Recognition of deaf gestures based on a bio-inspired neural network

M G Grif¹, A V Kugaevskikh²,³

1. Novosibirsk State Technical University, 20, Karla Marksa ave., Novosibirsk, 630073, Russia
2. Institute of Automation and Electrometry SB RAS, Academician Koptyug ave. 1, Novosibirsk, 630090, Russia
3. Novosibirsk State University, 1, Pirogova str., Novosibirsk, 630090, Russia

E-mail: grifmg@mail.ru, a-kugaevskikh@yandex.ru

Abstract. In this paper discusses the current situation in Russia and the world in the field of development of sign languages translation system. The main problems are formulated, and ways to solve them are given. One of the most important unresolved tasks is the task of recognizing the gestures of the deaf. To effectively solve it, an approach based on the development of bio-inspired neural networks is proposed. The architecture of a bio-inspired neural network, including four types of neurons, is described. New simpler MT neuron model proposed.

1. Introduction

The development of machine translation systems of Russian sign language is long overdue and has now moved from the field of theory to the practical plane. According to estimates of a number of sources, the number of consumers of machine translation systems of RSL in Russia reaches 600 thousand people (Deaf, their parents, students, social workers) Need to develop systems for machine translation of sign languages (From hearing to deaf and vice versa) consists of both insufficient number of sign language interpreters and not always desirable mediation (Medicine, personal relations) in the communication of deaf and hearing citizens.

An example of a sign language translation system available in the world can be mentioned: HandTalks (Brazil), ProDeaf (Brazil), ViSiCAST (Europe), The American Sign Language Avatar Project at DePaul University (USA) Surdofon (Russia) [1]. For feedback - from deaf to hearing, a set of text, selection of phrases from a thematic talkative, or a pictographic system can also be used for constructing a phrase followed by voicing with a computer speech synthesizer. It can also be noted that none of the systems considered has full functionality to support communications between deaf and hearing.

The analysis showed that there are five main problems in the development of Computer Sign Language Translation (RSL) systems:

1. Lack of effective RSL recognition systems;
2. Translation of Russian text mainly into calculating sign language (CSL), which makes it difficult to perceive it as deaf;
3. Low quality of the translation of Russian text into RSL (CSL);
4. Lack of a recognized grammatical system of Russian sign language;
5. Low ergonomics of sign language visualization system.

One of the most important unresolved tasks is the task of recognizing the gestures of the deaf. Such sign display phenomena complicate accurate recognition of the sign language as dynamism, overlap. The solution to this problem may be the development of a bio-inspired neural network. For traditional neural networks, including those trained by deep learning, if the network is not trained on some vector of motion, it will not detect it.

2. Features of sign languages of the deaf
One possible classification of gestures is gesticulation, language-like gestures, pantomimes, emblems and the actual sign languages [4]. Linguistic elements of sign languages are signs (gestures), each of which can be described by as many as five parameters (Fig. 1): a) handshape, b) hand orientation, c) location, d) movement and e) non-manual component. A change in a single component generally alters the meaning of the whole sign.

![Figure 1. The five components of signs in sign languages.](image)

Typically, deep learning is designed to handle large amounts of tagged data (datasets) and uses sophisticated algorithms to teach the model. On large datasets, deep learning shows higher accuracy of results compared to other machine learning techniques.

In machine learning tasks, the quality of models depends very much on the dataset. However, datasets themselves in real tasks are rarely ideal for the following main reasons:
- Few real data are marked;
- Low availability;
- Presence of noise and gaps in data;
- The problem of getting the correct data in the correct format.

For the sign language recognition task, it is even more difficult to select the correct structured datasets, many different phonetic transcripts (classes) of sign languages (SL), very few SL datasets, and scientific articles on machine translation of LJ, compared to sounding language. There are also no SL datasets in Russia, so we decided to collect various existing datasets from different countries, for example [5].

3. Design of the bio-inspired neural network
In the visual cortex, motion analysis begins back in the primary visual cortex. In particular, the complex cells that served as the basis for the pulling neurons perform the detection of elementary motion as a basic function. The motion is then analyzed sequentially in V3 and V5 (MT) areas. Eventually, a general map of movement within the visual field is constructed.

Most often, motion detection in computer vision is accomplished through optical flow computation. In neural networks, training is applied to isolate motion again through optical flow computation spread over several layers of neurons. For bio-inspired models, the most common model is HS [2, 3], the analysis of which is well shown in [6]. The HS model is based on Gabor energy, taking into account the real and imaginary parts of the Gabor filter. The use of the imaginary part is justified in signal processing, but unnecessary in image convolution. In the proposed model, we not only get rid of Gabor energy but also build a convolution differently in the neuron model to more conform to the optical flow equation.

Again, motion detection begins with edge detection. The neural network we design also receives information from edge detection neurons based on Gabor filters and hyperbolic tangent [7]. The L* frame channel in the CIE L*a*b* space is used as inputs for edge and motion detection. A two-layer network is used for edge detection, Figure 2. Lines of a specific orientation are detected on the first layer. The second layer is responsible for detecting combinations of lines, including angles.

Each layer contains four types of neurons, differing in the receptive field configuration. Each second-layer neuron (U_C2) is only connected to two first-layer neurons (U_S1). Thus, the second-layer neurons allow for line and angle detection (in the case of a Gabor filter) and quadrangles (in the case of a hyperbolic tangent). Receptive fields of the first two neuron’s types are formed using the Gabor filter with different phase.

\[ G_{1,2} = \exp \left( -\frac{x'^2 + y'^2}{2\sigma^2} \right) \cos \left( 2\pi \frac{x' + \phi}{\lambda} \right), \]  

\[ G_{3,4} = \tanh \left( (-1)^{p} \frac{2x'}{\lambda} \right), \]

where \( y = 0.1 \) – spatial aspect ratio; 
\( \theta \in [0,350] \) – orientation; 
\( \lambda = 3 \) – wavelength; 
\( \sigma = 0.56 \times \lambda \) – filter scale; 
\( \phi = \{0,\pi\} \) – phase offset;

\[ \begin{cases} 
  x' = x \cos \theta + y \sin \theta \\
  y' = -x \sin \theta + y \cos \theta 
\end{cases} \]

– rotation matrix.

