Dariusz Jacek Jakóbczak
Zakład Podstaw Informatyki i Zarządzania
Wydział Elektroniki i Informatyki
Politechnika Koszalińska

Reconstruction of High-dimensional Data using the Method of Probabilistic Features Combination

Keywords: image retrieval, pattern recognition, data modeling, vector interpolation, PFC method, feature reconstruction, probabilistic modeling

1. Introduction

The problem of multidimensional data modeling appears in many branches of science and industry. Image retrieval, data reconstruction, object identification or pattern recognition are still the open problems in artificial intelligence and computer vision. The paper is dealing with these questions via modeling of high-dimensional data for applications in 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. Handwriting and signature recognition and identification represents a significant problem. In the case of biometric writer recognition, each person is represented by the set of modeled letters or symbols. So high-dimensional data interpolation in handwriting identification [20] is not only a pure mathematical problem but important task in pattern recognition and artificial intelligence such as: biometric recognition, personalized handwriting recognition [3-5], automatic forensic document examination [6,7], classification of ancient manuscripts [8]. Also writer recognition [9] in monolingual handwritten texts is an extensive area of study and the methods independent from the language are well-seen [10-13]. Writer recognition methods in the recent years are going to various directions [14-18]: writer recognition using multi-script handwritten texts, introduction of new features, combining different types of features, studying the sensitivity of character size on writer identification, investigating writer identification in multi-script environments, impact of ruling lines on writer identification, model perturbed handwriting, methods based on run-length features, the edge-direction and edge-hinge features, a combination of codebook and visual features extracted from chain code and polygonized representation of contours, the autoregressive coefficients, codebook and efficient code extraction methods, texture analysis with Gabor filters and extracting features, using Hidden Markov Model [19] or Gaussian Mixture Model [1]. So hybrid soft computing is essential: no method is dealing with writer identification via N-dimensional data modeling or interpolation as it is presented in this paper [2]. Proposed method relies on nodes combination and functional modeling of curve points situated between the basic set of key points. The functions that are used in calculations represent whole family of elementary functions with inverse functions: polynomials, trigonometric, cyclometric, logarithmic, exponential and power function. These functions are treated as probability distribution functions in the range [0;1]. Nowadays methods apply mainly polynomial functions, for example Bernstein polynomials in Bezier curves, splines [25] 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). Numerical methods [21-23] for data interpolation are based on polynomial or trigonometric functions, for example Lagrange, Newton, Aitken and Hermite methods. These methods have some weak sides and are not sufficient for curve interpolation in the situations when the curve cannot be build by polynomials or trigonometric functions [24].

This paper presents novel Probabilistic Features Combination (PFC) method of high-dimensional interpolation and 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 $i=1,2,\ldots N-1$. So this chapter wants to answer the question: how to retrieve the image using N-dimensional feature vectors?

2. 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 $\alpha_i \in [0;1]$ ($i = 1,2,\ldots N-1$) between two successive values of feature. PFC method uses the combinations of nodes (N-dimensional feature vectors) $p_1=(x_1,y_1,\ldots,z_1)$, $p_2=(x_2,y_2,\ldots,z_2)$,…, $p_n=(x_n,y_n,\ldots z_n)$

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:

$c_1 = \alpha_1 \cdot x_k + (1-\alpha_1)\cdot x_{k+1}, \ldots\ldots c_{N-1} = \alpha_{N-1}\cdot z_k + (1-\alpha_{N-1})\cdot z_{k+1}$, $k = 1,2,\ldots n-1$,

$c = (c_1,\ldots, c_{N-1})$, $\alpha = (\alpha_1,\ldots, \alpha_{N-1})$, $\gamma_i = F_i(\alpha_i) \in [0;1]$, $i = 1,2,\ldots N-1$

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

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

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). The example 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:

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

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.

2.1. N-dimensional probability distributions in PFC modeling

Unknown values of features, settled between the nodes, are computed using PFC method as in (1). Key question is dealing with coefficient γ. The simplest way of PFC calculation means h=0 and $\gamma_i=\alpha_i$ (basic 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: $y=2^x$.

