Kinect-Sign: Teaching Sign Language to "Listeners" through a Game

João Gameiro, Tiago Cardoso, Yves Rybarczyk

▶ **To cite this version:**

João Gameiro, Tiago Cardoso, Yves Rybarczyk. Kinect-Sign: Teaching Sign Language to "Listeners" through a Game. 9th International Summer Workshop on Multimodal Interfaces (eNTERFACE), Jul 2013, Lisbon, Portugal. pp.141-159, 10.1007/978-3-642-55143-7_6 . hal-01350744

HAL Id: hal-01350744
https://inria.hal.science/hal-01350744

Submitted on 1 Aug 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Kinect-Sign: Teaching Sign Language to "Listeners" through a Game

João Gameiro, Tiago Cardoso, and Yves Rybarczyk
Universidade Nova de Lisboa,
Faculdade de Ciências e Tecnologia
Monte da Caparica, Portugal
j.gameiro@campus.fct.unl.pt, tomfc@uninova.pt, yr@uninova.pt

Abstract. Sign language is the hearing impaired form of communicating with other people, including listeners. Most cases, impaired people have learned sign language form childhood. The problem arises when a listener comes in contact with an impaired person. For instances, if a couple has a child which is impaired, the parents find a challenge to learn the sign language. In this article, a new playful approach to assist the listeners to learn sign language is proposed. This proposal is a serious game composed of two modes: *School-mode* and *Competition-mode*. The first offers a virtual school where the user learns to sign letters and the second offers an environment towards applying the learned letters. Behind the scenes, the proposal contains a sign language recognition system, based on three modules: 1 – the standardization of the Kinect depth camera data; 2 – a gesture library relying on the standardized data; and 3 – the real-time recognition of gestures. A prototype was developed – Kinect-Sign – and tested in a Portuguese Sign-Language school and on eNTERFACE'13 resulting in a joyful acceptance of the approach.

Keywords: Kinect Sensor, Sign Language, Serious Game, Gesture Recognition.

1 Introduction

Communication is the most vital mean for understanding and being understood. Nevertheless, hearing impaired persons are faced with barriers that prevents communication. According to the World Health Organization, 360 * 10^6 persons around the globe suffer from disabling hearing loss. From those, 328 * 10^6 are adults, where the rest are children. According to the same organization, approximately one-third of the population over 65 years of age suffer from disabling hearing loss [1].
Most of the hearing impaired knows sign language, being developed by this group through time. But a problem emerge when non-impaired persons try to communicate with an impaired person, for instances, in the case of family, friends and colleagues, of the impaired. This happens since there is almost no learning mechanism designed for listeners.

As summarized in [2], the research community has placed a great effort in developing a sign language recognition (SLR) system, but, as the authors say, the problem is far from being solved. Most implementations rely on image processing, and, despite the current advances on matching algorithms, still exists a need for great processing power in order to support real-time recognition. Nevertheless, the research community has been focused on showing that it is possible to make SLR, but limiting their studies in terms of lexicon or relying on the usage of special gloves, as presented in [3].

In what concerns the commercial aspect of teaching this form of communication, there are not many forms one can learn sign language. Nevertheless, specialized schools and videos exist, but are still limited.

The Kinect device, introduced by Microsoft, was intended to revolutionize the gaming industry [4], with the removal of the joystick/controller from the game. Alongside this revolution, the Kinect sensor also started appearing in other research areas. One of this is the SLR area, where the research community started changing from traditional camera-based approaches to Kinect-based approaches [5], [6].

The main goal of this article is to present an extension to the Kinect SDK, which provides gesture handling support, and proposes the creation of a serious game for listeners to learn sign language.

2 State of the Art

Teaching sign language through a serious game is a project that involves distinct areas. The first area of interest, a cornerstone of the proposal, is sign language, where there is a need to evaluate features, algorithms for recognition and existing games. The second area is the Kinect device, from which it is required to understand how it works and supported games.

2.1. Sign Language

Sign language shares, as much, similar aspects with spoken languages, as differences. For example, while spoken languages uses sound patterns to convey meaning, sign language is based on the usage of hand patterns and body language. Despite this difference both languages are considered natural languages and some of the sign languages have even achieved legal recognition [7], such as the Portuguese Sign Language (PSL). The Portuguese sign language alphabet is shown in Figure 1.

Fig. 1. Manual alphabet of the Portuguese Sign Language

Sign Language Games. As stated before, according to the World Health Organization (2013), the hearing impaired community is relatively small and, as a result, there is not much offer in what concerns games using sign language. Despite this, it is possible to find some games, in various types of platforms.
Three examples, which are available on the market are:

- *Sign Language Bingo*, a board game with 201 basic sign vocabulary words [8].
- *Sign-O*, a computer game, also based on bingo, with 12 categories and where each board contains 25 words [9]. The Figure 2b represents the cover for the CD.
- *Sign the Alphabet*, an internet game that is played by identifying the letters and numbers that are shown [10].

