MATRIX FEATURE VECTORS IN SPEECH AND GESTURE RECOGNITION

Volodymyr Donchenko, Andrew Golik

Abstract: *This paper draws parallels between speech recognition of one speaker on a limited set of words and recognition of tactile sign language. The paper also provides variant of formation of feature vectors in matrix form for both problems. It is suggested to use ellipsoidal and orthogonal compliance distances.*

Keywords: *speech and gesture recognition, orthogonal projectors, ellipsoidal distance, pseudoinverse, SVD – decomposition.*

ACM Classification Keywords: *I.2 Artificial Intelligence, I.4 Image Processing and Computer Vision, I.5 Pattern Recognition, G.1.3 Numerical Linear Algebra.*

Introduction

Among many researches that are considered within such direction as Computer Vision, problems of speech and gesture recognition take an important place. This article draws parallels between these tasks with further highlighting of sub-problems that are similar for both tasks and can be solved using the same tools and approaches. Specific cases of the above mentioned tasks were chosen for implementation and testing: speech recognition of one speaker on a limited set of words and finger recognition of sign language.

Consideration of matrices as "natural" feature vectors: representatives of the object which are analyzed, is suggested in the article. Such approach is natural, when speech signals or gestures are analyzed. Speech signals can be exhaustively presented by spectrogram. Gestures are presented with images (or sequence of images) that in early stages of processing of input data are captured from a webcam or other recording device.

Recognition algorithms that use both standard notion of compliance distances and so-called "ellipsoidal distance" are suggested. Ellipsoidal distances are based on a "minimal ellipse" that "covers" learning sample of class. Furthermore, this paper suggests another alternative, namely, usage of orthogonal distances that are based on Cartesian grouping operators and orthogonal projectors.

Clustering with usage of compliance distances that are based on pseudoinverse and SVD-decomposition can be successfully applied to numeric vectors. However, there is urgent case, when learning sample consists not from vectors but matrices. One of the main purposes of the research is to transfer properties of pseudoinverse and SVD-decomposition to the space of matrix feature vectors.

Recognition algorithm that is named "comparison with an etalon" is considered. Results of recognition programs that were implemented in Java + CUDA and C # + EmguCV environments justify an introduction of mentioned approaches, especially, compliance distances that are based on orthogonal projectors.

Results demonstrate effectiveness of the proposed compliance distances and emphasize prospectivity of further work on their improvement.

Spectrogram as representative of a speech signal in mathematical model

Representation of audio signals as a mixture of harmonics is a standard in analysis of speech signals. Standard mathematical apparatus which is used for realization of this approach for digital processing is the discrete Fourier transform in a form that is denoted by the term fast Fourier transform (FFT).

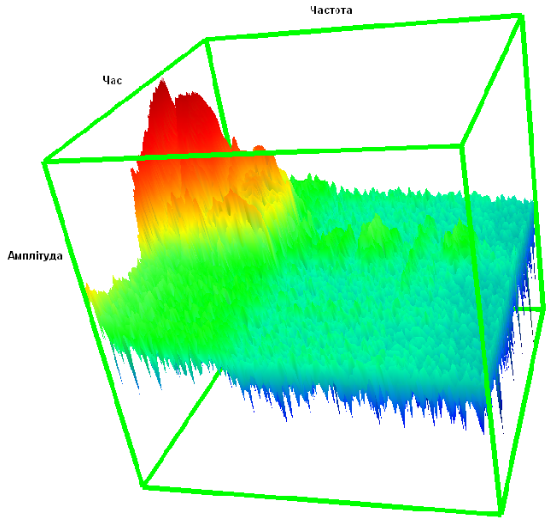

Figure 1. Three dimensional representation of spectrogram

According to this approach, a numeric vector, which represents all discrete audio signal or part of it, is associated with a vector of the same length, but with complex components - FFT of a signal. FFT can be described in standard way by a vector of modules of corresponding components or by a vector of squares of modules of these components and vector of phases of corresponding components. A vector of squares of modules of FFT components, which is bound to a discrete set of frequencies which is specified by FFT, is called an energy spectrum of an analyzed signal.

In practice of usage of harmonic analysis instead of considering FFT of the whole signal which corresponds to a spoken word, a set of parts of FFT in a moving window of fixed length which moves with a certain step is used. In accordance with this, matrix of FFT of a signal in each of the windows is associated with researched audio signal. Matrix that is composed of vectors of an energy spectrum for each of the windows is called spectrogram of a signal.

Available matrix of spectrogram which describes a signal with an accuracy of phase characteristics, in fact, is a natural representative of an audio signal in mathematical description of such signal. Therefore, a question of construction of effective speech recognition algorithms is the question of how the matrix should be used to identify words.

Contour as representative of a gesture

First stage of gesture recognition problem consists of capturing images from a webcam or other recording device, followed by finding and highlighting on the resulting image hand and its contour. This contour gives fairly complete information that can be used for gesture identification.

