Proc. of The Sixth Intl. Conf. On Advances In Economics, Social Science and Human Behaviour Study - ESSHBS 2017

ISBN: 978-1-63248-120-7 doi: 10.15224/ 978-1-63248-120-7-41

Theoretical Frames for Sign and Dactyl Recognition System (GESL data)

Tamar Makharoblidze

Abstract: This paper discusses the theoretical frames for sign language machinery recognizing using the data of Georgian Sign Language (GESL). The best practical achievements are the elaborated computer systems recognizing the static signs and dactyls.

Keywords - **Sign Languages, computer linguistics, Georgian Sign Language, GESL**

I. Introduction

Georgian Sign Language (GESL) is a native language for about 2500 Deaf and Hard of Hearing people (DHH) in Georgia. These people are the linguistic minority in the country, whose main problem a lack of communication with hearing people. The same problems have the other DHH-s in the other countries word-wide. The engineers of modern technologies try to resolve this communication problem. For today creating the computer translator from sign languages (SL) into spoken languages and vice versa is a challenge for scientists.

Research on technological solution for SL recognition already has its history as during the last two-three dozen years a number of virtual studios worldwide carried out the various experiments (among others Starner et al 1998, Imagawa et al 1998, Liang et al 1998, Gao et al 2002). In this paper we will discuss the theoretical frames for the abovementioned task taking into consideration the latest trends, new technologies, algorithms and approaches.

The truth is that Text to Sign translator is easier to elaborate then vice versa engine. There is a number of software and web sites working on desktops or/and mobiles. These systems have prerecorded SL samples for the each word (from the certain dictionary) in video or animation format. The system can display on screen the video or animated sign equivalent of the text, after converting the verbal data into textual one, via Speech to Text APIs (Application Programming Interface). American Sign Language (ASL) online Dictionary is a good example of it. (http://www.handspeak.com/word/)

In spite of such an intenssive interest towards this issue for today Sign-Spoken language translator (SSLT) device is not elaborated yet. We offer the theoretical frames and a new theory for SSLT.

II. The sign-processing paradigms

In spite of different devices and methods, the environment in this field is almost identical for the all researches in general. The various devices were used for sign recognition. In order to get information Human Computer Interaction (HCI) devices follow these steps below:

• Data is written in computer from the device;
• Database primary learning/elaboration/development is performed;
• New data is compared with the existing one.

The results appear with a various approximation depending on the approaches and methods, processing principles and algorithms and on the combinations of these factors.

We observed two basic technical paradigms to solve the problem:

- **Graphic processing** is the eldest paradigm historically (Tang 2011, Ying 2014). In this case, the processing and/or comparing of photo or video material is performed by histogram, outline/contour or graphic identification means. This method can have good results, but it is strongly dependent on the primary source of the resolution, on the quality of the entries in the environment and on the well-visible hand structure.
- **Non graphical processing** – in this case the data is "descriptive" (i.e. not graphical /visual format) the data depends on the particular device and on its technical characteristics. It can be as a custom made device as well as any other existing devises, like Microsoft Kinect (Marin at al 2014, Yi Li 2012, Murata & Shin 2014; Tang 2011), or Leap-Motion (Marin at al 2014, Marin at al 2015, McCartney, at al 2015, Nowicki, et al. 2014). These devices are cheap and do not require additional supplies or a special studio environment. These devises have the well-defined application programming interfaces (API) and strong users' communities. They are easy to integrate into any computer.

III. The sign-recognizing methods

A. Used methods

Based on our research of existing literature on the topic we can conclude that there are roughly two approaches for sign-recognizing.

- **SVM (Support Vector Machine).** It is a supervised machine learning algorithm which can be used for classification problems. It uses a

Proc. of The Sixth Intl. Conf. On Advances In Economics, Social Science and Human Behaviour Study - ESSHBS 2017

ISBN: 978-1-63248-120-7 doi: 10.15224/ 978-1-63248-120-7-41

technique called the kernel trick transforming the given data, and basing on these transformations it can find optimal boundaries between the possible outputs. SVM does some extremely complex data transformations, then figures out how to separate the proper data based on the labels or on the already defined outputs. http://www.yaksis.com/posts/why-use-svm.html

