A Proposed Standardization for Arabic Sign Language Benchmark Database

A. S. Elons[*1], M. F. Tolba[*2]

[*] *Scientific Computing Department- Faculty of Computers and
Information Sciences- Ain Shams University-Cairo-Egypt*

[1]`ahmed.new80@hotmail.com`
[2]`fahmytolba@gmail.com`

Abstract- *This The lack of a visualized representation for standard Arabic Sign Language (ArSL) makes it difficult to do something as commonplace as looking up an unknown word in a dictionary. The majority of printed dictionaries organize ArSL signs (represented in drawings or pictures) based on their nearest Arabic translation; so unless one already knows the meaning of a sign, dictionary look-up is not a simple proposition. In this paper we introduce the ASL database, a large and expanding public dataset containing video sequences of thousands of distinct ArSL signs. This dataset is being created as part of a project to develop an Arabic sign language translator. At the same time, the dataset can be useful for benchmarking a variety of computer vision and machine learning methods designed for learning and/or indexing a large number of visual classes especially approaches for analyzing gestures and human communication.*

Key words: *Arabic Sign Language (ARSL),Arabic Sign Language Database, Database Benchmark.*

1 INTRODUCTION

Arabic Sign language is different in each Arab region or/and country with many dialects. This difference gives the difficulty of communicating and dealing between deaf people in different Arabian countries. A need appeared to unify Arabic sign language in all Arabian countries. This derived the Council of Arab Ministers of Social Affairs (CAMSA) to take a decision of developing a unified Arab sign language dictionary and publish it to all countries, in an attempt to help Arab deaf people to have a common language in addition to their local language [1]. This dictionary is mostly used in education and in common communication such as sign language interpreters in television. Arabic sign language like other known sign languages depends on three basic factors that are used to represent the manual features: hand shape, hand location and orientation. In addition to the non-manual features that are related to head, face, eyes, eyebrows, shoulders and facial expression like puffed checks and mouth pattern movements. ASL is limited to represent nouns, adjectives and verbs. Prepositions and adverbs are represented in the context of articulation by specifying locations, orientations and movement. Intensifiers represented by iteration [1]. Signs forming and sequencing in the articulation, are done depending on the Arabic sign language grammar and rules.

Arabic sign languages (ARSLs) are still in their developmental stages. Only in recent years has there been an awareness of the existence of communities consisting of individuals with disabilities; the Deaf are not an exception. Arab Deaf communities are almost closed ones. Interaction between a Deaf community and a hearing one is minimal and is basically concentrated around families with deaf members, relatives of the deaf, and sometimes play friends and professionals [2]. As in other communities, communication with a deaf person is polarized within such circles. This situation has led to the emergence of many local means of sign communication. Until recently, such signs have not been gathered or codified. Signs are starting to spread, forming acknowledged sign languages. By and large, the view held vis-a-vis disability, including hearing, in the Arab society is still one of accommodation rather than assimilation [2].

Sign languages all over the world are not a new invention. They existed on par with the spoken languages. Their invention cannot be attributed to any person. Rather, they developed naturally just as other verbal languages. Similarly, ARSLs have been developing naturally. In their ''natural context,'' ARSLs developed as in-dependent systems of communication. They are not interpretations of standard Arabic or spoken vernaculars [2].

2   ARABIC SIGN LANGUAGE

ARSLs share many similarities and manifest certain features of difference. After all, this is true for all languages; indeed, trace features of universality can be traced among the sign languages of the world. Basically, ARSLs developed independently, although some have benefited from the pioneer experience of the others. The possible sources of ARSLs could be traced to the following:

- Borrowings, especially European and American.
- Creations, which are initialization of conceptual signs usually by gestural repertoire of spoken varieties.
- Miming actions, shapes, and things in nature.
- Expanding means, such as compounding and blending.
- ''Dumb'' regional signs, which are basically signs inherited over centuries, used by ''mute'' people, and of a local nature.

