MULTISTAGE APPROACH TO IMAGE PROCESSING

Y.V. Baybak[1], *N.O. Shweiki*[1], *Y.F. Kutaev*[2], *A.A. Gertsiy*[1], *L.I. Timchenko*[1].

[1]Vinnitsa State Technical University, Khmelnitskoye Shosse 95, Vinnitsa, 21021, Ukraine,
E-mail: timchen@svitonline.com
[2]State Scientific Enterprise "Astrophisika", Volokolamskoye shosse 95, Moscow, 129820, Russia

ABSTRACT

We present a novel three-dimensional network and its application to pattern analysis. In this paper, a novel approach is proposed, which allows for an efficient reduction of the amount of visual data required for representing structural information in the image. This algorithm is tolerant to minor structural changes and can be used for automatic face recognition.

The approach is based on a multistage architecture, which investigates partial clustering of structural image components. The initial grey-scale representation of the input image is transformed into a structural representation, so that each image component contains information about the spatial structure of its neighbourhood. The output result is represented as a pattern vector, whose components are computed one at a time to allow the quickest possible response. The input pattern is identified as the best match between the output pattern vector and the model vectors from the database. This approach may be employed for the biomedical image processing, such as ultrasonics and X - ray image.

1. INTRODUCTION

Efficient representation of visual information is one the most important issues in the automatic recognition of human faces. In the case of a feature-based representation, each face is stored as a pattern vector [1]. Yet the connectionist models generally operate on image-based representations by using distributed image representations.

Local feature detectors, which are often employed in these models, make direct representations much more effective in terms of the preservation of both texture and shape information [2]. However, they put excessive requirements on the amount of memory to store all possible descriptors [3]. Most of the attempts to reduce the amount of data necessary to store faces are based either on the "winner takes all" rule [4] or on simple statistical calculations [5]. Those models do not attempt to exploit the vast redundancy of the visual representation [6], nor take into account configural information.

2. METHODOLOGY

The aim of the proposed approach is to reduce the amount of visual data whilst preserving structural information in the image. For this task, a novel three-dimensional multistage network architecture is used (Figure 1). The first dimension of the network manifests itself as the parallel channels; the second dimension shows as the ranking of the clustered image components depending on their information content. The third network dimension is specified by a flexible hierarchical structure of further processing and is illustrated in Figure 1 as processing levels. The output of the network is represented as a pattern vector.

The purpose of the processing in the multistage network is to discover the structural regularities and to represent them as a smaller number of more complex units. No data is discarded; instead, data with the higher information content is processed first. The term "information content" is based on the degree of correlation between the data components. This feature allows the computation of the output vector in stages - the first components of the vector contain the larger proportion of the structural output, and the following components contain less correlated information. At each stage, a single output component is computed. Thus, it is possible to analyse the output result after each stage.

Consider a transformation $G(M)=\{a_j \mid j=1,2,...,m\}$ of the input image M, which performs mapping of the initial set of image components into a new set of components. Possible transformations are discussed in Section 3. If the initial image is partitioned into n segments then this transformation G may be applied in each processing channel i, $i=1,...,n$, simultaneously. The i-th image segment may be denoted as $M_i^1=\{a_{ij}^1 \mid j=1,2,...,m^1\}$, where m^1 is the cardinality of a new set. The new components are graded according to their information content.