Receptive fields of other types are formed using hyperbolic tangent.

First-layer neurons use the linear activation function

\[ U_{S1}(x,y,p) = \sum_{i,j} L_{ij} G_{p}(k' - i, l' - j) \]
Figure 2. The neural network of edge detection

First-layer neurons use the logistic sigmoid, synaptic weights \( w_{x,y} \) are learning with help backpropagation. The second layer neurons function on the winner-take-all principle within the receptive field. This determines the orientation of the selected edge or region

\[
U_{C2}(x, y, \theta, p) = \max_{\theta} \left( \frac{1}{1 + \exp(\sum_{x,y} U_{S1}(x,y,p)w_{x,y})} \right), \quad (4)
\]

Selecting edges and regions works with standalone images. Motion detection has space-time organization of neural network. Motion, in this case, is the sequential activation of several edge detection neurons in some neighborhood over time, i.e. with frame change. Thus, it is possible to learn the direction of motion \( \alpha \) and its velocity \( v \). The MT-neuron, as well as previous neurons, is created for the corresponding type of receptive field. Its connected edge detection neurons determine the receptive field of the MT-neuron. For linear motion detecting, edge detection neurons arranged in series in a direction \( \alpha \) are connecting to the MT-neuron \( U_{MT}^{(l)} \). To rotation detect neurons located at the same point but with different orientation \( \theta \) are attached to the MT-neuron \( U_{MT}^{(r)} \).

\[
U_{MT}^{(l)}(x, y, p, v, \alpha) = \sum_{x,y,t} U_{C2}(x, y, \theta, p) \ast w_{xy}(t),
\]

\[
U_{MT}^{(r)}(x, y, p, v, \alpha) = \sum_{\theta,t} U_{C2}(x_0, y_0, \theta, p) \ast w_{\theta}(t),
\]

\[
w_{xy}(t) = \exp \left( -\frac{x^2 + y^2 \gamma^2}{2\sigma^2} \right) \ast \exp \left( -\frac{t^2}{2\sigma^2} \right),
\]

\[
w_{\theta}(t) = \exp \left( -\frac{\theta^2}{2\sigma^2} \right) \ast \exp \left( -\frac{t^2}{2\sigma^2} \right).
\]

The motion is determined by calculating the weighted sum of the \( U_{C2} \) signals sequentially activated in the direction \( \alpha \) within the receptive filed.

Figure 3 shows the operation of such a neuron with the example of linear motion extraction.
The speed of movement is monitored through activation of neurons at times. If neurons in the direction are activated conditionally after 2, it is possible to talk about a higher speed of movement.

In the initial stage, motion neurons allow for highlighting minor displacement of lines and regions. Sequentially, from layer to layer, one can talk about building a general motion map. Based on this, the classification of gestures in dynamics, that is, the class becomes not a static set of characteristics, but the trajectory of segments.

4. Conclusion

The current situation in Russia and the world on work in the field of development of systems of computer translation of sign languages of the deaf is considered. One of the most important unresolved tasks is shown to be the task of recognizing the gestures of the deaf. To effectively solve it, an approach based on the development of bio-inspired neural networks is proposed. The architecture of a bio-inspired neural network, including four types of neurons is described. Motion selection begins with boundary selection. The projected neural network also receives information from boundary allocation neurons based on Gabor filters and hyperbolic tangent [7]. The L* frame channel in the CIE L*a*b* space is used as inputs for edge and motion detection.

A two-layer network is used for highlighting edges. Lines of a specific orientation are detected on the first layer. The second layer is responsible for detecting combinations of lines, including angles. In the future, it is assumed to use a bio-inspired neural network in the system of computer sign translation of Russian and Indian sign languages.

Acknowledgements

This paper was financially supported by the Russian Foundation for Basic Research (Grant No. 19-57-45006).

References

[1] Grif M G, Prihodko A L 2018 Approach to the Sign language gesture recognition framework based on HamNoSys analysis. In Proc. of Actual problems of electronic instrument engineering (APEIE–2018). DOI:1109/APEIE.2018.8545086.

[2] Heeger D J 1987 Model for the extraction of image flow. Journal of the Optical Society of America. A, Optics and image science. 4(8) 1455-1471 DOI:10.1364/JOSAA.4.001455.

[3] Simoncelli E P, Heeger D J 1998 A model of neuronal responses in visual area MT Vision Research 38(5) 743-761. DOI: 10.1016/S0042-6989(97)00183-1.

[4] Kendon A 1986 Current Issues in the Study of Gesture. The Biological Foundation of Gestures: Motor and Semiotic Aspects (Lawrence Erlbaum Associate) pp. 23-47.

[5] Ronchetti, F. et al. LSA64: A Dataset of Argentinian Sign Language. XXII Congreso Argentino de Ciencias de la Computación (CACIC), 2016.

[6] Chessa M, Sabatini S P, Solari F 2016 A systematic analysis of a V1–MT neural model for motion estimation. Neurocomputing 1731 811–1823 DOI: 10.1016/j.neucom.2015.08.091.
[7] Kugaevskikh A V, Sogreshilin A A 2019 Analyzing the efficiency of segment boundary detection using neural networks. *Optoelectronics Instrumentation and Data Processing*. 55(4) 414-422. DOI: 10.3103/S8756699019040137.