

# SignInstructor: An Effective Tool for Sign Language Vocabulary Learning

Xiujuan Chai, Zhuang Liu, Yongjun Li, Fang Yin, Xilin Chen

Key Lab of Intelligent Information Processing of Chinese Academy of Sciences(CAS), Institute of Computing Technology, CAS, Beijing, 100190, China

University of Chinese Academy of Sciences, Beijing, 100049, China

Cooperative Medianet Innovation Center, China

{xiujuan.chai, zhuang.liu, yongjun.li, fang.yin, xilin.chen}@vip1.ict.ac.cn

## Abstract

*Sign language (SL) is an efficient and important communication means for deaf community. To grasp sign language, many elementary schools set the SL course for both deaf and hearing students. However, the tutors are still limited resource in most of the countries. This paper presents a natural and facilitate substitute, SignInstructor, to help self-learning of SL vocabularies with an inexhaustible tutor. Briefly speaking, the system is mainly composed of three modules: 1) multimodal playing of the standard SL materials, including sign video, key postures, illustration figure and text; 2) online capturing action when the user performs sign words; 3) and an automatic evaluation module by scoring the user's action. Aiming at the effective learning, there designed three basic education functions, including teaching, practice and testing. A prototype of this SignInstructor is developed with Kinect input and deeply user study on it convincingly shows the effectiveness in SL learning, especially compared with other two traditional ways.*

## 1. Introduction

Deaf person has a large social community. There are about 360 million people who have disabling hearing loss around the world<sup>1</sup>. Therefore, Sign Language (SL) is very important for the communication between deaf and hearing communities. However, SL learning is difficult. Aside from some SL inherent problems (such as complex action, abstract meaning, relatively large vocabulary), it also confronts with other challenges, such as limited teaching resources and environment.

In recent years, rather than direct learning by textbooks or videos, types of computer-aided software have been developed to assist the sign language learning. There exist some systems taking advantage of the traditional multimedia demonstration to promote sign language education. Suzuki presented a bilingual sign language dictionary that enables learners to obtain target sign language directly by looking up dictionary [16]. Ellis created a rich multimedia application that allows children in

Primary school to learn Australian sign language [5]. Besides these basic multimedia materials, in some SL learning applications, the sign language recognition technology is further involved for the matching and evaluation. Sagawa developed a sign language teaching system applying sign language recognition and sign language generation technologies to realize effective sign language study with data glove [14]. Similarly, another kind of sensor is instrumented AcceleGlove, which was used in the work of English dictionary retrieval by interactive American Sign Language input [8]. By using the same sensor, AcceleSpell game was designed to help users to learn and practice continuous finger-spelling with a recognition algorithm based on decision trees [9]. Madeo developed a prototype educative and inclusive teaching tool with colored glove [12]. This Brazilian Sign Language Multimedia Hangman Game aimed to help deaf people to learn sign language and improve their vocabulary. Another work is CopyCat, an ASL game, which also helped young deaf practice ASL by using gesture recognition technology [2]. In this system, user needed to wear the colored gloves with the wireless accelerometers on the back of their wrists. It can also assist young deaf children's language acquisition by interactive tutoring and real-time evaluation [7]. Aran etc. built an interactive system to help the user learn sign language by continuous practice and corrective feedback with a pair of colored gloves [1].

Besides the above-mentioned work, there were also some other typical sign language learning systems. Ellis aimed to teach Australian Sign Language to hearing people with abundant learning material [6]. The biggest characteristic is the rich multimedia materials, which can be created and customized by the deaf community. In the database construction, Brashear collected a special database from user studies of deaf children playing a Wizard game at the school for the deaf [3]. The target was to improve the accuracy of the system by characterizing and modeling disfluencies found in the children's signing. Some related tools also developed mobile applications. Weaver etc. intended to create a mobile based system to help hearing parents learn sign language [19]. They presented the user study results of evaluation on the ability to ascertain the details of particular signs based on video presented on a

<sup>1</sup> <http://www.who.int/mediacentre/factsheets/fs300/en/>

mobile device. Further, they also conducted the study to quantify the game's impact on expressive language development, especially in their receptive, expressive and sentence repetition abilities [20]. In literature [21] and [10], the systems for automatic learning of sign and gestures are provided respectively.