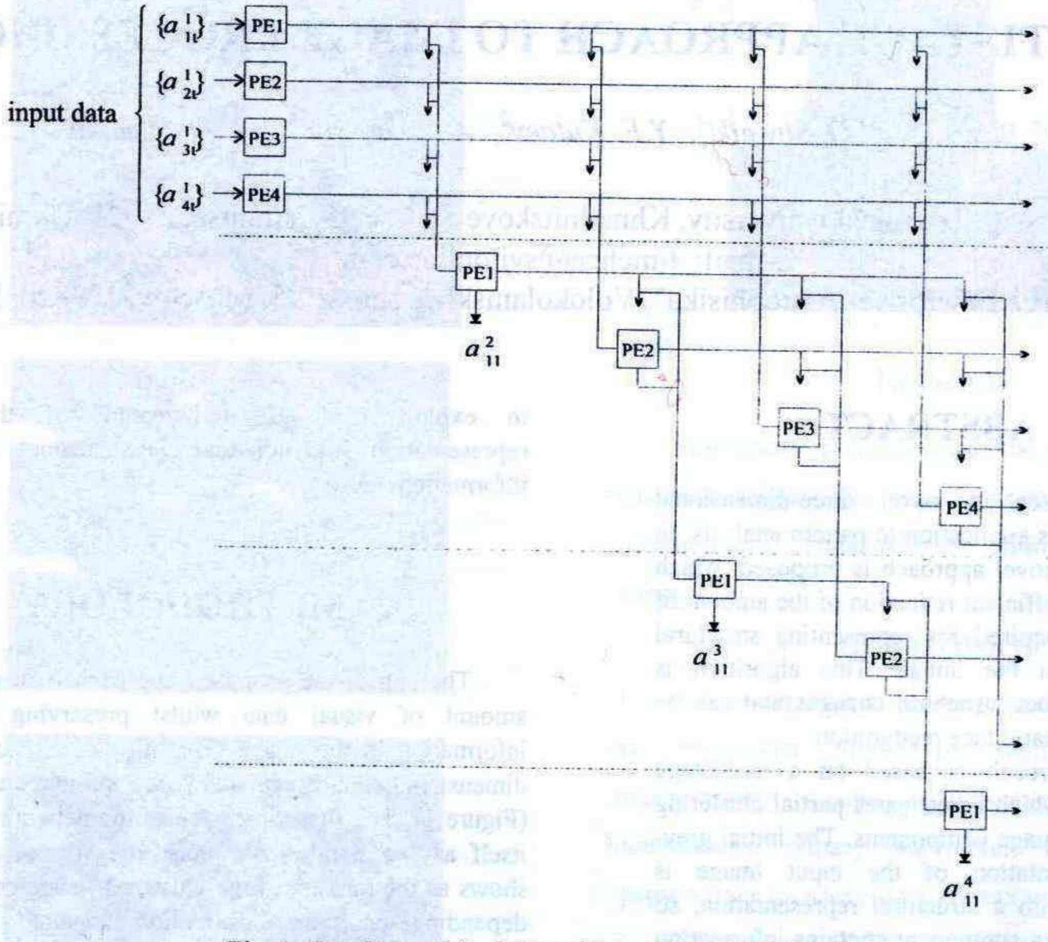

Fig. 1. Topology of the 3-D multistage hierarchical network.

They are extracted to the second processing level with respect to their grades, in order to further investigate similarities between them. At the first level, all channels can be considered as a matrix A_1 of the form:

$$A_1 = \begin{bmatrix} M_1^1 \\ M_2^1 \\ \vdots \\ M_i^1 \\ \vdots \\ M_n^1 \end{bmatrix} = \begin{bmatrix} a_{11}^1 & a_{12}^1 & \cdots & a_{1m}^1 \\ a_{21}^1 & a_{22}^1 & & a_{2m}^1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1}^1 & a_{i2}^1 & & a_{im}^1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}^1 & a_{n2}^1 & \cdots & a_{nm}^1 \end{bmatrix}.$$

The column index indicates the stage of the component extraction and the row index indicates the channel from which that component is extracted. In algebraic form transmission of the decomposed data A_1 to the second level is represented as transposition T, $A_1^T = T(A_1)$.

On the second level the components are placed into the processing channels according to their grades on the first level. The transformation G is applied again, this time to the rows of A_1^T.

$$A_2 = G(A_1^T) = G\Big(T\big(G(M)\big)\Big) = \begin{bmatrix} a_{11}^2 & a_{12}^2 & \cdots & a_{1m^2}^2 \\ a_{21}^2 & a_{22}^2 & \cdots & a_{2m^2}^2 \\ \vdots & & & \vdots \\ a_{n^2 1}^2 & a_{n^2 2}^2 & \cdots & a_{n^2 m^2}^2 \end{bmatrix}$$

where $n^2 = m^1$ is the number of parallel channels at the second level, and m^2 denotes the number of components (processing stages) at the second level.

