Maria GRZEGOREK[*]

VARIABLES APPLIED IN A NMR IMAGE DECOMPOSITION WITH THE AID OF GCCA

The paper describes the influence of variable selection on an image decomposition. A NMR image is a source of a set of variables describing pixels of the image: gray level, gradient magnitude and seven variables derived from gradient magnitudes of neighbouring pixels. A selection of the variables is the essence of the matter at this stage of the image processing. Two suggestions are proposed and tested: a normalization of gradient magnitude of the pixel by dividing it by a value of the gray level, and development of a nonlinear sequence of thresholds which are used in comparison of adjacent pixels.

1. INTRODUCTION

The efforts are made to apply Grade Correspondence Cluster Analysis (GCCA) to image processing, especially to NMR image exploration. The intention is to discover hidden structures in such images. GCCA seems to be an excellent method of analysis due to the fact that it is a modern and widely applied method of the multivariate data analysis which is helpful in many problems, also in important medical application ([5],[6]).

GCCA is based on a grade transformation of data. It has been developed in Institute of Computer Science Polish Academy of Sciences, Warsaw (extensive description of the theory and applications is contained in [4]) and was implemented in program GradeStat ([3]).

The source of data is a NMR image. Gray level g_l is attached to every pixel and originates from the image. Gradient magnitude g_m can be calculated using the gray level of the adjacent pixels. Next seven variables $n_1, ..., n_7$ are developed on the basis of g_ms of adjacent pixels which are compared with g_m of the pixel and thresholded with suitable values $t_1, ..., t_7$. Dataset consisting of rows corresponding to image pixels and columns corresponding to variables is processed by procedure GCA. GCA rearranges rows and columns in such way that most similar rows (and columns) are grouped possibly near each other while rows (and columns) which are different are distant in a data matrix. Then the set of rows is divided by clustering procedure to a sequence of disjoint sets of similar pixels. Each group of pixels is visualized in a separate subimage with the same resolution as the source image. The subimage displays areas of the pixels provided with approximately near features.

In the next section description of variables is placed whereas in Section 3 performed experiments are presented. Section 4 includes conclusions and outlines future work.

2. VARIABLES OF THE DATASET

In the dataset originated from the image there are some variables, two simple and obvious, e.g. gray level g_l and gradient magnitude g_m, and a set of seven variables constructed on the basis of adjacent pixels. Generally, there is a sequence of i variables n_i, $i = 1,...,k$. However, earlier investigation [1] permits to restrict the number of variables to a reasonable limit $k = 7$.

The variable n_i is constructed involving gradient magnitudes of pixels in a neighbourhood with extent of 3×3. Value of the variable n_i is equal to the number of pixels in the neighbourhood which g_m values are distant from g_m of the central pixel less than a threshold value t_i.

The parameter of the fundamental importance is a base parameter b due to b joins a content of the image with a sequence of t_i values. The characteristic feature of the image is a maximal value of the gradient magnitude g_{m_max} which is generally different in various images. Then, the sequence of thresholds values t_i can be established by equation [2]

$$t_i = i * b * g_{m\_max} /1000 \qquad (1)$$

Fixed factor 1000 induces that b is measured in one thousandth of g_{m_max}.

Now for every pixel in the image values of variables are calculated, then the procedure GCA reorders rows of the data matrix and the clustering procedure divides reordered sequence of pixels onto preliminary fixed number of groups. Such group is a subset of successive pixels which are strong similar to pixels belonging to this subset and are less similar to pixels in other subsets. Generally, subsets consist of different numbers of pixels. This results in a considerable different density of visualized pixels in each subimage of the ordered sequence of subimages [6].

The value of the gradient magnitude g_m, no matter which kind of gradient is applied, is sensitive to mean value of gray level g_l in the pixel neighbourhood. Differences of g_l of adjacent pixels in light image regions might be higher than in darker ones and pixels are shifted to different subsets while the importance of examined g_ms might be similar. The influence of such scatter is reduced by normalization of the g_m value of every pixel

---

[*] Institute of Computer Science PAS, Ordona 21, 01-237 Warsaw.

$$g_{m\_n} = g_m / g_l \qquad (2)$$

Normalized values g_{m_n} are used in the data matrix and to obtain the threshold values.

In the previous tests the values of thresholds were proportional. Value t_i was equal to t_1 multiplied by integers $1,...,k$. Now thresholds values are

$$t_{i+1} = t_i * r = t_1 * r^i \qquad (3)$$

where r is a factor which can be established if first and last thresholds are defined as in equation 1

$$t_1 = b * g_{m\_max} / 1000; \; t_k = k * b * g_{m\_max} / 1000 \qquad (4)$$

When $r > 1$ values of thresholds with lower i are closer to each other then those with higher i. It implies that variable with lower i (lower value of threshold) is more sensitive to changes of g_m, and variable with higher i requires more dynamic changes of g_m to increase its value. For $k = 7$ the value of the factor r is near 1.3.

