CHAPTER 2

LITERATURE SURVEY

[1] Low Rank Approximation

This paper proposes a framework based on the Hidden Markov Models (HMMs) benefited from the low rank approximation of the original sign videos for two aspects. First, under the observations that most visual information of a sign sequence typically concentrates on limited key frames, we apply an online low rank approximation of sign videos for the first time to select the key frames. Second, rather than fixing the number of hidden states for large vocabulary of variant signs, we further take the advantage of the low rank approximation to independently determine it for each sign to optimise predictions. With the key frame selection and the variant number of hidden states determination, an advanced framework based on HMMs for Sign Language Recognition (SLR) is proposed, which is denoted as Light-HMMs (because of the fewer frames and proper estimated hidden states). With the Kinect sensor, RGB-D data is fully investigated for the feature representation. In each frame, we adopt Skeleton Pair feature to character the motion and extract the Histograms of Oriented Gradients as the feature of the hand posture appearance. The proposed framework achieves an efficient computing and even better correct rate in classification. The widely experiments are conducted on large vocabulary sign datasets with up to 1000 classes of signs and the encouraging results are obtained. This paper proposes a HMMs based framework (Light HMMs) benefited from low rank approximation for two aspects. First, low rank approximation removes redundant frames and selects key frames for the training and test of HMMs to be faster. Second, the segmentations generated by low rank approximation contribute to determine the number of hidden states for training HMMs. Under the novel Light HMMs, the encouraging results are obtained in widely tests by using the fused posture appearance features (HOG) and the motion skeleton pair features (SP). Compared with the baseline HMMs method, our Light-HMMs framework costs around 1/3 time while maintaining or even promoting the high recognition rate. From the experiments, it can be seen that the performance is decreased dramatically when facing the signer independent situation. To make SLR more robust to different signers, further exploration

on the deep fusion among hand posture, body skeleton, and the depth map should be the focuses of our future work

[2] Sign Language Recognition System

This project is designed to improve the recognition rate of the alphabet gestures, (A-Z), in previously done works. Six alphabets are chosen for this purpose. These are alphabet A, alphabet D, alphabet J, alphabet O, alphabet P, and alphabet Q. Alphabet A has been chosen as it is recognized easily at 100% rate to show that this project not only improves the recognition rates of lacking alphabets, but also maintains the 100% recognition rate of other alphabets. The recognition time has also been improved significantly. It was observed that the recognition rate was fairly improved and recognition time was reduced significantly. This has been achieved by using knn search instead of contour method as is done before. This result was achieved by following simple steps without the need of any gloves or any specifically coloured backgrounds. This work may be extended to recognizing all the characters of the standard keyboard by using two hand gestures. The recognized gesture may be used to generate speech as well as text to make the software more interactive. This is an initiative in making the less fortunate people more independent in their life. Much is needed to be done for their upliftment and the betterment of the society as a whole.

[3] Continuous Sign Language Recognition

In this work an end-to-end embedding of a CNN into a HMM is introduced, while interpreting the outputs of the CNN in a truly Bayesian fashion. Most state-of-the-art approaches in gesture and sign language modelling use a sliding window approach or simply evaluate the output in terms of overlap with the ground truth. While this is sufficient for data sets that provide such training and evaluation characteristics, it is unsuitable for real world use. For the field to move forward more realistic scenarios, such as those imposed by challenging real-life sign language corpora, are required. A hybrid CNN-HMM frame-work is presented that combines the strong discriminative abilities of CNNs with the sequence modelling capabilities of HMMs, while abiding to Bayesian principles**.** It is believed to be the first work to embed a deep CNN in an HMM framework in the context of sign language and gesture recognition, while treating the outputs of the CNN as true Bayesian posteriors

and training the system end-to-end as a hybrid CNN-HMM. Sign language is highly multimodal and makes heavy use of manual components (hand shape, orientation, place of articulation, movement) and also non-manual components (facial expression, eyebrow height, mouth, head orientation, upper body orientation). For a fair comparison, they only listed competitors that also just focus on the single hand. The previously best hand only results also relied on CNN models, but did not employ the hybrid approach end-to-end in recognition. In terms of future work, approaches will be extended to cover all relevant modalities.

[4] Real Time Hand Gesture Recognition

The scale invariant feature transform (SIFT) algorithm, developed by Lowe [1,3,4], is an algorithm for image features generation which are invariant to image translation, scaling, rotation and partially invariant to illumination changes and affine projection. SIFT algorithm can be used to detect distinct features in an image. Once features have been detected for two different images, one can use these features to answer questions like “are the two images taken of the same object?” and “given an object in the first image, is it present in the second image?”. Computation of SIFT image features is performed through the four consecutive phases which are briefly described in the following. In order to get a reliable recognition, it is quite important that the features extracted from the training image are detectable even under changes in image scale, noise and illumination. Such points generally lie on high-contrast regions of the image, for example object edges. Gesture recognition is initially performed by matching each key point independently to the database of key points extracted from training images. Many of these initial matches will be incorrect due to ambiguous features or features that arise from background clutter. Therefore, clusters of some features are first identified that agree on an object and its pose, as these clusters have a much higher probability of being correct than individual feature matches. Then, each cluster is checked by performing a detailed geometric fit to the model, and the result is used to accept or reject the interpretation. With the help of this algorithm, we were able to decode gestures successfully. The feature extraction was done efficiently using SIFT. The SIFT features described in our implementation have been computed at the edges which are invariant to scaling, rotation, addition of noise. These features are useful due to their distinctiveness, which enables the correct match for key points between different hand

gestures. The proposed approach was tested on real images. Also, computation time was found to be lesser for grey scale images than with colour images.