Finger spelling is fairly new and is mostly a combination of creation and miming source. It is used to spell out proper nouns and words that do not have sign correspondence. Finger spelling, however, is not used to read out or communicate the standard form of Arabic. Therefore, there is no ''manual Arabic'' yet; perhaps such form of signed standard Arabic might develop if the deaf are to be educated through sign language and if need arises to have a signed Arabic that corresponds to the standard. Further, there has been no attempt so far to write down ARSLs (sign writing). ASL, for example, has established writing systems, but these have not been widely used to record ASL literature; however, there is a large body of ASL literature available in movies, videotapes, and compact disks [3].
Arabic, on the other hand, has a considerable body of signed literature mainly in movies, TV series, and news bulletins; this body has been neither recorded nor utilized for the development of Arabic sign vernaculars.
Arabic sign languages are not particularly different from other known sign languages, such as BSL. In fact, the Arabic varieties in use have undergone some lexical influence from other sign languages [4]. ARSLs are basically manual languages made from cheremes that involve the three recognized elements: configuration of hands (hand shape), placement/space (position of hand in relation to body), and movement (directions and contacts within space). In addition to these manual shapes, ARSLs make use of other non-manual features, like those of the face, mouth, and tongue.
Arabic sign languages also exhibit similar forms to other established sign languages, such as links between form and meaning that may be iconic, pictorial, conventional, or arbitrary [5]. Arabic sign languages' word correspondence (i.e., signs) is limited to two basic classes, nouns/adjectives and verbs, and lacks, unlike standard Arabic, many of the particles (e.g., prepositions and some adverbs or intensifiers). However, the relationships and concepts represented by prepositions and intensifiers, for example, can be expressed by other means. This could be done by the position and direction of one sign in relation to another in the case of prepositions and by repetition of sign regarding intensifier [6]. Other vocabulary items can be explained under the following categories: synosigns, antosigns, homosigns, and compounds.

- Synosigns: usually two different signs with one meaning are not common in ARSLs. However, they do exist and mostly evolve as a result of shifting from one sign to another, and when the first sign is not totally abandoned, the two signs continue to coexist for some time until one, usually the second, dominates. Examples from Jordanian Sign Language are girl and rich.
- Antosigns: The type of antosigns present in ARSLs is mostly complementary pairs, which is different only in one element: movement. This makes an to signs in sign language different from antonyms in spoken languages, in which the sounds and meaning are different
- Homosigns: Arabic sign languages use some homo-signs. There is no difficulty in understanding the referential meaning of such signs, which is usually clear from the context
- Compounds: A very important method to expand vocabulary is through compounding. This is also true for sign languages, including Arabic. Whenever two signs can give the meaning of another concept when combined, they are employed to do so, especially in developing sign languages such as those of Arabic. Indeed, it is much easier to understand a concept in relation to another rather than to invent one; consider these examples: dentist, internist, vet, and dream.

Arabic sign languages are similar to other sign languages of the world in that they are basically spatial–gestural languages. This makes it difficult to compare sign languages with their spoken counter-parts; Arabic in this regard is not an exception. As a matter of fact, many concepts used to describe spoken languages are inadequate for the description of sign languages. Nevertheless, inevitably, one system should be mapped practically into the other.
Generally, ARSLs do not follow the same order of their spoken or written counterparts. Usually, a reversed order is used. This is because sign languages are highly schematized and indeed more pragmatic than the spoken ones. In Arabic, emphasis is given to content signs, those representing nouns and verbs. The nominal ''sentence'' is usually made up from a subject and a predicate, such as ''she/he deaf '' [6]. And, unlike spoken and written varieties, there is no singular, dual, or plural agreement in ARSLs.
Signed sentences, on the other hand, do not make use of tense/aspect as in spoken and written varieties. Tense is simply and practically used. Past, present, and future times are indicated at beginnings of conversation chunks and only shifted

when there is need to indicate a different tense (e.g., worked). Negatives and interrogatives have more than one way of expression. While in some cases non manual gestures are important (e.g., raised eyebrows, head and shoulders leaning forward, signed question mark), in other cases signs are used, for instance, ''red not''.

