Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Technology 00 (2014) 000–000

Procedia Technology

www.elsevier.com/locate/procedia

Conference on Electronics, Telecommunications and Computers – CETC 2013

Kinect-Sign, Teaching sign language to "listeners" through a game

João Gameiro[a], Tiago Cardoso[a], Yves Rybarczyk[a]

[a]*Universidade Nova de Lisboa*

Abstract

The sign language is widely used by deaf people around the globe. As the spoken languages, several sign languages do exist. The way sign language is learned by deaf people may have some details to be improved, but one can state that the existing learning mechanisms are effective when we talk about a deaf child, for example. The problem arises for the non-deaf persons that communicate with the deaf persons – the so-called listeners. If, for example, one couple has a new child that turns to be deaf, these two persons find a challenge to learn the sign language. In one hand, they cannot stop their working life, especially because of this sad news turns to be more costly, on the other hand, the existing mechanisms target the deaf-persons and are not prepared for the listeners. This paper proposes a new playful approach to help these listeners to learn the sign language. The proposal is a serious game composed of two modes: *School-mode* and *Competition-mode*. The first provides a school-like environment where the user learns the letter-signs and the second provides the user an environment used towards testing the learned skills. Behind the scenes, the proposal is based on two phases: 1 – the creation of a gestures library, relying on the Kinect depth camera; and 2 – the real-time recognition of gestures, by comparing what the depth camera information to the existing gestures previously stored in the library. A prototype system was developed – the Kinect-Sign – and tested in a Portuguese Sign-Language school resulting in a joyful acceptance of the approach.

Keywords: Kinect Sensor; Sign Language; Serious Game; Gesture Recognition

1. Introduction

Earing is an "acquired sense" to which not much importance is given by most persons. Nevertheless, earing-impaired persons naturally find a barrier to understand and be understood. This problem gains particular importance if we take into account the numbers: according to [1], 650*10^6 persons around the globe are deaf and from this group 470*10^6 are in their working age. Statistics from Europe in 2002 state that 16% of the active population has some sort of earing problems, according to the same author [1].

Nevertheless, the majority of the countries have already a settled learning system for the corresponding sign language. The problem arises in what concerns non-deaf persons that need to learn sign-language: friends and colleagues of the deaf persons – the so-called listeners. This problem arises because the learning mechanisms are designed for the deaf persons and not for the listeners.

In order to ease the barrier between deaf and non-deaf, the research community has put a big effort in Sign Language Recognition (SLR), as summarized in [2], but as the authors say, the problem is far from being solved. Current implementations rely on image processing, which requires a great processing power, on one hand, and on special gloves, as presented in [3], on the other hand.

Other than computer science or image processing research communities, the Natural User Interfaces (NUI) groups are also contributing to these challenges, as presented in [4] [5]. Nevertheless, the problem complexity grows even more if one consider the non-manual aspects of sign-language, as stated by [2].

In what concerns the commercial devices that may help, namely in the SLR issues, the game consoles are playing an important role, like for example for rehabilitation purposes [6]. In terms of provided features, the Kinect device, introduced by Microsoft, has a depth camera and a Software Development Kit (SDK) that provides a skeleton representation composed of 20 joints with (x,y,z) information. Soon after the release of this device, the research community started changing from traditional-camera based approaches to Kinect-based approaches, [7].

This paper extends the Kinect SDK, providing gesture handling support, and proposes the creation of a serious game for listeners to learn sign-language. The remaining of the article is organized as follows: section 2 analyzes related work and the state of the art, in section 3 a framework is proposed for the creation of a sign-language teaching game; in section 4 the validation of this research work is presented and section 5 concludes the paper, including some guidelines for the future work.

2. Related Work

The creation of a serious game aiming to teach sign-language stands in a multi-dimension research area. First, the evolution of input devices, and their features, provide a growing improvement in the so-called NUIs. Second, the algorithms based on this features also contribute as a base element. Third, the serious game, which constitute a base element of the proposal presented in the research community. Finally, there are already some research groups targeting sign-language through games.

