Review Article

Journal of Sensor Networks and Data Communications

Modeling and Reconstruction of Data with Applications as Matrix of Pixels

Dariusz Jacek Jakóbczak

Department of Electronics and Computer Science, Koszalin University of Technology Sniadeckich 2, 75-453 Koszalin, Poland

***Corresponding author**
Dariusz Jacek Jakóbczak, Department of Electronics and Computer Science, Koszalin University of Technology Sniadeckich 2, 75-453 Koszalin, Poland

Submitted: 24 Nov 2021; **Accepted**: 30 Nov 2021; **Published**: 24 Dec 2021

Citation: *Dariusz Jacek Jakóbczak (2021) Modeling and Reconstruction of Data with Applications as Matrix of Pixels. J Sen Net Data Comm, 1(1): 01-08.*

Abstract
Probabilistic Features Combination method (PFC), which is proposed by the author, is the approach of multi-dimensional data modeling, extrapolation and interpolation using the set of high-dimensional feature vectors. This method is a hybridization of numerical methods and probabilistic methods. Identification of faces or fingerprints need modeling and each model of the pattern is built by a choice of multi-dimensional probability distribution function and feature combination. PFC modeling via nodes combination and parameter γ as N-dimensional probability distribution function enables data interpolation for feature vectors. Multi-dimensional data is modeled and interpolated via nodes combination and different functions as probability distribution functions for each feature treated as random variable.

Keywords: Image Retrieval, Pattern Recognition, Data Modeling, Vector Interpolation, Pfc Method, Feature Reconstruction, Probabilistic Modeling

Introduction

Multidimensional data modeling appears in science and industry. Image retrieval, data reconstruction, object identification or pattern recognition are still the open questions. The paper is dealing with these questions via modeling of high-dimensional data for applications of image segmentation in image retrieval and recognition tasks. This paper is concerned with image retrieval. Image retrieval is based on probabilistic modeling of unknown features via combination of N-dimensional probability distribution function for each feature treated as random variable. The sketch of proposed Probabilistic Features Combination (PFC) method consists of three steps: first handwritten letter or symbol must be modeled by a vector of features (N-dimensional data), then compared with unknown letter and finally there is a decision of identification. Author recognition of handwriting and signature is based on the choice of feature vectors and modeling functions. So high-dimensional data interpolation in handwriting identification is not only a pure mathematical problem but important task in pattern recognition and artificial intelligence such as: personalized handwriting recognition, automatic forensic document examination, classification of ancient manuscripts [1,2-8]. Also, writer recognition in monolingual handwritten texts is an extensive area of study and the methods independent from the language are well-seen [9-12]. Writer recognition methods in the recent years are going to various directions, also based on Hidden Markov Model or Gaussian Mixture Model [13-19]. So, hybrid soft computing is essential: no method is dealing with writer identification via N-dimensional data modeling or interpolation and multidimensional points comparing as it is presented in this paper. The paper wants to approach a problem of curve interpolation and shape modeling by characteristic points in handwriting identification [20].

Current methods apply mainly polynomial functions, for example Bernstein polynomials in Bezier curves, splines and NURBS. But Bezier curves don't represent the interpolation method and cannot be used for example in signature and handwriting modeling with characteristic points (nodes) [21]. Numerical methods for data interpolation are based on polynomial or trigonometric functions, for example Lagrange, Newton, Aitken and Hermite methods [22-24]. These methods are not sufficient for curve interpolation in the situations when the curve cannot be build by polynomials or trigonometric functions [25].

This paper presents novel Probabilistic Features Combination (PFC) method of high-dimensional interpolation in hybrid soft computing and takes up PFC method of multidimensional data modeling. The method of PFC requires information about data (image, object, curve) as the set of N-dimensional feature vectors. Proposed PFC method is applied in image retrieval and recognition tasks via different coefficients for each feature as random variable: polynomial, sinusoidal, cosinusoidal, tangent, cotangent, logarithmic, exponential, arc sin, arc cos, arc tan, arc cot or power. Modeling functions for PFC calculations are chosen individually for every task and they represent probability distribution functions of random variable $\alpha_i \in [0;1]$ for every feature i=1, 2, N-1. So, this paper wants to answer the question: how to retrieve the image using N-dimensional feature vectors and to recognize a handwritten letter or symbol by a set of high-dimensional nodes via hybrid soft computing?