(a) Bingo (b) Sign-O (c) Sign the Alphabet

Fig. 2. Image with three different games available for teaching Sign Language

"Serious games have become a key segment in the games market as well as in academic research" [11]. Therefore, the games presented in Figure 2, can be useful when learning sign language.

Nevertheless, there is still room for improvements, for instance, on one hand, most games are based on Bingo and, on the other hand, there is an inability to correctly evaluate the signs made. This makes the computer games not truly interactive and, therefore, makes some people say that a truly sign language interactive game would be a good game to buy [12].

Sign Language Algorithms. SLR has been of great interest on the research community for the past 20 years, for example the work of Starner in 1995 [13]. This is due to the characteristics of sign language, but, despite this interest, the improvements made on SLR were still very limited, generally recognizing up to ten signs and where not able to reach the hearing impaired community.

Until recently, this field of study have been losing importance in the research community. But with the appearance of the Kinect sensor a change occurred in this field. This is justified by the development of a depth camera capable of detecting distances in relation to the sensor. Other contributing fields are the machine learning and the computer vision fields, which allowed the detection of the human skeleton, with 20 joints, through the depth camera.

Until recent years, most of the SLR systems were based on Hidden Markov Models, a statistical model in which the system to be modelled is assumed as a Markov process with hidden states, as shown in [14]. The work of Starner is based on this models, where he uses a single camera for data acquisition and the signing is tracked with the assistance of solid colour gloves. The prototype developed achieved a 97 percent accuracy while sing a forty word lexicon [13].

Recently, also due to the development on the machine learning and computer vision fields, Correia proposed two different algorithms for SLR: 1 – a K-Nearest Neighbour (KNN); and 2 – a Support Vector Machine (SVM). Just like in Starner' work, Correia was able to achieve good accuracy results, 96.58 percent for the KNN and 92.06 percent for the SVM, but he's work was limited to just four letters [15].

One of the most recent studies in SLR was developed in cooperation between Key Lab of Intelligent Information Processing, from China, with Microsoft, where the implemented prototype is based on a three dimensional trajectory matching, assisted by the Kinect sensor. This trajectory matching is made by acquiring the trajectory of the user hand, normalizing this trajectory, by linear resampling, and comparing with the trajectories gallery they possess. The study achieved a 96.32 percent accuracy using a gallery with 239 [5].

Based on this three systems it's possible to determine that high accuracy rates have been achieved in those studies, but the great majority of the studies are very limited in terms of amount of elements recognizable and there is still the matter of transition between academic study and market product.

2.2. Kinect Sensor

Code named *Project Natal*, the Kinect sensor was released on November 4, 2010, born from the collaboration between Microsoft and, the Israeli company, Prime Sense [4]. Their intent, for Kinect sensor, was to revolutionize the world of games, opening what they called new horizons and domains. To achieve it they alienated the standard controller and joystick. Instead, through the Kinect sensor, the natural movements of the body are used to control the form in which the game is played.

The introduction of the third dimension on the computer vision field, through a depth camera, is one of the key elements that assist in the removal of the game controller/joystick. This third dimension, or distances from the sensor, rely on the eco of infrared lights, or in other words, the measurement of distortions between distinct infrared beams, emitted by the sensor. Moreover, at a firmware level, other key element, the Kinect sensor is capable of providing skeleton and facial tracking. One last element, which allowed the first two to work properly, is the development of machine learning techniques.

Fig. 1. Composition of the Kinect sensor: (1) - Infrared optics; (2) - RGB camera; (3) - Motorized tilt; (4) - Multi-array microphone

The Kinect sensor is mainly composed, as shown in Fig. 1, of the following four components:

1) The *depth camera*, or infrared optics, responsible for understanding the 3D environment in front of the Kinect sensor.
2) The *RGB camera*, which besides showing the user on screen it's also responsible to make the facial recognition of the user.
3) A *motorized tilt*, mechanical gear that let the sensor follow the user.
4) A *multi-array microphone*, composed of four microphones embedded inside the Kinect sensor, is capable of isolate the user voice from the rest of the noise. It's also capable of pinpointing the user location in relation to the Kinect sensor.

Besides this four visible components, the Kinect sensor possesses its own processor and firmware. As stated before, this is one of the key elements for the working of the Kinect sensor, since they are responsible for identifying the information acquired by the infrared optics. In other words, they have the ability to "build the world" visible by the sensor and, from that "world", it's capable to extrapolate 20 joints in the human body [4].

Kinect Games. Kinect sensor was initially design to support games, and, since it's been released, many games have been created, for example *Kinect Sports* and *Star wars*, using this innovative device [16].

Another game, that appeared using the Kinect sensor, is *Minute To Win It*. In this game the user has 60 seconds to perform a simple task, and for every task he completes, he will earn a certain amount of money which can lead up to $1 million [17], shown in Fig. 2.