Figure 2. Image of gesture that is captured from a webcam. Contour of hand is found and highlighted

This information cannot be considered as exhaustive, since in most cases it does not include data about positions of individual fingers. Of course, for open hand it is not a problem, but if it is partially closed, that is a typical situation for gesture recognition; it is really tricky to get the data. In general, this problem can be partially solved by applying algorithms of skeletonization. Solution for a problem of gesture recognition by analyzing a contour of hand on an image is proposed in the article. There are several ways to analyze a contour of hand, for example, as series of interrelated points. In addition, there are a number of numerical characteristics that can be calculated for the contour: moments, Freeman chains etc. We are going to talk about representation of gesture contour in matrix form. Transition to matrix form begins with finding the smallest rectangle covering a contour of hand on an image.

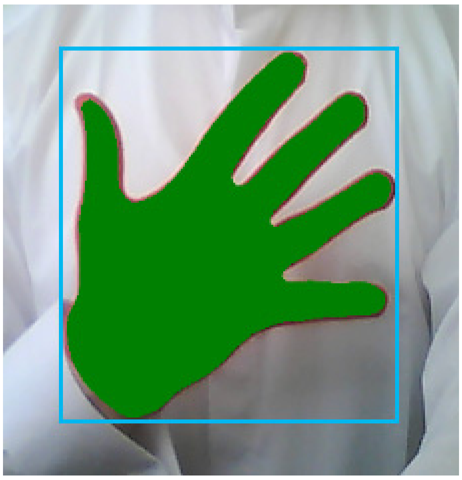

Figure 3. The smallest rectangle covering a contour of hand

Having its coordinates, we can cut it from an image and convert into binary matrix. Conversion process will be described in detail in the next section of this article.

"Characteristic" matrices

Identification of words in speech recognition can be accomplished both by constructing certain numerical feature vectors and direct analysis of spectrogram in matrix form. A similar statement holds for gesture recognition, because, as mentioned in the previous section, a contour of gesture can also be presented in a form of matrix. However, in both cases, standardization problem of dimension of such matrices is urgent. Possible solution of this problem is a construction of characteristic matrix: capturing images (two-dimensional, color) of spectrogram (or contour of hand, respectively) and its subsequent compression or stretching to standard size with conversion into matrix form according to certain rules.

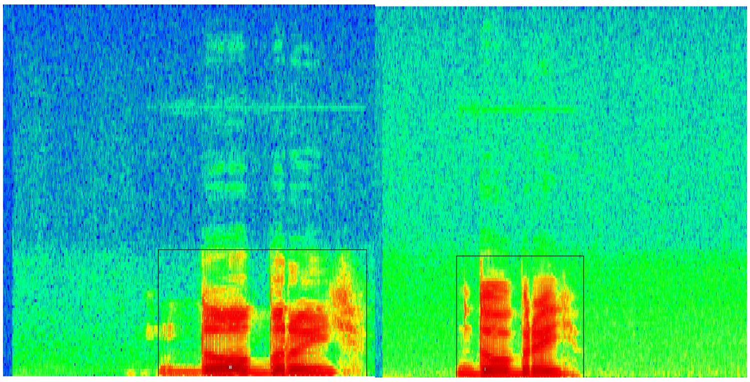

Figure 4. Examples of minimal covering rectangles that are built for an image of spectrogram

Note, that you can see similarity in sub problems for speech recognition and gesture recognition tasks, namely, a fact that at this stage a question of standardization of images - a matrix of spectrogram and a minimum rectangle covering a contour - has a place with subsequent conversion into matrix form.

The proposed approach can be called "a method of characteristic matrices", i.e. signal analysis based on matrices that are constructed in a certain way for spectrogram as an image (or an image of minimal rectangle that covers a contour) using formats of image converting. These formats include, in particular, scaling, stretching etc.

A process of formation of characteristic matrix for speech recognition problem begins with construction of two-dimensional color spectrogram. For the purpose of allocation of informative component on an image we separate rectangle that covers a part of an image that corresponds to a spoken word. Search of similar rectangle for gesture recognition problem was mentioned in the previous section. This approach allows discarding everything before and after spoken word, leaving only those frequencies which are expedient for analyzing. Regarding an image of hand it keeps only an area that covers a contour. A part of an image that falls into the rectangle is being saved for further processing. Note that size of mentioned rectangles for the same words spoken by different narrators, at different paces, with different intonation will be different. Minimal covering rectangles for an image of gesture will have different sizes too, because they depend on many factors: size of hand of a person, a distance from hand to recording device, etc. Therefore, there is a problem of standardization, which can be solved using image scaling. In other words, size of the saved image is being converted to certain standard values of height and width. Applying of smoothing and other operations on images also can be appropriate.

However, there is a difference in the technique of scaling for speech recognition and gesture recognition problems. When it comes to scaling of axis of time for an image of speech signal, it is sufficient to simply compress an image or stretch it, as width of the whole image corresponds to speed and pronunciation of narrators, so such scaling will not cause losing of information. If we consider a gesture recognition problem, more complex variant of scaling is required.