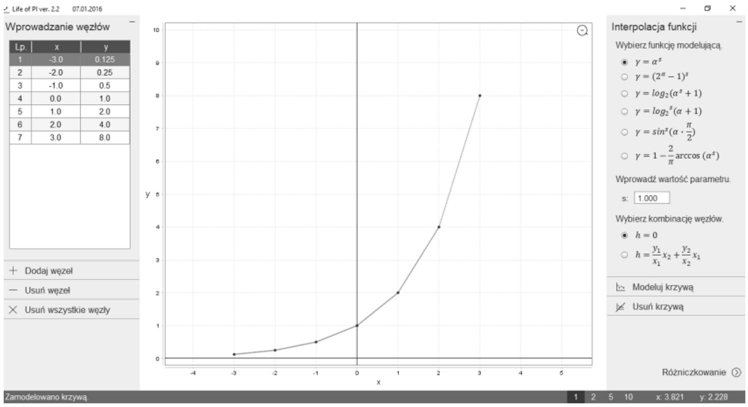

Fig. 1. PFC linear 2D modeling of function $y=2^x$ with seven nodes (in left window) and options in right window (modeling functions γ and nodes combination h).

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

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

and F is strictly monotonic for each random variable α_i 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 $s>0$, $i = 1,2,\dots N-1$:

$\gamma_i=\alpha_i^s$, $\gamma_i=sin(\alpha_i^s\cdot\pi/2)$, $\gamma_i=sin^s(\alpha_i\cdot\pi/2)$, $\gamma_i=1-cos(\alpha_i^s\cdot\pi/2)$, $\gamma_i=1-cos^s(\alpha_i\cdot\pi/2)$, $\gamma_i=tan(\alpha_i^s\cdot\pi/4)$, $\gamma_i=tan^s(\alpha_i\cdot\pi/4)$, $\gamma_i=log_2(\alpha_i^s+1)$, $\gamma_i=log_2^s(\alpha_i+1)$, $\gamma_i=(2^\alpha-1)^s$, $\gamma_i=2/\pi\cdot arcsin(\alpha_i^s)$, $\gamma_i=(2/\pi\cdot arcsin\alpha_i)^s$, $\gamma_i=1-2/\pi\cdot arccos(\alpha_i^s)$, $\gamma_i=1-(2/\pi\cdot arccos\alpha_i)^s$, $\gamma_i=4/\pi\cdot arctan(\alpha_i^s)$, $\gamma_i=(4/\pi\cdot arctan\alpha_i)^s$, $\gamma_i=ctg(\pi/2-\alpha_i^s\cdot\pi/4)$, $\gamma_i=ctg^s(\pi/2-\alpha_i\cdot\pi/4)$, $\gamma_i=2-4/\pi\cdot arcctg(\alpha_i^s)$, $\gamma_i=(2-4/\pi\cdot arcctg\alpha_i)^s$

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).

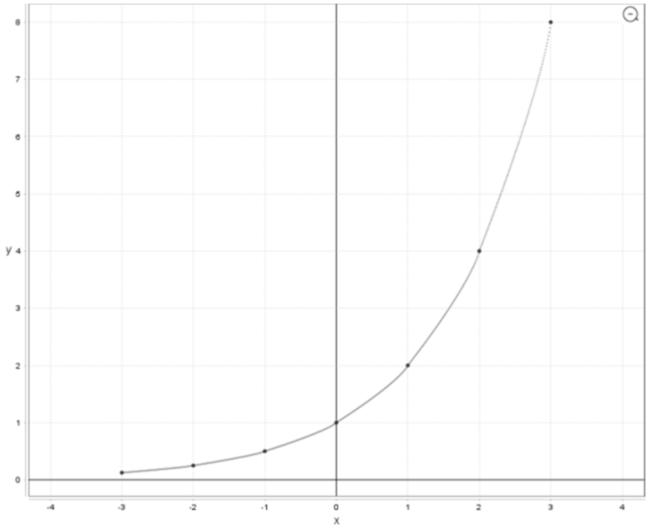

Fig. 2. PFC two-dimensional modeling of function $y=2^x$ with seven nodes as Fig.1 and h=0, $\gamma=\alpha^{0.8}$.

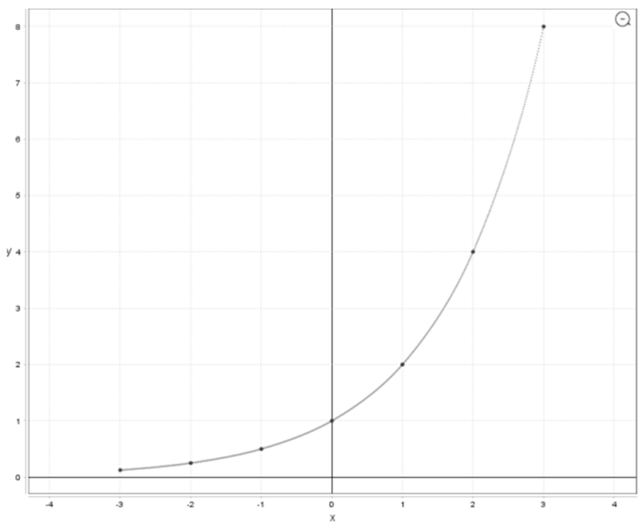

Fig. 3. PFC two-dimensional reconstruction of function $y=2^x$ with seven nodes as Fig.1 and h=0, $\gamma = log_2(\alpha+1)$.