As for other grammatical features like emphasis and adverb position, emphasis is done by repetition, longer signing time, and facial expressions and dramatization; adverbs are explained manually, by one hand's position in relation to the other. Other features, such as passivization, declension, and indeclension, are nonexistent. Conditional expressions, sentence boundaries, and turn taking are usually achieved by non-manual features of facial expressions and context.

Sign languages show greater link between form and meaning than spoken languages [5]. Arabic word order is so flexible that it allows for one meaning to be expressed in different formal structures, such as V-S-O (verb-subject-object), S-V-O, O-V-S, V-O-S. This makes the structure of ARSLs familiar, especially to hearing learners, and easily comprehensible to the uneducated (most deaf people in the Arab countries are) because of their grammatical simplicity, which does not exist in standard Arabic. All this in my opinion makes sign language in general and Arabic in particular more ''pragmatic'' than the spoken varieties of language, which adds to the advantages of sign language more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

3   RELATED WORK

For evaluation and benchmarking of automatic sign language recognition, large corpora are needed. Recent research has focused mainly on isolated sign language recognition methods using video sequences that have been recorded under lab conditions using special hardware like data gloves. Such databases have often consisted generally of only one speaker and thus have been speaker-dependent, and have had only small vocabularies. Most databases used in sign language processing so far do not provide or include what is important for the evaluation of sign language processing algorithms.

The National Center for Sign Language and Gesture Resources at Boston University has published an expanding database of American Sign Language (ASL). Dreuw and colleagues from the RWTH Aachen University created several subsets for the evaluation of isolated and continuous sign language recognition: RWTH-BOSTON-50 [7, 8], RWTHBOSTON- 104 [9], and the new RWTH-BOSTON-400.

The new RWTH-BOSTON-400 is the largest publicly available benchmark corpus for video-based continuous sign language recognition. It contains 843 sentences, several speakers, and separate splits for training, development, and testing of automatic sign language recognition systems.

The RWTH-BOSTON-400 database is created from a subset of the larger data set available through Boston University. The BU ASL corpus has been used previously in evaluation of computer vision and pattern recognition methods, including detection of head gestures [10], recognition of facial expressions [11], hand tracking and recognition of hand shapes and movements [12, 13, 14].

The National Center for Sign Language and Gesture Resources (NCSLGR) at Boston University has been engaged in the collection of ASL data (including sets of individual utterances, narratives, and dialogues) from Deaf native signers.

The NCSLGR makes available high-quality video files showing the signing from multiple angles, including a close-up of the face, in a variety of video formats, along with linguistic annotations that have been carried out in conjunction with the American Sign Language Linguistic Research Project (ASLLRP) at Boston University, using SignStream [15, 16].

4   ARSL DATABASE BENCHMARK

This benchmark database is being created as part of a project to develop an Arabic sign language translator. This project aims to provide a credible tool of communication between the deaf sector and the community. The first milestone in the project is to build a national digital database for standard Arabic sign language.

The Standard Arabic sign language dictionary can be categorized in 27 categories of words. The number of signs is: 1216 signs including the alphabets and numerical. Four sign language experts have captured the complete dictionary and two different experts have reviewed the signs validity. The four sign experts are deliberately picked as: two are a left-handed person and the other are a right-handed. For each captured person:

- Videos are captured from 4 different angles (Fig. 1).
- Recording signs' videos using different viewing angles (0, 270 then 315, 225).We recorded videos of 5, 10, 30 and 50 frames per second. The recording has been done for 2 different persons; each sign is recorded 3 times.

a) 45° degree orientation   b) 135° Orientation

c) 90 ° Orientation   d) 0° Orientation

Figure 1: The word (one) captured by different orientation angles.

This concludes the database size as: 4*(1216*4) = 19,456 signs videos. The facial expressions represent a major non-ignorable key feature in identifying the meaning. (Ex: divorce and marriage have the same hand signs but differ in the facial expressions). A study has been conducted on the captured signs database conclude that nearly 72% of the signs mainly depend on facial expressions and body language to deliver the right meaning to the recipient .

The database benchmark has several orientations:
- The performance benchmark.
- The data validity benchmark.
- The data variation and generalization benchmark.