Hybrid Multidimensional Modeling of Feature Vectors

The method of PFC is computing (interpolating) unknown (unclear, noised or destroyed) values of features between two successive nodes (N-dimensional vectors of features) using hybridization of probabilistic methods and numerical methods. Calculated values (unknown or noised features such as coordinates, colors, textures or any coefficients of pixels, voxels and doxels or image parameters) are interpolated and parameterized for real number $\epsilon i \in [0;1]$ (i = 1, 2, N-1) between two successive values of feature. PFC method uses the combinations of nodes (N-dimensional feature vectors) $p_1=(x_1, y_1, z_1)$, $p_2=(x_2, y_2, z_2)$, $p_n=(x_n, y_n, z_n)$ as $h(p_1, p_2, \ldots, p_m)$ and m=1,2,…n to interpolate unknown value of feature (for example y) for the rest of coordinates:

$$c_1 = \alpha_1 \cdot x_k + (1-\alpha_1) \cdot x_{k+1}, \ldots\ldots$$

$$c_{N-1} = \alpha_{N-1} \cdot z_k + (1-\alpha_{N-1}) \cdot z_{k+1}, \qquad k = 1,2,\ldots n-1,$$

$$c = (c_1, \ldots, c_{N-1}), \quad \alpha = (\alpha_1, \ldots, \alpha_{N-1}),$$

$$\gamma_i = F_i(\alpha_i) \in [0;1], \quad i = 1,2,\ldots N-1$$

$$y(c) = \gamma \cdot y_k + (1-\gamma) y_{k+1} + \gamma(1-\gamma) \cdot h(p_1, p_2, \ldots, p_m) \tag{1}$$

$$\alpha_i \in [0;1], \ \gamma = F(\alpha) = F(\alpha_1, \ldots, \alpha_{N-1}) \in [0;1].$$

The basic structure of eq. (1) is built on modeling function $\gamma = F(\alpha)$ which is used for points' interpolation between the nodes. Additionally, for better reconstruction and modeling there is a factor with function $\gamma=F(\alpha)$ and nodes combination h.

Then N-1 features $c_1, \ldots, c_{N-1}$ are parameterized by $\alpha_1, \ldots, \alpha_{N-1}$ between two nodes and the last feature (for example y) is interpolated via formula (1). Of course, there can be calculated $x(c)$ or $z(c)$ using (1). Two examples of h (when N = 2) computed for MHR method [26] with good features because of orthogonal rows and columns at Hurwitz-Radon family of matrices that origins from some calculations with orthogonal matrices:

$$h(p_1, p_2) = \frac{y_1}{x_1} x_2 + \frac{y_2}{x_2} x_1 \tag{2}$$

or

$$h(p_1, p_2, p_3, p_4) = \frac{1}{x_1^2 + x_3^2}(x_1x_2y_1 + x_2x_3y_3 + x_3x_4y_1 - x_1x_4y_3) +$$

$$+ \frac{1}{x_2^2 + x_4^2}(x_1x_2y_2 + x_1x_4y_4 + x_3x_4y_2 - x_2x_3y_4).$$

The simplest nodes combination is

$$h(p_1, p_2, \ldots, p_m) = 0 \tag{3}$$

and then there is a formula of interpolation:

$$y(c) = \gamma \cdot y_i + (1-\gamma) y_{i+1}$$

Formula (1) gives the infinite number of calculations for unknown feature (determined by choice of F and h) as there is the infinite number of objects to recognize or the infinite number of images to retrieve. Nodes combination is the individual feature of each modeled data. Coefficient $\gamma = F(\alpha)$ and nodes combination h are key factors in PFC data interpolation and object modeling.

N-Dimensional Probability Distribution Functions in PFC Modeling

Unknown values of features, settled between the nodes, are computed using PFC method as in (1). The simplest way of PFC calculation means $h = 0$ and $\gamma_i = \alpha_i$ (uniform probability distribution for each random variable α_i). Then PFC represents a linear interpolation. Fig.1 is the example of curve (data) modeling when the formula is known: y=2x.

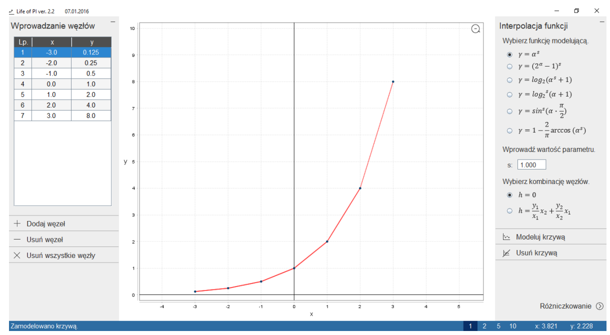

Figure 1: Example of PFC modeling for function y=2x with linear version and seven nodes (own sources).