Fig. 2. Screenshot of the game "*Minute To Win It*"

3 Innovation - Multimodal

This project is developed around the Kinect sensor, so through the usage of this sensor a great level in innovation on the multimodal interfaces is already achieved. The reason for that is the lack of games and applications that uses this type of device. Also recognizing sign language through the Kinect sensor can be viewed as innovative, since we are recognizing the entire alphabet.

In the project, the main innovation is the creation of a serious game used to teach and, also, to be played with sign language. This is a brand new area, where there is nothing to compare, as it was visible in chapter 2.1. So it's possible to affirm that this types of games will develop into a brand new dimension.

This project is composed, mainly, by two components, the SLR system and the sign language game. The first can be expressed as the core for the entire project, since it's one of the most important systems that are embedded in the game. The SLR system will be studied in great detail so that it might be achieved the best results possible, before the usage of this system in the game. The second component, the sign language game, offers the players lessons where one can learn letters in sign language and offers some games where those lessons can be applied.

This project has a wide range of applications, starting from a classroom, where the teacher is assisted by the game to teach and to correctly evaluate the sign the student do. Or, in a more familiar environment, this game can be played between family members, assisting in the creation and development of bonds.

4 Proposal

The main goal on this section is to propose a serious game – Kinect-Sign – to teach sign language to listeners and to enrol the users in games on the sign language taught. For that reason, the game is designed with two different modes: *School-mode* and *Competition-mode*. Prior to the development of the serious game itself, there was the need to create a SLR algorithm that would work in the simplest and fastest way. Therefore, despite the already existing algorithms, it is proposed a very simple recognition algorithm to be used in the game.

4.1. Sign Language Recognition

The proposed SLR is divided in three components: 1 – the data standardization, that consists on the standardization of the acquired data through the Kinect sensor; 2 – the data storage, responsible for the creation of a sign language gestures library; and 3 – the data recognition, where the acquired depth data, from the Kinect sensor, is matched with the existing data in the sign language gesture library.

Data Standardization. The first step, prior of storing or matching any data, is the standardization of the acquired data. In other words, set the acquired data into a specific format for which standardizes the recognition process. This selected format is a 144×144 grayscale bitmap. In order to acquire this format the data acquired from the Kinect sensor goes through the following 6 steps:

1. Acquire the raw depth data from the Kinect sensor.
2. Obtain, from the skeleton frame (provided by the Kinect SDK), the right hand point. This point is the centre of the region of interest (ROI).
3. From the right hand point, determine the size of the ROI, according to the distance of the hand to the Kinect sensor.
4. After acquiring the size of the ROI, determine the corners for this ROI and then the depth image is split to get just the ROI.
5. Convert the depths in the ROI into a grayscale.
6. Concluding with scaling the grayscale bitmap to 144×144, through a nearest neighbour interpolation.

Data Storage. Data storage is the component responsible for storing the sign language library. This library is composed of the grayscale bitmaps, stored with two different purposes: 1 – a source data to use in the recognition process; and 2 –a test data, used in the validation of the SLR algorithm. Also, this data can be used for future developments.

According to the proposed purposes the data is divided into two categories. First, there is single bitmaps, or images, used both on the source data and test data. For the source data it will be stored 300 bitmaps per letter of the alphabet, while for the test data it is only stored 6 distinct bitmaps. Second, there is videos, composed of 300 bitmaps, stored in the test data, as to simulate real-time recognition. So, there will be three videos per letter of the alphabet, for recognition purposes.

Data Recognition. The last component is responsible for the matching between the gesture library and the Kinect acquired data. So, the proposed is an algorithm responsible for SLR. This occurs by acquiring a new image, with the Kinect sensor, and comparing it with all the images in the library. In order to implement this algorithm, the centrepiece for making the recognition is used the distance between depths, or in the prototype case, since the depths are converted into a grey scale, the distance between the greys.

Even though this evaluation of distances might be a good idea, looking to the standard format of the images and how the images are, Fig. 5, it is possible to get to two conclusions: 1 – the recognition achieves great accuracies, due to the great amount of black pixels in the image acquired; and 2 – being the images set to 144×144 pixels, there will be 20.736 pixels to match, which can be too expensive for most computers available.

Bearing this in mind, it is devised five different conditions to make the recognition and to assist in determining the most accurate algorithm possible, also developing a lighter algorithm. Those five conditions are shown in Table 1, and can be viewed as algorithms to ignore the black pixels of the images.

Table 1. Different conditions to ignore black pixels when using the recognition system

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

According to Table 1, the five conditions are:

- *None* – where none of the black pixels is ignored on any image,
- *And* – this condition ignores only when both pixels, in the matching, are black,
- *Or* – similar to the *And* condition, the *Or* condition ignores the matching when either pixel is black,
- *Kinect/Library* – skips the matching when the pixel is black, in the respective image, *Kinect* or *Library*.

Through extensive experimentation it is expected to achieve the best condition and, on one hand side, decreasing the computing power and increasing the accuracy of the system, on the other hand side.

(a) Image from Kinect (b) Image from library

Fig. 5. Example of images used in a recognition. Same sign "A".