2.1. Natural User Interface

Multi-modal Interfaces is a research area that gathered much effort from the research community in the last two decades. In a simple way, one can state that this area tackles the improvement of the computer interfaces towards reducing their distance to the Natural User Interfaces. In other words, create the means for a human user to interact with the machines in the same way he or she is used to interact with the surrounding environment.

This research effort has had particular enthusiasm in the health care area, where several examples can be found. One example can be found in the usage of NUI in training activities for physician novice [8].

Another economy area that has put a big effort in the NUI research is the game industry as one can see in the game console devices. The Kinect sensor, used in this research work is one example of this effort.

2.2. Serious Games

“Serious games have become a key segment in the games market as well as in academic research” [9]. In fact, the usage of games in a serious purpose has been gaining the interest of several research groups. In what concerns the “serious” applications, examples can be found ranging from the military to the education.

This last research area has also been associated to the so called EduTrainment, which can be defined as “a continuous and innovative brain-training, which stimulates, in an interactive way, the capacity to combine attention and motivation to explore and learn” [10].

2.2.1. Sign Language Games

Due to the fact that the deaf community is relatively small in numbers, there is not much offer in terms of games that use sign language. Despite this, it is possible to find games, in various types of platforms, which target this group.

One can state that the existing games are quite basic. For example, the *Sign Language Bingo* is a board game with a 201 basic lexicon [11]. On the computer world, the *Sign-O* is another game also based on bingo and supports 12 categories of boards containing 25 words each [12]. The last platform, where sign language games can be played, is the internet and one example is the *Sign the Alphabet* where just two levels do exist: in the first the user has to answer correctly to a question, by selecting one sign between four possible answers. In the second level, the user must correctly reply to the sign visible in the screen with the keyboard [13].

Fig. 1. Sign Language Games: (a) *Sign Language Bingo*; (b) *Sign-O*; (c) *Sign the Alphabet*

Despite the usefulness of the games to teach sign language, they lack the ability to correctly validate if the user express correctly the signs or not. This creates a gap where none of the computer games can be truly interactive, which has made several people state that if such type of game existed it would be a good game to buy [14].

2.3. Sign Language Recognition Algorithms

SLR is a theme that has been subjected to many studies throughout the past 20 years, for example the work of Starner in 1995 [15]. Despite this great amount of studies none of those where able to reach the market and they were still very limited in terms of SLR, generally recognizing up to ten signs.

With the appearance of the Kinect sensor, this field of study regained importance. This is justified by the appearance of the depth camera and the possibility to detect distances in relation to the sensor. Another reason that helped to increase importance of this research area is the contribution of the machine learning and computer vision fields.

Until most recent years, the majority of the SLR algorithms were based on Hidden Markov Models, as shown in [15]. Recently other types of recognition have appeared, for instance in the work of [16] where two different algorithms have been proposed to make the SLR: 1 – a K-Nearest Neighbour; and 2 – a Support Vector Machine. Despite the fact that both give a good accuracy rate they are still limited to a very short lexicon.

Other of the most recent algorithms implemented for SLR was proposed by the cooperation of the Key Lab of Intelligent Information Processing, from China, with Microsoft, where they implemented a 3D trajectory matching algorithm capable of recognizing a sign from a library of 239 signs with an accuracy rate of 96.32 percent [7]. To the contrary of most of the existing SLR algorithms this is one algorithm that possesses a good lexicon.

From the current algorithms it is possible to verify that most are not validated for an extensive lexicon, but there is already some interesting solutions, that can be explored, for the proposal.

3. Proposal

Kinect-Sign is a serious game with the objective of teaching sign language and to enroll the users in games based on the sign language taught. For that reason the game is developed with two different modes: *School-mode* and *Competition-mode*.