MHR method is the example of PFC modeling for feature vector of dimension N = 2 [26]. Each interpolation requires specific distributions of random variables αi and γ in (1) depends on parameters $\alpha_i \in [0;1]$:

$$\gamma = F(\alpha),\ F:[0;1]^{N-1} \to [0;1],$$

$$F(0,\ldots,0) = 0,\ F(1,\ldots,1) = 1$$

and F is strictly monotonic for each random variable αi separately. Coefficient γ_i are calculated using appropriate function and choice of function is connected with initial requirements and data specifications. Different values of coefficients γ_i are connected with applied functions $F_i(\alpha_i)$. These functions $\gamma_i=F_i(\alpha_i)$ represent the examples of probability distribution functions for random variable $\alpha i \in [0;1]$ and real number s>0, i = 1,2,…N-1:

$$\gamma_i=\alpha_i^s,\ \gamma_i=\sin(\alpha_i^s \cdot \pi/2),\ \gamma_i=\sin^s(\alpha_i \cdot \pi/2),$$

$$\gamma_i=1-\cos(\alpha_i^s \cdot \pi/2),\ \gamma_i=1-\cos^s(\alpha_i \cdot \pi/2),$$

$$\gamma_i=\tan(\alpha_i^s \cdot \pi/4),\ \gamma_i=\tan^s(\alpha_i \cdot \pi/4),\ \gamma_i=\log_2(\alpha_i^s+1),$$

$$\gamma_i=\log_2^s(\alpha_i+1),\ \gamma_i=(2^\alpha-1)^s,\ \gamma_i=2/\pi \cdot \arcsin(\alpha_i^s),$$

$$\gamma_i=(2/\pi \cdot \arcsin\alpha_i)^s,\ \gamma_i=1-2/\pi \cdot \arccos(\alpha_i^s),$$

$$\gamma_i=1-(2/\pi \cdot \arccos\alpha_i)^s,\ \gamma_i=4/\pi \cdot \arctan(\alpha_i^s),$$

$$\gamma_i=(4/\pi \cdot \arctan\alpha_i)^s, \quad \gamma_i=ctg(\pi/2-\alpha_i^s \cdot \pi/4),$$

$$\gamma_i = ctg^s(\pi/2-\alpha_i \cdot \pi/4),\ \gamma_i = 2-4/\pi \cdot arcctg(\alpha_i^s),$$
$$\gamma_i = (2-4/\pi \cdot arcctg\alpha_i)^s$$

or any strictly monotonic function between points (0;0) and (1;1) – for example combinations of these functions.

Interpolations of function $y=2^x$ for $N=2$, $h = 0$ and $\gamma = \alpha^s$ with s = 0.8 (Fig.2) or $\gamma=log_2(\alpha+1)$ (Fig.3) are quite better then linear interpolation (Fig.1).

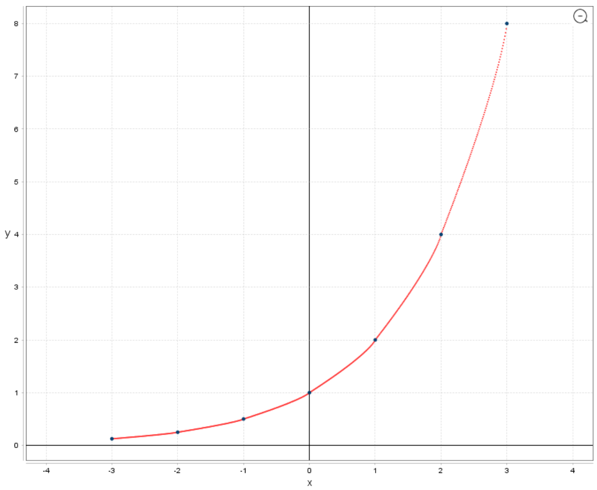

Figure 2: PFC two-dimensional modeling of function $y=2^x$ with seven nodes as Fig.1 and h=0, $\gamma=\alpha 0.8$ (own sources).

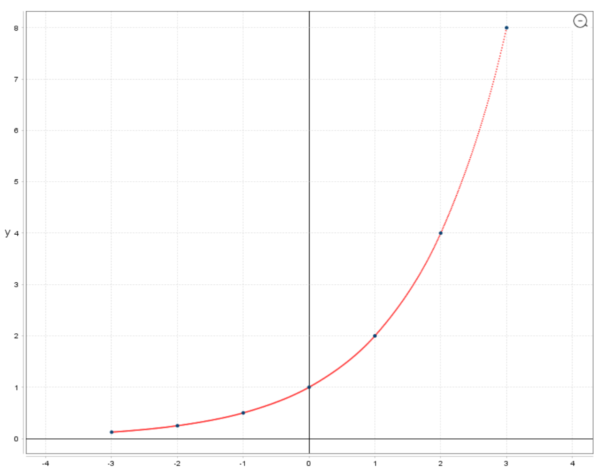

Figure 3: PFC two-dimensional reconstruction of function $y=2^x$ with seven nodes as Fig.1 and $h=0$, $\gamma=log_2(\alpha+1)$ (own sources).

