Interactive System for Polish Signed Language Learning

K. Nurzyńska, A. Duszeńko
Silesian University of Technology, Gliwice, Poland

***Abstract*— The aim of this study is to present an overview about computer singed language course with module for automatic signed language recognition as a part of language acquisition test. The idea to create an interactive sign language learning system seems to be a new one. We hope that this solution helps to overcome the barrier between the silent and hearing world. On the other hand, we concentrate our efforts to create a system for a home use that will not need any sophisticated hardware. Moreover, we put pressure on utilization of already proposed and popular description scheme. The MPEG-7 standard formally called the Multimedia Content Description Interface has been chosen. This standard provides a rich set of tools for complete multimedia content description. The most important application for sign language is the possibilities to describe static and dynamic features of objects in image sequences both. This description schema gives the opportunity to create description of signing person on required level of granularity.**

In the article a brief description of many suggested solutions for semiautomatic or automatic sign language recognition systems is given. Besides, there are described some implemented learning application which aim was to learn sign languages. The main groups, which could be distinguished are: animated avatars observation, messenger for deaf people, testing progress in learning sign languages by using education platforms.

Index Terms— **computer aided learning, Internet learning, image recognition, signed language recognition**

I. INTRODUCTION

Knowledge is the most important issue in a development of civilizations. Due to that fact modern communities pay special attention to the level of education. The broad researches are conveyed about the ways of learning and their efficacy. It is important to find ways for effective transfer of information nowadays. Hence the students memorize as much information as possible. It is not only crucial for science development but also it was proven that well educated society implies an economical growth. On the other hand, fast progress in each domain of life demands from people to enhance their qualification, and that makes them to learn all the time.

During centuries studying was only associated with learning from the books or with a teacher. The most recent technical solutions are found to be very useful in the process of knowledge acquisition. Consequently there are created many programs for learning. Those programs to exploit its additional possibilities need only a computer or also the Internet access. According to this fact companies present many programs for learning different subjects. The most common domain of interest is foreign language computer or Internet courses. One has seen software for words translation with additional tests to check proper words usage. To keep pace with growing market demands newer versions are equipped with speech recognition modules for checking the proper pronunciation of foreign words or even whole sentences. The next step is to employ image processing modules to improve those systems. For foreign languages it may seem useless, however spoken languages constitute only a part of all languages existing in the world. There exist also sign languages, where the information is delivered with visual signs.

The aim of this studies is to present an overview about computer sing language course with module for automatic sign recognition as a part of language acquisition test. The world of deaf people, however noticed, is still in the background of whole problems, which disabled people have. Nevertheless the deaf people are getting better and better prepared for living in a world of hearing ones. Most of them learn speaking and lip-reading, but this is not enough. To overcome the barrier between the silent and hearing world the hearing ones should also learn the language of the other part of society. Therefore we hope that presented system could help both worlds.

II. BRIEF DESCRIPTION OF SIGN LANGUAGE RECOGNITION SYSTEMS

In the world literature there are many suggested solutions [1, 2, 3, 4, 5, 6, 7] for semiautomatic or automatic sign recognition systems using vision based approach. The variety of solution derives from the fact that the input information may by acquired in many ways as well as the level of complexity of those system differ a lot. The basic distinction is the number of cameras used for image acquisition:

- one camera systems – the input information consists only of the camera image, thus the visual information of signs/hands position is described in 2D space.

- two or more camera systems – utilization of more cameras helps to reconstruct the 3D world, which helps to achieve exact signs/hands position description.

Although, the one camera solution gives less information, it is easier to construct. For support system performance authors suggest the usage of special markers. Among them can be distinguished:

- sensor data gloves [5] – a special device for describing position of the hand.
- colour data gloves [4] – used for improving hand detection in a colour images. In some examples each finger has different colour, which helps extract exact shape.

Certainly there also exist systems, which does not require from user any special interactions. They concentrate mostly on skin colour detectors. This problem is widely researched as the performance depends between other from the colour space (eg. RGB, HSI, YCbCr) [1, 2] used for colour description. Those solutions have also some limitation. Skin colour detection algorithms are very sensitive on changes in illumination condition as well as have problems to cope with complex background. Despite the results are satisfactory.