Even many efforts have been done, computer-aided sign language learning still confronted with many challenges, e.g. the effectiveness, the interesting and the naturality. Motivated from these factors, we give a natural and facilitate solution to the requirements of effective SL vocabulary learning for both deaf and hearing students. Firstly, the system adopts Kinect as the input device and realized pure vision-based sign language recognition, which avoids wearing any kind of external devices or sensors. Secondly, besides the accessible multimedia materials, such as the videos, illustration figures and texts, key hand postures are also extracted from the videos automatically and listed as the reference for the users' learning. Finally, to overcome the tedious practice problem, the scoring mechanism is introduced which will stimulate the users to play better and better. At the same time, the students feel more comfortable slowly studying and repeating signs with the developed SignInstructor than they would with a human teacher, who might get bored for repeating the same material continuously.

According to the automatic SL learning tool, there are three research questions we should investigate:

Q1: Whether the computer can comprehend or understand the user's signing?

Q2: Whether the developed SignInstructor has the education ability?

Q3: Whether the system has positive affect for the self-learning of SL?

In this paper, first we will verify the computer can understand the signing well based on sign language recognition algorithm. According to the education ability of this tool, a test is designed to evaluate whether the scoring is consistent with the human intuitive cognition. In terms of learning effect, widely user studies are carried out and the results show that SignInstructor is significantly superior to traditional face-to-face learning and self-learning with sign videos.

## 2. System Overview

Our envisioned system is called SignInstructor, which is an interactive sign language learning tool as shown in Figure 1. Inspired by the typical steps of learning procedure, our system is designed to be composed of three modes: teaching mode, practice mode and testing mode. Figure 2 gives the interfaces corresponding to the three modes. In this work, we take the Chinese Sign Language (CSL) vocabulary as example.



Figure 1. The SL vocabulary learning with SignInstructor.

In the teaching mode, through the observation on the given materials, the user can imitate the corresponding sign activity following the standard video. In our system, the rich multimedia materials are embedded, which includes illustration figure, plain text, corresponding sound and videos. With these multi-modal materials, the user will understand and memorize the corresponding signs impressively and then play the sign accordingly.

It is well known that practice is very important to enhance the learning of a language, especially the practice with objective evaluation and feedback. Therefore, in our designed practice mode, the system will give an evaluation by automatic scoring after each time of signing. The score is a direct feedback for the signing, which reflects the similarities between the signing and the standard one in terms of both motion trajectory and hand postures. With the given score, one can easily transform it to the evaluation on the correctness of the sign from the perspective of motion trajectory and hand posture. Then in his next practice, the sign will be performed more specifically. This just shows that the scoring satisfies the cognitive demands [13, 15] of the signer in his learning process and the evaluation mechanism makes the whole learning procedure more smooth and effective.

To verify the study effect, the system has the testing function. The user can select the test in random manner or in specific chapter. The test results can be recorded for the later comparison.

## 3. Sign Evaluation

So far, our work mainly focuses on the evaluation and feedback component. The modules of the SignInstructor mainly consists feature extraction, sign verification by sign language recognition, and the final sign scoring. Figure 3 shows the workflow of the sign evaluation. As shown in this figure, the trajectory and hand posture features are considered in our verification and scoring procedures. When an action is fed into the system, first it should determine whether the action is the appointed sign or a meaningless act. The hand segmentation and sign verification are realized on our previous works [4] and [18].

