**Chapter 8**

**Conclusion and Future work**

**8.1 Conclusion**

One of the main purposes of the interactive virtual environments is to provide natural, efficient, and flexible communication between the user and the computer. Human gestures including positions and movements of the fingers, hands, and arms represent one of the richest non-verbal communication modalities allow human users to interact naturally with the virtual environment. Sign gesture can be static, where the human takes on a specific pose.

Real-time vision-based sign recognition is one of the most challenging research areas in the human computer interaction area. Vision-based sign gesture recognition can rely on generic video cameras already available on a large variety of computers, tablets, smart phones, etc to recognize hand gestures.

This thesis proposed sign gesture recognition system, which works under different lightning conditions with changed transformations and cluttered background. The Sign gesture recognition systems will run a more efficient and natural interaction, modality for artistic applications. Another important application will be in the sign language recognition for the deaf people.

A First real-time visual hand gesture recognition system proposed in Chapter 5 contains of two parts: one for hand detection and tracking using face subtraction, skin detection and contours comparison algorithms proposed in Chapter 5, and the second one which performs gesture recognition using Conventional Neural Network (CNN). The developed CNN algorithm consists of two stages: a training stage where hand postures training images are processed to calculate the M highest eigenvectors and weights for every training image, and a testing stage where the weights of the detected hand gestures are matched with the weights of the training images for hand gesture recognition.

To conclude, in this thesis a real-time system was proposed that consists of three modules: hand detection and tracking using face subtraction, skin detection and contour comparison algorithm, posture recognition using bag-of-features and multiclass SVM, and a grammar that generates a large number of gesture commands by monitoring the scale of the detected hand posture, its movement direction, and the transitions among postures. In the training stage, after extracting the keypoints for all the training images using SIFT algorithm, VQ was performed on the keypoints for every training image using a k-means clustering to map them into a unified dimensional bag-of-words vector, which is used as input vector for building the multiclass SVM classifier model. Then, the multiclass SVM classifier will be used in the testing stage to classify the detected hand posture captured from a webcam after constructing visual words vector for keypoints of the small image (50 × 50 pixels) that contains the detected hand gesture only. The testing stage proves the effectiveness of the proposed scheme in terms of accuracy and speed as the keypoints extracted represent the detected hand gesture only. Experiments show that the system can achieve satisfactory real-time performance regardless of the frame resolution size as well as high classification accuracy of 96.23% under variable scale, orientation and illumination conditions, and cluttered background. Three important factors affect the accuracy of the system, which are the quality of the webcam in the training and testing stages, the number of the training images, and choosing number of clusters to build the cluster model.

**8.2 Future Work**

Sign gesture recognition still has a long way to go in the research path, especially for 2D systems. This study offers fascinating ideas for future research. Some of these possibilities are defined in this section.

**Dynamic Sign Gesture Recognition**

As this thesis focused only on static sign gesture recognition, one next step forward is to recognize the dynamic sign gesture for the ASL.

**Sign Gesture Recognition from video:**

Now a day’s videos are generally found on the internet. The idea of categorizing single frames is a start to classifying frames in videos. This can be applied in real time classifications. Extending the algorithms proposed in this thesis to video and building an automatic transcript system is an important step onward. For this purpose, it might be fascinating to explore sequential models that study the time dimension, such as recurrent neural networks and or a neural architecture combining CNNs and RNNs.

**Apply to 3 Dimension** **technique**

Now days 3Dimension cameras and sensors are very easy to available in market and less expensive. This different type of sensor can provide much more information about the hand, making it possible to create more accurate systems for sign language real time recognition.

**Add more gesture in dataset:**

Even though that study introduces a self generated new dataset with a rather more gesture for American Sign Language, it still does not offer all the possible movements for American Sign Language. Videos with rotation in 3Dimension, words and expressions are examples of how this dataset can be extended.