An example that clearly demonstrates a need for changes in the above mentioned algorithm of standardization is shown in Figure 5.

There are 3 parts of Figure 6: the first - minimal rectangle covering a gesture, the other two - variants of its standardization. Suppose that square of certain size was chosen as a standard. In this case, after stretching an image we will get results that are presented in the second part of Figure 6. It is not difficult to see that in this case an image of a gesture largely lost its informative value, because the ratio of width and height, which is important in this problem, was changed. More correct approach is illustrated in the third part of Figure 6. In this case additional empty areas were placed on the left and right from the image. The size of these areas is identical and found in such way that a resulting image conforms with the standards.

Figure 5. An example of gesture

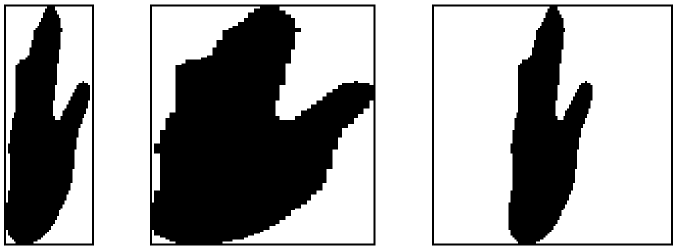

Figure 6. Different variants of standardization of an image

It is suggested to do a reverse transition from an image to the matrix on the next stage. We remind that RGB is a format of presentation of color, as a combination of red, green and blue colors. Having results of experiments we set the legitimate values of RGB, which allow to make decision: whether a pixel should be examined as meaningful or not. The transformation of an image consists of replacement of pixels which satisfy the set of legitimate values of RGB by 1 and all other by 0. Finally we get a matrix that consists of 0 and 1. The matrix can be called a "characteristic" matrix. Its image can be reproduced in a black-an-white form which is natural for binary matrices.

Figure 7. Images of different characteristic matrices for the same word

Transformations of spectrograms as matrices through the use of format transformations of images have some advantages. In particular, such approach allows solving the problem of standardization of spectrogram according to length of speech signal and obtaining suitable for analysis objects. In the same way the standardized «characteristic» matrices for the images of gestures can be used for recognition of signs of tactile language. The only difference of process of converting for this task is the circumstance that an image of gesture can be considered in a black-and-white variant and establishment of legitimate values of RGB is not a necessity because in fact every pixel simply is black or white. All points which got in a contour marked as black, all other – as whites. Therefore, a binary characteristic matrix is obtained as the result converting process.

Compliance distances: ellipsoidal distance

After forming feature vectors on the stage of clustering there is a necessity for comparison of the vectors, establishment of the so-called compliance distance between them. Possibility of usage of ellipsoidal and orthogonal distances is considered in the article.

The main feature of the mentioned distances is that while training the system, they work not with one etalon, but with a set of etalons (for the different environmental conditions).

Ellipsoidal distance is built by facilities of pseudoinverse for different variants of linear operators. Such distance leans against conception of «minimum ellipses of grouping». Actually, we talk about ellipses that «cover» each of training sets by a «minimum» and «optimum» rank. Ellipsoidal distance is built for matrices as matrices of linear operators between matrix Euclidian spaces by facilities of pseudoinverse for the mentioned spaces. They are implemented, as well as in the case of vector Euclidian spaces, through the so-called «groupings operators» of theory of pseudoinverse. Such operators are determined after the matrix of operator A that is operator between vector Euclidian spaces, and is defined by expressions:

$$R(A) = A^+ A^{+T}, R(A^T) = (A^T)^+ (A)^{T+T} = A^{+T} A^+$$

The principle role of grouping operators is that they allow us to build the «minimum ellipses of grouping»: ellipsoids which contain all vectors of set $a_k, k = \overline{1,n}$ and are optimum in certain sense. Optimum lies in following: all axis of the ellipse are formed by the orthonormal set of vectors, sum of squares of projections on which is maximal, and the squares of lengths of proper axis coincide with the proper sums of squares of projections. More precisely next four theorems have place [4].

Theorem 1 For an arbitrary set of vectors $a_k \in R^m, k = \overline{1,n}$, solution of optimization problem of search of maximum sum of squares of projections on subspace that is formed by the normalized vector $u \in R^m : \|u\| = 1$ is a vector u_1 from singularity (u_1, λ_1^2) of singular decomposition of matrix $A = (a_1 \vdots \ldots \vdots a_n)$:

$$u_1 = \arg \min_{u \in R^m : \|u\|=1} \sum_{k=1}^{r} \| \Pr_u a_k \|^2$$

$$\min_{u \in R^m : \|u\|=1} \sum_{k=1}^{r} \| \Pr_u a_k \|^2 = \lambda_1^2$$

Theorem 2 For arbitrary set of vectors $a_k \in R^m, k = \overline{1,n}$, solution of optimization problem of search of maximum sum of squares of projections on subspace that is formed by normalized vector $u \in R^m : \|u\| = 1$ is a vector u_1 from singularity (u_1, λ_1^2) of singular decomposition of matrix $A = (a_1 \vdots \ldots \vdots a_n)$:

$$u_k = \arg \min_{u \in R^m : \|u\|=1, u \perp L(u_1,\ldots,u_k)} \sum_{k=1}^{r} \| \Pr_u a_k \|^2$$

$$\min_{u \in R^m : \|u\|=1, u \perp L(u_1,\ldots,u_k)} \sum_{k=1}^{r} \| \Pr_u a_k \|^2 = \lambda_{k+1}^2$$

$$k = \overline{1, r-1},$$

where $(u_k, \lambda_k^2), k = \overline{1,r}$ as well as in the previous theorem of singularity of singular decomposition of matrix which is formed from the elements of the researched set of vectors.

