Abstract

Communication is an integral part of life and can be either verbal or non-verbal. Non-verbal communication is generally achieved with hand gestures, body poses, or facial expressions which also find utility in human computer interaction (HCI). It is interesting to observe that classical dance forms across the world also use non-verbal communication to engage with the audience. Practitioners of Indian classical dance (ICD) enact tales from epics/mythology using complex non-verbal cues defined in classical texts such as *Natyashastra*.

Understanding ICD depends on the recognition of hand gestures, poses and facial expressions associated with it. The recognition of hand gestures is hindered by the presence of clutter, occlusions, varying sizes of hand, presence of jewellery and color in hands, blur, illumination variations, etc. Hence, initially in the thesis we attempt to recognize static hand gestures robustly in unconstrained scenarios of ICD. We propose datasets captured under controlled laboratory conditions as well as in real-world for single and double hand gestures in ICD. Extensive experiments conducted over these datasets with the proposed CNN demonstrate its superiority over shallow learning approaches.

Next, we proposed a deep learning approach for robust recognition of static body postures in ICD. Furthermore, we addressed the problem of robust recognition of dynamic hand gestures and body poses in unconstrained real-world dance performances. For detecting static hand gestures in video frames an adaptive boosting (AdaBoost) approach and region based CNN (R-CNN) are used. A skin detector is also proposed for refining the localization of hands in the images. Since no public datasets are available for ICD, we propose dataset for dynamic hand gestures captured under controlled laboratory settings. We also create a separate dataset for dynamic body poses, namely, *Adavus* of *Bharatnatyam*. Extensive experiments for identification of dynamic hand gestures and dynamic body poses are performed using a three dimensional-convolutional neural network (3D-CNN). In order to classify dynamic hand gestures in ICD, we also utilized motion history images derived from the video frames along with a CNN.

The emotions/affect associated with *Bhratnatyam* (*Navarasas*) are mainly enacted with the aid of facial expressions which may also be augmented with appropriate upper body postures and hand gestures. Initially, we propose a deep learning based approach using CNNs for recognition of emotions associated with typical ICD forms which convey semantic meaning in accordance with the context. Since image sequences of facial expressions contain rich spatio-temporal information, the thesis proposes to utilize local binary patterns (LBP) and robust principal component analysis (RPCA). We train two separate CNNs using LBP features and the sparse component of RPCA derived from frames of a video corresponding to a facial expression. The deeply learnt features from the two CNNs are combined to train a long short-term

memory (LSTM) for dynamic facial expression recognition. We experimentally show that by concatenating CNN features extracted from both sparse matrices of RPCA and LBP feature maps results in a performance boost in contrast to the case when these CNN features are used separately for classification.

Next, we address the task of recognition of hand gesture images in standard publicly available datasets. For this purpose, we propose a framework to exploit the deep features obtained from our previously proposed two convolutional layered CNN and derive more discriminative features using dictionary learning. We use the approaches of K singular value decomposition (K-SVD), discriminative K-SVD (D-KSVD), and label consistent K-SVD (LC-KSVD) for dictionary learning. Deeply learnt features are extracted from the proposed CNN having a small architecture with random weight initialization as well as after pre-training with CIFAR-10 dataset. Interestingly, we observe that classification performance using dictionary learning with LC-KSVD on fine-tuned features from the proposed two convolutional layered CNN is at par with that obtained by using dictionary learning in conjunction with the deep fine-tuned VGG19 network.

Keywords - Deep learning; convolutional neural network; hand gesture recognition; body pose estimation; histogram-of-gradients; motion history image; adaptive boosting; facial expression recognition; emotion recognition.