Main advantage and superiority of PFC method comparing with known approaches are that there is no method connecting all these ten points below together (see Section 3):

1. Interpolation of some complicated functions using combinations of a simple function;
2. Only local changes of the curve if one node is exchanged;
3. No matter if the curve is opened or closed;
4. Data extrapolation is computed via the same formulas as interpolation;
5. Object modeling in any dimension N;
6. Curve parameterization;
7. Modeling of specific and non-typical curves: signatures, fonts, symbols, characters or handwriting (for example Fig.4-6 or 12-13);
8. Reconstruction of irregular shapes (Fig.4-6 and 12-13);
9. Applications in numerical analysis because of very precise interpolation of unknown values;
10. Even for only two nodes a curve can be modeled.

Functions γ_i are strictly monotonic for each random variable $\alpha_i \in [0;1]$ as $\gamma = F(\alpha)$ is N-dimensional probability distribution function, for example:

$$\gamma = \frac{1}{N-1}\sum_{i-1}^{N-1}\gamma_i \qquad \gamma = \prod_{i=1}^{N-1}\gamma_i$$

and every monotonic combination of γi such as
$\gamma = F(\alpha)$, $F:[0;1]^{N-1}\rightarrow[0;1]$, $F(0,\ldots,0) = 0$, $F(1,\ldots,1) = 1$.
For example when N = 3 there is a bilinear interpolation:

$$\gamma 1 = \alpha_1,\ \gamma_2 = \alpha_2,\ \gamma = ½(\alpha_1+\alpha_2) \quad (4)$$

or a bi-quadratic interpolation:

$$\gamma_1 = \alpha^{12},\ \gamma_2 = \alpha_2^{\ 2},\ \gamma = ½(\alpha_1^{\ 2}+\alpha_2^{\ 2}) \quad (5)$$

or a bi-cubic interpolation:

$$\gamma_1 = \alpha_1^{\ 3},\ \gamma_2 = \alpha_2^{\ 3},\ \gamma = ½(\alpha_1^{\ 3}+\alpha_2^{\ 3}) \quad (6)$$

or others modeling functions γ. Choice of functions γi and value s depends on the specifications of feature vectors. What is very important in PFC method: two data sets (for example a handwritten letter or signature) may have the same set of nodes (feature vectors: pixel coordinates, pressure, speed, angles) but different h or γ results in different interpolations (Fig.4-6). Here are three examples of PFC reconstruction (Fig.4-6) for N=2 and four nodes: (-1.5;-1), (1.25;3.15), (4.4;6.8) and (8;7). Formula of the curve is not given.

Algorithm of PFC retrieval, interpolation and modeling consists of five steps: first choice of nodes pi (feature vectors), then choice of nodes combination $h(p_1,p_2,\ldots,p_m)$, choice of distribution (modeling function) γ = F(α), determining values of $\alpha_i \in [0;1]$ and finally the computations (1).

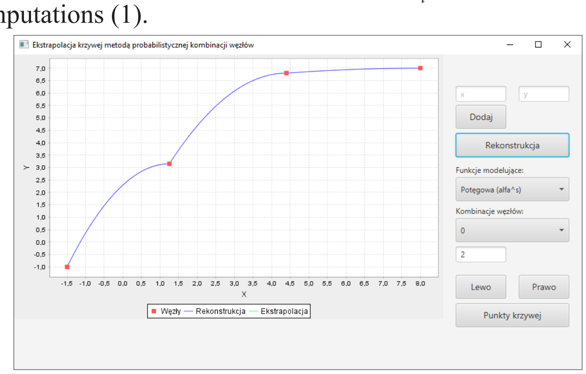

Figure 4: A curve in PFC modeling for $\gamma = \alpha^2$ and $h = 0$ (own

sources).

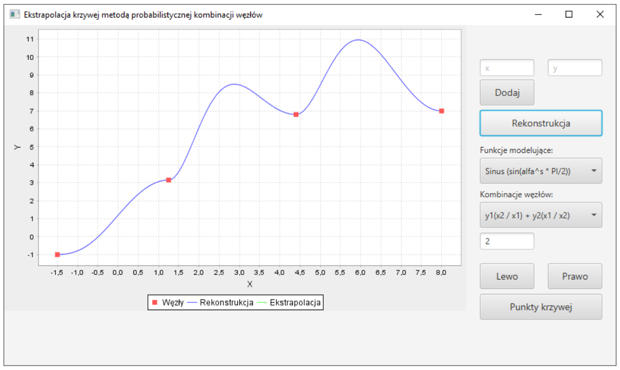

Figure 5: The example of PFC reconstruction for *γ=sin(α²•π/2)* and *h* in (2) (own sources).

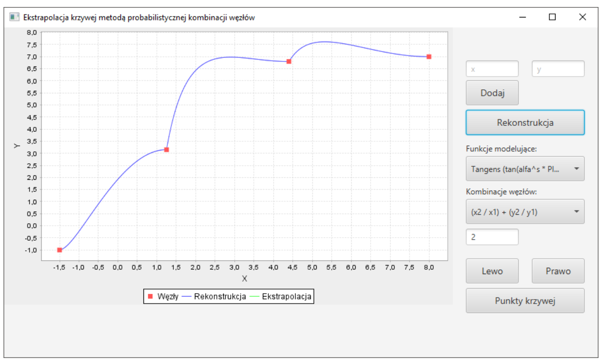

Figure 6: Data PFC interpolation for γ = tan(α2•π/4) and h=(x2/x1) + (y2/y1) (own sources).