Theorem 3 For arbitrary set of vectors $a_k \in R^m, k = \overline{1,n}$

$$a_k^T R(A^T) a_k \le r_{\max}^2 < r$$

$$r_{\max}^2 = \max_{k=\overline{1,n}} a_k^T R(A^T) a_k ,$$

Where, as well as in two previous theorems, A is a matrix that is formed from the vectors of a set as its columns.

Ellipsoid of theorem 3 groups the vectors of set according to the central location of the ellipse of grouping: based on an ellipse which has center at origin. In practical applications center of ellipse is mean value $\bar{a}$ of elements from the set:

$$\bar{a} = \frac{1}{n}\sum a_k$$

In this case a grouping operator is built based on a matrix $\tilde{A}$ which is formed from centered average vectors from the set $\tilde{a}_k : \tilde{a}_k = a_k - \bar{a}, k = \overline{1,n}$. Consequently following theorem has place.

Theorem 4 For arbitrary set of vectors $a_k \in R^m, k = \overline{1,n}$ we have following inequalities

$$(a_k - \bar{a})^T R(\tilde{A}^T)(a_k - \bar{a}) \le \tilde{r}^2_{\max} \le r, k = \overline{1,n}$$

$$r^2_{\max} = \max_{k=\overline{1,n}} \tilde{a}_k^T R(\tilde{A}^T)\tilde{a}_k$$

As a set of vectors the training sets of classes are used $Kl_l, l = \overline{1,L}$. As compliance distances (namely their squares): functionals $\rho^2(x, Kl_l), x \in R^m, l = \overline{1,L}$ according to minimum value of which sorting is performed, - it is possible to use the minimum ellipses of grouping. It means that compliance distances are determined as following:

$$\rho^2(x, Kl_l) = (x - \bar{a}_l)^T \frac{R(\tilde{A}_1^T)}{\tilde{r}^2_{1\max}}(x - \bar{a}_l), x \in R^m,\ l = \overline{1,L}$$

Such ellipsoidal distance is used for characteristic matrices.

Compliance distances: orthogonal distance

Together with ellipsoidal compliance distance orthogonal distance is offered in the article. It gives ability to carry properties of pseudoinverse and SVD– decomposition in case of matrix feature vectors.

$R^{(m\times n),K}$ is Euclidian space $m \times n$ of matrix corteges of length K $\alpha = (A_1 \vdots \ldots \vdots A_K) \in R^{(m\times n),K}$ with «natural» component-wise scalar multiplication:

$$(\alpha, \beta) = \sum_{k=1}^{K}(A_k, B_k)_{tr} = \sum_{k=1}^{K} trA_k^T B_k$$

$$\alpha = (A_1 \vdots \ldots \vdots A_K), \beta = (B_1 \vdots \ldots \vdots B_K) \in R^{(m\times n),K}$$

$\wp_\alpha : R^K \to R^{m\times n}$ linear operator between corresponding Euclidian spaces, that is set by a matrix cortege $\alpha = (A_1 \vdots \ldots \vdots A_K) \in R^{(m\times n),K}$ and determined by matrix cortege operations according to expression:

$$\wp_\alpha y = \sum_{k=1}^{K} y_k A_k, \alpha = (A_1 \vdots \ldots \vdots A_K) \in R^{(m\times n),K}, y = \begin{pmatrix} y_1 \\ \cdots \\ y_K \end{pmatrix} \in R^K$$

Theorem 5 [5] Conjugate $\wp_\alpha^*$ of the operator $\wp_\alpha : R^K \to R^{m\times n}$ is a linear operator, which obviously, operates in reverse to $\wp_\alpha$ direction: $\wp_\alpha^* : R^{m\times n} \to R^K$ and is determined by expression:

$$\wp_\alpha^* X = \begin{pmatrix} trA_1^T X \\ \cdots \\ trA_K^T X \end{pmatrix}$$

Proof

Indeed,

$$\left(\wp_\alpha y, X\right)_{tr} = \left(\sum_{k=1}^{K} y_k A_k, X\right)_{tr} = \sum_{k=1}^{K} y_k \left(A_k, X\right)_{tr} = \sum_{k=1}^{K} y_k \left(trA_k^T X\right) = \left(y, \begin{pmatrix} trA_1^T X \\ \cdots \\ trA_K^T X \end{pmatrix}\right)$$

This proves the theorem.

