHARD - Industrial Human Action Recognition Dataset was produced. It contains several online videos of different operators assembling an industrial system assisted by a cobotic arm. The robot must be able to anticipate the needs and guarantee the safety of the operator by recognizing his actions.

Problem statement

These interactions, in real or virtual environment, and the associated decision-making between the different agents of the system (operators, collaborative robots, etc.) require operators' gestures recognition. Gestures classification using AI methods and in particular recent work based on Deep Learning approaches have shown their effectiveness (Liu and Wang, 2018) (Sijie Yan, 2018). This automatic learning step for action recognition performed in industrial systems is mandatory to allow collaboration between humans and robots in real or virtual environment.

However, these approaches require segmenting the action sequences in order to be able to classify them. It is therefore necessary to study approaches allowing to make "online" recognition, in order to be able to classify these actions on the fly during the robot human collaboration. Several approaches based on CNN, RNN, LSTM (Singh et al., 2017) (Li et al., 2016) (Liu et al., 2019) and or on "Memory attention" (You and Jiang, 2019) (Chunyu Xie, 2018) seem relevant.

The objective of this Study is to focus on this both papers (Singh et al., 2017) (You and Jiang, 2019) and therefore to study these approaches and apply them on the InHARD dataset created by the LINEACT CESI Laboratory.

Missions:

- Participation in the state of the art on AI in the context of Man-Robot collaboration and more specifically on the recognition of online actions by deep learning methods and studies of associated architectures.
- Getting started with the InHARD dataset and data-pre-treatment.
- Identification of the interesting approach(es) according to the InHARD dataset.
- Implement the chosen method on the dataset used.
- Adapt the method to fit the InHARD dataset and study the result.

This work will take place within the framework of a PIA (Programme d'Investissement d'Avenir) project entitled DEFI&Co (Développer l'Expertise Future pour l'Industrie et la Construction) which aims to develop training in response to the new professions of tomorrow. The state of the art and the use case developed on CESI's industry of the future platform, integrating AI algorithms to enable collaboration between humans and robots, will feed the reflection on the evolution of industrial trades impacted by AI.

Required profile:
Skills:

- Master 2 or AI-oriented computer engineer, machine learning or deep learning
- Have writing skills,
- English proficiency

Social skills:

- Be autonomous, have a spirit of initiative and curiosity,
- To be rigorous

Contacts:

- Vincent HAVARD (vhavard@cesi.fr), lecturer, LINEACT CESI Rouen.
- Mejdi DALLEL (mdallel@cesi.fr), PhD in Human Action Recognition, LINEACT CESI Rouen.
- David BAUDRY (dbaudry@cesi.fr), Head of Engineering and digital tools research team, LINEACT CESI Rouen.

Duration: 4 to 6 months.

References

Chunyu Xie, C.C., Ce Li, Baochang Zhang, 2018. Memory Attention Networks for Skeleton-based Action Recognition. International Joint Conference on Artificial Intelligence (IJCAI-18).

Li, Y., Lan, C., Xing, J., Zeng, W., Yuan, C., Liu, J., 2016. Online Human Action Detection using Joint Classification-Regression Recurrent Neural Networks.

Liu, H., Wang, L., 2018. Gesture recognition for human-robot collaboration: A review. International Journal of Industrial Ergonomics 68, 355–367.

Liu, J., Li, Y., Song, S., Xing, J., Lan, C., Zeng, W., 2019. Multi-Modality Multi-Task Recurrent Neural Network for Online Action Detection. IEEE Transactions on Circuits and Systems for Video Technology 29, 2667–2682. https://doi.org/10.1109/TCSVT.2018.2799968

Pilat, Z., Varga, J., Hajduk, M., 2017. Development of the Human-Robot Communication in Welding Technology Manufacturing Cells, in: Szewczyk, R., Kaliczyńska, M. (Eds.), Recent Advances in Systems, Control and Information Technology, Advances in Intelligent Systems and

Computing. Springer International Publishing, Cham, pp. 387–395. https://doi.org/10.1007/978-3-319-48923-0_42

Sijie Yan, D.L., Yuanjun Xiong, 2018. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. AAAI 2018.

Singh, G., Saha, S., Sapienza, M., Torr, P., Cuzzolin, F., 2017. Online Real-time Multiple Spatiotemporal Action Localisation and Prediction. arXiv:1611.08563 [cs].

You, Q., Jiang, H., 2019. Action4D: Online Action Recognition in the Crowd and Clutter, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, pp. 11849–11858. https://doi.org/10.1109/CVPR.2019.01213