A Brain-Like Approach to Multistage Hierarchical Image Processing

Leonid I. Timchenko[1], Yuri F. Kutaev[2], Maxim A. Grudin[3], Serge V.Chepornyuk[1], David M. Harvey[3], Alexander A. Gertsiy[1]

[1] Department of Optoelectronic Devices, Vinnitsa State Technical Univ., Ukraine
[2] State Scientific Enterprise "ASTROPHISICA", Moscow, Russia
[3] School of Electrical Eng., Electronics and Physics, Liverpool John Moores Univ., United Kingdom

Multistage integration of visual information in the brain allows people to respond quickly to most significant stimuli while preserving the ability to recognize small details in the image. Implementation of this principle in technical systems can lead to more efficient processing procedures. The multistage approach to image processing, described in this paper, comprises both main types of cortical multistage convergence. One of these types occurs within each visual pathway and the other between the pathways. This approach maps input images into a flexible hierarchy which reflects the complexity of the image data. The multistage system highlights spatial regularities, which are passed through a number of transformational levels to generate a coded representation of the image which encapsulates, in a computer manner, structure on different hierarchical levels in the image. At each processing stage a single output result is computed to allow a very quick response from the system. The result is represented as an activity pattern, which can be compared with previously computed patterns on the basis of the closest match.

1. Introduction

Creating effective data analysis with efficient representation of complex hierarchical architectures is one of the key issues when developing novel computer systems. Therefore investigating structures with flexible, complexity dependent hierarchies is of great importance.

The idea of using the effective processing methods developed by nature has resulted in mathematical models which implement certain features of the cerebral cortex.

The latest neurophysiological research, primarily that of Prof. Zeki, points towards a multistage structure for cortical processing [1]. His important assumption is that it is not necessary to pass the data through all hierarchical transformational levels. Certain *pro tem* information about the input can be extracted after each of the early processing stages. In that way further processing involves less data and so becomes faster. In synthetic systems this approach is not sufficiently developed and not attached to real time. In the case of artificial neural networks (ANN), most of the structures implement fixed hierarchical mapping with numerous redundant connections, which slow down the process of the output computation.

The main distinguishing feature of the proposed approach is the introduction of a dynamic of successive transformations applied to spatial regions. This dynamic enables to extract specific temporally uncorrelated components after each transformation.

Parameters of the elements, which participate in correlational and decorrelational processes, can have various physical properties, such as signal magnitude, frequency, phase, power spectrum, image brightness, texture,

connectivity, etc. These are defined by the type of transformation, which must be chosen according to the nature of the problem.

The paper presents the multistage methodology applied to image processing. The concept of parallel multistage image processing can be considered as a consecutive transformation of image components, which consist of elements grouped under a certain criterion. At each stage one component from each parallel channel is extracted and passed to the next processing level, thus reducing the entropy of the remaining image data on that level. In case of the simultaneous extraction from different parallel channels, the extracted components are passed onto the next hierarchical level for further processing. Otherwise, processing of these components is already completed, and they reflect the output result of the processing at the current stage. This procedure of image processing is applied on every level of the hierarchy iteratively until all components on that level are extracted. At each processing level some output information about the input data is extracted, representing a *pro tem* result.

This enables a new view of real-time image analysis with the final result being obtained from the principal components uncorrelated in the spatio-temporal domain. The multistage output object description is represented as an activity pattern rather than as a set of connection weights.

2. Mathematical description of multistage image processing

The procedure of the multistage processing is hierarchical, and on the first level it includes temporal decomposition of the initial image data and transmission of the obtained components to the second level. On each following processing level consecutive operations of temporal decomposition, temporal aligning of the decomposed data, extraction of a principal component, and transmission of the remaining data to the next level are applied once. The processing terminates on the level which receives a single input component.

