1. Sign Based Approach

Sign based approaches use the sign as the smallest unit and model each sign with

a different temporal model.

* 1. Isolated Sign Recognition

Isolated sign recognition can be thought as lexicon search. It can be search of known word or finger spelled special word.The importance of the research on isolated sign recognition is that it enables the finding of better mathematical models and features that represent the performed sign.

* 1. Continous Sign Recognition

Continous sign recognition can be thought as translating sentences from one language to another.Meaning can change the according to previous or next sign. Additional movements or shapes may occur during transition between signs. These movements are called movement epenthesis (ME). ME has to be considered in sign model.

1. Sub-Unit based Approach

When sign number increases, model complexity and scalability problems arise. For solving this issue signs are considered as spoken language so that they have phonemes/subunits like syllable.The advantage of identifying phonemes is to decrease the number of units that should be trained. The number of subunits is expected to be much lower than the number of signs. This will enable large-vocabulary sign language recognition with tractable computation times. But there is no certain sub unit definition. Phonemes for sign language recognition are not clearly defined.

**Kuznetsova, A., Leal-Taixé, L., & Rosenhahn, B. “Real-time sign language recognition using a consumer depth camera”, ICCV 2013 Workshop paper**

**[1.1]**

Most of sign language recognition works difference in viewpoint (or alternatively, different rotation, position and scaling of hand), enviroment and subject appearence, represents significant difficulty. In this work they adress the problems mentioned above by using rotation, position and scale invariant features. By using these features intra class variation is reduced. Such features exist for 3D point clouds which can be derived from the depth data. Ensenmble of shape function (ESF) descriptors which is suggested in state of art 3D object classification work are used. ESF consists of a set of histogram concetaned together:

1. The first histogram describes the distribution of distances between two random points in the point cloud (funtion D2).
2. The second histogram describes the distiribution of angles enclosed by three randomly sampled points (function D3).
3. For each line connecting two points it is determined if it is on the point cloud surface, off the surface or intersects the surface.

Then they use multi layered random forest (MLRF) for classification.MLRF significantly reduces the training time and memory consumption. MLRF is a random forest consisting of two layers.

This method demonstrates low training time, low memory usage and high accuracy in one subject tests.

**Cooper, H., Ong, E. J., Pugeault, N., & Bowden, R. (2012). Sign language recognition using sub-units. *The Journal of Machine Learning Research*, *13*(1), 2205-2231.**

**[2]**

Classic sign language recognition tecniques often builds classifier per gesture. This approach becomes intractable when recognising large lexicons of signs. Like speech recognition sub-unit based SLR uses two stage recognitoin. In the first stage sign linguistic sub-units are identified. In the second stage these sub units are combined together to create a sign level classifier. This paper discusses sign language recognition using linguistic sub-units. It presents three types of sub-unit extraction for consideration; those learnt from apperance data as well as those inferred from both 2D or 3D tracking data. Then sub-units are combined two different types for a sign level classifier. The first uses Markov Models to enconde temporal changes, second uses the sequential pattern boosting to apply discriminative feature selection at the same time as encoding temporal information. In 2D apperance based sub unit extraction location features, motion and hand arrangment moment features, motion binary patterns and additive classifiers are used. In 2D tracking sub unit extraction 2D tracking data as motion features with appereance based handshapes are used.HoGs are calculated for hand shape classification. In 3D sub unit extraction kinect’s tracking ablity and body joint points are used.

This paper gets best results when 3D tracking and a new learning method Sp-Boosting is used. They get 99.9% accuracy subject dependent and 85.1% accuracy subject independent tests. For using sub-unit based SLR more linguastically annotated data is required.

**Ong, Eng-Jon, et al. "Sign language recognition using sequential pattern trees."*Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012.**

**[2]**

This paper presents a novel classifier which is called Sequential Pattern Trees. It is efficient to learn compared to other Sequential Pattern methods and scalable for use with large classifier banks. This paper focuses on classification stage of SLR. They use same features and sub-units with Cooper et al.( paper above). In classification stage firstly they improve existing SP based approaches in the following ways: They introduce a novel SP tree that is multi class in nature. Allowing them to tackle the binary limatation of existing SP classifiers. Importantly the tree structured model facilitates spatiotemporal feature sharing across different classes. The classifiers scale very well with number of classes. Run time of this classifier is independent of the number of classes.

**Ren, Zhou, et al. "Robust part-based hand gesture recognition using kinect sensor." IEEE Trans. Multimedia 15.5 (2013): 1110-1120.**

**[1.1]**

To handle the noisy hand shapes obtained from Kinect sensor, they propose a novel distance metric, Finger-Earth Mover’s Distance (FEMD), to measure the dissimilarity between hand shapes. As it only matches the finger part not whole hand it can better distinguish the hand gestures of slight differences. This approach is robust to hand articulations, distortions and orientation or scale changes.