Theorem 6 [5] Multiplication of two operators is a linear operator $\wp_\alpha^* \wp_\alpha : R^K \to R^K$ which is given by a matrix (we will identify it with the operator), which is determined by expression:

$$\wp_\alpha^* \wp = \begin{pmatrix} trA_1^T A_1, ..., trA_1^T A_n \\ \cdots \\ trA_n^T A_1, ..., trA_n^T A_n \end{pmatrix} \quad (1)$$

Notice that matrix that is defined by expression (1) is the matrix of Gramm of elements $A_1, ..., A_K$ of matrix cortege $\alpha = (A_1 \vdots ... \vdots A_K)$, that specifies operator $\wp_\alpha$.

Proof

Indeed,

$$\wp_\alpha^* \wp_\alpha y = \wp_\alpha^* (\wp_\alpha y) = \begin{pmatrix} trA_1^T \sum_{i=1}^{n} A_i y_i \\ \cdots \\ trA_n^T \sum_{i=1}^{n} A_i y_i \end{pmatrix} = \begin{pmatrix} tr \sum_{i=1}^{n} A_1^T A_i y_i \\ \cdots \\ tr \sum_{i=1}^{n} A_n^T A_i y_i \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} trA_1^T A_i y_i \\ \cdots \\ \sum_{i=1}^{n} trA_n^T A_i y_i \end{pmatrix} = \begin{pmatrix} trA_1^T A_1, ..., trA_1^T A_n \\ \cdots \\ trA_n^T A_1, ..., trA_n^T A_n \end{pmatrix} \begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix} = (tr(A_i^T A_j))y$$

This proves the theorem.

A singular decomposition for a matrix (1) is obvious: it is symmetric and non-negatively defined matrix. It is determined by the set of singularities $(v_i, \lambda_i^2), i, j = \overline{1,r}$: by the orthonormal set of vectors $\| v_i \| = 1, v_i \perp v_j, i \neq j; i, j = \overline{1,r}; \lambda_1 > \lambda_2 > ... > \lambda_r > 0$ which are own for an operator $\wp_\alpha^* \wp_\alpha : R^K \to R^K$: $\wp_\alpha^* \wp_\alpha v_i = \lambda_i^2 v_i, i = \overline{1,r}$. Defined by singularities $(v_i, \lambda_i^2), i = \overline{1,r}$ matrices $U_i \in R^{m \times n} : U_i = \frac{1}{\lambda_i} \wp_\alpha v_i, i = \overline{1,r}$ are the elements of set of singularities $(U_i, \lambda_i^2), i = \overline{1,r}$ of the operator $\wp_\alpha \wp_\alpha^*$. Singular decomposition of cortege operator: singularities of two operators: $\wp_\alpha^* \wp_\alpha, \wp_\alpha \wp_\alpha^*$, determine the singular decomposition of operator $\wp_\alpha$.

Theorem 7 [5] (singular decomposition of cortege operator)

$$\wp_\alpha = \sum_{k=1}^{K} \lambda_k U_k v_k^T.$$

Variant of singular decomposition: taking into consideration the expression $U_i \in R^{m\times n} : U_i = \frac{1}{\lambda_i}\wp_\alpha v_i, i = \overline{1,r}$ n and its investigation, we have

$$\wp_\alpha = \sum_{k=1}^{K}\lambda_k U_k v_k^T = \sum_{k=1}^{K}\left(\wp_\alpha v_k\right)v_k^T$$

Remark of general character: the general variant of the theorem about singular decomposition is needed. This statement should touch general Euclidian spaces. It needs to be formulated for linear operators on general Euclidian spaces.

Theorem 8 [5] For an arbitrary linear operator $\wp_E : E_1 \to E_2$ on the pair of Euclidian spaces $(E_i,(,)_i), i = 1,2$ there is a set of singularities $(v_i,\lambda_l^2),(u_i,\lambda_l^2) i = \overline{1,r}, r = rank\wp_E$ of operators $\wp_E^*\wp$, $\wp\,\wp_E^*$ accordingly with the general set of own numbers $\lambda_l^2, i = \overline{1,r}$ that

$$\wp_E x = \sum_{i=1}^{r} u_i\lambda(v_i,x)_1, \wp_E^* y = \sum_{i=1}^{r} v_i\lambda(u_i,y)_2$$

In addition following expressions have place:

$$u_i = \lambda_l^{-1}\wp v_i, i = \overline{1,r}$$

$$v_i = \lambda_l^{-1}\wp_E^* u_i, i = \overline{1,r}$$

Basic operators of PDO theory are for cortege operators: a pseudoinverse by svd-decomposition.

According to svd-determination, PDO of cortege operator is set by following expression [5]:

$$\wp_\alpha^+ = \sum_{k=1}^{K}\lambda^{-1}v_k\left(U_k,\cdot\right)_{tr} = \sum_{k=1}^{K}\lambda^{-2}v_k\left(\wp_\alpha v_k,\cdot\right)_{tr}$$

The orthogonal projectors of base subspaces of operator and, accordingly, - grouping operators are determined after svd-presentation of cortege operator in standard way.