Besides the input data distinction the systems work on different data granulation levels. Where the data is understood as a part of sign language. The easiest systems limit its functionality to sign alphabet recognition. In this case the crucial role takes very accurate hand fingers shape description, thus high resolution images are necessary. The second level systems are capable of recognition of signs from the sign language wordbook [3, 6]. Some of them go a little bit further and try to restore whole sentences basing only on the word given as a signs. Finally, the researchers try to create a system, which recognizes whole sentences [7]. At each level of functionality the systems work well, although they need many training examples, which makes the process of system learning a time consuming task. Moreover, the systems are prone to changes in the way of signing between people.

III. LOOK INTO A SIGNED LANGUAGE LEARNING SYSTEMS SOLUTIONS

The idea of computer system for sign language learning is not new. There were created many applications, which aim to facilitate firstly sign languages learning as well as to introduce more complex knowledge to the deaf user. The authors of these systems have chosen an Internet or computer application to assess large population of people. Some examples of the system are described below.

The beginnings of computer science are connected with raw interfaces based on text. The progress in computer technology enabled creation of graphical interfaces but still this interfaces use text as a basic form of information representation. For most of the people this is natural representation of the language which they use every day. But for people using sign language this is not natural. They have other form of representation of the word, which are gestures. The natural effect of this situation is a rogress in the interface of so popular messenger systems. As described in [10] the MakMessenger is the example of application using sign language gestures as the basis of the communication interface. This application uses predefined symbols representing gestures of the sign language. This is very basic support for sign language, which we can call "passive support" because in fact this application does not render actual gestures just transfers user predefined pictures.

In the article [8] is described an Internet Test Battery (BALS On-line) system which tries to overcome problem of reading and spelling in Portuguese by hearing and deaf students as well. This application tends also to develop the ability to comprehend Brazilian Sign Language for deaf users. It was constructed as a Web application to allow usage to large population. Additionally, it uses commonly used Internet technology, like Flash for animation. The system can be divided into four groups of tasks:

- picture matching with written words,
- sign language matching with spelling words,
- sign language mathching with written words,
- picture matching with sign language.

Advancement in graphic interfaces enables creation of the interfaces accommodated to specific needs. The very good example is [9]. Kids Sign Online is in fact application in which interface uses all possible media to represent information. Text, pictures video and sign marks are used to present information in very natural form. This system not only transfers gesture selection from one user to another but it verifies it to determine user progress in education. This system gives very wide possibilities to present sign language even for so specific users like children, but still there is no feedback. System still does not have the ability to verify actual gestures presented by users.

Paper [11] presents very interesting idea for interactive learning arithmetic for deaf children. This application is not only addressed to these children but also to their parents. For both groups are designed parts responsible to learn signed language gesture needed to deal with mathematical problems and the testing parts. To show the information a 3D animation was chosen as it gives the control of the orientation, location, zoom and speed of the animated avatar.

IV. PROPOSED SYSTEM OVERVIEW

The proposed system task is not only the signed language learning but also the verification of signed language acquisition process. We address a problem of signed language learning creation system on the basis of reviewed systems described above. Additionally, the goal is to broaden its functionality of ability testing module for students. In our understanding this additional function is a novelity in the signed language learning process. The

testing consists of user image acquisition, showed sign recognition and finally of verification whether it is the learned sign or not. Moreover, during the researches for signed language recognition algorithms was paid a special attention for its hardware demands. It tended to find such a solution, which is not based on any sophisticated hardware. We claim that optimally the application should work on average personal computer with additional Internet camera only.

The system consists of two parts. First part is responsible for learning, second for testing the acquired language abilities. Most probably the first part solution will be taken from the ideas described above. However, what concerns us most is the way of gesture learning, which may be realized in various ways: starting from image sequences showing how to move hands, finishing with more complicated solutions as movies with a signing teacher or the utilization of animated avatar to show the gesture. The second part of the application will consist basically of automatic signed language recognition system. Its aim is to recognize the gesture and to decide whether it is shown correctly. Its is worth to mention, that the system knows what the user wants to show, therefore the recognition process in first step of system implementation may be limited to checking only whether it was done properly or not. And only this information sends back to the user. In the next step of the program development the recognition will be realized with utilization of the full possible information, hence the user will be given more detailed information like for instance, with which other sign what he/she shows is confused, if any. Very important issue is the face and hands object detection in an image. To assure easy use of the system we do not want to make the user to wear any coloured gloves or sign on the uniform background. The object detection system is based on skin colour detection algorithm. To choose a proper method for this module we have carried out broad researches, which resulted in an article “in press” [16].