- **HMM (Hidden Markov Model).** The Hidden Markov Model is a finite set of *states*, each of which is associated with a (generally multidimensional) probability distribution. Transitions among the states are governed by a set of probabilities called *transition probabilities.* In a particular state an outcome or *observation* can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are 'hidden" to the outside; hence the name Hidden Markov Model." http://jedlik.phy.bme.hu/~gerjanos/HMM/node4.html

  This statistical learning theory has the ability to absorb both the variability and the similarity between patterns. It is based on the empirical risk minimization (ERM) principle, which is the simplest of induction principles, where a decision rule is chosen. The decision rule is based on a finite number of known examples (training set). (Justino, et al 2005)

This second method is used for recognizing dynamic signs. It gives the possibility to compare the data in time series with HMM. The approximation varies at some researches (McCartney et al 2015) it is about 50%, while others have 73% (Chuan et al 2014). On average the approximation rates about 50% -75%.

B. New Hybrid method of Bottom-up and Top-down approach

One common drawback of bottom-up approach is that it can be a highly dependent on accurate hand segmentation. However, a top-down approach requires a complete understanding and prior knowledge of the domain and its constraints. Also, top-down approach is less adaptable to revise the state transition matrix (when computing with HMM) in the presence of new states. Whereas, bottom-up approach starts with fine-scale representation and sequentially clusters the states that are similar. Therefore, it is optimal to use a hybrid approach of bottom-up and top-down approach. A S. Lu (2012) proposed a hybrid approach of using motion and color cues to select multiple hand candidates as bottom-up approach and information form th model is used to select a single optimal sequence among the many possible sequences of hand candidate as top-down approach. This hybrid approach reduced the requirement of accurate segmentation and the system in much more robust.

IV. Leap-Motion device

Our attention was drawn to Leap-Motion. Unlike Kinect (http://www.xbox.com/en-US/xbox-360/accessories/kinect) it is fully oriented to the hand motion detection and it has its gesture recognizing system – default gesture system. This device is compact. Its dimensions are 13 × 30 × 76 mm. It can receive more accurate signal about 50 centimeters in radius. Leap-Motion can be joined the computer by USB port and it gives the raw data of 120fps frequency. Leap-Mpotion has a well-developed API (Application Programming Interface) for almost all of the popular programming languages C #, JavaScript, Python, etc. It is mostly used in video games, as well as a-la mouse, or a virtual presentation device. Leap-Motion gives the structured data in JSON (JavaScript Object Notation) format with the described coordinates (positions and rotations) for the wrist, hand and fingers.

With Leap-Motion we receive the data submitted in a structured JSON (JavaScript Object Notation). Its format describes the coordinates for wrist, hand and fingers (positions and rotations). Along with the computed tracking data, one can get the raw sensor images from the Leap-Motion cameras. leap-Motion is limited with its working area, but some signs are out of the limited area (out of the chest and head area), although the number of such signs is not big, but still it is important to have wider working-area. Leap-Motion has a problem to detect the both hand together. Additionally mimic is very important for SL texts and it may carry the important information. (For more detailed comments on limitation of Leap-Motion see Potter et al 2013). Hopefully in near future the updated versions of Leap-Motion will able to overcome these limitations.

Consequently, the combined devises which will be able to perform the data processing of full body infrared 3D and video face recognition data can resolve the problem of sign recognizing in the signing process. Despite the term that LeapMotion isn't suitable for SL deep exploration, anyway it might be useful to use as a data provider and a format.

Modern system more and more are AI (Artificial Intelligence) based, and working on big data collections can give some results. It's possible to record SL data with LeapMotion integrating it to the systems like Wiktionary.com in order to create a crowd-sourcing platform. Data collected by LeapMotion is simple JSON. Thus it can be stored as usual text and then it is possible to make sign animation also to use it for machine learning: finding similarities and differences; describe NS and perform any type of experiments. This approach is more convenient comparing to video format as it is more straightforward and less depends on the recording environment.

V. Sign classification

Signs can be static or dynamic, one or two-handed. Two-handed signs may be symmetric or asymmetric. Besides, among two handed signs either both hands are

Proc. of The Sixth Intl. Conf. On Advances In Economics, Social Science and Human Behaviour Study - ESSHBS 2017

ISBN: 978-1-63248-120-7 doi: 10.15224/ 978-1-63248-120-7-41

producing dynamic or static signs, or one hand produces a static sign while another one does a dynamic sign.