Theorem 9 Operators marked as $P(\wp_\alpha^*), P(\wp_\alpha)$ and determined by expressions:

$$P(\wp_\alpha^*) = \sum_{k=1}^{r} U_k\left(U_k,\cdot\right)_{tr}$$

$$P(\wp_\alpha) = \sum_{k=1}^{r} v_k\left(v_k,\cdot\right) = \sum_{k=1}^{r} v_k v_k^T$$

are orthogonal projectors $P_{L_{\wp_\alpha}}, P_{L_{\wp_\alpha^*}}$ on subspaces $L_{\wp_\alpha}$, $L_{\wp_\alpha^*}$ of possible values of operators $\wp_\alpha$, $\wp_\alpha^*$ accordingly:

$$P(\wp_\alpha^*) = P_{L_{\wp_\alpha}}, P(\wp_\alpha) = P_{L_{\wp_\alpha^*}}$$

These subspaces are the linear shells of the corresponding orthonormal sets:

$$L_{\wp_\alpha} = L(U_1,\ldots,U_r),\ \ L_{\wp_\alpha^*} = L(v_1,\ldots,v_r)$$

Proof

Proof is the same as in the case of linear operators between Euclidian spaces of numerical vectors: symmetry and idempotence is simply checked up for both operators. Similarly obvious are assertions that

$U_k \in L_{\wp_\alpha}$, $v_k \in L_{\wp_\alpha^*}$, and consequently from reasoning of dimension $L_{\wp_\alpha} = L(U_1,...,U_r)$, $L_{\wp_\alpha^*} = L(v_1,...,v_r)$. In addition, as follows from determination $P_{L_{\wp_\alpha}}, P_{L_{\wp_\alpha^*}}$ the last spaces are spaces of possible values for them accordingly. Finally, note, that subspace on which an orthogonal projector carries out the orthogonal projection can be described, in particular, as a space of possible values for it.

Theorem 10 Operators $Z(\wp_\alpha^*), Z(\wp_\alpha)$ which are complements to the identical operator of orthogonal projectors $P(\wp_\alpha^*), P(\wp_\alpha)$ accordingly:

$$Z(\wp_\alpha^*)X = X - P(\wp_\alpha^*)X, \quad Z(\wp_\alpha) = E_K - P(\wp_\alpha),$$

are orthogonal projectors on the kernels of operators accordingly.

Proof

Firstly, proof follows from the fact that for $\wp_\alpha^*$, $\wp_\alpha$ each of operators $Z(\wp_\alpha^*), Z(\wp_\alpha)$ is symmetric and idempotent. In addition they are orthogonal projectors on the orthogonal adding to subspaces $L_{\wp_\alpha} = L(U_1,...,U_r)$, $L_{\wp_\alpha^*} = L(v_1,...,v_r)$ accordingly. Namely, these orthogonal complements are the kernels of operators $\wp_\alpha^*, \wp_\alpha$ accordingly.

Theorem 11 Square of distance $\rho^2(X, L_{\wp_\alpha})$ from arbitrary $m \times n$ matrix X to linear subspace $L_{\wp_\alpha}$ that is the set of possible values of cortege operator $\wp_\alpha$ is given by formula:

$$\rho^2(X, L_{\wp_\alpha}) = (X, Z(\wp_\alpha^*)X)_{tr} = \| X \|^2 - \sum_{k=1}^{r} (X, U_k)_{tr}^2$$

Proof

Indeed,

$\rho^2(X, L_{\wp_\alpha}) = \| X_{L_{\wp_\alpha^\perp}} \|^2$ in decomposition $X = X_{L_{\wp_\alpha}} + X_{L_{\wp_\alpha^\perp}}$ by decomposition $R^{m\times n} = L_{\wp_\alpha} + L_{\wp_\alpha^*}$.

Obviously, $X_{L_{\wp_\alpha^\perp}} = Z(\wp_\alpha^*)X$ so:

$$\rho^2(X, L_{\wp_\alpha}) = \| X_{L_{\wp_\alpha^\perp}} \|^2 = \| Z(\wp_\alpha^*)X \|_{tr}^2 = \left(Z(\wp_\alpha^*)X, Z(\wp_\alpha^*)X_{tr} \right) = \left(X, Z(\wp_\alpha^*)Z(\wp_\alpha^*)X_{tr} \right)_{tr} = \left(X, Z(\wp_\alpha^*)X_{tr} \right)_{tr}$$

As an orthonormal set $U_i, i = \overline{1,r}$ is an orthonormal base in $L_{\wp_\alpha} = L(U_1,...,U_r)$ and $(X, U_i)_{tr}, i = \overline{1,r}$ is the co-ordinates of decomposition $X_{L_{\wp_\alpha}}$ by this orthonormal base, then $\| X_{L_{\wp_\alpha}} \|^2 = \sum_{i=1}^{r} (X, U_i)_{tr}^2$.