V. MPEG-7 STANDARD AS A MEDIUM FOR DATA DESCRIPTION

A set of tools for describing hands and face location and movements of a signing person is required for a realization of a sign language automatic system. This tools, according to general process of description methods standardization, should refer to some already proposed and popular description scheme. This approach makes them much more attractive and prepared for extension as well as for integration with other systems. Among many different multimedia resources description methods the most popular one is developed by Moving Picture Experts Group (MPEG) and presented as ISO/IEC International Standard 15938. This document defines the MPEG-7 standard formally called the Multimedia Content Description Interface. This standard provides a rich set of tools for complete multimedia content description. Its popularity is also associated with broad extension possibilities. The most important in sign language applications are possibilities to describe static and dynamic features of objects in image sequences both. MPEG-7 standard well suits its demands.

To distinguish from other description methods, the MPEG-7 temporal object position relation is treated equally well as spatial description. This makes easy to take into consideration not only spatial, but also temporal relations between visual objects like hands and face.

Using MPEG-7 Description Schema standard to describe signing person gives the opportunity to create description on required level of granularity. MPEG-7 Description Definition Language specifies the DDL [12,13] to describe wide variety of relations between visual object. This description is defined as a graph, not like in other solution through treelike structures.

In the case of sign language recognition systems image preprocessing is performed to extract information about hands and face relative position. This can be achieved by skin colour models utilization. Image areas containing hands and face should be after detection described and stored for future usage (in next analysis stages). Such a description makes it possible to analyze hands and face location. Relative positions and recognized hands gesture contain sufficient information to perform sign language recognition. The illustration below shows an example of such a description.

MPEG-7 Visual part specifies the tools dealing with visual-only description. It is possible to describe content of a picture collecting all interesting visual objects features with these tools. The MPEG-7 Visual Descriptors can be classified into generic and high-level (application-specific) description tools. The generic Visual Descriptors describe color, texture, shape, and motion features. The high-level descriptors provide description tools for face-recognition applications.

In presented application of MPEG-7 the most important Descriptors are:

- Basic Elements (used by the other Visual Descriptors): grid layout, time series, 2D–3D multiple view, spatial 2D coordinates, and temporal interpolation.
- Color Descriptors: Color Space, Color Quantization, Scalable Color, Dominant Color, Color Layout, Color Structure, and Group-of-Frames/Group-of-Pictures Color.
- Texture Descriptors: Homogeneous Texture, Non-Homogeneous Texture (Edge histogram), and Texture Browsing.
- Shape Descriptors: Region-Based, Contour-Based, and 3D Shape.
- Motion Descriptors (for video): Motion Activity, Camera Motion, Parametric Motion, and Motion Trajectory.
- Location Descriptors: Region Locator and Spatio–Temporal Locator.