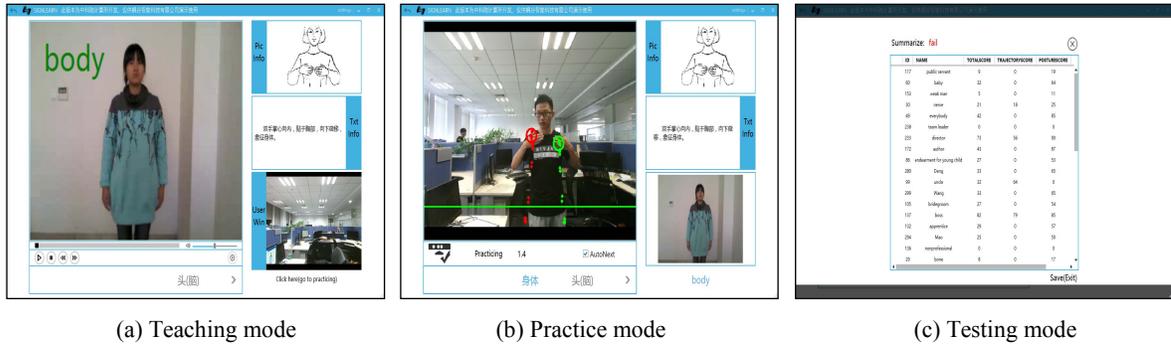


Figure 2. Main Interfaces of our SignInstructor.

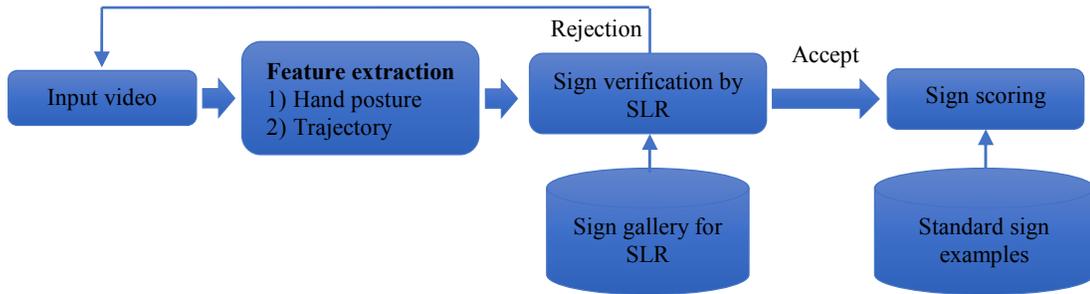


Figure 3. The workflow of the sign evaluation algorithm.

For the input sign which passes the verification, the system will further give an evaluation score as a feedback. In previous work related with sign evaluation, the most common way is to determine the score by transforming directly from the similarity obtained by recognition [1, 10]. Different to the recognition-dependent scoring, we think that the score should only relate to the sign word itself, while has nothing to do with other signs in recognition gallery. Inspired by this idea, a novel scoring scheme is designed and implemented with only several samples for each sign. The detail of the sign scoring method is described below.

For each sign, a standard example  $S$  is taken as the demo video. Moreover, we can also collect several good samples  $\{S_1, S_2, \dots, S_m\}$  for this sign. Given an input sign  $S_{in}$  which has been verified as the appointed sign, our target is to measure the score between this input sign  $S_{in}$  and the standard one  $S$ . the algorithm can be decomposed into the following steps:

- 1) Computing the similarities  $\{s_1, s_2, \dots, s_m\}$  between standard sign and its good samples through cosine distance measurement.
- 2) Finding the largest similarity  $s_{max}$  among the set  $\{s_1, s_2, \dots, s_m\}$  and computing the mean  $\bar{s}$  of the similarity set.
- 3) Deriving a score mapping from the multiple samples. Concretely speaking, the largest similarity corresponds

to the score of 100 and the mean of similarities corresponds to the score of 95.

- 4) Computing the similarity  $s$  between input sign and the standard one.
- 5) Getting the final score by the following equation:

$$score = 100 - (s_{max} - s) \times 5 / (s_{max} - \bar{s})$$

If  $s$  is larger than  $s_{max}$ , the score will set to be 100. Here, steps (1) to (3) are off-line computed and steps (4) to (5) are calculated online according to the input signing. In our evaluation method, the scores corresponding to hand motion and hand posture are computed separately and then fused into a total score. The weight used for fusion is predetermined according to the importance of the trajectory and posture for each sign vocabulary.