There is also a need to implement some algorithm to make the SLR in a simple and fast way. Therefore, despite the already existing algorithms, it was designed a very simple recognition algorithm to be used in the game.

Both of this elements are extensible explained in the following sections.

3.1. Sign Language Recognition

The proposed SLR is divided in three phases: 1 – the data acquisition that consists in the acquisition and standardization of the Kinect sensor images; 2 – the data storage, where the depth data acquired is stored, in order to create a library of sign-language gestures; and 3 – the data recognition, that is responsible to match the acquired depth data from the Kinect sensor in real time and match it with the existing sign-language gestures.

3.1.1. Data Acquisition

Data acquisition consists in the phase that each Kinect depth frame must go through in order to be standardized and ready to be used for SLR. The first step in the data acquisition phase is the definition of the format to use towards storing the depth camera info. A 144×144 grayscale bitmap was selected for this purpose.

In order to make the data acquisition of one frame, this frame goes through the following four step process:

1. Acquire the raw depth data from the Kinect sensor.
2. Use the skeleton data (provided by the Kinect SDK) and obtain the user right hand point. Define the ROI, according to the distance of the user to the Kinect sensor.
3. Split the ROI, from the rest of the image, and convert the depths of the ROI into a grayscale. This conversion, depth to grayscale, is done by changing a buffer of depths that fit the distance, according to the Kinect SDK, of the right hand and convert this depths to grayscale. Any distance outside of this buffer is converted to black.
4. The final step consists in scaling the resulting bitmap to 144×144.

An example of a final image is visible in Fig. 2.

3.1.2. Data Storage

This phase represents the storage of bitmaps, resulting from the data acquisition, towards creating a library of sign-language gestures. This library will be extensive enough to support all the required gestures, in the current proposal this means the entire alphabet, from A to Z. For each letter of the alphabet the data stored will be divided into two categories: 1 – the source data, composed of 300 bitmaps, which refers to the signs used to make the recognition, in other words, the images used to validate the correct sign; and 2 – the test data, used in validating the SLR algorithm. The test data is composed of six bitmap images and three videos of 300 bitmaps.

3.1.3. Data Recognition

The data recognition is the most “heavy” phase of the problem, as several comparisons have to be made. In order to decrease this complexity towards turning the overall algorithm effective, five condition masks were identified for each point. The results obtained from the condition masks reflect the number of pixels that are ignored when making the matching.

The findings based on the experiments carried out showed that the selection of the best algorithm / condition mask depends on the application area and the available processing power.

Table 1. Different condition masks for ignoring black pixels in the recognition system

Condition	Kinect	Library
None	No	No
And	Yes	Yes
Or	Yes	Yes
Kinect	Yes	No
Library	No	Yes

The five identified condition masks, from Table 1, are:

- *None* – where none of the pixels is to be ignored;
- *And* – if both pixels, from the Kinect and Library bitmaps, are black, then the comparison is skipped;
- *Or* – skips the comparison if one of the pixels is black;
- *Kinect/Library* – when the corresponding bitmap has a black pixel, then that matching is ignored.

Using all these condition masks in extensive experimentation leads to the best selection towards decreasing the needed computing power, on one hand side, and increasing the accuracy of the system, on the other hand side.

Figure 2 shows two example images: one from the Kinect depth-camera (left-side) and other from the library (right-side). Both images correspond to the sign of letter “a” in the Portuguese sign-language.

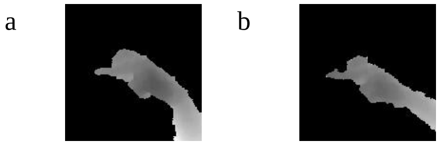


Fig. 2. Examples of bitmaps used during recognition, for the same sign “a”: (a) Kinect bitmap; (b) Library bitmap.

Using the conditions expressed in Table 1on the images in Fig. 2, it is possible to obtain the different matching areas, as visible in Fig. 3.