Discussion In Details Over Pfc Approach

What are the unique features of PFC method comparing with other methods of function interpolation, curve modeling and data extrapolation? This paragraph is answering this question.

Interpolation of Some Complicated Functions using Combinations of a Simple Function

Some mathematical formulas of functions are very complicated and have very high complexity of calculations. Then there is necessity of modeling via any simple function. Of course, one can take a linear function between two nodes but this is non-effective approach. The idea of PFC formula

$$y(c) = \gamma \cdot y_k + (1-\gamma) y_{k+1} + \gamma(1-\gamma) \cdot h(p_1, p_2, ..., p_m)$$

is to calculate unknown value or coordinate as follows: take another modeling function (not linear) between two nodes. This function $\gamma = F(\alpha)$ is a probability distribution function of random variable $\alpha \in [0;1]$: for example, uniform distribution means linear interpolation when $\gamma = \alpha$. Random variable $\alpha \in [0;1]$ is a parameter for known coordinate or value between two nodes:

$$c = \alpha \cdot x_k + (1-\alpha) \cdot x_k + 1.$$

Additionally, there is nodes combination *h* for better modeling. The simplest nodes combination is $h = 0$ and then PFC formula is

$$y(c) = \gamma \cdot y_i + (1-\gamma) y_{i+1}$$

Only Local Changes of the Curve if One Node is Exchanged

Nodes combination h is responsible for the range of changes if one node is exchanged. For example h = 0 means changes between two nodes whereas

$$h(p_1, p_2, p_3, p_4) = \frac{1}{x_1^2 + x_3^2}(x_1 x_2 y_1 + x_2 x_3 y_3 + x_3 x_4 y_1 - x_1 x_4 y_3) +$$

$$\frac{1}{x_2^2 + x_4^2}(x_1 x_2 y_2 + x_1 x_4 y_4 + x_3 x_4 y_2 - x_2 x_3 y_4)$$

No Matter if the Curve is Opened or Closed

PFC formulas require the order and numbering of nodes exactly like on the curve, for example a graph of function. The only assumption for closed curve is that first node and last node are the same.

Data Extrapolation is Computed via the Same Formulas as Interpolation

Extrapolation is computed for real parameter $\alpha \epsilon [0;1]$. Then modeling function $\gamma = F(\alpha)$ has to be chosen for the situation when $\alpha < 0$ or $\alpha > 1$. Sometimes one can take a parallel version of PFC formulas:

$$y(c) = \gamma \cdot y_{k+1} + (1-\gamma) y_k + \gamma(1-\gamma) \cdot h(p_1, p_2, ..., p_m)$$

$$c = \alpha \cdot x_k + 1 + (1-\alpha) \cdot x_k.$$

when for example calculations for $\alpha < 0$ are impossible.

Object Modeling in Any Dimension N

Section four is dealing with this subject.

Curve Parameterization

Parameterization of the curve between each pair of nodes is connected with random variable α.

Modeling of Specific and Non-Typical Curves: Signatures, Fonts, Symbols, Characters Or Handwriting

Figures 4-6 and 12-13 show the examples of PFC modeling. In the individual cases one can take for each pair of nodes different functions $\gamma = F(\alpha)$ and different nodes combinations h.

Reconstruction of Irregular Shapes

Very important matter is dealing with closed curves. PFC reconstruction of the contour or shape is done with the same formulas.

Applications in Numerical Analysis Because of Very Precise Interpolation of Unknown Values

All numerical methods for numerical analysis (quadratures, derivatives, non-linear equations etc.) are based on the values of function given in the table. PFC method enables precise interpolation of the function (for example Fig.2).

Even for Only Two Nodes a Curve Can Be Modeled

Thankfully that PFC is modeling the curve between each pair of nodes, even two nodes are enough in some cases for interpolation and reconstruction.

This section was concerned on some aspects and features of PFC approach from mathematical and computational points of view.

Image Retrieval via PFC High-Dimensional Feature Reconstruction

Having monochromatic (binary) image which consists of some objects, there is only 2-dimensional feature space (x_i,y_i) – coordinates of black pixels or coordinates of white pixels. No other parameters are needed. Thus, any object can be described by a contour (closed binary curve). Binary images are attractive in processing (fast and easy) but don't include important information. If the image has grey shades, there is 3-dimensional feature space (x_i,y_i,z_i) with grey shade zi. For example, most of medical images are written in grey shades to get quite fast processing. But when there are color images (three parameters for RGB or other color systems) with textures or medical data or some parameters, then it is N-dimensional feature space. Dealing with the problem of classification learning for high-dimensional feature spaces in artificial intelligence and machine learning (for example text classification and recognition), there are some methods: decision trees, k-nearest neighbors, perceptrons, naïve Bayes or neural networks methods. All of these methods are struggling with the curse of dimensionality: the problem of having too many features. And there are many approaches to get a smaller number of features and to reduce the dimension of feature space for faster and less expensive calculations.