## 4. Experiment

### 4.1. Q1 - Understanding Sign Language

An important presupposition of this SL learning tool is that the system can understand the user's signing. Simply speaking, the system can determine whether the input is a sign word. Further, if the input is a sign, then the system should distinguish whether it is the appointed sign. Therefore, the system must have the powerful recognition ability.

Table 1. Results of the 5-folds Top-1 recognition rate for all the three methods

Probe-ID	HMM[18]	SDTW[11]	Our method
Fold-1	<b>84.0%</b>	77.2%	<b>84.0%</b>
Fold-2	85.1%	81.9%	<b>93.0%</b>
Fold-3	88.1%	86.5%	<b>93.2%</b>
Fold-4	81.6%	85.7%	<b>94.0%</b>
Fold-5	79.1%	84.9%	<b>93.0%</b>
Ave.	83.6%	83.2%	<b>91.5%</b>
Std.	0.034	0.038	0.042

Table 2. p-values given by the student's distribution comparing with our method.

Baseline/our method	HMM/our method	SDTW/our method
<i>p</i> -values	0.035	0.0005

To validate the sign language recognition ability, a dataset consisting 370 daily used signs is constructed by using Kinect. The data are signed by one girl deaf student by 5 times. Thus totally the dataset has  $370 \times 5$  sign sequences. We compare our method with the classical HMM and statistical DTW (SDTW) algorithms. To be fair to all the methods, the same leave-one-out strategy for cross validation is adopted in all recognition tests.

We report the 5-folds top-1 recognition performance for our used recognition method and also the comparison with HMM[17] and SDTW[11] in Table 1. Table 2 shows the p-values given by the student's distribution for the significance testing. The statistical tests convincingly show that comparing with HMM and SDTW methods, the performance improvement of our used method is statistically significant ( $p < 0.05$ ).

#### 4.2. Q2 - Education Ability

For a SL self-learning tool, a key problem is whether it has the education ability. Our SignInstructor provides the instruction by giving the evaluation and feedback to the user for his signing. Thus, a question is whether the score is consistent with the intuitive cognition. If the score is reasonable, then the system can give the correct feedback for the signing of the user like a human teacher.

In this test, we take 10 signs as examples, which include 5 trajectory dominant signs and 5 hand shape dominant signs. Each sign is played by 2 signers 15 times, which includes normal signing and the signing with some disturbances on their own initiative. One sign expert will observe the signings and give a subjective score ranging from 0 to 100 for each signing.

Having obtained all the data pairs, the correlation between scores from system and expert can be analyzed through the correlation coefficients computing. With the 150 data points ( $2 \text{ signers} \times 5 \text{ trajectory dominant signs} \times 15 \text{ times}$ ) on trajectory scoring, the correlation coefficient is

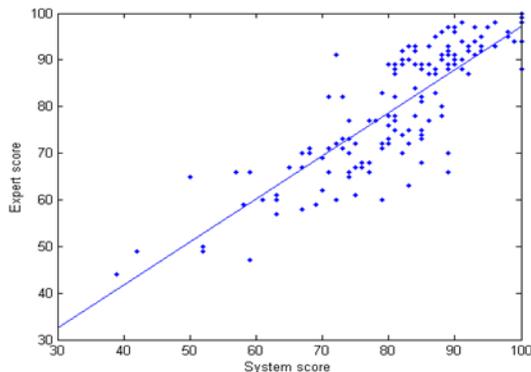


Figure 4. The data correlation distribution in terms of trajectory.

0.85 and similarly the correlation coefficient for hand posture is 0.75. The data correlation distribution in terms of trajectory is illustrated in Figure 4. It can be seen that the score given by this tool is significantly correlative with the expert evaluation ( $\rho_{150} = 0.264$ ,  $p = 0.001$ ). In other words, the automatic scoring can reflect the quality of the signing.