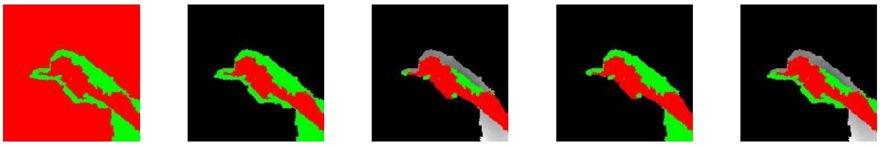
From the conditions stated before it is possible to see that each condition gives a different output of pixels to compare. To verify how they work, it will be used the images in Fig. 5, where both images are a representation of sign A, to display an example on how each condition works. The Fig. 5a contains the image to be compared, therefore the image that is going to be changed, and the Fig. 5b contains the image that is supposedly from the gesture library displayed.

(a) None (b) And (c) Or (d) Kinect (e) Library

Fig. 6. Overlay of two images according to the condition, the green pixels shows the pixels used on recognition

In the Fig. 6. , it is shown how the various conditions affect an image, where green represents the pixels available for comparison. With the definition of this conditions for recognition, it's possible to make a matching between two images. This matching gives the accuracy of distances. In other words, when a pixel in the image to recognize is close enough to the equivalent pixel in the library image it flags that pixel as a valid pixel.

After comparing all the pixels, according to a pre-set condition from Table 1, it's determined the accuracy of the matching by dividing the number of valid pixels by the total number of pixels.



(a) None (b) And (c) Or (d) Kinect (e) Library

Fig. 7. Overlay of valid pixels, in red, according to the distance, over Fig. **6.**

From Fig. , its possible see the pixels that gave a positive match, represented in red, when the distance between each figure is smaller than the variance V_D, in this case the distance, in grey scale, is 25, or in other words, approximately 2cm. In this matching is achieved an 88% approximation rate using the *None* condition and 52% approximation rate using the *Kinect* condition.

4.2. Serious Game

One of the main objectives of this work is to build a serious game where the user can learn and play with sign language. Kinect-Sign, the proposed serious game, is divided into two main modes: *School* mode and *Competition* mode, as stated before. The idea behind this modes is to help the families to interact with their hearing impaired loved ones. So, the *School* mode will assist in the teaching of sign language and the *Competition* mode is intended to help strengthen the connection inside the family, done through playful games.

School Mode. The *School* mode is designed so that any user can learn sign language in the most easy and playful way, but at the same time in a serious environment. In order to meet this parameters, the environment was designed as a classroom where the user will be enrolled in short lessons. For that reason, the lesson have at most five elements, also so that the user don't lose focus while learning.

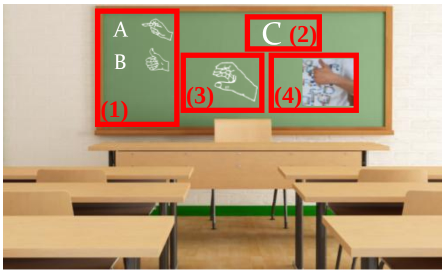


Fig. 8. Lesson design: (1) –Elements already learnt; (2) – Element being learnt; (3) – Representation of the element being learnt; (4) – Hand image acquired from the Kinect sensor.

The elements learnt in the lessons, as expressed in chapter 1, are letters and so the lessons will be continuous, separated by evaluations after every two lessons. An example of lesson can be viewed in Fig. , the user must reproduce with its hands the symbol the game presents and when the image get a good accuracy rate percent sets that element as correct and passes for the next element in the lesson.

After every two lessons, we should verify if the various elements have been learnt. To do so, there were created some evaluations. These evaluations appear whenever there is a need to verify the knowledge of the user. Normally, an evaluation will focus the two previous lessons, taken by the user, and they verify if the user needs or not to remake one certain lesson.

In order to pass an evaluation and unlock the next lesson the user must have at least 60 percent of the evaluation correct, this means that three out of five elements must be right.

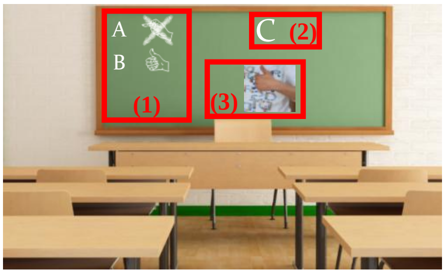


Fig. 9. Evaluation design: (1) –Elements answered and if they are correct or not; (2) –Element being evaluated; (3) – Hand image acquired from the Kinect sensor.

The displaying of the evaluation is very similar to a lesson, Fig. , where the main differences are that the user cannot see the representation of the element he must reproduce, and the elements that the user already answered wrongly will be crossed on the table represented in Fig. (1). This differences can be seen in Fig. .

***Competition* Mode.** The Competition mode is the more playful area on this serious game. Therefore it was design so that it may transport the user to a fun environment, while playing games that could relate to sign language. Bearing this in mind, it was developed a TV Show environment.