This paper aims at inverse problem to the curse of dimensionality: dimension N of feature space (i.e. number of features) is unchanged, but number of feature vectors (i.e. "points" in N-dimensional feature space) is reduced into the set of nodes. **So, the main problem is as follows: how to fix the set of feature vectors for the image and how to retrieve the features between the "nodes"?** This paper aims in giving the answer of this question.

Grey Scale Image Retrieval using PFC 3D Method

Binary images are just the case of 2D points (x,y): 0 or 1, black or white, so retrieval of monochromatic images is done for the closed curves (first and last node are the same) as the contours of the objects for N = 2 and examples as Fig.1-6. Grey scale images are the case of 3D points (x,y,s) with s as the shade of grey. So the grey scale between the nodes p1=(x_1,y_1,s_1) and $p_2=(x_2,y_2,s_2)$ is computed with $\gamma = F(\alpha)=F(\alpha_1,\alpha_2)$ as (1) and for example (4)-(6) or others modeling functions γi. As the simple example two successive nodes of the image are: left upper corner with coordinates $p_1=(x_1,y_1,2)$ and right down corner $p_2=(x_2,y_2,10)$. The image retrieval with the grey scale 2-10 between p1 and p2 looks as follows for a bilinear interpolation (4):

2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9	10

Figure 7: Reconstructed grey scale numbered at each pixel (own sources).

Or for other modeling functions γ_i:

2	2	2	2	2	2	2	2	2
2	3	3	3	3	3	3	3	3
2	3	4	4	4	4	4	4	4
2	3	4	5	5	5	5	5	5
2	3	4	5	6	6	6	6	6
2	3	4	5	6	7	7	7	7
2	3	4	5	6	7	8	8	8
2	3	4	5	6	7	8	9	9
2	3	4	5	6	7	8	9	10

Figure 8: Grey scale image with shades of grey retrieved at each pixel (own sources).

The feature vector of dimension N = 3 is called a voxel.

Color Image Retrieval via PFC Method

Color images in for example RGB color system (r,g,b) are the set of points (x,y,r,g,b) in a feature space of dimension $N = 5$. There can be more features, for example texture t, and then one pixel (x,y,r,g,b,t) exists in a feature space of dimension $N = 6$. But there are the sub-spaces of a feature space of dimension $N_1 < N$, for example (x,y,r), (x,y,g), (x,y,b) or (x,y,t) are points in a feature sub-space of dimension N_1=3. Reconstruction and interpolation of color coordinates or texture parameters is done like in section 3.1 for dimension $N = 3$. Appropriate combination of α_1 and α_2 leads to modeling of color r,g,b or texture t or another feature between the nodes. And for example (x,y,r,t), (x,y,g,t), (x,y,b,t)) are points in a feature sub-space of dimension N_1=4 called doxels. Appropriate combination of α_1, α_2 *and* α_3 leads to modeling of texture t or another feature between the nodes. For example color image, given as the set of doxels (x,y,r,t), is described for coordinates (x,y) via pairs (r,t) interpolated between nodes $(x_1,y_1,2,1)$ and $(x_2,y_2,10,9)$ as follows:

2,1	3,1	4,1	5,1	6,1	7,1	8,1	9,1	10,1
2,2	3,2	4,2	5,2	6,2	7,2	8,2	9,2	10,2
2,3	3,3	4,3	5,3	6,3	7,3	8,3	9,3	10,3
2,4	3,4	4,4	5,4	6,4	7,4	8,4	9,4	10,4
2,5	3,5	4,5	5,5	6,5	7,5	8,5	9,5	10,5
2,6	3,6	4,6	5,6	6,6	7,6	8,6	9,6	10,6
2,7	3,7	4,7	5,7	6,7	7,7	8,7	9,7	10,7
2,8	3,8	4,8	5,8	6,8	7,8	8,8	9,8	10,8
2,9	3,9	4,9	5,9	6,9	7,9	8,9	9,9	10,9

Figure 9: Color image with color and texture parameters (r,t) interpolated at each pixel (own sources).

So, dealing with feature space of dimension N and using PFC method there is no problem called "the curse of dimensionality" and no problem called "feature selection" because each feature is important. There is no need to reduce the dimension N and no need to establish which feature is "more important" or "less important". Every feature that depends from N_1-1 other features can be interpolated (reconstructed) in the feature sub-space of dimension $N_1 < N$ via PFC method. But having a feature space of dimension N and using PFC method there is another problem: how to reduce the number of feature vectors and how to interpolate (retrieve) the features between the known vectors (called nodes).