#### 4.3. User Study on Q3 - Effectivity in SL learning

In this section, a formal user study is conducted to give a comprehend evaluation on our developed SignInstructor.

##### 4.3.1 Conditions & Participants

To evaluate the instruction ability objectively, we take other two traditional learning methods as baselines. One is learning with sign videos and the other is face-to-face learning. For the convenience of writing, we denote the three learning methods as SI (learning with SignInstructor system), SV (learning with sign video) and FF (face-to-face learning) respectively.

In this test, we set a learning task for 18 CSL signs, which ranges from easy to hard. These signs are classified into three subsets with almost equal difficulty and denoted as Voc\_1, Voc\_2 and Voc\_3 respectively. The grouping is in line with the group learning style of SignInstructor.

The test is carried out for the community who usually contacts with deaf students but has little knowledge with sign language. These subjects have strong requirements and willingness to learn CSL. We recruited 15 volunteers (13 girls and 2 boys) whom range in age from 20 to 24 for this study. They all come from Special Education School of Beijing Union University. However, they haven't any experience in using computer-aided sign language learning tool.

##### 4.3.2 Procedure & Design

To eliminate the effect of subjects and vocabularies, an elaborate grouping test scheme is designed and shown in Table 3. As has been stated above, the signs are classified into three subsets with almost equal difficulty, which ensures that learning different groups of vocabularies is fair.

Table 3. Grouping test scheme in the formal user study.

	G1(P1~P5)	G2(P6~P10)	G3(P11~P15)
T1	SV / Voc_1	SI / Voc_1	FF / Voc_1
T2	SI / Voc_2	FF / Voc_2	SV / Voc_2
T3	FF / Voc_3	SV / Voc_3	SI / Voc_3

Similarly, the participants are randomly divided into three groups, which are denoted as G1, G2 and G3. The random partition scheme implies each group has roughly same learning ability. For each participant, the learning and test are carried out according to the predetermined scheme, i.e. in the order of T1, T2 and T3. No matter using what kind of learning method, the learning procedure will stop until the subject thinks that he has learnt all the signs. Then the test for this vocabulary subset begins. The same two experts evaluate the study by scoring for each signing. The scores ranging from 1 to 5 are recorded for the final analysis.

In the questionnaire survey, the participants are asked to answer several questions, which include: like or dislike about each method; advantages and disadvantages of each method; ranking for the preference of three learning ways to assist his CSL learning. In respect of the SignInstructor, the questions include: ranking for the proficiency in using computer; ranking for the simplicity for the using of SignInstructor; and additional comments or suggestions.

Averagely, the whole learning procedure for using three learning methods roughly lasts 30 minutes. The questionnaire survey and interview usually lasts 20 minutes. The total test for one subject lasts about one hour.

#### 4.3.3 Results

All participants finished the study as planned. In each test for one group of 6 vocabularies, the time costs for three methods were 4 minutes (FF), 5 minutes (SV) and 9 minutes (SI) respectively. Although SignInstructor consumes more time than the other two ways, it indeed achieved best learning effect. The learning procedure is not so boring as the traditional face-to-face learning and the simple study with the videos.

**Quantitative Analysis:** With the recorded scores from the ordered experiments, we analysis the results statistically and get the final scores of three kinds of learning methods, which are 3.68 (0.75), 3.81 (0.63) and 4.28 (0.72) for SV, FF and SI respectively.

From the results, it can be seen that three kinds of learning methods are all effective for CSL learning. Among them, SignInstructor has the highest score and the learning with sign video has the lowest score. Statistically, score with SignInstructor is significantly higher than face-to-face learning ( $t = 4.7 > t_{.005} = 2.878, p < 0.005$ , two-tailed) and learning with sign videos ( $t = 5.47 > t_{.005} = 2.878, p < 0.005$ , two-tailed). While the score with face-to-face learning is not significantly higher than learning with sign videos

( $t = 1.26 < t_{.005} = 2.878, p > 0.005$ , two-tailed). The results reflect that SignInstructor has the best instructive effect for CSL vocabulary learning.