The formal Arabic sign language dictionary approved by the league of Arab states is used to build Arabic sign language database .In order to validate the created database; a survey was conducted with a sample of 80 deaf students who were participating in the various user studies. The sample of deaf people is chosen from 4 different certified NGOs with different levels of education.

The main objective of interface development is to ease the access of the required video(s) or sign of a specific word. The Database Access Interface is implemented in ASP.Net and runs in any modern browser. As discussed before, the procedure of building ASL at Boston University is used to build our database. The Interface provides high-quality video files showing the signing from multiple signers, multiple angles, including a close-up of the face, different lighting conditions,

in a variety of video formats as shown in figure 2. Summary of the number of recorded videos available in the interface with different acquisition techniques is outlined in table 1.
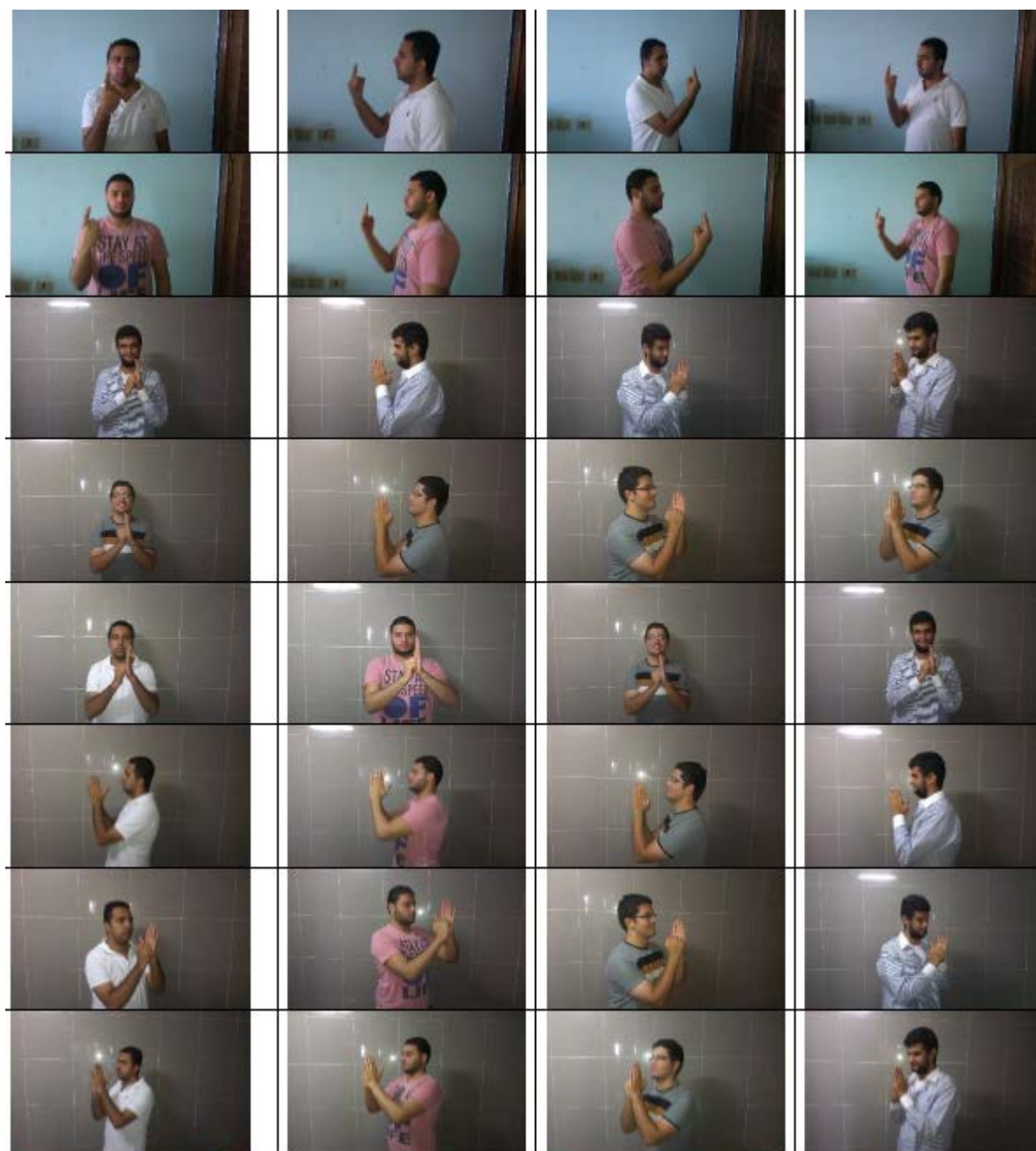
Figure 2: Different samples of the database with multiple signers, positions and lighting