It remains to notice that according to the theorem of Pythagoras in an abstract variant $\| X \|^2 = \| X_{L_{\wp_\alpha}} \|^2 + \| X_{L_{\wp_\alpha^\perp}} \|^2$, and consequently:

$$\| X_{L_{\wp_\alpha^\perp}} \|^2 = \| X \|^2 - \| X_{L_{\wp_\alpha}} \|^2 = \| X \|^2 - \sum_{k=1}^{r} (X, U_k)_{tr}^2$$

The theorem is well-proven.

Theorem 12. A square of distance $\rho^2(X, L)$ of arbitrary $m \times n$ matrix X to linear subspace $L = L(A_1,...,A_K)$, which is the linear hull of set $m \times n$ matrices $A_1,...,A_K$ is determined by formula:

$$\rho^2(X,L)=\rho^2(X,L_{\wp_\alpha})=(X,Z(\wp_\alpha^*)X)_{tr}=\|X\|^2-\sum_{k=1}^{r}(X,U_k)_{tr}^2$$

for a cortege operator $\wp_\alpha$, formed by a set $A_1,\dots,A_K$: $\wp_\alpha=(A_1,\dots,A_K)$.

Proof

Proof follows from the fact that subspaces $L=L(A_1,\dots,A_K)$ and $L_{\wp_\alpha}$ coincide between itself.

Theorem 13 A square of distance $\rho^2(X,\Gamma(a,L))$ of arbitrary $m\times n$ matrix X to the hyperplane $\Gamma(\bar{a},L)$:

$$\bar{a}=\frac{1}{K}\sum_{k=1}^{K}A_k, L=L(\tilde{A}_1,\dots,\tilde{A}_K), \tilde{A}_k=A_k-\bar{a}, k=\overline{1,K},$$

formed by set of $m\times n$ matrices $A_1,\dots,A_K$ is given by the formula:

$$\rho^2(X,\Gamma(\bar{a},L))=(X-\bar{a},Z(\wp_{\tilde{\alpha}}^*)(X-\bar{a}))_{tr}=\|X-\bar{a}\|^2-\sum_{k=1}^{r}(X-\bar{a},\tilde{U}_k)_{tr}^2$$

where cortege operator $\wp_{\tilde{\alpha}}$ is determined by expression $\wp_{\tilde{\alpha}}=(\tilde{A}_1,\dots,\tilde{A}_K)$, and $\tilde{U}_i, i=\overline{1,r}$ orthonormal set of eigenmatrices of operator $\wp_{\tilde{\alpha}}^*$.

Proof

Proof is obvious because of $\rho^2(X,\Gamma(\bar{a},L))=\rho^2(X-\bar{a},L)$ and previous theorem.

Algorithm of recognition

Algorithm of recognition called «comparing with an etalon» is suggested in the article. After dictionary of words or gestures are formed, for each element from the dictionary a set of characteristic matrices is formed and kept. The matrices correspond to different records of words or images of gestures under different environmental conditions. This dictionary is used in the process of clustering. After converting initial signal into the characteristic matrix, this matrix, using one of the compliance distances considered in the article, is checked for closeness to every element from the dictionary. Element that appears to be "the nearest" in the terms of the compliance distance is accepted as a result.

Testing and results

Arbitrary set of 32 words (a dialog) was selected for testing. On the base of this set experimental researches were conducted. Testing of gesture recognition was conducted on the set of dactyls.

Specially developed program module formed a training set (base of standards) for every element from the dictionary, using characteristic matrices. Depending on system configuration, ellipsoidal or orthogonal distance was used. For comparative analysis Euclidian distance was tested too.

For implementation of programmatic part of the problem of speech recognition Java environment was chosen. It is an object-oriented programming language which has rich set of tools for development of software. In addition, Java supports CUDA- technologies and has a number of libraries, namely JCublas, JCufft, in which basic mathematical operations are implemented with application of parallel computing.

For implementation of programmatic part of task of gesture recognition C# was chosen. There is a shell of library "Open CV" for this environment, called Emgu CV. It includes a rich toolkit which allows working with the data flow that is obtained from recording device in real-time. In addition, it contains a number of functions and classes which can be effectively applied to recognition.

Although the proposed compliance distances require a subsequent study and optimization, even on current stage for the tasks of speech recognition of one narrative on the limited set of words and gesture recognition of tactile language, mathematical results that are illustrated in the applications shows the capacity, especially, in the task of gesture recognition.

Conclusion

Disadvantage of using of characteristic matrices for the problem of speech recognition lies in sensitiveness to the choice of RGB parameters, in fact, for this task converting pixel to element of a 0\1 matrix is not such a trivial task as for gesture recognition, where initial picture is black-and-white and it can be easily converted into a binary matrix. Therefore, with correct configuration and a number of additional stages of processing it is possible to achieve considerable improvement.

Prospective direction is usage of multilevel clustering, where different technics and algorithms are applied stage-by-stage. Suggested matrix feature vectors and the compliance distances can make a basis of one of such stages.