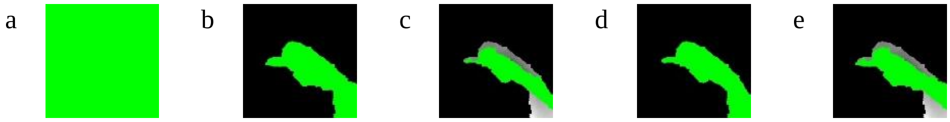


Fig. 3. Overlay, in green, of the two images from Fig. 2, according to the condition mask: (a) None; (b) And; (c) Or; (d) Kinect; (e) Library

After acquiring the matching area, green pixels of Fig. 3, between two images it is possible to conclude the recognition process. To do so, there will be a match between the valid pixels according to a predefined distance.

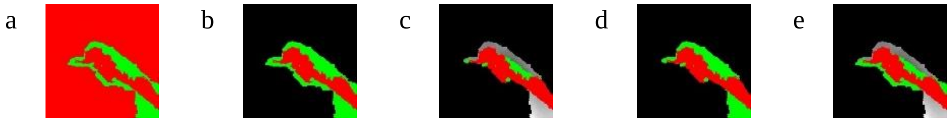


Fig. 4. Overlay, in red, of valid pixels from Fig. 3, according to the condition mask: (a) None; (b) And; (c) Or; (d) Kinect; (e) Library

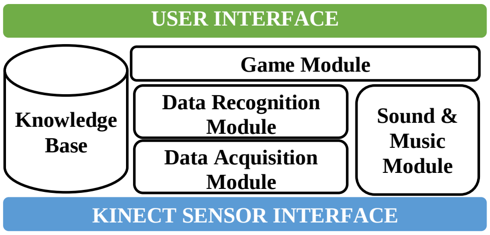
In Fig. 4 it is possible to view the matching using a distance of 25, in grayscale, which is approximately 2 cm. The experiment resulted in 88% accuracy for the *None* condition mask and 52% accuracy for the *Kinect* condition mask.

3.2. Serious Game

As mentioned before, the serious game is composed of two modes: the *School-mode* and the *Competition-mode*. The first has the objective to teach sign language while the second is where the user will play with sign language, testing the learned skills.

Fig. 5. Game architecture

Fig. 5 shows the proposed system architecture for the game. There are two different interface layers, the Kinect sensor and the user interface. The first is responsible for acquired the user gestures and the second, composed by the display and speakers, is responsible for displaying the proposal design. The knowledge base exists to store all the data, for instance, the results obtained by the user and the data necessary to make the SLR. There is also a module responsible for reproducing the sound and music in the speakers, the modules to treat the data from the Kinect sensor, data acquisition and recognition modules, and finally the game module that controls all the modules and interfaces.



3.2.1. School-mode

The *School-mode*, as the name suggests, is where the user will learn sign language. Therefore, it is proposed to design a classroom-like environment where the user would be enrolled in short lessons. In this level, the user should repeat the displayed signs.

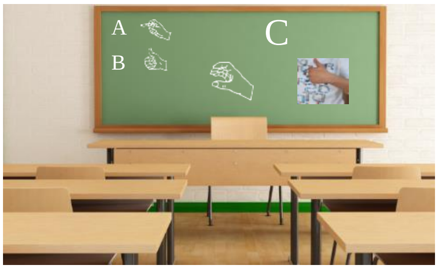


Fig. 6. Classroom design during a lesson

Also, for every two lessons this mode verifies if the user acquired the knowledge through tests. This tests unblock new lessons for the user to continue learning.

3.2.2. Competition-mode

In order to offer a more playful area in the game the *Competition-mode* is also proposed. The idea is to create a playful environment where the user experiments a competitive scenario. There, for example, following the style of a well-known TV Show, the user would test the learned skills.