The matrix A_2 represents the second processing level after the transformation in the way that each component a_{ij}^2 is extracted at the stage $t=i+j-1$. If the matrix A_2 is aligned so that the column index illustrates the stage of the component extraction, the same processing procedure can be applied on each following level of the hierarchy. This alignment is implemented as horizontal shift of each row of A_2 in order to position the first nonempty element on the leading diagonal. The resultant matrix is denoted A_2', and $A_2' = P(A_2)$, where P is the aligning transformation.

$$A_2' = \begin{bmatrix} a_{11}^2 & a_{12}^2 & \cdots & \cdots & a_{1m^2}^2 & x & x & \cdots & x \\ x & a_{21}^2 & a_{22}^2 & \cdots & a_{2m^2-1}^2 & a_{2m^2}^2 & x & \cdots & x \\ x & x & a_{31}^2 & \cdots & a_{3m^2-2}^2 & a_{3m^2-1} & a_{3m^2}^2 & x & x \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x & \cdots & \cdots & x & x & a_{n^21}^2 & a_{n^22}^2 & \cdots & a_{n^2m^2}^2 \end{bmatrix}$$

where x denotes empty positions.

The component a_{11}^2 is the only one to be extracted at the first stage of the processing at the second level. Therefore, it is unrelated with all the other second-level components in time and effectively is an output component. It represents an intermediate result of the processing, and is the only component extracted as a result of the multistage processing at the second level. This component is extracted from the matrix by removing the first matrix column.

The remaining clustered components are stored in the matrix A_2''. This extraction is performed by applying a transformation L, that is, $L(A_2') = a_{11}^2 + A_2''$. The initial matrix of the third level A_2^T is obtained after transposition T of the A_2'', that is $A_2^T = T\big(L\big(P(A_2)\big)\big)$.

The described procedure is appled at each of the following hierarchical levels until the $k_{\max}$-th level, where the matrix $A_{k_{\max}}$ contains a single element. Then $T\big(L\big(P(A_{k_{\max}})\big)\big) = \varnothing$.

3. SYSTEM PERFORMANCE

Several transformations can be applied in the multistage network. Some possible examples include principal component analysis of local image patches [7], partial clusterisation of the Gabor coefficients [8], and the cosine transform. In this paper, analysis of the facial structure is performed using partial clusterisation of pixel connectivity. A three-level representation [9] of grey-scale facial image is used to specify a three-level pixel connectivity The three connectivity levels are then processed using three multistage networks, each producing a pattern vector (Figure 2).

In each network, the image components are clustered using the ISODATA algorithm [10]. Then, the newly organised clusters are graded on the basis of the uniformity of data in those clusters.

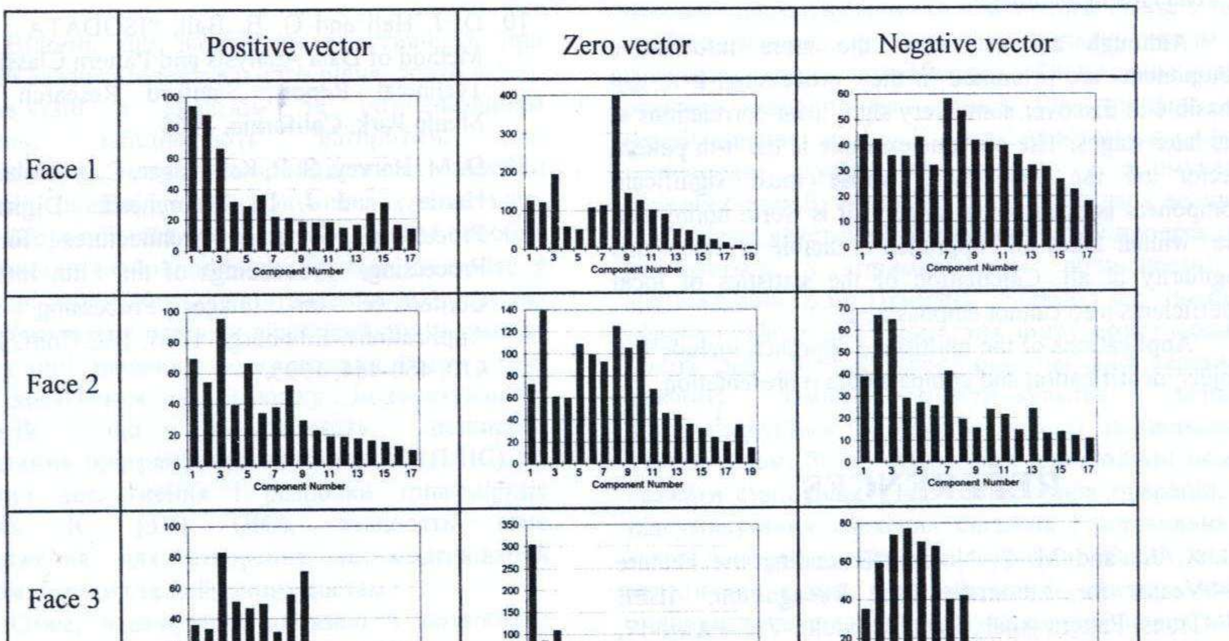

Fig. 2. Pattern vectors generated by the multistage algorithm for three faces.