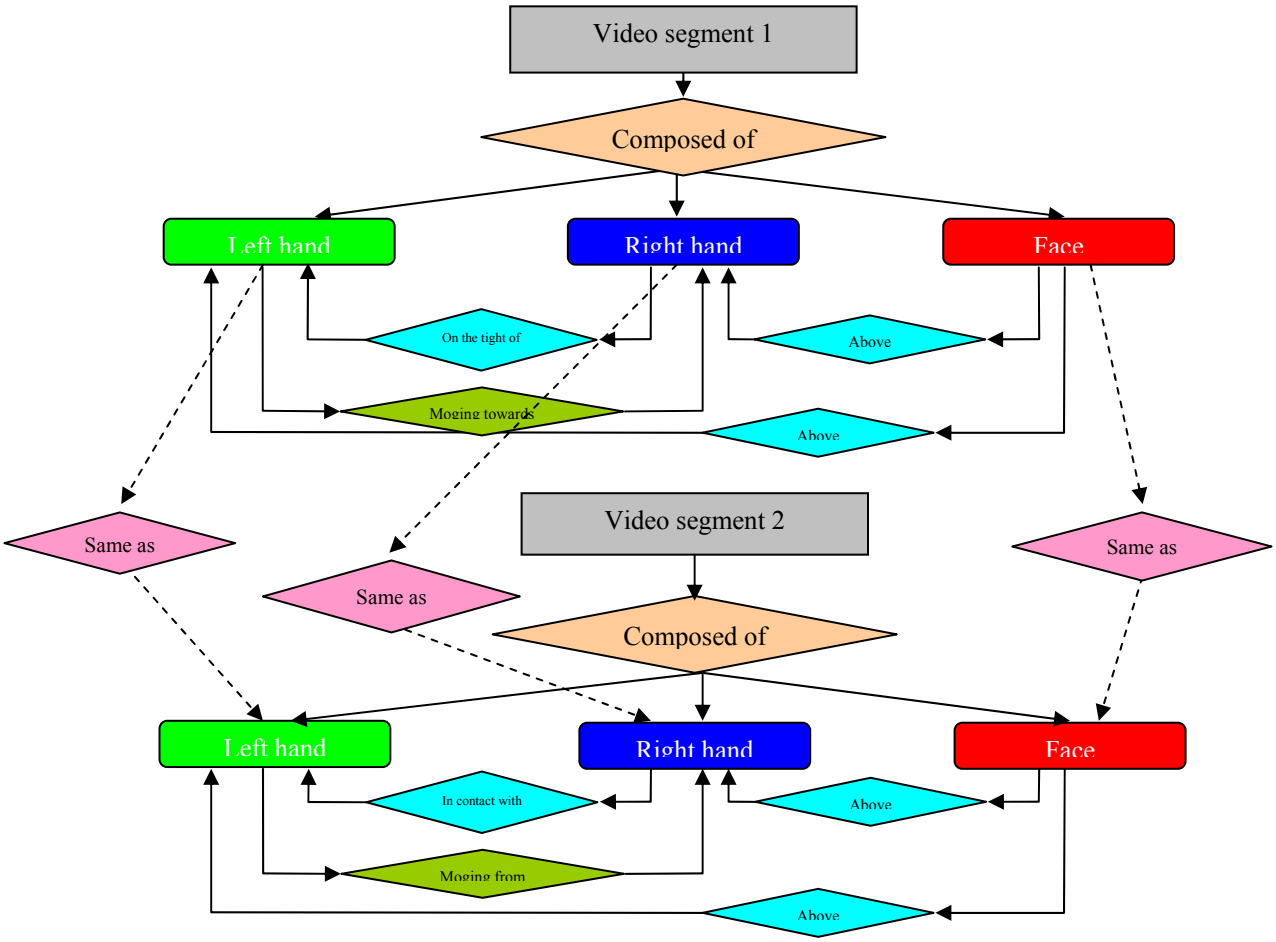

Figure 1. Temporal relations between objects composed in the form of graph.

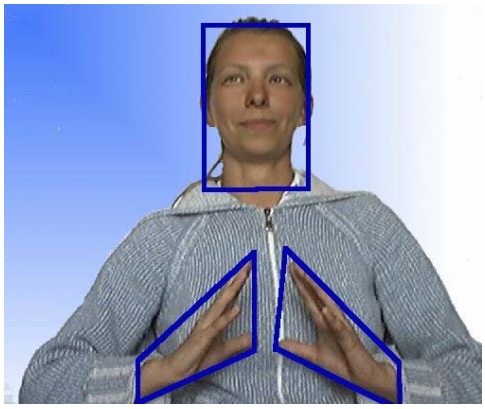

Figure 2. Example of system work 1.

The main area of the image containing signing person that should be described is a background. In this application every object except hands and face of a signing person is considered as a background.

```
<Mpeg7>
…
<StillRegion id = “sign language symbol”>
<SpatialDecomposition>
        <StillRegion id = “background”>
                <VisualDescriptor
                xsi:type=“DominantColorType”>
                110 108 140
                </VisualDescriptor>
```

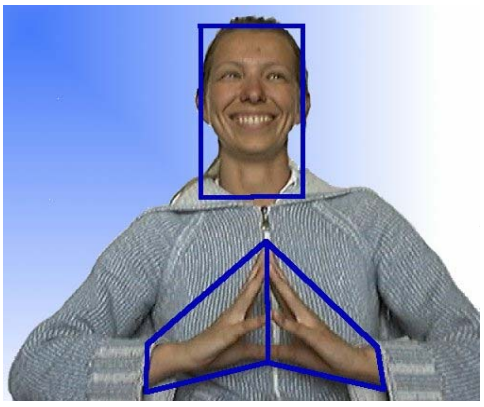

Figure 3. Example of system work 2.

```
        <StillRegion id = “speaker”>
                ...
</SpatialDecomposition>
</StillRegion>
…
</Mpeg7>
```

The main and the most important in this application part of this document is description of hands and face area location. These areas (shown on the Fig. 2 and Fig. 3) are usually represented by polygonals or ellipses. Its shape and location can be retrieved automatically by methods mentioned above.

```
<Mpeg7>
…
<StillRegion id = “speaker”>
<TextAnnotation>
        <FreeTextAnnotation>
                Lector Karolina
        </FreeTextAnnotation>
</TextAnnotation>
<Mask xsi:type="SpatialMaskType">
        <SubRegion>
        <Poly>
                <Coords>  80  288,100  200,…,352
288
                </Coords>
        </Poly>
        <SubRegion>
</Mask>
</StillRegion>
…
</Mpeg7>
```

The possibility to describe temporal relations between objects composed in the form of graph is presented on the Fig. 1.