Intuitively, we usually think that the face-to-face learning should be most effective. However, in our experiment, face-to-face learning gets a lower score than SignInstructor. The reasons may lie in the following three aspects: 1) short learning time; 2) possible incompleteness in the education procedure; 3) occasional inefficient communication between the teacher and the participants. It is known that the interaction is very important in sign education. However, the long-time teaching will make the tutors very tired, especially for the fresh students. Moreover, some introverted students wouldn't like to interact with human tutors, which will lead to a bad learning result. Considering on this point, a machine with infinite patience (just like our SignInstructor) maybe much better than a human teacher, especially in early stages of learning. Another conclusion is that SignInstructor is obviously better than learning with sign videos. That maybe because more multimedia materials, the additional audio information, the clear key hand postures, the interesting feedback mechanism and the effective teaching, practice and testing functions.

We have found an interesting instance in this study. For the vocabulary "Seek", the signer usually learnt the action of hand while ignoring the facial expression when using the two traditional learning methods. However, there are 2 (2 out of 5) participants who captured the eyes of this vocabulary correctly in the test by using SignInstructor. Actually, the audio voiceover gives the hint for this point. Another phenomenon is that the signing is more standard if the participant learns the vocabulary through SignInstructor. We think that the display of the real-time signing contributes to this point because the user can compare and revise the action and movement range according to the standard videos. So, it can be concluded that SignInstructor performs well in some details.

**Qualitative Analysis:** From the feedback of the questionnaire survey and interview, we also get some useful information. In terms of the proficiency in using computer, 7 participants rated "proficient", 3 participants rated "very proficient" and the remaining 5 rated "general". While on the using of the SignInstructor, most participants rated "very simple" (3 out of 15) or "simple" (9 out of 15) and the remaining rated "general" (3 out of 15). The ratings show that the SignInstructor can be easily grasped by common users without highly demand for computer techniques.

The statistic for the preference of three learning methods is also considered. Most participants liked the face-to-face learning best (9 out of 15) and there are 5 participants liked our SignInstructor best. I think the potential reason for like of face-to-face learning maybe lie in the high efficiency of the interaction on the test condition, which only has small number of signs to learn. So the teacher won't be tired and

can maintain the enthusiasm to communicate with students, which is a good experience to these subjects. However, if the learning content is the large vocabulary of signs, the result may be totally different. This is also one important point needed to be verified in our near future.

Further, we conclude the comments on the advantage and disadvantage of three learning methods. For our SignInstructor, the users think that it is “simple”, “facilitate” and “effective”, but “time consuming” is its disadvantage.

In terms of the open-ended question on the general impression of SignInstructor, the feedbacks from the interview are also general positive:

- 1) “I like SignInstructor. Rather, I love it. I hope it can be popularized widely.”(P13)
- 2) “The SignInstructor makes CSL learning funny. The user can watch and learn in an intuitive way, which makes the sign learning even more impressive.”(P3)

## 5. Conclusion

As a computer-aided tool, SignInstructor provides a novel solution for effective SL vocabulary learning. Aiming to break the constraints of the traditional learning methods, SignInstructor tries to give an engaging and effective leaning means, where the intervention from human teacher is no longer needed. The system provides abundant multimedia materials and shows them in both audio and visual channels. To satisfy the cognitive demands in learning process, a quick evaluation by scoring is given automatically after each time’s signing. Widely user studies also show the effectiveness for SL learning with the outstanding score, which is even higher than that of face-to-face learning.

## 6. Acknowledgements

This work was partially supported by the Microsoft Research Asia, and Natural Science Foundation of China under contracts Nos. 61390511, 61472398, 61532018, and the Youth Innovation Promotion Association CAS.