Let a still 2D image M of the size $l \times n$ be symbolized as a set of elements $M = \{h_{ij} \mid i = 0, 1, ..., l-1; \quad j = 0, 1, ..., m-1\}$, where h_{ij} is a parameter of the image element. It can represent brightness, colour, etc. The image components are grouped from elements with similar parameter values. Then the image M can be represented as a set of components $M = \{a_k \mid k = 0, 1, ..., R-1\}$, where R is the number of the components. A possible expression of temporal decomposition of still image can be shown to be [1]:

$$\sum_{i=0}^{l-1}\sum_{j=0}^{m-1} h_{ij} = \sum_{t=1}^{R}\left(\left(n - \sum_{k=0}^{t-1} n_k\right)\left(a_t - a_{t-1}\right)\right),$$

where $n = l \times n$ is the number of elements in the image M, n_k is the number of elements in the k-th component, t is the step of temporal decomposition, and a_t is the arbitrary element of the component a_t, with $a_0 = 0$ and $n_0 = 0$.

The decomposed image M^1 consists of temporal components of the form:

$$M^1 = \left\{\left(n - \sum_{k=0}^{t-1} n_k\right)\left(a_t - a_{t-1}\right)\middle| t = 1, 2, ..., R\right\} \tag{1}$$

Let the transformation (1) be denoted as $G(M) = M^1 = \left\{a_t^1 \mid t = 1, 2, ..., R\right\}$. The initial image M is segmented into S segments each of them being allocated in a parallel channel. The temporal decomposition G is applied to each segment simultaneously. Then the decomposed s-th image segment is $M_s^1 = \left\{a_{st}^1 \mid t = 1, 2, ..., R_s^1\right\}$, where $s = 1, 2, ..., S$, and R_s^1 is the number of the components in the s-th segment (length of the decomposed segment sequence).

Now the whole decomposed image can be represented as a matrix A_1 of the form:

$$
A_1 = \begin{bmatrix} M_1^1 \\ M_2^1 \\ ... \\ M_i^1 \\ ... \\ M_S^1 \end{bmatrix} = \begin{bmatrix} a_{11}^1 & a_{12}^1 & ... & a_{1R_1^1}^1 & x & x & ... & x \\ a_{21}^1 & a_{22}^2 & ... & ... & a_{2R_2^1}^1 & x & ... & x \\ ... & ... & ... & ... & ... & ... & ... & x \\ a_{i1}^1 & a_{i2}^1 & ... & ... & ... & ... & ... & a_{iR_i^1}^1 \\ ... & ... & ... & ... & ... & ... & ... & x \\ a_{S1}^1 & a_{S2}^1 & ... & a_{SR_S^1}^1 & x & x & ... & x \end{bmatrix}
$$

where x denotes an empty position. The column index indicates the stage at which the component was extracted and the row index indicates the channel. In mathematical terms transmission of the decomposed data A_1 to the second level is represented as transposition T. That is, the initial matrix of the second level $A_1^T = T(A_1)$.

On the second level the temporal decomposition G discovers correlation between the components, which were simultaneously extracted from different channels of the first level. The transformation G is applied to the rows of A_1^T to obtain the matrix A_2:

$$
A_2 = G\left(A_1^T\right) = G\left(T\left(G\left(M\right)\right)\right) = \begin{bmatrix} a_{11}^2 & a_{12}^2 & ... & ... & a_{1R_1^2}^2 & x & ... & x \\ a_{21}^2 & a_{22}^2 & ... & a_{2R_2^2}^2 & x & x & ... & x \\ ... & ... & ... & ... & ... & ... & ... & x \\ a_{j1}^2 & a_{j2}^2 & ... & ... & ... & ... & ... & a_{jR_j^2}^2 \\ ... & ... & ... & ... & ... & ... & ... & x \\ a_{R^11}^2 & a_{R^12}^2 & ... & a_{R^1R_{R^1}^2}^2 & x & x & ... & x \end{bmatrix},
$$

where R^1 denotes the maximal length of the segments, $R^1 = \max\left\{R_1^1, R_2^1, ... R_S^1\right\}$.