To achieve maximum comfort of MPEG-7 description usage it is stored in XML documents. Storing signing person description in the form of XML document ensure wide data integration and exchange possibilities but also forces the concern about data storage model used to store this documents. In this situation it is necessary to consider documents storage in XML enabled databases [15] or in hybrid XML data model systems [14].

VI. Conclusion

Computer systems operating with signed language become a center of interest among other applications lately. Most of them focus only on this language visualization, although, without the realization of reflexive interaction with a user. The additional prospect of user-computer interaction creates a novel spectrum of possibilities. In particular, it allows to realize signed language learning systems, which will have a tools for learners skills verification.

The presented analysis of the concept of signed language system organization allow for its realization presently. The further realization of this system could be based on the cooperation with existing gesture presentation systems. In this aspect, the most important is the presented analysis of information storage data schema, including the option of work in the distracted environment due to the adaptation of MPEG-7 standard.

References:

[1] Soontranon, N; Aramvith, S.; Chalidabhongse, T.H.: Face and Hands Localization and Tracking for Sign Language Recognition, ISCIT 2004, Sapporo, Japan

[2] Soontranon N, Aramvith S., Chalidabhongse T.H.: Improved Face and Hand Tracking for Sign Language Recognition, Proc ITCC 2005

[3] Bauer B, Hermann H.: Relevant Features for Video-Based Continuous Sign Language Recognition, Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition, March 2000, Page(s):440 - 445

[4] Wu J., Gao W., Song Y., Liu W., Pang B.: A Simple Sign Language Recognition System Based on Data Glove, Proc. ICSP 1998

[5] Culver V. R.: A Hybrid sign Languge RECOGnition YSTEm, Proc8th ISWC 2004

[6] Jianjun Y., Honqxun Y., Feng J.: Based on HMM and SVM Multilayer Architecture Cliassifier for Chinese Sign Language Recognition with Large Vocabulary, Proc. ICIG 2004

[7] Bowden R., Windridge D., Kadir T., Zisserman A., Brady M: A Linguistic Feature for the Visual Interpretation of Sign Language,

[8] Coutinho E., Capovilla F. C., Diana C., Orsati F., Nikeado C.: Development of a Test Battery to Assess Deaf Language Skills via WWW, LA-Webmedia 2004

[9] Ohene-Djan J., Naqvi S.: An Adaptive WWW-based System to teach British Sign Language ICALT 2005

[10] Ohene-Djan J., Zimmer R., Bassett-Cross J., Mould A., Cosh B.: Mak-Messenger and Finger Chat, Communications Technologies to Assist in the Teaching of Signed Languages to the Deaf and Hearing, ICALT 2004

[11] Adamo-Villani N., Doublestein J., Martin Z.: The MathSigner: an Interactive Learning Tool for American Sign Language K-3 Mathematics, 2004

[12] Overview of MPEG-7 Description Tools, Part 2, IEEE Computer Society, July-September 2002,

[13] Overview of MPEG-7 Description Tools, Part 1, IEEE Computer Society, April-June 2002,

[14] Duszenko A.,Nurzynska K.: XML hybryd database model, Bazy Danych, Aplikacjie i Systemy, May 2006.

[15] SteegmansB., Bourret r., Cline O., Guyennet O., Kulkarni Sh., Priestley S., Sylenko V., Wahli U.: XML for DB2 Information Integration International Technical Support Organization, July 2004,

[16] Duszenko A.,Nurzynska K.: Skin colour detection method comparison (as an initial step for face and gesture recognition), Systemy Czasu Rzeczywistego, September 2006

Authors

K. Nurzyńska is with the Departament of Computer Science, Software Division, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland (email: Karolina.Nurzynska@polsl.pl)

A. Duszeńko is with the Departament of Computer Science, Computer Science Teory Division, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland (email: Adam.Duszenko@polsl.pl)