

Abstract for
10th Joint Symposium on Neural Computation

**ATRV-I: A Dataset for Development and Evaluation
of Pose Estimation and Expression Recognition Systems**

Javier R. Movellan^{1,2}, Takayuki Kanda¹ Ian Fasel^{1,2}, Joel Chenu^{1,2}, Hiroshi Ishiguro¹

Intelligent Robotics and Communication Laboratories¹

ATR, Kyoto, Japan.

&

Institute for Neural Computation²

University of California, San Diego, CA 92093

One of the most important functions of the human brain is to enable interaction with other human beings, a problem that can be analyzed from the point of view of real time systems identification and control. In order to develop computer animated agents and robots designed to communicate with people it is important to have a precise understanding of the time delay distributions and levels of uncertainty typically found in social interactions. Unfortunately up to date there are very few publicly available video datasets with frame by frame labeling that could support this type of statistical analysis. In addition the field of automatic face tracking and expression recognition is in critical need for datasets to standardize evaluation of the wide variety of algorithms proposed in the literature. In the absence of such databases it is extremely difficult to compare different systems and to evaluate progress towards realistic applications.

In this document we introduce a new dataset (ATRV-I) designed to facilitate progress on statistical analysis of social interactions and for development and evaluation of 3D face tracking and expression recognition systems. The dataset, which is being made available to the research community, contains an hour of video captured with 3 directional cameras and an omnidirectional camera. The video has been automatically labeled frame by frame, using a 16 camera Vicon system, for head pose with millimeter accuracy. The dataset will help standardize training and evaluation of head tracking and expression recognition systems. It will also allow a statistical characterization of the levels of predictability and time delay distributions typically found in social interactions.