By designing the TV Show scenario, some games appeared to be more natural, for this type of environment, than others, one example is the *Quiz*. The proposed game is based, as the name implies, on making questions. So, when playing *Quiz*, the user will be asked five questions in order to make the best score possible.

Fig. 10. *Quiz* game: (1) - Question made to the user; (2) - Possible answers for the question; (3) – User (4) – Countdown timer to count how much time is left to end the round/game.

As it's possible to see in the Fig. , this game gives four possible answers for the question asked. This is a way to help the user, answering the question, and to help the recognition system, since it's just needs to compare with four elements in the library, making it a faster recognition. Also, in Fig. , it's visible that the user has responded to the question correctly, so the border of the answer is green. If the answer was wrong then the border would be red.

Other game proposed is the *Lingo*. Very similar to *Mastermind*, this game was created for the user to discover five letter words. Therefore, the idea of the game is for the user to spell, in sign language, the five characters that compose one word and find out if he was right, wrong or if there is correct letters, in the wrong place.

In order to ease the game there is a limit of five tries before giving the word as a wrong word. In the Fig. it's possible to see how this game is going to work.

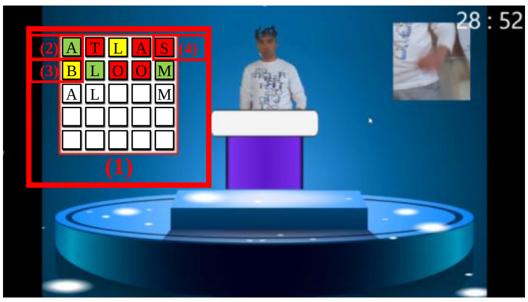


Fig. 11. *Lingo* Game: (1) - Lingo board; (2) - Right character in the right place; (3) - Right character in the wrong place; (4) - Wrong character.

From the figure above, Fig. , its visible the various states for each character. For instance when a character is right and in the right place, Fig. element (2), it's shown in green, when a character is right but in the wrong place, Fig. element (3), it's shown in yellow and if the character does not exist in the word, Fig.   element (4), appears in red.

5 Validation

The validation chapter is where the explanation on how is the work is tested takes place. In other words, it will be referred what where the experimental work done over the project, including a detailed description on the work, and what were the experimental results obtained from this project. After obtaining all the results, there will be a brief analysis and discussion on this results, where they will be explained and carefully considered their viability in order to continue the project development.

5.1. Sign Language Recognition

SLR is a simplified system where the Kinect sensor acquires an image, with the depth sensor, transforms that image, so that it may return only the right hand of the user in a grey scale, and then uses that image to compare with the gesture library, according to the condition and the distance between images. To validate the SLR system, there are two types of validations where the results cooperate to find the best values for matching.

The first type of validation is the validation through images, in other words, one image is matched with the entire gesture library, using all the conditions and distances (1 to 50 in grayscale). From this validation it is determined the three best conditions and five best distances to use apply in the algorithm. The second type of validation is made through videos, simulating a real-time recognition. During this validation it is going to be used the three conditions and five distances determined on the first validation and it is obtained the condition and distance that will be used in the serious game.

Prior to the validation process it is the constructed a strong dataset to support the recognition system. Therefore, it is acquired 300 images per character, from just one person. Also, from the same person, it is constructed the validation dataset, composed, per character, of six images and three videos, of 300 frames each.

There is two types of values obtained from the SLR validation: 1 – the approximation rate, which refers to the approximation between images, in other words, the ratio between valid and used pixels, as shown in red and green, respectively, in Fig. ; and 2 – the accuracy rate, or in other words, the ratio of signs acquired for each character.

Image Recognition. Image recognition is the process where one image is compared with the entire gesture library and from that values it is possible to extrapolate the three best conditions and five best distances. To implement this recognition it is used the five conditions, described in chapter 4.1: *None*, *And*, *Or*, *Kinect* and *Library*; and distances between images ranging from 1 to 50 in grayscale, equivalent of 0,08 cm to 3,92cm.

As an example, it is shown the results acquired when making the recognition, only with the *None* condition, for the six images that contains the letter 'L', this images can be seen in Fig. .

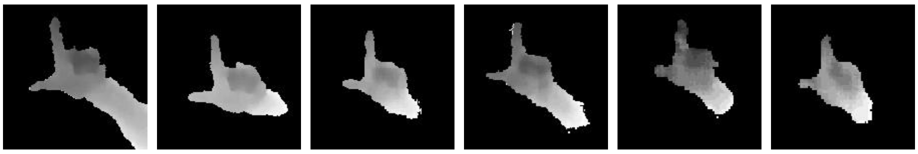


Fig. 12. Images of 'L' to be compared with the gesture library

None Condition. As expressed before, the *None* condition is where none of the pixels are ignored when matching two images. Therefore, it is expected to be achieved great values of approximation, in one hand, and, in the other hand, a better accuracy rate.