Correspondence-based shape matching algorithm and skeleton matching methods cannot robustly recognize shape contour with severe distortions. This proposed FEMD is robust to orientation, scale, articulation changes as well as local distortions of hand shapes. The signer must wear a black belt for robust hand segmentation. Kinects depth information and black belt are used to hand segmentation. After detecting the hand shape they represent it as time series curve. Earth mover’s distance (EMD) is a general and flexible metric used for measuring distance between signature or histograms. EMD is widely used in content based image retrieval and pattern recognition. In FEMD they represent the input hand by global features (the finger clusters) and they add a penalty on empty holes to alleviate partial matches on global features. In order to measure FEMD between hand shapes hand shapes must be represented as signature with each finger as a cluster. Finger Detection plays the most important role in this work.

**Yang, Hee-Deok, and Seong-Whan Lee. "Robust sign language recognition with hierarchical conditional random fields." *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010.**

**[1.1]**

In this paper, a novel method for spotting signs and fingerspellings simultaneously is proposed, which can distinguish signs, fingerspellings, and nonsign patterns. This is achieved through a hierarchical framework consisting of three steps; (1) Candidate segments of signs and fingerspellings are discriminated with a two-layer conditional random field (CRF). (2) Hand shapes of detected signs and fingerspellings are verified by BoostMap embeddings. (3) The motions of fingerspellings are verified in order to distinguish those which have similar hand shapes and differ only in hand trajectories.

The two-layer CRFs only discriminates candidate signs, fingerspellings, and non-sign patterns using both hand motion and hand location as features. The BoostMap method, which is an embedding method, is applied in order to recognize hand shapes in candidate sign and fingerspelling segments.

**Uebersax, D., Gall, J., Van den Bergh, M., & Van Gool, L. (2011, November). Real-time sign language letter and word recognition from depth data. In*Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* (pp. 383-390). IEEE.**

**[1.1]**

This work presents a system for recognizing letters and finger-spelled words fo the ASL in real time. They segment the hand and estimate the hand orientation from captured depth data. The letter classification is based on average neighborhood margin maximization and relies on the segmented depth data of the hands. For word recognition, the letter confidences are aggregated. Furthermore, the word recognition is used to improve the letter recognition by updating the training examples of the letter classifiers on-line.

For letter classification three methods have been evaluated. The first method relies on a codebook of hand gestures where each codebook entry contains only one single training example. The second method is based on average neighborhood margin maximization (ANMM) [26] that is more suited for classification of hand gestures in a multi-user environment. The third method estimates the hand orientation and uses the orientation as additional cue for letter recognition. For letter recognition firstly they estimate the palm and orientation of the hand.They find the mass center as hand center. Then they find the finger tips which is the point has largest distance from hand center and is inside the hand region. Then they find the orientation by calculating the angle between the hand center and finger tips.

For the word recognition, we have used a lexicon where new words can be added or removed without requiring an additional off-line training step. The results have shown that an accurate recognition of all letters is not necessary for reliable word recognition. A letter with low confidence can be updated based on the high confidences of the other letters within the same word.

**Tang, Matthew. "Recognizing hand gestures with microsoft’s kinect." Palo Alto: Department of Electrical Engineering of Stanford University:[sn] (2011).**

**[1.1]**

In this work they propose and implement a novel method for recognizing hand gestures using rgb and depth data. This approach looks at spesific hand motions in addition to full body motions. They only recognized two gestures like grasp and drop. Because of low resolution fo kinect depth camera hand occupies 64x64 pixel the most and resulting depth images suffer from IR occlusion and other effects. In this work skin color is fitted to Mixture of Gaussian (MoG) model. After model fitting color balancing is done for elimination of false positive pixel such as neck and background pixel in the skin color. Then depth information is integrated with color image and hand pixels are segmented. They normalize segmented image for scale and rotation invariance. They use SURF like feature and call it modified SURF. They crop 64x64 images into 8x8 sub regions and calculate integral of these blobs. After feature extraction they demonstrate a simple method of gesture recognition using forward recursion to make their results more robust to classification noise.