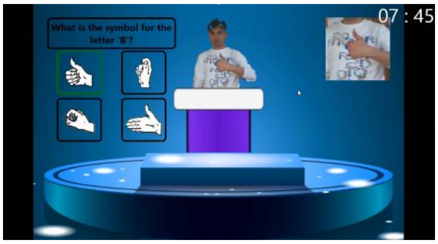


Fig. 7. TV Show scenario during a Quiz game

With a TV Show scenario it’s required to implement games that suit this environment. Because of that reason, the creation of the *Quiz* game is proposed, where the user is presented with a question, for instance “What is the symbol for the letter ‘B’?” and he or she must answer reproducing the correct letter from the presented 4 possibilities. An example of this game is visible in Fig. 7.

Other possible game for the *Competition-mode* is *Lingo*. In this game the user must find the correct five letter word in five different tries, for example. So, for each try the user must spell one word in sign language and then the game verifies if the word is correct or not, what letters are in the right and wrong positions, and so on.

4. Validation

The validation of this proposal was made through the development of the Kinect-Sign prototype. Afterwards, a validation of the prototype itself was also made.

The validation process took place in two different stages. The first stage was defined as the validation of working. In other words, it was verified if the current recognition algorithm and serious game worked for a single user. The second stage was the validation of usage, where the algorithms and game were tested with a broad number of users.

4.1. Sign Language Recognition

To make the validation of the SLR algorithm, the five condition masks presented in the proposal section were extensively tested in terms of accuracy rate and acquisition rate. The “distance” between images and the effect that increasing the distance makes on the recognition also tuned the algorithms.

The validation was based on the source data, composed of 300 bitmaps per letter, acquired from just one person. From those 300 bitmaps, each 100 were acquired using different distances from the Kinect sensor. Also, two images and one video, from the same person in the same situations (distance from the Kinect sensor), were also acquired to make the validation, this composes the test data.

The first process of validation was made using the images from the test data and making the recognition for those images. In this process, all the condition masks were used and the distance between bitmaps ranged from 1 to 50. From this process it was determined the condition masks that give the best overall accuracy rate - the *None*, *And* and *Library* conditions. Also, from the approximation rate, it was possible to determine five distances (10, 16, 19, 27 and 39) to implement in the second process of validation.

The second process for validation was done using the videos acquired for each sign. As stated before, each video is composed of 300 bitmaps and is used to simulate the system during a real time recognition. This process uses the five distances and three condition masks determined from the first process. From those it was determined that *Library* is the best condition mask and 19 the best distance.

4.2. Serious Game

The validation of the prototype serious game was made through real-time recognition during the game play. The response-time and the accuracy of the recognition were the success indicators.

The prototype game was also extensively tested by students of two classes from the “Instituto Jacob Rodrigues Pereira”: A beginner’s class and a more experienced level class. This institute is a Portuguese school that teaches deaf kids.

The results were quite satisfactory in terms of turning all the process transparent for the player. The enthusiasm of the kids and their professors was also rewarding.

Nevertheless, the game works much better when the player is the one that saved sign-language library. This fact points out to further tuning in this near future work effort.

5. Conclusion

This paper corresponds to the first steps of a work in progress. The usage of a serious game was proposed to be used as a learning mechanisms for the “listeners” of deaf persons – family, friends, colleagues, etc.

The proposal is based on the depth-camera from the Kinect device and works in two phases: 1st the users learn sign language through lessons; 2nd after learning sign language, the users are able to play gesture based games, *Quiz* and *Lingo*, provided in the proposal.

The work was validated through the development of a prototype – the Kinect-Sign – that was tested locally in terms of recognition rate and accuracy. The prototype was also tested in a Portuguese deaf school: the “Instituto Jacob Rodrigues Pereira”.

5.1. Future Work

After the completion of the current work it will be time to expand the validation of the SLR algorithm. In other words, it will be designed a knowledge base with multiple users and with that knowledge base the algorithm is going to suffer the same type of validation described in this article. The usage of distinct inputs towards improving the reliability of this teaching approach will also tackle the research area of Collaborative Networks, both in terms of support infrastructures, as mentioned in [17], and in terms of teaching approaches, as for example putting the students to work collaboratively, as mentioned in [18].