Table 2. Signs acquired using the *None* condition in a sign 'L'

	G	L	Q	R	U	X
1	10	34	1	5	0	0
2	0	46	0	0	4	0
3	0	50	0	0	0	0
4	0	50	0	0	0	0
5	0	47	0	0	0	3
6	0	48	0	0	0	2

Table shows the dispersion of acquired signs. From the obtained values is possible to conclude that the images used during the recognition are good 'L' images, since most of the results are correctly determined.

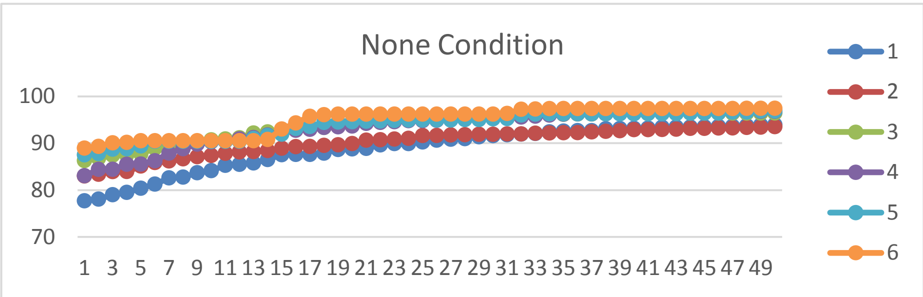


Fig. 13. Graphical representation of the approximation rates

On the other hand, Fig. shows the approximation rates obtained for each image according to the distance used in the recognition. Despite some of this rates not belonging to the wanted sign, it is fairly possible to verify that the rates don't suffer great variations when the recognition distance is increased. In the *None* condition this is explained by the fact that there is no ignored pixel and most of those pixels are black.

Global Results. Now, that it has been seen how the recognition works and the results acquired using the None condition, it is time to make the validation for the entire test set and obtain the global results for the image recognition. With these results it will be possible to extrapolate the distances and conditions in which the algorithm works best, as required for the video recognition.

In order to obtain the three conditions used on the video recognition, the accuracy rates are used. This is due to the fact that, as expected, the images recognized have a great tendency to refer to it' own sign. This accuracy rates are shown in Fig..

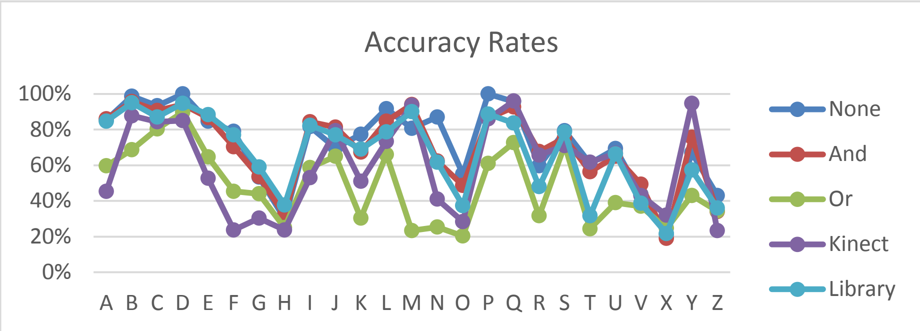


Fig.14. Average of accuracy rates for all the letters

From Fig. it is clearly visible that the *Or* condition is the worst condition in terms of accuracy rates. Other key element to be retained, from Fig., is the fact that not all the letters have good accuracy rates, this is explained for the fact that not all the letters possess "good" images for recognition and some signs are very similar between images, for instance the signs 'A', 'K' and 'T'.

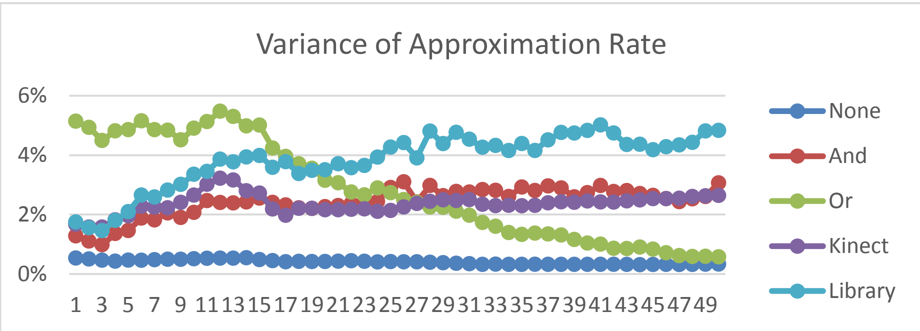


Fig. 15. Average of the variance on approximation rates

The figure above, Fig. , is the graphical representation of the average of variances on approximation rates. In other words, as seen in Fig., not all the acquired signs represent the correct sign, so the variance of approximation is the difference between the approximation rate acquired when matching with the entire library and the approximation rate acquired when matching just with the correct sign.

