Abstract

Humans want to communicate with the computers in the same way as they communicate with other humans. Speech is the most natural and spontaneous form of communication. Speech is bimodal in nature and it combines audio and visual information to enhance speech recognition rate especially under poor audio conditions. This paper proposes novel computer vision architecture using fusion technique. This architecture combines or fuses more than one modality using multi-agents. In this we have used two modalities- audio and video. The audio part extracts the speech of a person and the video part extracts the face and lip information of the person. Here, different agents process the modalities and the fusion agent fuses these modalities for effective and efficient automatic speech recognition.

References

- Elfriede I. Krauth, Jos van Hillegersberg, Steef L. van de Velde, "Agent-based


- Simon C. Lynch (University of Teesside, UK) and Keerthi Rajendran (University of Teesside, UK); A multiagent approach to teaching complex systems development a hand book; 2011


- Prof. S. Qamar Abbas, Nidhi Srivastava; Development of Framework for Automatic Speech Recognition; IJCSE, Vol. 4 No. 05 May 2012.


- Erno Makien; Face Analysis Techniques for Human-Computer Interaction; Tampere 2007.


- Nidhi Srivastava, Dr. Harsh Dev, Dr. Qamar Abbas; Framework for Face Recognition; IJCA, Vol. 58, No. 17, November 2012.


- Tieyan Fu, Xiao Xing Liu, Lu Hong Liang, Xiaobo Pi, Ara V. Nefian; Audio-Visual Speaker Identification Using Coupled Hidden Markov Models; Proceedings of


- Nidhi Srivastava, Dr. Harsh Dev, Dr. Qamar Abbas, &quot;Speech recognition using MFCC and Neural Network&quot;, National Conference on Challenges & Opportunities for Technological Innovation in India, AIMT, February, 2013.

**Index Terms**

Computer Science

Artificial Intelligence

**Keywords**

Computer vision; fusion; agent; architecture; multimodal; multi-agent