All in all in this article were considered problems of classification of speech signals and tactile language. Note, that matrices are natural representatives of objects in the mentioned tasks. Development of mathematical apparatus of pseudoinverse for analysis of such objects on the basis of theory of pseudoinverse for matrix Euclidian spaces was proposed.

Acknowledgement

The paper is published with partial support by the project ITHEA XXI of the ITHEA ISS (www.ithea.org) and the ADUIS (www.aduis.com.ua).

Bibliography

1. Golik A.O. Designing the feature vectors: implementation of speech recognition algorithms using parallel computing capabilities. Visn., Ser. Fiz.-Mat. Nayky, Kyiv Univ. im. Tarasa Shevchenka. – 2012. – N 3. – P. 141-150. (in Ukrainian).
2. Donchenko V.S., Golik A.O. Implementation of algorithms of speech recognition using parallel computing capabilities // XVI International Conference “Problems of Decision Making under Uncertainties”(PDMU -2010),– October 4-8, 2010.– Yalta, Ukraine.
3. Donchenko V.S., Golik A.O. Using of parallel computing technology in speech recognition problems // XVII International Conference “Problems of Decision Making under Uncertainties”(PDMU -2011),– May 23-27, 2011.–Skhidnytsia, Ukraine.
4. Donchenko V.S. Euclidian space of vectors and matrices: constructive methods of description of basic structures and their usage // International Journal “Information technologies & Knowledge”.- 2011.- Vol. 5.- Number 3.-P.203-216.
5. Donchenko V., Zinko T., Skotarenko F.’Feature vectors’ in grouping information problem in applied mathematics: vectors and matrices /Problems of Computer Intellectualization..-Editors Vitali Velichko, Aleksey Voloshin O, Krassimir Markov.- Kyiv,Ukrain- Sofia Bulgaria: NASU, V.M.Glushkov Institute of Cybernetics, ITHEA : 2012- p.111-124.
6. Donchenko V., Krivonos U., Omardibirova V. Basic structures of Euclidian space: methods of description and usage / Information Models of Knowledge.- Editors: Krassimir Markov, Vitalii Velichko, Oleksy Voloshin - ITHEA.- Kiev, Ukraine – Sofia, Bulgaria. Number 16.– P. 155-169.- ISBN 978-954-16-0048-1.
7. Kirichenko N.F., Donchenko V.S. Pseudoinverse in problems of grouping of information // X International Conference “System analyses and information technologies”– 20-24 May 2008 y. Kyiv. – p.86.

8. Kirichenko N.F., Donchenko V.S. Linear structures in problems of clustering and pseudoinverse // Abstracts: International Conference ‘Problems of Decision making under uncertainties (PDMU-2008)’.– May, 13-17, 2008. – Kyiv-Rivne, Ukraine. – 2008. – P.72-75.

9. Kirichenko N.F., Donchenko V.S. Hyperset clustering and pseudoinverse // Abstracts: International Conference ‘Problems of Decision making under uncertainties (PDMU-2007)’.– May, 21-25, 2007. – Chernivtsi, Ukraine.– 2007.– P.143 -144.

10. Kirichenko N.F., Donchenko V.S. Pseudoinverse in clustering // Cybernetics and System Analyses– 2007 – №4.– p.73-92.

11. Kirichenko N.F. Analytical representation of pseudoinverse matrices // Cybernetics and System Analyses - №2. –1997.– p.98-122.

12. Kirichenko N.F., Donchenko V.S. Pseudoinverse and linear structures in problems of prediction and clustering // Abstracts/ International Conference “Problems of Decision Macing under Uncertainties”(PDMU -2009),– April 27-30, 2009.–Skhidnyysia, Ukraine.– p.114.

13. Kirichenko N.F., Donchenko V.S. Pseudoinverse in clustering // Cybernetics and System Analyses- №4, 2007– p.98-122.

14. Bondhugula, V., Govindaraju, N. and Manocha, D. 2006. Fast Singular Value Decomposition on Graphics Processors. Technical report. University of North Carolina.

15. Moore E.H. On the reciprocal of the general algebraic matrix.//Bull. Amer. Math. Soc. – 26, 1920. – p. 394-395.

16. Penrose R. A generalized inverse for matrices.// Proc. Cambr. Philosophical Soc.- 51, 1955.– p.406-413.

17. William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery. 2.6 Singular Value Decomposition // Numerical Recipes in C. — 2nd edition. — Cambridge: Cambridge University Press.

Authors' Information

Volodymyr Donchenko – *professor, Taras Shevchenko National University of Kyiv, Department of System* Analysis and Decision Making Theory,

e-mail: voldon@bigmir.net

Major Fields of Scientific Research: statistical methods in mathematical modeling of objects, Hough transform.

Andrew Golik – *post-graduate student, Taras Shevchenko National University of Kyiv, Department of System Analysis and Decision Making Theory,*

e-mail: analytical@ukr.net

Major Fields of Scientific Research: computer vision (especially speech and gesture recognition), artificial intelligence, pattern recognition, algorithms.