Difference between two given approaches (the curse of dimensionality with feature selection and PFC interpolation) can be illustrated as follows. There is a feature matrix of dimension N x M: N means the number of features (dimension of feature space) and M is the number of feature vectors (interpolation nodes) – columns are feature vectors of dimension N. One approach: the curse of dimensionality with feature selection wants to eliminate some rows from the feature matrix and to reduce dimension N to $N_1 < N$. Second approach for PFC method wants to eliminate some columns from the feature matrix and to reduce dimension M to $M_1 < M$.

Result Analysis and Conclusions

PFC method is interpolating a curve between each pair of nodes using modeling function $\gamma=F(\alpha)$ and nodes combination h. The simplest way of comparing PFC with another method is to see the example. Here is the application of PFC method for function $f(x)=1/x$ and nine nodes: $y = 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8$. PFC represents (Fig.10) much more precise interpolation than Lagrange or Newton polynomial interpolation (Fig.11).

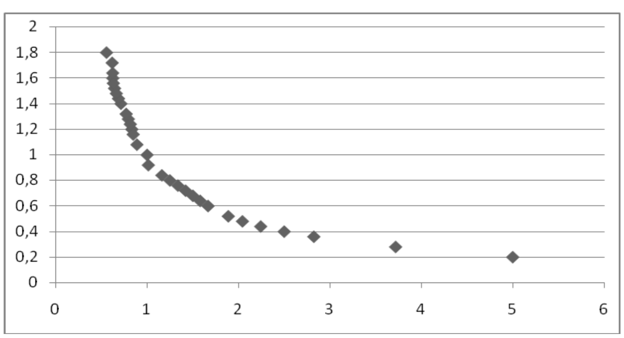

Figure 10: Points of function $f(x) = 1/x$ using PFC method with 9 nodes – better than polynomial interpolation (own sources).

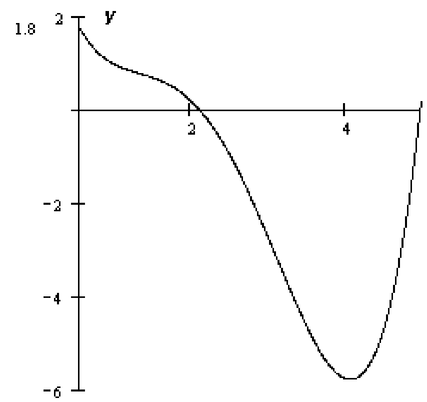

Figure 11: Interpolation polynomial of function $f(x)=1/x$ is completely wrong (own sources). Also Fig.2-3 are the examples of PFC interpolation much more accurate than polynomial interpolation by Newton or Lagrange. Very important matter is dealing with closed curves. PFC reconstruction of the contour or shape is done with the same formulas. Another important problem is connected with extrapolation. PFC method gives the tool of data anticipation or prediction.

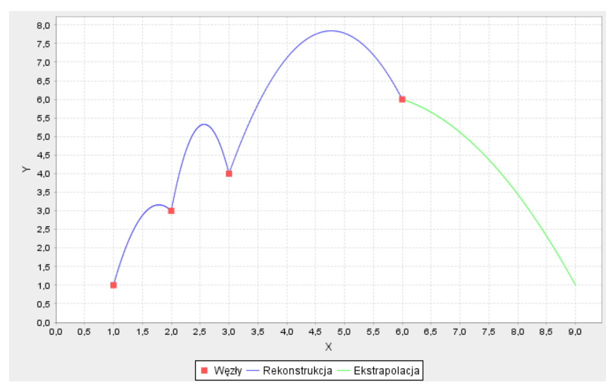

Figure 12: Extrapolation of data right of the last node (own sources).

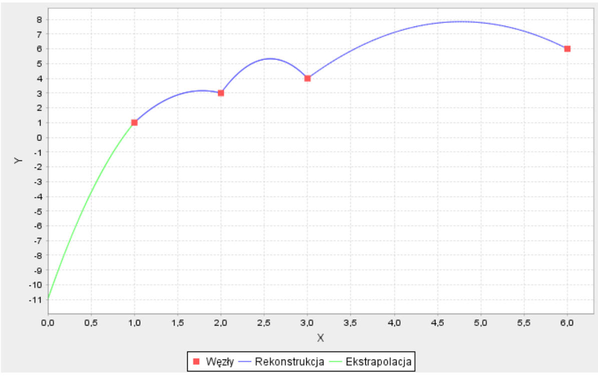

Figure 13: Prediction of values left of first node (own sources).

PFC is a novel approach to the matter of data modeling, reconstruction and extrapolation.