The matrix A_2 represents the process of temporal decomposition on the second level, in the way that each component a_{ij}^2 is extracted at the stage $t = i + j - 1$. For further processing it is more convenient to adjust the image representation so that for each component the column index illustrates the stage t when this component was extracted. Then on the next level the correlation between the components, which are simultaneously extracted from different channels, can be discovered in the same way as at the previous level. This adjustment is implemented as horizontal offset of the A_2 rows, so that in each row the first non-empty element has its column index equal

to the row index. The resultant matrix is denoted A_2', and $A_2' = P(A_2)$, where P is the offset transformation.

$$A_2' = \begin{bmatrix} a_{11}^2 & a_{12}^2 & \ldots & \ldots & \ldots & a_{1R_1^2}^2 & x & \ldots & \ldots & \ldots & \ldots & x \\ x & a_{21}^2 & a_{22}^1 & \ldots & a_{2R_2^2}^2 & x & x & \ldots & \ldots & \ldots & \ldots & x \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & x \\ x & x & x & a_{j1}^2 & a_{j2}^2 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & a_{jR_j^2}^2 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & x \\ x & x & x & x & x & a_{R^21}^2 & a_{R^22}^2 & \ldots & a_{R^2R_{R^2}^2}^2 & x & \ldots & x \end{bmatrix}$$

According to the principle of multistage organization of the result, the component a_{11}^2, extracted at the first stage of the processing on the second level, is a principal component. It represents a *pro tem* result of the processing on that level. This extraction is performed by extracting the first matrix column. The remaining data is stored in the matrix A_2''. In algebraic form this procedure appears as $A_2'' = L(A_2')$. The initial matrix of the third level A_2^T is obtained by transposing A_2'', that is $A_2^T = T\left(L\left(P(A_2)\right)\right)$.

The described procedure is applied on each of the following hierarchical levels until the k-th level, where the matrix A_K contains a single element. Then $T\left(L\left(P(A_k)\right)\right) = \varnothing$. Let the consecutive use of the transformations G, P, L, T be denoted by F: $F\left(A_j^T\right) = T\left(L\left(P\left(G\left(A_j^T\right)\right)\right)\right) = A_{j+1}^T$. Successive applications of F are denoted as $F^j\left(A_1^T\right) = F^{j-1}\left[F\left(A_1^T\right)\right]$. Then the whole multilevel process can be represented below in algebraic form:

$$F^{k-1}\left[T\left(G\left(\bigcup_{s=1}^{S} M_s\right)\right)\right] = \left\{a_{11}^j \mid j = 2, 3, \ldots, k\right\}, \tag{2}$$

where M_s are the segments of the input image M, k is the number of processing levels, and a_{11}^j is the principal component extracted on the j-th level, starting from the second one. Thus the multistage image processing procedure can be considered as network processing which on the first level incorporates the consecutive use of the temporal decomposition G and the transposition T applied to each of the image segments. On each of the $(k-1)$ successive levels the complex transformation F is applied once. It involves consecutive operations of decomposition, offset, extraction and transposition.

Structural objects of the real world are reconstructed during the latter stages of cortical sensor information analysis by means of combining maximally purified visual data [2]. In the authors' terminology this is presented by the process of extracting the resulting principal components a_{11}^j from the matrices A_j' in the

separate structures of k levels. This methodology utilizes certain neurophysiological principles, such as parallel channels for handling colour, shape and motion, multistage processing in different cortical areas, spatio-temporal correlation between neural coalitions, and the mechanisms of associative learning [3].

3. Functional organization of the multistage network

The concept central to the multistage processing is parallel-consequential interaction of data streams [13]. As shown in Figure 1, on the first processing level the multiprocessor network decomposes the input data into temporal intermediate sequences $\left\{a_{st}^1\right\}$. On the second level the processor elements (PE's) form the second series of intermediate signals. Such intermediate series are generated on all subsequent levels if different PE's are activated simultaneously, similarly to the associative coincidence in the cerebral cortex. An output result a_{11}^j is produced on each level on condition that no simultaneous activations of different PE's occur. This mechanism is analogous to the formation of cortical coalitions which are isolated from the others in space and time [4].