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:

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

and every monotonic combination of γ_i such as

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

For example when $N = 3$ there is a bilinear interpolation:

$$\gamma_1 = \alpha_1\,,\ \gamma_2 = \alpha_2\,,\ \gamma = \tfrac{1}{2}(\alpha_1 + \alpha_2) \qquad (4)$$

or a bi-quadratic interpolation:

$$\gamma1 = \alpha12\,,\ \gamma2 = \alpha22\,,\ \gamma = \tfrac{1}{2}(\alpha12 + \alpha22) \qquad (5)$$

or a bi-cubic interpolation:

$$\gamma1 = \alpha13\,,\ \gamma2 = \alpha23\,,\ \gamma = \tfrac{1}{2}(\alpha13 + \alpha23) \qquad (6)$$

or others modeling functions γ. Choice of functions γ_i and value s depends on the specifications of feature vectors and individual requirements. 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 p_i (feature vectors), then choice of nodes combination $h(p_1,p_2,\ldots,p_m)$, choice of distribution (modeling function) $\gamma = F(\alpha)$, determining values of $\alpha_i \in [0;1]$ and finally the computations (1).

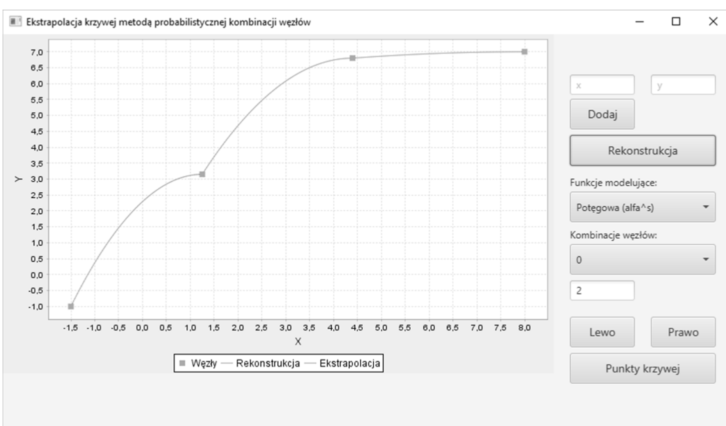

Fig. 4. PFC 2D modeling for $\gamma = \alpha^2$ and $h = 0$.

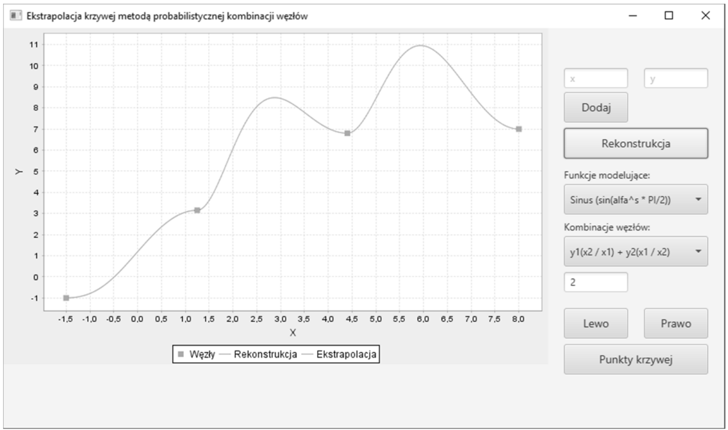

Fig. 5. PFC 2D reconstruction for $\gamma = sin(\alpha^2 \cdot \pi/2)$ and *h* in (2).