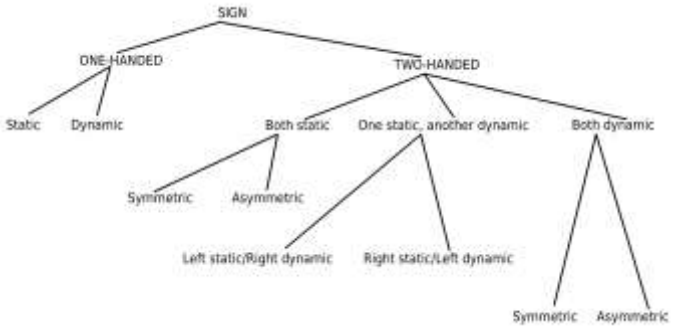
For sign classification we used the combination approaching:

- Dynamical gradation (with space and time parameters) - The signs are statistic or/and dynamic. Dynamic signs may have one, two or more movement - phases;

- Composition of a sign / sign structure – the signs may have one, two, three or four (very rarely five) elements or the independent signs with (sometimes totally different) meanings. Signs may be as following A=a; A=a+b, A=a+b+c, etc.;

- For our description one-handed and two handed signs can be described in the same way, although there can be a significant difference between the sign producers and their moving/sign producing kinetics.

Classification of signs schematically looks as follows:



The signs may be simple or compound. Compound signs may have two or more (up to five as maximum) meaningful signs in the strict sequence.

{ MSa+NSab+MSb=MSc
MSb+ NSba+MSa ≠ MSc

For example in GESL the sign “agricultural” is the sum of three MS “village”, “variety” and “function”.

VI. Theory of Neutral Signs

A. The types of signs in signing process

To elaborate SSLT from SL into spoken languages is more difficult comparing with vice versa version - translation from spoken language into SL. Usually SL texts are performed smoothly and there are no spaces between meaningful signs (MS). The biggest problem for elaborating a good engine of SL machinery translating is a lack of sign separators or spaces. In SL texts it is hard to understand where is the beginning or ending of a proper sign. To overcome this obstacle we offer a new theory - Theory of neutral signs (TNS).

There are two types of manual signs:

- Sign with meaning – MS (meaningful sign). These are the signs with lexical content (like words) or with morpho-semantical meanings (such as particles or morphemes of different grammar categories), and
- Sign without any meaning, who serves as a connection for manual positions of two neighbor meaningful signs (MSs). It is a neutral sign (NS). NS could be also named as a garbage sign. NSs are inter-signs between MSs.

MS can be static or dynamic, one- or two-handed, simple or compound with two or three (and rarely more) signs in the specific sequence. The compound signs can be described as A+B(+C+D)=S.

NS is a dynamic signs between MS-s (static or dynamic). Unlike MS, NS is always dynamic. Every MS has three steps of sign production:

- The first step is preparation or excursion - MSe;
- ;The second step is a top MSt - the moment of sign exposition, and
- The last third step is post production or recursion (or disposition) - MSr.

The first and third steps are usually mixed with the parts of neighbor signs. At the beginning of the signing process there is a neutral sign beginner - NSb and it brings the hand(s) from zero position to MSe. (1. Zero position is the position hands hanging down and maybe slightly bent in the elbows.) NSf – is the final neutral sign in the signed text, bringing the hand(s) to zero position from MSr.

In SL text sequence, the signing dynamics of the two signs is Sa+Sab+Sb. In real signing time there are the three signs, where Sab is NS between these two MSs (Sa and Sb). This type of NS is a middle or intermediate. It connects two MSs having the mixed characteristics from the ending part of the first (MSr) and the beginning part of the second sign - MSe.

Thus, there are three types of NS:

- NS connecting (Sab, MSr+MSe);
- NSb – the first, beginning sign, and
- NSf – last, finishing sign.