In general, interactive characteristics of the multiprocessor network, which implements the multistage structure, can be described by six types of functional series. These series represent interactions between the parallel channels on different levels of the hierarchical network.
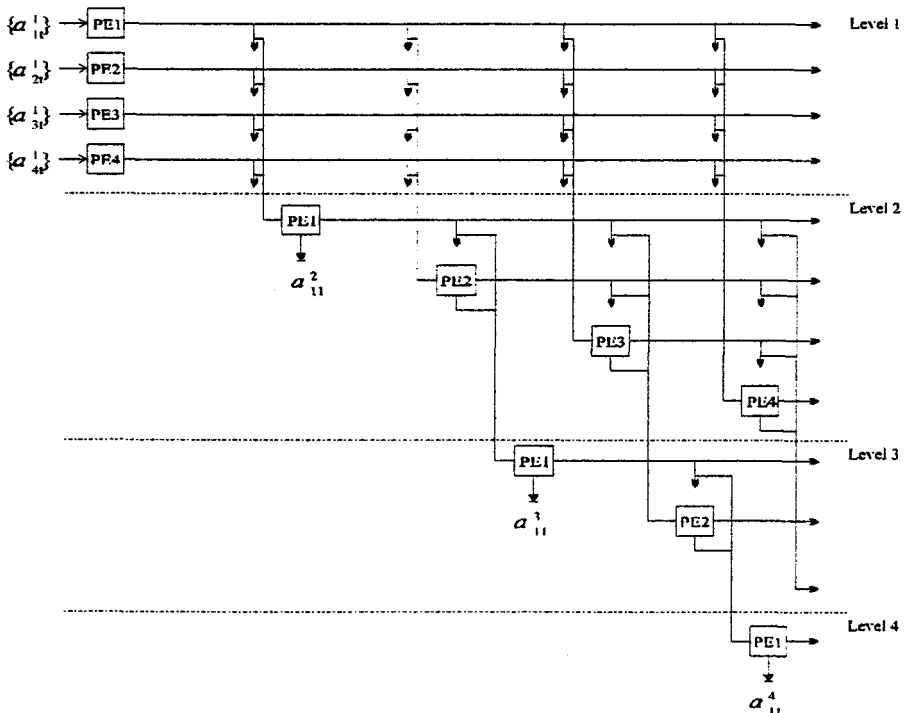

Figure 1: Multistage interactions of data streams in a parallel network.

The first series is defined as:

$$N_1(i',j') = \bigcup_{i'=1}^{n} \bigcup_{j'=1}^{n} \left(A \left] \frac{n-(i'-1)}{j'} \right[\right) \tag{3}$$

where n denotes the number of the stages on the first processing level. Notion i' represents the stage of the process, j' represents the level of the hierarchy, and $][$ denotes the equal or least bigger integer function.

The easiest way to describe all the subsequent series is by arranging them in pairs.

$$N_2(i',j') = \bigcup_{i'=n+1}^{2n-1} A(2n-i'), \quad j'=1; \quad N_3(i',j') = \bigcup_{i'=n+1}^{2n-1} \bigcup_{j'=2}^{i'} \left(A \left] \frac{i'}{j'} \right[\right) \tag{4}$$

The second series is defined by the first part of the expression (4) for $j'=1$ and by the second part of (4) for $j'=2, 3, ..., 2n-1$. It partially describes the first level of the network by $N_2(i',j')$ at each i'-th stage. On the higher levels of the hierarchy the second series is represented by $N_3(i',j')$. This pairwise description also applies to the other series, given in the expressions (5)-(8).

The third series is:

$$N_4(i',j') = \bigcup_{i'=2n}^{3(n-1)} A(3n-2-i'), \quad j'=3; \quad N_5(i',j') = \bigcup_{i'=2n}^{3(n-1)} \bigcup_{j'=4}^{i'} \left(A \left] \frac{i'}{j'} \right[\right) \tag{5}$$

The fourth series is:

$$N_6(i',j') = A(1), \quad i'=3n-2, j'=3,4 \quad N_7(i',j') = \bigcup_{j'=5}^{i'} \left(A \left] \frac{i'}{j'} \right[\right), \quad i'=3n-2 \tag{6}$$

The fifth series is:

$$N_8(i',j') = A(1), \quad (3n-1) \le i' \le (3n-2)+(n-3)^2, \quad 5 \le j' \le 4+(n-3)^2;$$

$$N_9(i',j') = \bigcup_{i'=3n-1}^{(3n-2)+(n-3)^2} \bigcup_{j'=i'-(7-3n)}^{i'} \left(A \left] \frac{i'}{j'} \right[\right) \tag{7}$$

The sixth series is:

$$N_{10}(i',j') = A(1), \quad (3n-1)+(n-3)^2 \le i' \le n^2, (n-3)^2+6 \le j' \le (n-3)^2+2(3n-5);$$

$$N_{11}(i',j') = \bigcup_{i'=(3n-1)+(n-3)^2}^{n^2} \bigcup_{j'=2i'-(n-3)^2+9-6n}^{i'} \left(A \left] \frac{i'}{j'} \right[\right) \tag{8}$$

These six series organize a functional parallel-hierarchical structure and describe the spatio-temporal interaction in the multiprocessor network.