TABLE 1
NUMBER OF RECORDED VIDEOS FOR EACH SIGNERS AND DIFFERENT RECORDING CONDITIONS

	Signer 1	Signer 2	Signer 3	Signer 4
Number of Videos	1216	1216	1216	1216
Top-Side orientation	1216	1216	1216	1216
45-135 degrees	1216	1216	1216	1216
Close up face	980	682	733	822
Body Only	1216	720	1216	1216
Normal Lighting	1216	1216	1216	1216
Low Lighting	912	941	891	951

In addition, the interface provides with linguistic annotations that have been carried out in XML format. The interface allows users to query the data (or some user-specified subset of the data) in search of specific signs (or types of signs, e.g. finger-spelled signs), non-manual behaviors, or combinations thereof, while facilitating transfer of video files and annotations from the web site to the user's computer without the need of third-party software.

The annotations are available as XML files. Video files are available in a variety of formats that offer different trade-offs between file size and video quality. The original, uncompressed video sequences have resolution of 600x800 pixels, and were recorded at 60 frames per second. Grayscale and color cameras were used for recording the sequences. Each sequence was captured simultaneously by multiple (two to four) synchronized cameras: one or two cameras showing a front view of the upper body of the signer, one camera zooming in on the face from the front, and in many cases a camera showing the signer's upper body from the side. Calibration sequences are available for most of the recording sessions. The calibration sequences show a chessboard-like calibration pattern at a variety of 3D orientations, as seen from multiple cameras.

5   NEW TRENDS IN SIGN LANGUAGE APPLICATIONS

Different researchers in Sign language recognition all over the world investigated different sensors to capture input signs. The rapid technology growing facilities the appearance for more accurate sensors for capturing the signs features. This section illustrates the usage of new sensors such as Kinect and Leap motion that captures the hands and body poses.

A. Kinect

Microsoft Kinect is one of the modern sensors used to recognize sign language. It produces a live stream with depth information, body motion and skeletal movements. As Kinect produces a live stream of the tracked object, it provides information about the full body parts while moving such as: the parts position, velocity and direction. And the main advantage of this sensor is that the data is independent of light condition, as it uses infrared light to detect surrounding objects. Many researchers used kinect to recognize sign language. In 2013 Billiet et al [17] built a rule based system on 8 different hand postures Kinect images and accomplished 96% recognition rate.J. Shotton [18, 19] used Kinect to collect input data for his system, this data consists of 10 American Sign Language symbols collected from 17 participant's males and females and the participants drew the symbols using their right hand. The resulted images are 256x256 grayscale depth images with a rate of ten frames each second . J. Shotton then made some preprocessing on the data to resulting in 32x32 pixels of the hands with no background to eliminate noise. This data used two Artificial Neural Networks (ANN). The first ANN used is the Random forests network [18], with each pixel in the 32x32 image regarded as a feature. This network achieves an average accuracy of 76%. Dropout nets used with 32x32 input layer, assigned each pixel in 32x32 input image to an input neuron. Although Kinect facilitates body and hands tracking, it does not support hand shape recognition, and since sign language generally features different hand-shapes, similar signs cannot be distinguished. Figure. 3 illustrates an image of the kinect produced data [18].