Components with equal grades are extracted from different channels in order to be processed in the same way at the higher hierarchical levels. At each processing level, starting from the second one, an output component is extracted. Thus, analysis of the output can be performed after the early stages of processing.

The output pattern vectors highlight different structural organisation of each particular face. At the same time, minor structural changes do not affect the first vector components, thus making the algorithm tolerant to slight changes in facial expression.

In the current realisation, the system may be viewed as performing a multistage sorting of the changing data. Even if the components are not correlated at the low levels of the multistage network, they can be found correlated at some higher levels. Any correlations affect the values of the remaining data and influence the output result.

4. DISCUSSION AND CONCLUSIONS

There are several factors, which contribute to the fast speed of this multistage processing:

1. Image data is segmented and fed into a number of parallel channels. Components from different segments may be only processed together at the higher levels, if they are assigned the same rank. Thus, the number of possible combinations is significantly reduced.
2. Calculations in the channels can be performed in parallel, for example, on our custom TMS320C40 image processing computer [11];
3. Components of the output vector are extracted one at a time, with priority given to the more informative ones. Thus, the algorithm is robust to minor structural changes in the image.

Although at each level the more informative components are promoted in their processing, it is still possible to discover some very significant correlations at the later stages. The obvious example is the 0-th pattern vector of the first face, whose most significant component is extracted at stage 8. It is worth noting that the "winner takes all" approach is unable to detect this regularity at all. Calculation of the statistics of local coefficients also cannot emphasize it.

Applications of the multistage approach include fast object identification and compact data representation.

REFERENCES

1. X. Jia and M. S. Nixon, "Extending the Feature Vector for Automatic Face Recognition," IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-17, No. 12, December 1995, pp. 1167-1176.
2. J. G. Daugman, "Uncertainty Relations for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters," J. Optical Soc. Amer. A2, Vol. 7, pp. 1160-1169, 1985.
3. M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. v. d. Malsburg, R. P. Wurtz and W. Kohen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," IEEE Trans. Computers, Vol. 3, pp. 300-311, 1993
4. N. Petkov, "Biologically Motivated Computationally Intensive Approaches to Image Pattern Recognition," Future Generation Computer Systems, Vol. 11, pp. 451-465, 1995.
5. J F. Goudail, E. Lange, T. Iwamoto, K. Kuyma, and N. Otsu, "Fast Face Recognition Using High Order Autocorrelations", Proceedings of 1993 Joint Conference on Neural Networks, pp. 1297-1300, 1993
6. J. M. Tenenbaum and A. P. Witkin, "Perceptual Organization as Building-Blocks for Vision," Journal of the Optical Society of America, A-Optics and Image Science, Vol. 1, No. 12, pp. 1216, 1984.
7. R. P. N. Rao and D. H. Ballard, "Natural Basis Functions and Topographic Memory for Face Recognition," Proc. Intl. Joint Conf. on Artif. Intell. (IJCAI), pp. 10-17, 1995.
8. J. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by 2D Visual Cortical Filters," J. Opt. Soc. Amer. (A), Vol. 2, No. 7, pp. 1160-1169, 1985.
9. J. M. White and G. D. Rohrer, "Image Thresholding for Optical Character Recognition and Other Applications Requiring Character Image Extraction," IBM J. Res. Develop., Vol. 27, No. 4, pp. 400-411, 1983.
10. D. J. Hall and G. B. Ball, "ISODATA : A Novel Method of Data Analysis and Pattern Classification," Technical Report, Stanford Research Institute, Menlo Park, California, 1965.
11. D. M. Harvey, S. P. Kshirsagar, C. A. Hobson, D. A. Hartley, and J. D. Moorehead, "Digital Signal Processing Systems Architectures for Image Processing," Proceedings of the Fifth International Conference on Image Processing and Its Applications, Edinburgh 1995, IEE Conf. Publ. No. 410, 460-464, 1995.