In SL phrase / sentence “I paint” looks as follows:
MR(I)e+MR(I)t+MR(I)r+MS(paint)e+MS(paint)t+MS(paint)r
MR(I)r+MS(paint)e=NS(I+paint)
But this description is still incomplete as MR(I)e and MS(paint)r will be bordering with the other signs in longer sequence creating specific NSs, or if this it is a separate text, then before MS(I)e there will be NS(I)b and MS(paint)r will be followed by NS(paint)f. This SL text will be described as

NS(I)b+MR(I)e+MR(I)t+NS(I+paint)+MS(paint)t+MS(paint)r+NS(paint)f

Proc. of The Sixth Intl. Conf. On Advances In Economics, Social Science and Human Behaviour Study - ESSHBS 2017

ISBN: 978-1-63248-120-7 doi: 10.15224/ 978-1-63248-120-7-41

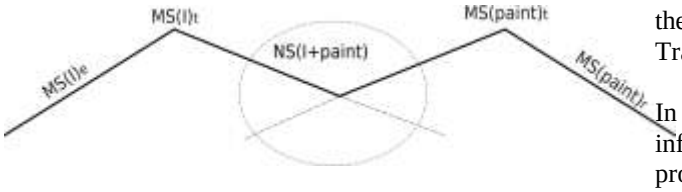


Figure 1. *SL text fragment scheme*

B. NS recognition methods

Thus, NS can be considered as a space between MSs or in the other words, NS is a sign separator. The question is how NS can be recognized by the engines. We revealed the four methods to identify NS in SL texts:

- Synergistic method for NS recognition – This method can work after analyzing a big number of SL texts of different SLs, having a big common SL textual base that will be NS base at the same time. Such data base can be filled only with common effort using open sources and word wide collaboration of the area specialists;

- The approximate parameters of NS can be defined depending on its neighboring; NS can be characterized with less tension of the manual muscles and skin and without any accompanied mimic; non-prosodic element; with freely and lightly curved/hanging, resting fingers; with transitional palm orientation and hand location depending on the proper neighborhood in the sign sequence. NS may look like a MS, or it can be MS in another SL, but the abovementioned general parameters (such as less tension and non-prosody) helps to separate any type of NS from MS.

- The combined identification of NS could be performed with Leap-Motion and Myo armband. The pause converged with Leap-Motion's minimal activity can be considered as a sign-separator in SL process, if it is not a static sign. Meanwhile, there is a limited number of static signs in any SL and they can be described in the proper SL corpora, or data-base, or learned by the neuro-nets. The engine can identify static signs and distinguish them from NSs and pausing.

VII. Recognizing of static signs

Our first experiments were connected with static signs. A few signs were recorded and we tried to complete data abstraction introducing the signs as separate objects and identifying the correlation between them by Pearson product-moment correlation method. Surprisingly, the results varied between 60-70% for dozens of signs. Taking into consideration the work of our colleagues (Nowicki et al 2014, Marin et al 2014). We tried SVM method and improved the results.

We recorded the Georgian dactyls (Makharoblidze 2013) and the signs from the GESL dictionary (Makharoblidze 2015) using the existing library (O'Leary), which nicely fits the static sign recognizing extending with the MIT license. Training and recognizing takes place by SVM method.

In spite the fact that in general, Leap-Motion gives the exact information, even in case of static signs the following problem may appear - the device cannot identify the coordinates of the closed hands or fingers crossing each-other, or two fingers together, fingers with certain angles, and it is often depends on the signer. Australian dactyl recognizing system had the same problem. (Potter et al 2013)

We created the micro-corpora of GESL recording the signs from the GESL dictionary (Makharoblidze 2015) with a few Deaf persons – this recorded dictionary was oriented to Leap-Motion data. We tried to use sign-to-word recognizing method and we noticed that the increasing the number of signs reduces the quality of sign recognizing process. In addition more signs are not static, but they are dynamic, the problems were deeper in case of the combined or composed dynamic signs.

On the next step we performed the data processing taking into consideration the existing experiences of HMM methodology. The results were rather miscellaneous and unsatisfactory, from 40% to 90%. The main reason is that Leap-Motion data while hand-moving is quite noisy. This especially occurs at the rapid dynamics of the hands. In some cases the data of finger positions can be missing or may be interpreted incorrectly.