Figure 3: Kinect image in sign language recognition

B. Leap Motion

Leap Motion is another modern sensor used to track user movements, it is regarded as the most easy to use and handy in everyday situations due to its small size and low cost Fig 4. Unlike Kinect, it tracks hand movements only, also it has the ability to distinguish the fingers' joints and track their movements and this information is supported by Leap Motion Vendors "https://www.leapmotion.com". Although the Kinect produces information about other body parts that help in

distinguishing the closed signs, it lacks the details about the hand's fingers. Leap motion tracking data is a series of snapshots called frames; each frame contains the measured positions, velocities, and other information about the detected fingers. Such information is promising and could be useful in the recognition area [20, 21].
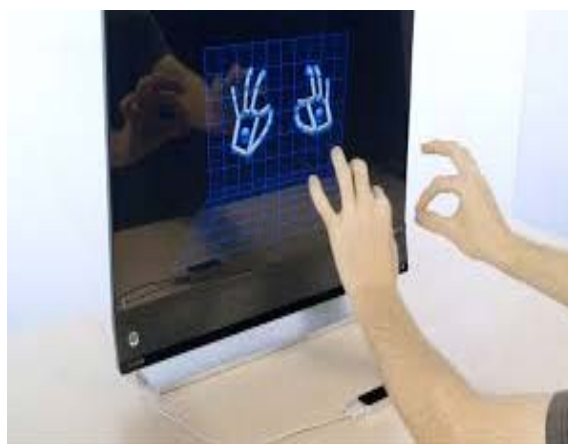

Figure 4: Leap Motion Device

Stability of images acquired and used by Leap Motion to provide the information about hands and finger can be different. Both slow and fast movements may result in obtaining data that cause invalid frames. Sometimes, the invalid input information processed by Leap Motion result short-time lost tracking of fingers, or non-existing objects appear in a view. These short-time invalid frames happen usually last less than 10 frames. The preprocessing principle is based on a median filter – it uses a window with defined size (w) to minimize the noisy information. The step of data capturing is performed with 50 frame/second rate and median filter window size equal 4.

6   CONCLUSIONS

The In this paper, we described the recording of a new sign language corpus which meets the requirements for an Arabic sign language translator. The database is based on a vocabulary of 1216 basic signs in Arabic sign language and comprises 531 sentences each articulated by 4 different signers. The whole database will be made available for interested researchers in order to establish the first benchmark. The currently extracted features produce good recognition performance for a single trained signer. The experimental results reveal that they are robust enough for signer-independent sign language recognition.

REFERENCES

[1] S. Samreen and M. Benali,"قواعد لغة الإشارة العربية القطرية الموحدة ", 2009.
[2] M. A. Abdel-Fattah, "Arabic Sign Language: A Perspective". Journal of Deaf Studies and Deaf Education, vol. 10 no. 2, 2005.
[3] Wilcox, S. P., & Kreeft, J. "American Sign Language as a foreign language". ERIC Digest. Retrieved April 20, 2004, from http://www.ericfacility.net/databases/ERIC_Digest/ed49464.htm. (Accessed in Jan 2015)
[4] Miller, C. "Disc: Arabic Sign Language", Re: 7.1101. Retrieved March 14, 2004, from http://www.linguislist.org/issues/7/7–1110.html. (Accessed in May 2014)
[5] Brennan, M. "British Sign Language: The Language of the Deaf Community". In T. Booth and W. Swann (Eds.), Including Pupils with Disabilities: Curricula for All. Milton Keynes, UK: Open University Press, 1987.
[6] Suwed, A. A. " لغة الإشارة العربية الصم – لغة لغة [Arabic Sign Language, Deaf language as in Libya]". Tripoli, Libya: AI-Mansha'ah AI Aamah Lin-Nasher wal I'lam, 1984.
[7] M. Zahedi, D. Keysers, and H. Ney. "Pronunciation clustering and modeling of variability for appearance-based sign language recognition". In International Gesture Workshop 2005, volume 3881, Vannes, France, May 2005.

[8] M. Zahedi, P. Dreuw, D. Rybach, T. Deselaers, and H. Ney. "Continuous sign language recognition approaches from speech recognition and available data resources". In Second Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios, pages 21–24, Genoa, Italy, May 2006.