The method of Probabilistic Features Combination (PFC) enables interpolation and modeling of high-dimensional N data using features' combinations and different coefficients γ as modeling function. Functions for γ calculations are chosen individually at each data modeling and it is treated as N-dimensional probability distribution function: γ depends on initial requirements and features' specifications. PFC method leads to data interpolation as handwriting or signature identification and image retrieval via discrete set of feature vectors in N-dimensional feature space. So PFC method makes possible the combination of two important problems: interpolation and modeling in a matter of image retrieval or writer identification. PFC interpolation develops a linear interpolation in multidimensional feature spaces into other functions as N-dimensional probability distribution functions. Future works are going to applications of PFC method in biometric recognition, computer vision and artificial intelligence.

References

1. Marti UV, Bunke H (2002) The IAM-database: an English sentence database for offline handwriting recognition. Int J Doc Anal Recognit 5: 39-46.
2. Djeddi C, Souic Meslati L (2010) A texture-based approach for Arabic writer identification and verification. In: International Conference on Machine and Web Intelligence 2010: 115-120.
3. Djeddi C, Souici Meslati L (2011) Artificial immune recognition system for Arabic writer identification. In: International Symposium on Innovation in Information and Communication Technology 2011: 159-165.
4. Nosary A, Heutte L, Paquet T (2004) Unsupervised writer adaption applied to handwritten text recognition. Pattern Recogn. Lett 37: 385-388.
5. Van EM, Vuurpijl L, Franke K, Schomaker L (2005) The WANDA measurement tool for forensic document examination. J Forensic Doc Exam 16: 103-118.
6. Schomaker L, Franke K, Bulacu M (2007) Using codebooks of fragmented connected-component contours in forensic and historic writer identification. Pattern Recogn Lett 28: 719-727.
7. Siddiqi I, Cloppet F, Vincent N (2009) Contour based features for the classification of ancient manuscripts. In: Conference of the International Graphonomics Society 2009: 226-229.
8. Garain U, Paquet T (2009) Off-line multi-script writer identification using AR coefficients. In: International Conference on Document Analysis and Recognition 2009: 991-995.
9. Bulacu M, Schomaker L, Brink A (2007) Text-independent writer identification and verification on off-line Arabic handwriting. In: International Conference on Document Analysis and Recognition 2007: 769-773.
10. Ozaki M, Adachi Y, Ishii N (2006) Examination of effects of character size on accuracy of writer recognition by new local arc method. In: International Conference on Knowledge-based Intelligent Information and Engineering Systems 2006: 1170-1175.
11. Chen J, Lopresti D, Kavallieratou E (2010) The impact of ruling lines on writer identification. In: International Conference on Frontiers in Handwriting Recognition 2010: 439-444.
12. Chen J, Cheng W, Lopresti D (2011) Using perturbed handwriting to support writer identification in the presence of severe data constraints. In: Document Recognition and Retrieval 2011: 1-10.
13. Galloway MM (1975) Texture analysis using gray level run lengths. Comput Graphics Image Process 4: 172-179.
14. Siddiqi I, Vincent N (2010) Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recogn. Lett 43: 3853-3865.
15. Ghiasi G, Safabakhsh R (2013) Offline text-independent writer identification using codebook and efficient code extraction methods. Image and Vision Computing 31: 379-391.
16. Shahabinejad F, Rahmati M (2007) A new method for writer identification and verification based on Farsi/Arabic handwritten texts. Ninth International Conference on Document Analysis and Recognition ICDAR 2007: 829-833.
17. Schlapbach A, Bunke H (2007) A writer identification and verification system using HMM based recognizers. Pattern Anal Appl.10: 33-43.
18. Schlapbach A, Bunke H (2002) Using HMM based recognizers for writer identification and verification, 9th Int. Workshop on Frontiers in Handwriting Recognition 2004 :167-172.
19. Schlapbach A, Bunke H (2006) Off-line writer identification using Gaussian mixture models. In: International Conference on Pattern Recognition 2006: 992-995.
20. Bulacu M, Schomaker L (2007) Text-independent writer identification and verification using textural and allographic features. IEEE Trans Pattern Anal Mach Intell 29: 701-717.
21. Schumaker LL (2007) Spline Functions: Basic Theory. Cambridge Mathematical Library 2007.
22. Collins II G W (2003) Fundamental Numerical Methods and Data Analysis. Case Western Reserve University 2003.
23. Chapra SC (2012) Applied Numerical Methods McGraw-Hill 2012.
24. Ralston A, Rabinowitz P (2001) A First Course in Numerical Analysis – Second Edition. Dover Publications, New York 2001.
25. Zhang D, Lu G (2004) Review of Shape Representation and Description Techniques. Pattern Recognition 1: 1-19.
26. Jakóbczak DJ (2014) 2D Curve Modeling via the Method of

Probabilistic Nodes Combination - Shape Representation, Object Modeling and Curve Interpolation-Extrapolation with the Applications. LAP Lambert Academic Publishing, Saarbrucken 2014.