The expressions (3)-(8) are validated on a simple example of calculating the processor interactions in the multiprocessor network with network dimensions $S \times n = 4 \times 4$, where S denotes the number of initial image segments and n denotes the number of processing stages on the first level.

Hierarchical level →

$$
\begin{aligned}
&N_1(\bar{r},\bar{j})\begin{cases}1\\2\ \ 1\\3\ \ 2\ \ 1\\4\ \ 2\ \ 2\ \ 1\end{cases}\\[4pt]
&N_2(\bar{r},\bar{j}),N_3(\bar{r},\bar{j})\begin{cases}3\ \ 3\ \ 2\ \ 2\ \ 1\\2\ \ 3\ \ 2\ \ 2\ \ 2\ \ 1\\1\ \ 4\ \ 3\ \ 2\ \ 2\ \ 2\ \ 1\end{cases}\\[4pt]
&N_4(\bar{r},\bar{j}),N_5(\bar{r},\bar{j})\begin{cases}2\ \ 2\ \ 2\ \ 2\ \ 2\ \ 1\\1\ \ 3\ \ 2\ \ 2\ \ 2\ \ 2\ \ 1\end{cases}\\[4pt]
&N_6(\bar{r},\bar{j}),N_7(\bar{r},\bar{j})\{1\ \ 1\ \ 2\ \ 2\ \ 2\ \ 2\ \ 2\ \ 1\}\\[4pt]
&N_{10}(\bar{r},\bar{j}),N_{11}(\bar{r},\bar{j})\begin{bmatrix}1\ \ 2\ \ 2\ \ 2\ \ 2\ \ 2\ \ 1\\ \ \ \ \ 1\ \ 2\ \ 2\ \ 2\ \ 2\ \ 1\\ \ \ \ \ \ \ \ \ 1\ \ 2\ \ 2\ \ 2\ \ 1\\ \ \ \ \ \ \ \ \ \ \ \ \ 1\ \ 2\ \ 2\ \ 1\\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ 1\ \ 2\ \ 1\\ 1\ \ 1\end{bmatrix}
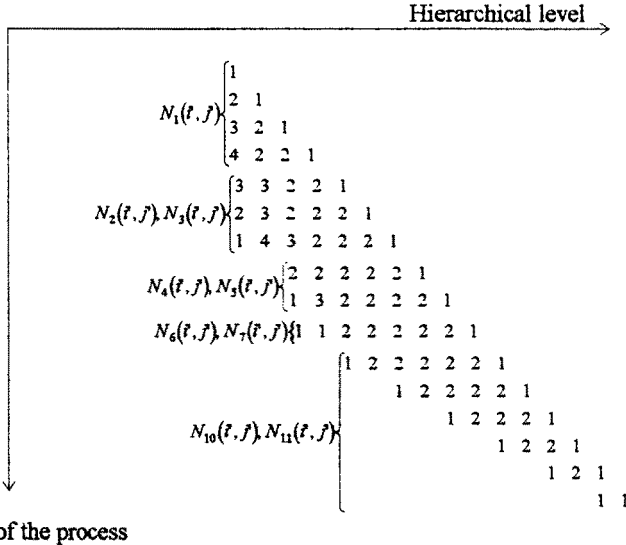\end{aligned}
$$

Stage of the process

Figure 2: Interactions between the four parallel channels of the network (Fig. 2) on different hierarchical levels. Processing on the first level involves four stages.

The expressions (1)-(2) and (3)-(8) describe the spatio-temporal interaction of certain parameters of the image elements. In Figure 2, the numerical series $N_1(i,j) \div N_{11}(i,j)$ illustrate the dynamic correlation on different levels of the multistage hierarchy.