[9] P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney. "Speech recognition techniques for a sign language recognition system". In Inter speech 2007, pages 2513–2516, Antwerp, Belgium, August. ISCA best student paper award of Interspeech 2007.

[10] U. M. Erdem and S. Sclaroff. "Automatic detection of relevant head gestures in American Sign Language communication". In International Conf. on Pattern Recognition (ICPR), volume 1, pages 460–463, 2002.

[11] C. Vogler and S. Goldenstein. "Facial movement analysis in ASL". Springer Journal on Universal Access in the Information Society, 2007.

[12] C. Vogler and D. Metaxa. "Handshapes and movements: Multiple-channel ASL recognition". Springer Lecture Notes in Artificial Intelligence, (2915):247–258, 2004.

[13] Q. Yuan, S. Sclaroff, and V. Athitsos. "Automatic 2d hand tracking in video sequences". In IEEE Workshop on Applications of Computer Vision, 2005.

[14] P. Dreuw, T. Deselaers, D. Rybach, D. Keysers, and H. Ney. "Tracking using dynamic programming for appearance-based sign language recognition". In IEEE Intl. Conf. on Automatic Face and Gesture Recognition, pages 293–298, Southampton, April 2006.

[15] C. Neidle, S Sclaroff, and V. Athitsos. "Signstream: A tool for linguistic and computer vision research on visual-gestural language data". Behavior Research Methods, Instruments, and Computers, 3(33):311–320, 2001.

[16] C. Neidle. "Signstream: A database tool for research on visual-gestural language". Journal of Sign Language and Linguistics, 1/2(4):203–214, 2002.

[17] L. Billiet, J. A. Oramas Mogrovejo, M. Hoffmann, W. Meert, and L. Antanas. Rule-based hand posture recognition using qualitative finger configurations acquired with the kinect. In Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods, pages 1–4, Feb.2013.

[18] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore, "Real-time human pose recognition in parts from single depth images", Communications of the ACM, 56(1), pp. 116–124, 2013.

[19] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, RichardMoore, Alex Kipman, and Andrew Blake. Real-Time Human Pose Recognition in Parts from Single Depth Images. CVPR, 2011.

[20] Leigh-Ellen Potter, Jake Anthony Araullo and Lewis Antony Carter, "The Leap Motion controller: A view on sign language", The 25th Australian Computer-Human Interaction Conference, 2013.

[21] Leap Motion, Inc. LEAP. 2012. 22 October 2012.

BIOGRAPHY

Prof. Dr. Mohamed Fahmy Tolba is a Professor of Scientific Computing, FCSIS (1996-Present). Dr. Tolba has more than 150 publications in the fields of AI, Image Processing, Pattern Recognition, OCR, Scientific Computing, Simulation and Modeling. Also Dr. Tolba has supervised more than 50 M.Sc. and 25 Ph.D. degrees in Ain Shams University and other Egyptian Universities.

Dr. Ahmed Samir is a Lecturer at the Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt. His research interests: Image Processing and, pattern recognition and AI. He worked in Arabic Sign Language Recognition field from 2004 till now.

توحيد قياسي مقترح لبناء قاعدة بيانات موحدة للغة الاشارة العربية

أحمد سمير - محمد فهمي طلبة

قسم الحسابات العلمية ـ كلية الحاسبات و المعلومات- جامعة عين شمس

خلاصة:

تعتبر الصعوبة في التمثيل الشكلي للغة الاشارة من أكبر العوائق و التحديات للبحث عن اشارة في القاموس مما يسبب منع انتشار مترجمات لغة الاشارة و وصول قاموس لغة الاشارة لمستخدميه. يهدف هذا البحث لاقتراح و تنفيذ أول قاموس للغة الاشارة العربية معتمدا علي فيديوهات متنوعة الزوايا, الاضاءة و منفذس الاشارات. نهدف من هذا وضع قاعدة بيانات علي أساس توحيد قياسي للبحث. يمكن الاستفادة من قاعدة البيانات في مجالات البحث الاكاديمي و الاستخدام العادي.