Other improvement on the project can be the introduction of a new algorithm to make the SLR and posterior validation and result comparison with the multiple user knowledge base.

In terms of serious game, there is also room for improvement. For that reason, on a first stage there will be implemented one or more games to be played with sign language and on a second stage the improvement of the game UI into a 3D environment, with the support of a game engine.

The research group will also make the system evolve towards its usage in classrooms.

References

[1] T. A. Delgado, “Surdez e participação no mercado de trabalho,” Lisboa, 2012.

[2] H. Cooper, B. Holt and R. Bowden, “Sign Language Recognition,” in *Visual Analysis of Humans*, Springer, 2011, pp. 539-562.

[3] Y. Song and Y. Yin, “Sign Language Recognition,” 2013.

[4] C.-H. Chao, Y.-C. Chen, T.-J. Yang and P.-L. Yu, “Intelligent Classroom with Motion Sensor and 3D Vision for Virtual Reality e-Learning,” in *The 2nd International orkshop on Learning Technology for Education in Cloud*, 2014.

[5] A. Angelopoulou, J. García-Rodríguez, A. Psarrou, M. Mentzelopoulos, B. Reddy, S. Orts-Escolano, J. A. Serra and A. Lewis, “Natural User Interface in Volume Visualisation Using Microsoft Kinect,” in *New Trends in Image Analysis and Processing - ICIAP 2013*, Naples, 2013.

[6] B. Lange, S. Flynn and A. Rizzo, “Initial usability assessment of off-the-shelf video game consoles for clinical game-based motor rehabilitation,” in *Physical Therapy Reviews*, vol. 14, Maney Publishing, 2009, pp. 355-363.

[7] X. Chai, G. Li, Y. Lin, Z. Xu, Y. Tang, X. Chen and M. Zhou, “Sign Language Recognition and Translation with Kinect,” Beijing, China, 2013.

[8] G. Carrasco, Y. Rybarczyk, T. Cardoso e I. Martins, “A Serious Game for Multimodal Training of Physician Novices,” em *ICER’13*, 2013.

[9] J. S. Breuer e G. Bente, “Why so serious? On the relation of serious games and learning,” *Eludamos. Journal for Computer Game Culture*, vol. 4, nº Serious Games, 2010.

[10] F. Corona, C. Cozzarelli, C. Palumbo e M. Sibilio, “Information Techology and Edutainment: Education and Entertainment in the Age of Interactivity,” *International Journal of Digital Literacy and Digital Competence*, vol. 4, 2013.

[11] T. Winnie and A. Drennan, *Sign Language Bingo*, 2008.

[12] *Sign-O*, 2007.

[13] “Sign the Alphabet,” Funbrain, [Online]. Available: http://www.funbrain.com/signs/. [Accessed 23 August 2013].

[14] “Kinect Could Reinvent Educational Gaming,” Winextra, 2013. [Online]. Available: http://www.winextra.com/tech/opinion/kinect-could-reinvent-educational-gaming/. [Accessed 23 August 2013].

[15] T. E. Starner, “Visual Recognition of American Sign Language Using Hidden Markov Models,” Cambridge, 1995.

[16] M. M. Correia, “Reconhecimento de Elementos da Língua Gestual Portuguesa com Kinect,” Porto, 2013.

[17] K. M. T. C. LM Camarinha-Matos, “ICT support infrastructures and interoperability for VOs,” *Virtual Organisations Cluster–VOSTER WP4 D*, 2003.

[18] T. C. L. C.-M. Edmilson Klen, “Teaching Initiatives on Collaborative Networked Organizations,” em *8th CIRP - International Seminar on Manufacturing* , Florianópolis-SC, Brazil, 2005.