This mechanism is very similar to various neurophysiological processes, such as associative learning [6], spatial summation of postsynaptic potentials [5], and various interconnection pathways between the specialized visual cortical areas [7]. Implementation of the spatio-temporal image processing and multistage integration of the results may be very promising when realizing real-time artificial neural structures [12].

4. Discussion and conclusions

The paper investigates mapping of the part-whole hierarchies into the connection's network in order to accomplish the multistage information processing. Image elements are grouped into structural components, which are then consequentially passed to the higher levels of the hierarchy to participate in correlational processes.

A possible criterion for discovering structural relationships in the image is connectedness [8], which is one of the most informative visual characteristics. It explains a powerful tendency of the human visual system to perceive any uniform connected area as a single unit.

The structure shown in Figure 1, includes temporal components as well as spatial ones. Such implementation may result in the appearance of new procedures for unsupervised learning. The structure implements five general properties, which are fundamental for the human visual perception:

- topographical character of calculations [9];

- simultaneousness (parallelism) of signal actions;
- coarse cortical hierarchy [10];
- temporally correlated spatial mechanism of perception;
- mosaic structure of the cortex.

The strategy of multistage processing has several analogies in the cortical organization. It conforms with the theory of multistage integration in the visual cortex [1], which states that integration does not occur simultaneously by virtue of the convergence of signals in a single master area. On the contrary, it is a parallel-hierarchical process in which each level directly contributes to the perception of incoming visual information. This process involves two major types of convergence operating on all levels. Topical convergence occurs within each visual pathway, accumulating information on a certain feature from different image locations. Confluent convergence occurs between the pathways and involves integration of the features of different nature. The proposed multistage architecture, shown in Figure 1, embodies both these convergence types.

A number of artificial neural structures with effective learning algorithms have been developed. However, the mechanisms of the formation of cortical coalitions and their joint actions are not fully understood. The task of experimentally observing interactions between different neural groups has proved to be extremely difficult [11]. Therefore the attempt of simulating such interactions using mathematical knowledge is very pertinent and can help discover the methods created by nature.

References

1. Zeki, S. *A Vision of the Brain.* Blackwell Scientific Publications, Oxford, 1993.
2. Blum, F. E., Lazerson, A., and Hofstadter, L. *Brain, Mind and Behavior.* W. H. Freeman and Co., New York, 1985.
3. Nicoll, R. A., Kauer, J. A., and Malenka, R. C. The Current Excitement in Long-Term Potentiation. *Neuron 1(2),* pp. 97-103, 1988.
4. Crick, F. and Koch, C. Towards Neurobiological Theory of Consciousness. *Seminars in the Neurosciences 2,* pp. 263-275, 1990.
5. Sherrington, G. E. *The Integrative Action of the Nervous System.* Yale University Press, New Haven, Conn., 1961.
6. Kandel, R. and Hawkins, R. The Biological Basis of Learning and Individuality. *Scientific American 67(3),* pp.53-60, 1992.
7. Livingstone M. S. and Hubel, D. H., Anatomy in the Primate Visual Cortex. *J. Neurosci. 4,* pp. 309-356, 1984.
8. Rock, I. and Palmer, S. The Legacy of Gestalt Psychology. *Scientific American 263(6),* pp. 48-61, 1990.
9. Churchland, P. and Sejnowski, T. J. *The Computational Brain.* The MIT Press/Bradford Books, 1992.
10. Kuffler, S. W., Nicholls, J. G., and Martin, A. R. *From Neuron to Brain.* Sinauer Associates, Sunderland, Mass., 1984.
11. Hubel, D. H. *Eye, Brain and Vision.* Scientific American Library, New York, 1988.
12. Kytaev, Y.F. Systems for Correlational-Extremal Co-ordinates Measuring with Generalised Q-transforming of Images. *Ph.D.Dissertation.* Vinnitsa Tech. Univ., Ukraine, 1989, (in Russian).
13. Timchenko, L. I., Grudin, M. A., Gel, V. P., Harvey, D. M. "A Multistage Hierarchical Approach to Image Processing,", *Proc. of IEE Colloquium on Mulitdimensional Systems,* pp. 2.1-2.6, January 1996.