References

- American Sign Language Dictionary *Retrieved from* http://www.handspeak.com/word/\
- Bernhard, H. P., and G. Kubin. "Speech production and chaos." *XI-Ith Int. Congress Phonetic Sciences*. 1991.
- Casdagli, Martin, and Stephen Eubank. *Nonlinear modeling and forecasting: proceedings of the Workshop on Nonlinear Modeling and Forecasting held September, 1990, in Sante Fe, New Mexico*. Vol. 12. Westview Press, 1992.
- Cowper, Mark R., Bernard Mulgrew, and Charles P. Unsworth. "Nonlinear prediction of chaotic signals using a normalised radial basis function network." *Signal Processing* 82.5 (2002): 775-789.
- Chuan, Ching-Hua, Eric Regina, and Caroline Guardino. "American Sign Language recognition using leap motion sensor." Machine Learning and Applications (ICMLA), 2014 13th International Conference on. IEEE, 2014.
- Faundez-Zanuy, M., et al. "Nonlinear speech processing: overview and applications." *Control and intelligent systems* 30.1 (2002): 1-10.
- Georgian Dactyl Font, *Retrieved from* http://sign.iliauni.edu.ge/font/ (last access: November 2016)
- Jonathan Harringthon, Contextual ambiguities in speech signals and their consequences for sound change. International Conference on Nonlinear Speech Processing. NOLISP 2015. Jointly organized with the 25th Italian Workshop on Neural Networks, WIRN 2015 https://sites.google.com/site/nolisp2015/home http://www.phonetik.uni-muenchen.de/~jmh/
- Jordan, Michael I., and Robert A. Jacobs. "Hierarchical mixtures of experts and the EM algorithm." *Neural computation* 6.2 (1994): 181-214.
- Kantz, Holger, and Thomas Schreiber. *Nonlinear time series analysis*. Vol. 7. Cambridge university press, 2004.