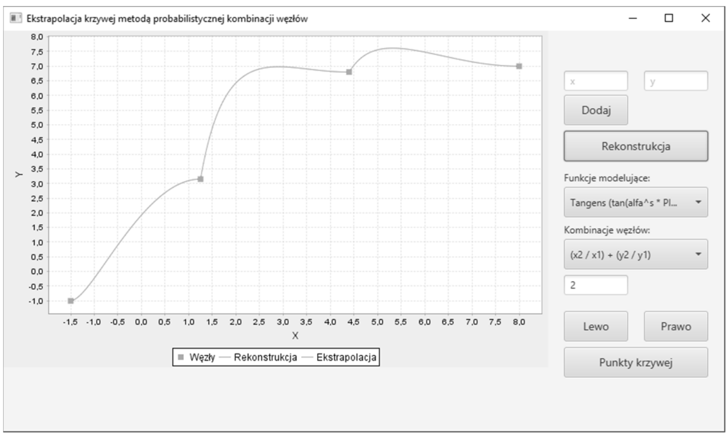

Fig. 6. PFC 2D interpolation for $\gamma = tan(\alpha^2 \cdot \pi/4)$ and $h = (x_2/x_1)+ (y_2/y_1)$.

3. Image Retrieval via PFC Reconstruction

After the process of image segmentation and during the next steps of retrieval, recognition or identification, there is a huge number of features included in N-dimensional feature vector. These vectors can be treated as “points” in N-dimensional feature space. For example in artificial intelligence there is a high-dimensional search space (the set of states that can be reached in a search problem) or hypothesis space (the set of hypothesis that can be generated by a machine learning algorithm). This paper is dealing with multidimensional feature spaces that are used in computer vision, image processing and machine learning.

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 z_i. 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 less 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.

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

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 $p_1=(x_1,y_1,s_1)$ 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 γ_i. 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 p_1 and p_2 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

Fig. 7. Reconstructed grey scale numbered at each pixel.

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

Fig. 8. Grey scale image with shades of grey retrieved at each pixel.

3.2. 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 chapter 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

Fig. 9. Color image with color and texture parameters (r,t) interpolated at each pixel.

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 (Fig.10): 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 (Fig.11) for PFC method wants to eliminate some columns from the feature matrix and to reduce dimension M to $M_1 < M$.

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

→

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

Fig. 10. The curse of dimensionality with feature selection wants to eliminate some rows from the feature matrix and to reduce dimension *N*.

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

→

2	2	2	2	2	2	2
2	3	3	3	3	3	3
2	3	4	4	4	4	4
2	3	4	5	5	5	5
2	3	4	5	6	6	6
2	3	4	5	6	7	7
2	3	4	5	6	7	8
2	3	4	5	6	7	8
2	3	4	5	6	7	8

Fig. 11. PFC method wants to eliminate some columns from the feature matrix and to reduce dimension *M*.

So after feature selection (Fig.10) there are nine feature vectors (columns): $M = 9$ in a feature sub-space of dimension $N_1 = 6 < N$ (three features are fixed as less important and reduced). But feature elimination is a very unclear matter. And what to do if every feature is denoted as meaningful and then no feature is to be reduced? For PFC method (Fig.11) there are seven feature vectors (columns): $M_1 = 7 < M$ in a feature space of dimension $N = 9$. Then no feature is eliminated and the main problem is dealing with interpolation or extrapolation of feature values, like for example image retrieval (Fig.7-9).

4. Conclusions

The method of Probabilistic Features Combination (PFC) enables interpolation and modeling of high-dimensional *N* data using features' combinations and different coefficients γ: polynomial, sinusoidal, cosinusoidal, tangent, cotangent, logarithmic, exponential, arc sin, arc cos, arc tan, arc cot or power 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. Main features of PFC method are: PFC interpolation develops a linear interpolation in multidimensional feature spaces into other functions as *N*-dimensional probability distribution functions; PFC is a generalization of MHR method and PNC method via different nodes combinations; interpolation of L points is connected with the computational cost of rank $O(L)$ as in MHR and PNC method; nodes combination and coefficients γ are crucial in the process of data probabilistic parameterization and interpolation: they are computed individually for a single feature. Future works are going to applications of PFC method in signature and handwriting biometric recognition: choice and features of nodes combinations h and coefficients γ.