**Elmezain, M., Al-Hamadi, A., Appenrodt, J., & Michaelis, B. (2008, December). A hidden markov model-based continuous gesture recognition system for hand motion trajectory. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on (pp. 1-4). IEEE.**

**[1.1], [1.2]**

An automatic system that recognizes both isolated and continuous gestures for Arabic numbers (0-9) in real-time based on Hidden Markov Model (HMM) is proposed. In this system they recognize Arabic numbers from color image sequences by the motion trajectory of a single hand using HMM. Their system depends upon the following main steps; using Gaussian Mixture Model (GMM) for skin color detection, the orientation between two consecutive points is extracted as basic feature, zero-codeword detection with static velocity motion, Baum-Welch (BW) algorithm for training and forward algorithm in conjunction with Viterbi path for testing. For hand segmentation 3D depth map and color information is used. YCbCr color space used for robust segmentation. There are three basic features: location, orientation and velocity. Orientation is the best in term of accuracy results. For the continuous gesture, this system is designed to segment and recognize the isolated gesture by zero-codeword detection. Each gesture ends by line segment, which is assigned a 0-codeword. The proposed system is suitable for real-time application and depends on novel idea of zero-codeword detection with static velocity to recognize the continuous gestures.

**Aran, O., & Akarun, L. (2010). A multi-class classification strategy for Fisher scores: Application to signer independent sign language recognition. Pattern Recognition, 43(5), 1776-1788.**

**[1.1], [1.2]**

Fisher kernels have been proposed as a method to map a variable length sequence to a new fixed dimension feature vector space . The mapping is obtained by the derivatives of the parameters of an underlying generative model. This new feature space is called the Fisher score space on which, any discriminative classifier can be used to perform discriminative training. The main idea of Fisher kernels is to combine generative models with discriminative classifiers to obtain a robust classifier which has the strengths of each approach. This study aims to use Fisher kernels to map the original variable length sign sequences based on HMMs to the fixed dimension Fisher score space, and to apply a discriminative multi-class classification on this new feature space. They use the discriminative power of the Fisher scores of one class to classify other classes.

They have concentrated on the manual component and extracted features only from the hand motion, shape and position. Hand motion analysis, the center of mass of each hand is tracked and filtered by a Kalman filter. The posterior states of each Kalman filter x; y coordinates of CoM and horizontal and vertical velocities, are the hand motion features. Hand shape features are appearance based shape features calculated on the segmented hand images. These features include the width, height and orientation parameters of an ellipse and seven Hu moments calculated on the binary hand image.

For a binary classification problem, one might have three different score spaces based on likelihoods: (1) LSS from the generative model of class 1. (2) LSS from the generative model of class 2. (3) LRSS from the generative models of classes 1 and 2. They propose a new strategy, MDLC, which applies a multi-class classification on the LSS of each class and then combines the decisions of each classifier.

Sequential floating forwards search (SFFS) is originally proposed as a feature selection algorithm that applies a top down search strategy. They use SFFS as a score space selection method and for each score space, obtained from the generative model of each class, they train a classifier.

As a summary in this study, they proposed a multi-class classification strategy for Fisher scores. The main idea of multi-class classification strategy is to use the Fisher score mapping of one model in the classification process for all of the classes. As a result, each mapping is able to discriminate all the classes up to some degree. When all of these mappings are combined, higher accuracies are obtained when compared to the existing multi-class classification approaches in the literature. Selection strategies and see that the SFFS strategy finds the smallest sized subsets with performance comparable to that of exhaustive search for all score spaces. The results with the score space selection show that without a significant decrease in the accuracy, they are able to reduce the computational cost.

**Kong, W. W., and Surendra Ranganath. "Towards subject independent continuous sign language recognition: A segment and merge approach."*Pattern Recognition* 47.3 (2014): 1294-1308.**

**[1.1], [1.2]**

This paper closely follows the sign language model defined by linguists, and includes all the four components, viz, handshape, movement, palm orientation and location.

In this work they propose a two-layer multichannel system for recognizing continuously signed ASL sentences. They used Cyberglove and magnetic trackers for data acquisition. They first segment the continuous input sequences using a segmentation algorithm based on minimum velocity and maximum directional angle change in the movement channel (segmentation algorithm is taken from their previous work). They adopt a segment-based approach to recognize the continuously signed sentences, rather than the usual frame-based approach. The SIGN/ME (movement epenthesis) labeling is done by a Bayesian network (BN) classifier that fuses the results of independent conditional random field (CRF) and SVM classifiers. Following this,the SIGN sub-segments are retained for sign decoding; the ME sub-segments are discarded, though their locations are recorded, as they provide useful information to the final decoding algorithm for reducing computations.

They then work out a strategy to merge the SIGN sub-segments and recognize the signs with a two-layer CRF model. The lower layer of the recognition module consists of four (linguistic sub units) independent linear CRF models to recognize and output sequences of phoneme labels in the component channels, from the corresponding SIGN sub-segment sequences. The corresponding phoneme labels from the four channels are then combined and input to the upper layer semi-Markov CRF for sign recognition.

The raw data obtained from the glove and trackers consists of handshape (16-D vectors), palm orientation (9-D vectors) and position (3-D vectors). In the movement channel, they used movement direction and trajectory shape as the basic descriptors rather than raw position vectors.