ISBN: 978-1-63248-120-7 doi: 10.15224/ 978-1-63248-120-7-41

- Kiraz, George Anton. *Computational nonlinear morphology: with emphasis on semitic languages*. Cambridge University Press, 2001.
- Kokkinos, Iasonas, and Petros Maragos. "Nonlinear speech analysis using models for chaotic systems." *IEEE Transactions on Speech and Audio Processing* 13.6 (2005): 1098-1109.
- Kumar, Arun, and S. K. Mullick. "Nonlinear dynamical analysis of speech." *The Journal of the Acoustical Society of America* 100.1 (1996): 615-629.
- Marin, Giulio, Fabio Dominio, and Pietro Zanuttigh. "Hand gesture recognition with leap motion and kinect devices." 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014.
- Marin, Giulio, Fabio Dominio, and Pietro Zanuttigh. "Hand gesture recognition with jointly calibrated Leap Motion and depth sensor." Multimedia Tools and Applications (2015): 1-25.
- Imagawa, Kazuyuki, Shan Lu, and Seiji Igi. "Color-based hands tracking system for sign language recognition." Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. IEEE, 1998.
- Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. Applied Linguistics, 18(2), 141-165.
- Larsen-Freeman, D. (2002). Language acquisition and language use from a chaos / complexity theory perspective. In C. Kramsch (Ed.), Language acquisition and socialization (pp.33-46). London: Continuum International Publishing Group.
- Leap Motion, http://www.leapmotion.com (last access: November 2016)
- Liang, Rung-Huei, and Ming Ouhyoung. "A real-time continuous gesture recognition system for sign language." Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. IEEE, 1998.
- Makharoblidze, Tamar. "The Georgian Dactyl Alphabet." *Disability studies Quarterly. DSQ*, Vol. 33, No. 3 2013 33 N. 3 (2013).
- Makharoblidze T. (2015) Georgian Sign Language Dictionary. Ilia State University; Shota Rustaveli National Scientific Foundation. Tbilisi. ISBN 978-9941-16-225-5 1368 pp.
- Makharoblidze, Tamar, *Retrieved from* http://gesl.iliauni.edu.ge/, 2015
- McCartney, Robert, Jie Yuan, and Hans-Peter Bischof. "Gesture Recognition with the Leap Motion Controller." Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2015.
- Nowicki, Michał, et al. "Gesture recognition library for Leap Motion controller." Bachelor thesis. Poznan University of Technology, Poland (2014)
- Rob O'Leary, *Retrieved from* https://github.com/roboleary/LeapTrainer.js/ (last access: March 2016)
- G. Pfurtscheller, D. Flotzinger, W. Mohl, M. Peltoranta Prediction of the side of hand movements from single-trial multi-channel EEG data using neural networks. Electroencephalography and Clinical Neurophysiology. Volume 82, Issue 4, April 1992, Pages 313–315
- Potter, Leigh Ellen, Jake Araullo, and Lewis Carter. "The leap motion controller: a view on sign language." Proceedings of the 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration. ACM, 2013.
- Riedmiller, Martin, and Heinrich Braun. "A direct adaptive method for faster back propagation learning: The RPROP algorithm." *Neural Networks, 1993., IEEE International Conference On*. IEEE, 1993.
- Mukherjee, Sayan, Edgar Osuna, and Federico Girosi. "Nonlinear prediction of chaotic time series using support vector machines." *Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop*. IEEE, 1997.
- Sato, Masa-Aki, and Shin Ishii. "On-line EM algorithm for the normalized Gaussian network." *Neural computation* 12.2 (2000): 407-432.
- Gregor Schöner, Hermann Haken, Scott Kelso A stochastic theory of phase transitions in human hand movement. Biological Cybernetics 53(4):247-57 · February 1986
- Starner, Thad, Joshua Weaver, and Alex Pentland. "Real-time American Sign Language recognition using desk and wearable computer based video." IEEE Transactions on Pattern Analysis and Machine Intelligence 20.12 (1998): 1371-1375.
- Nasrin HADIDI TAMJID, CHAOS / COMPLEXITY THEORY IN SECOND LANGUAGE ACQUISITION Novitas-ROYAL, Vol.: 1(1), pp.10-17. ISSN: 1307-4733 http://www.novitasroyal.org/tamjid.pdf access 1 Nov. 2016
- Tang, Matthew. "Recognizing hand gestures with microsoft's kinect." Palo Alto: Department of Electrical Engineering of Stanford University:[sn] (2011).
- Tobias Pistohl, Tonio Ball, Andreas Schulze-Bonhage, Ad Aertsen, Carsten Mehring. Prediction of arm movement trajectories from ECoG-recordings in humans. Journal of Neuroscience Methods · vol. 167. February 2008
- Murata, Tomoya, and Jungpil Shin. "Hand gesture and character recognition based on kinect sensor." International Journal of Distributed Sensor Networks 2014 (2014).
- Vapnik, Vladimir, Steven E. Golowich, and Alex Smola. "Support vector method for function approximation, regression estimation, and signal processing." *Advances in neural information processing systems* (1997): 281-287.
- Jakkula, Vikramaditya. "Tutorial on support vector machine (svm)." School of EECS, Washington State University (2006).
- Alessandro Vinciarelli, The social life of features extracted from speech (and from other interesting behaviours). International Conference on Nonlinear Speech Processing, NONLINEAR SPEECH PROCESSING, NOLISP 2015. Jointly organized with the 25th Italian Workshop on Neural Networks, WIRN 2015 https://sites.google.com/site/nolisp2015/home http://www.dcs.gla.ac.uk/vincia/
- Gao Wen and Chunli Wang. "Sign language recognition." SERIES IN MACHINE PERCEPTION AND ARTIFICIAL INTELLIGENCE 48 (2002): 91-120.
- Wichmann, S., 2008. *The emerging field of language dynamics*, Language and Linguistics Compass 2/3: 442.
- Yang, Jie and Yangsheng, Xu. Hidden markov model for gesture recognition. No. CMU-RI-TR-94-10. CARNEGIE-MELLON UNIV PITTSBURGH PA ROBOTICS INST, 1994.
- Li, Yi. "Hand gesture recognition using Kinect." 2012 IEEE International Conference on Computer Science and Automation Engineering. IEEE, 2012.
- Pei Yin, Thad Starner, Harley Hamilton, Irfan Essa, James M. Rehg, "Learning the basic units in American Sign Language using discriminative segmental feature selection." School of Interactive Computing. Georgia Institute of Technology, Atlanta. GA. USA Proceedings of ICASSP 2009
- Yin, Ying. Real-time continuous gesture recognition for natural multimodal interaction. Diss. Massachusetts Institute of Technology, 2014.