References

1. Schlapbach, A., Bunke, H.: Off-line writer identification using Gaussian mixture models. In: International Conference on Pattern Recognition, pp. 992–995 (2006)
2. Bulacu, M., Schomaker, L.: Text-independent writer identification and verification using textural and allographic features. IEEE Trans. Pattern Anal. Mach. Intell. 29 (4), 701–717 (2007)
3. Djeddi, C., Souici-Meslati, L.: A texture based approach for Arabic writer identification and verification. In: International Conference on Machine and Web Intelligence, pp. 115–120 (2010)
4. Djeddi, C., Souici-Meslati, L.: Artificial immune recognition system for Arabic writer identification. In: International Symposium on Innovation in Information and Communication Technology, pp. 159–165 (2011)
5. Nosary, A., Heutte, L., Paquet, T.: Unsupervised writer adaption applied to handwritten text recognition. Pattern Recogn. Lett. 37 (2), 385–388 (2004)
6. Van, E.M., Vuurpijl, L., Franke, K., Schomaker, L.: The WANDA measurement tool for forensic document examination. J. Forensic Doc. Exam. 16, 103–118 (2005)
7. Schomaker, L., Franke, K., Bulacu, M.: Using codebooks of fragmented connected-component contours in forensic and historic writer identification. Pattern Recogn. Lett. 28 (6), 719–727 (2007)

8. Siddiqi, I., Cloppet, F., Vincent, N.: Contour based features for the classification of ancient manuscripts. In: Conference of the International Graphonomics Society, pp. 226–229 (2009)
9. Garain, U., Paquet, T.: Off-line multi-script writer identification using AR coefficients. In: International Conference on Document Analysis and Recognition, pp. 991–995 (2009)
10. Bulacu, M., Schomaker, L., Brink, A.: Text-independent writer identification and verification on off-line Arabic handwriting. In: International Conference on Document Analysis and Recognition, pp. 769–773 (2007)
11. Ozaki, M., Adachi, Y., Ishii, N.: 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, pp.1170–1175 (2006)
12. Chen, J., Lopresti, D., Kavallieratou, E.: The impact of ruling lines on writer identification. In: International Conference on Frontiers in Handwriting Recognition, pp. 439–444 (2010)
13. Chen, J., Cheng, W., Lopresti, D.: Using perturbed handwriting to support writer identification in the presence of severe data constraints. In: Document Recognition and Retrieval, pp. 1–10 (2011)
14. Galloway, M.M.: Texture analysis using gray level run lengths. Comput. Graphics Image Process. 4 (2), 172–179 (1975)
15. Siddiqi, I., Vincent, N.: Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recogn. Lett. 43 (11), 3853–3865 (2010)
16. Ghiasi, G., Safabakhsh, R.: Offline text-independent writer identification using codebook and efficient code extraction methods. Image and Vision Computing 31, 379–391 (2013)
17. Shahabinejad, F., Rahmati, M.: A new method for writer identification and verification based on Farsi/Arabic handwritten texts, Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), pp. 829–833 (2007)
18. Schlapbach, A., Bunke, H.: A writer identification and verification system using HMM based recognizers, Pattern Anal. Appl. 10, 33–43 (2007)
19. Schlapbach, A., Bunke, H.: Using HMM based recognizers for writer identification and verification, 9th Int. Workshop on Frontiers in Handwriting Recognition, pp. 167–172 (2004)
20. Marti, U.-V., Bunke, H.: The IAM-database: an English sentence database for offline handwriting recognition, Int. J. Doc. Anal. Recognit. 5, 39–46 (2002)

21. Collins II, G.W.: Fundamental Numerical Methods and Data Analysis. Case Western Reserve University (2003)
22. Chapra, S.C.: Applied Numerical Methods. McGraw-Hill (2012)
23. Ralston, A., Rabinowitz, P.: A First Course in Numerical Analysis – Second Edition. Dover Publications, New York (2001)
24. Zhang, D., Lu, G.: Review of Shape Representation and Description Techniques. Pattern Recognition 1(37), 1-19 (2004)
25. Schumaker, L.L.: Spline Functions: Basic Theory. Cambridge Mathematical Library (2007)
26. Jakóbczak, D.J.: 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)

Abstract

Proposed method, called Probabilistic Features Combination (PFC), is the method 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 parameterization and 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: polynomial, sine, cosine, tangent, cotangent, logarithm, exponent, arc sin, arc cos, arc tan, arc cot or power function.

Streszczenie

Autorska metoda Probabilistycznej Kombinacji Cech - Probabilistic Features Combination (PFC) jest wykorzystywana do interpolacji i modelowania wielowymiarowych danych. Węzły traktowane są jako punkty charakterystyczne N-wymiarowej informacji, która ma być odtwarzana (np. obraz). Wielowymiarowe dane są interpolowane lub rekonstruowane z wykorzystaniem funkcji rozkładu prawdopodobieństwa: potęgowych, wielomianowych, wykładniczych, logarytmicznych, trygonometrycznych, cyklometrycznych.