## References

- [1] Aran, O., Ari, I., Akarun, L., Sankur, B., Benoit, A., Caplier, A., Campr, P., Carrillo, A., Fanarda, F. 2009. Signtutor: An interactive system for sign language tutoring. *MultMedMag*, (1):81–93.
- [2] Brashear, H., Henderson, V., Park, K-H., Hamilton, H., Lee, S., Starner, T. 2006. American Sign Language Recognition in Game Development for Deaf Children. *ASSETS*, 2006.
- [3] Brashear, H. Improving the Efficacy of Automated Sign Language Practice Tools. 2007. *SIGACCESS Accessibility and Computing*, (89):11-17.
- [4] Chai, X, Xu, Z, Li, Q, M, B., Chen, X. 2012. Robust Hand Tracking by Integrating Appearance, Location and Depth Cues. 4<sup>th</sup> International Conference on Internet Multimedia Computing and Service, pp.60-65.
- [5] Ellis, K. 2009. Multimedia for Primary School Children Learning Sign Language. *OZCHI*, pp.97-104.
- [6] Ellis, K., Ray, N., Howard, C. 2011. Learning a Physical Skill via a Computer: A Case Study Exploring Australia Sign Language. *OzCHI*, pp.98-103.
- [7] Henderson, V., Lee, S., Hamilton, H., Starner, T., Brashear, H., and Hamilton, S. 2005. A Gesture-based American Sign Language Game for Deaf Children. *CHI*, pp.1589-1592.
- [8] Hernandez, R. J. L. Mendez, E. 2004. Interactive American Sign Language Dictionary. *SIGGRAPH*, 26.
- [9] Hernandez, R. J.L., Elsakay, E., Alanis, U.J. 2008. *AcceleSpell*, A Gestural Interactive Game to Learn and Practice Finger Spelling. *ICMI*, pp.189-190.
- [10] Kamal, A., Li, Y., and Lank, E. 2014. Teaching Motion Gesture via Recognizer Feedback. *ACM IUI*, pp.73-82.
- [11] Lichtenauer, J., Hendriks, E., Reinders, M. 2008. Sign Language Recognition by Combining Statistical DTW and Independent Classification. *TPAMI*, 30(11): 2040–2046.
- [12] Madeo, R.C.B. 2011. Brazilian Sign Language Multimedia Hangman Game: A Prototype of an Educational and Inclusive Application. *ASSETS*, pp.311-312.
- [13] Owen, E., and Sweller, J. 1985. What do Students Learn While Solving Mathematics Problems? *Journal of Educational Psychology*, (77):272-284.
- [14] Sagawa, H. and Takeuchi, M. 2002. A Teaching System of Japanese Sign Language using Sign Language Recognition and Generation. *ACM MM*, (2002), 137-145.
- [15] Sweller, J. 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, (12):257-285.
- [16] Suzuki, E., Horikoshi, M., Kakihana, K. 2004. Bilingual Sign Language Dictionary for Learning a Second Sign Language without Learning the Target Spoken Language. *Workshop on Multilingual Linguistic Resources*, pp.93-96.
- [17] Wang, C., Gao, W., Shan, S. 2002. An Approach based on Phonemes to Large Vocabulary Chinese Sign Language Recognition. *FG*, pp.411-416.
- [18] Wang, H., Chai, X., Chen, X. 2016. Sparse Observation (SO) Alignment for Sign Language Recognition. *Neurocomputing*. (175), Part A, pp.674-685.
- [19] Weaver, K. A., Starner, T., Hamilton, H. 2010. An Evaluation of Video Intelligibility for Novoice American Sign Language Learners on a Mobile Device. *ASSETS*, pp.107-114.
- [20] Weaver, K. A. Hamilton, H., Zafrulla, Z., Brashear, H., Starner, T., Presti, P. and Bruckman, A. 2010. Improving the Language Ability of Deaf Signing Children through an Interactive American Sign Language-Based Video Game. *ICLS*, (2):306-307.
- [21] Weaver, K. A., Starner, T. 2011. We Need to Communicate! Helping Hearing Parents of Deaf Children Learn American Sign Language. *ASSETS*, pp.91-98.