Single Recognition Conclusions. From the values obtained in this validation it was decided to use the *None*, *And* and *Library* conditions on the video recognition, since they are the best conditions when it is refer to accuracy results. On the other hand, the values for distance are: 10, 16, 19, 27 and 39. This values where obtained through the approximation rates and the variances from the approximation rate of the acquired sign with the approximation rate of the correct sign.

Video Recognition. Video recognition is the process where a group of images, acquired in a consecutive way is compared, one by one, with the entire gesture library, determining the best match for each frame of the video. Just like in the single recognition, in this process it will be experimented all the five conditions, but the distances used will be just a small group of five values, determined from the single recognition process. For this process of validation it will be used, as expressed before, three videos per character to make the validation.

Single Letter. As written many times, a video is a simulation of a real-time recognition where there is a sequence of 300 images ready to be recognized and those images will be used to validate the recognition system. So, by passing the video through the recognition system it will be possible to acquire the values of Table .

Table 3. Acquired signs using the *Library* condition on all the videos for sign ‘K’

	A	B	E	F	I	J	K	L	M	R	T	U	V	W	Y	Z
39	106	129	0	1	25	3	507	7	8	20	30	5	4	4	19	21
16	108	126	0	6	24	5	495	0	7	5	58	4	4	5	18	27
10	104	110	9	9	23	4	452	3	7	5	100	9	8	4	14	18
19	110	96	1	3	21	3	570	0	8	5	20	7	3	7	20	21
27	95	160	1	2	35	4	486	0	10	6	41	8	2	3	20	21

This table shows that the sign ‘K’ appears mostly when the set distance is 19, using the *Library* condition. While the distance 10 is a worst distance for detecting ‘K’. Despite this, from the 900 frames, where some are noise, in all the distances more than 50% is detected as the correct sign.

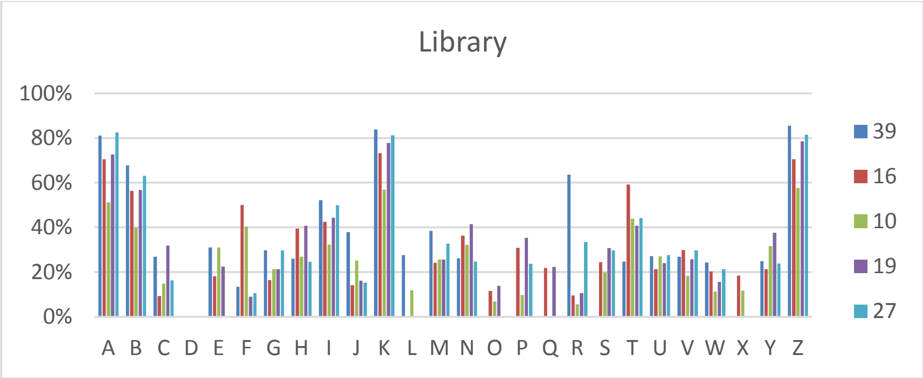


Fig. 16. Average of accuracy rates using the *Library* condition

Fig. shows the average accuracies for all the video frames when using the *Library* condition. From it, is visible that, for most distances, the difference between the wanted and the acquired sign is very short. Despite that, is also visible that the sign 'A', 'K' and 'Z' get the higher accuracies,

All Videos. Through the validation of all the videos shows what are the best parameters for recognition. Through the video validation it was possible to conclude that the best condition for recognition is the *Library* and the best distance is 19. This parameters are then to be used on the validation of the serious game prototype.

5.2. Serious Game

The validation on the serious game resulted of the testing of multiple users of the game. An empirical method was applied on the test subjects of two classes from the "Instituto Jacob Rodrigues Pereira" and on the participants of the eNTERFACE workshop. This resulted on three different test groups: a beginner, with no experience on sign language, a medium, where the group already knew something about sign language and an experienced group, the hearing impaired used to communicate with sign language.

The results were satisfactory, since it was possible to view the enthusiasm of the different participants, and the assistance provided by them in order to improve the current prototype. Still, a more in-depth analysis should be done (e.g. with questionnaires) to validate the developed prototype.

Nevertheless, the gesture library was reduced from 300 images per character to three images per character in order to the recognition process into a real-time process. This allowed the recognition of 4 frames per second. Also, the images used on the recognition where the images that appeared most in the validation process.

For this SLR algorithm it was verified that the serious game prototype works much better when the player and the creator of the gesture library are the same.

6 Concluding

In this chapter, it's pretended to conclude about the results obtained through the experimentation of the SLR system and the usage of this system in the sign language game and, therefore, testing its viability. Afterwards, there will be described on how the eNTERFACE workshop contributed for the development of this project and there will be presented a set of options for future development, subjected to the themes presented on this project.

6.1. General Conclusions

The existing documentation is rich in studies on SLR, including studies using the Kinect sensor. Despite this amount of studies, they lack, as much of the researchers' work, the step of bringing it into market. Therefore, the proposal of this project is to create a serious game, combining sign language and the Kinect sensor, which will lead to a product market-ready.

As seen throughout this report, this game is composed of two parts: the creation of a SLR system and the application of this system on a Kinect based serious game. For this article only the SLR algorithm was subjected only to a more extensive test. To validate the algorithm, tests were made evaluating the recognition rate and accuracy, as they will be detailed in the following sub-chapters.

Sign Language Recognition. In this system it was extensively studied the algorithm responsible for the workings of SLR. There were studied different methods of working and different parameters in order to acquire the best algorithm possible. From validation it was concluded that the *Library* condition and the distance 19 are the best parameters to work with this system.

Serious Game. The serious game has suffered a very simple validation, since it was just studied in terms of satisfying and pleasure of playing by multiple users. But, despite the simple study made it was possible to achieve very important conclusions about this project. The most important conclusion is that this is a viable project in terms of recognition system and really it's an enjoyable serious game.

6.2. eNTERFACE Contribution

The work presented in this project has been developed to serve as a master thesis and the passage through the eNTERFACE workshop was a step towards the conclusion of this thesis. This workshop served as a platform to discuss new ideas, also to revisit old ones and to understand features, about the project at hands, which were not possible to understand until then.

Clearly, one lesson that can be returned from this workshop is that every person has different body features. Applied the lesson to the project means that each is unique and that creates the necessity to implement some form of generic module capable of making the SLR. Since one user library will probably not make the right sign recognition for other user.

To conclude, the contribution provided by the eNTERFACE workshop was of great value for the improvement and expansion of the project Kinect-Sign.

6.3. Future Works

From the work developed for this project, it's possible to express that this game is not yet market ready. This means that there is still some work to be done, especially in the SLR system. Therefore, it is required the improvement of the SLR system, which might include machine learning or image processing algorithms. Other improvement is the serious game, particularly, the game UI which can be transformed into a 3D environment, for example using the Unity 4 game engine. Also, for the serious game, an increase in number of lessons and existing games can come to place.

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 [18], and in terms of teaching approaches, as for example putting the students to work collaboratively, as mentioned in [19].

Some other important work is required to continue the game development. First of all, a questionnaire about the serious game, for validation purposes. Second, a market study, to detect the needs of the target audience. And for last, the expansion of the gesture library so that the SLR algorithm might work more accurately.

References

[1] W. H. Organization, "Deafness and hearing loss," World Health Organization, February 2013. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs300/en/. [Accessed 2 October 2013].

[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. [Online]. Available: http://people.csail.mit.edu/yingyin/resources/doc/projects/6867term-project.pdf.

[4] "Kinect," Wikipedia, 13 August 2013. [Online]. Available: http://en.wikipedia.org/wiki/Kinect. [Accessed 22 August 2013].

[5] 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.

[6] J. Gameiro, T. Cardoso and Y. Rybarczyk: "Kinect-Sign - Teaching sign language to 'listeners' through a game," in *Conference on Electronics, Telecommunications and Computers*, Lisboa, Portugal, 2013.

[7] "Sign Language," Wikipedia, 16 August 2013. [Online]. Available: http://en.wikipedia.org/wiki/Sign_language. [Accessed 23 August 2013].

[8] T. Winnie and A. Drennan: “Sign Language Bingo,” 2008. [Online]. Available: http://www.superduperinc.com/products/view.aspx?pid=bgo133&s=sign-language-bingo#.UheEIxtwquI. [Accessed 23 August 2013].

[9] “Sign-O,” 2007. [Online]. Available: http://www.amazon.com/Sign-ASL-Bing-Game-Deaf/dp/B003AX6VO8.

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

[11] 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.

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

[13] T. E. Starner: “Visual Recognition of American Sign Language Using Hidden Markov Models,” 1995. [Online]. Available: http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA344219. [Accessed 24 August 2013].

[14] “Hidden Markov model,” Wikipedia, 4 August 2013. [Online]. Available: http://en.wikipedia.org/wiki/Hidden_Markov_model. [Accessed 24 August 2013].

[15] M. M. Correia: “Reconhecimento de Elementos da Língua Gestual Portuguesa com Kinect,” 2013. [Online]. Available: http://paginas.fe.up.pt/~ee06160/thesis/. [Accessed 19 August 2013].

[16] “Top 10 Kinect Games,” Squidoo, 2013. [Online]. Available: http://www.squidoo.com/top-10-kinect-games. [Accessed 22 August 2013].

[17] Zoo Games, “Minute To Win It,” October 2011. [Online]. Available: http://kinectplus.com/games/minute-to-win-it/. [Accessed 22 August 2013].

[18] L. Camarinha-Matos, K. Menzel and T. Cardoso: “ICT support infrastructures and interoperability for VOs,” *Virtual Organisations Cluster–VOSTER WP4 D*, 2003.

[19] E. Klen, T. Cardoso and L. Camarinha-Matos: “Teaching Initiatives on Collaborative Networked Organizations,” em *8th CIRP - International Seminar on Manufacturing*, Florianópolis-SC, Brazil, 2005.