3. EXPERIMENTS

Results of tests are presented with the aid of test image shown in Figure 1. Resolution of a fragment of the NMR image is 320×200 pixels, gray level range is 0-255 and gradient magnitude 0-167.7
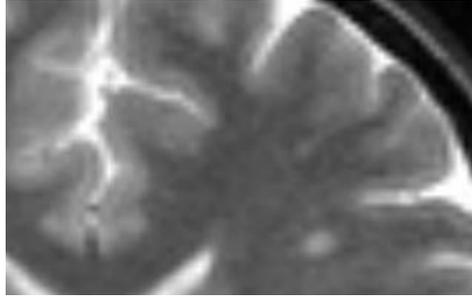

Fig.1 Test NMR image

Extent of variables n_i is 0-8. There are eight pixels in the neighbourhood of the central pixel. When none of g_ms of pixels in neighbourhood is closer to g_m of central pixel than corresponding threshold value t_i then n_i is equal to zero. On the other hand if all eight pixels g_ms are closer than t_i then variable n_i for this pixel is equal to 8.

Selection of the parameter b value is a challenge. At present, limits of the value can be established as follows. Upper limit is $b_{max} < 1000/k$ due to t_k should not be greater than g_{m_max} which could result in fact that variable n_k for all pixels would be equal to eight (the highest value of variable n_i) which cause that such variable is useless. Lower limit of b depends on changes of g_m in adjacent pixels. Threshold t_1 should be much lower than the highest difference of g_ms. This condition is assured by previous one. Too low value of b will result in equality of variables values to zero (in program GradeStat no column can have all values equal to 0). Practically $b = 0.0002$ occurs too small. In tests are fixed five values of b, $b = 0.002, 0.02, 0.2, 2, 20$.

In previous investigation [1] division onto ten subsets is used. At present number of clusters is predefined equal to 20.

Tests apply three sets of variables:
1. variables constructed according to equation 1 (old variables),
2. the same variables but used to data matrix with normalized g_m values according to equation 2,
3. variables constructed according to equations 3 and 4 (thresholds t_i that are not proportional).

In Figure 2 plots for variables from set 1 (left) and set 2 (right) for three values of parameter $b = 0.002, 0.2$ and 20 are shown. For lower b most of pixels are concentrated in a few subsets (8-11, solid line), for medium b most of pixels are in first subsets (2-3, dashed line), while for higher b pixels spreads more uniformly in all subsets (dotted line). Pixels distribution in subsets is similar for all kinds of variables. Maxima of plots for variable set 2 are smaller and for variable sets 1 and 3 are more sharp.
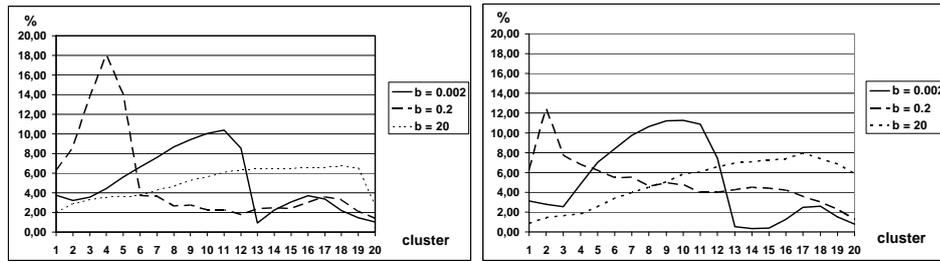
Fig.2 Pixels distribution in subsets for three chosen values of parameter b. Left chart is for variable set 1, right for variable set 2. On horizontal axis are consecutive subsets (subimages), on vertical axis is percentage of pixels belonging to the subset

Pixels belonging to each subset are visualized in a separate subimages as black points. Some of subimages contain distinct groups of pixels. In other subimages are scattered pixels. Figure 3 shows three subimages of cluster 2 for $b = 0.002$, 0.2 and 2 for variable set no. 1, whereas Figure 4 presents similar subimages for variable set no. 3. Thresholds in variable set 3 cause that structures in subimages are more distinct. The number of subimages containing dispersed pixels diminish a little.

Fig.3 Images visualizing pixels from subset 2, $b = 0.002$, 0.2, 2, set of variables no. 1

Fig.4 Images visualizing pixels from subset 2, $b = 0.002$, 0.2, 2, set of variables no. 3

It is difficult to compare subsets for variable sets 1 and 2. due to similar structures appearing in different subimages of the sequence for different values of parameter b. Moreover, two structures visible in one subimage originated from one variable set can be found in different subimages of the other variable set. Normalization of g_m is a considerable change of the variables and seriously rearranges resulting sequence of pixels and its division onto subsets. Other connections between pixels with new features are discovered. Figure 5 shows three out of 100 subimages resulting from variable set 2 (subimage 6 for $b = 0.002$, subimage 4 for $b = 0.02$, subimage 7 for $b = 20$).
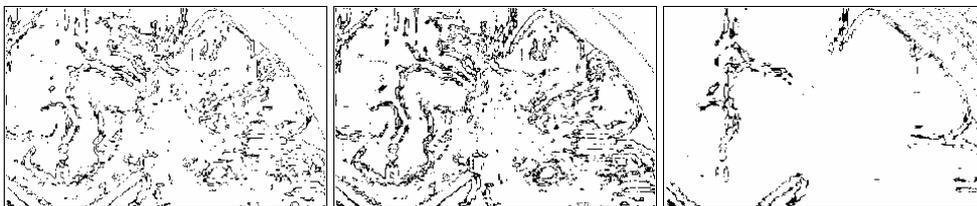

Fig.5 Subimages visualizing pixels from subset 6, $b=0.002$, subset 4, $b=0.02$, subset 7, $b=20$, variable set no. 2

Finally, segments of adjacent pixels are detected in each subimage. Then all segments having fixed number of pixels from whole sequence of subimages are gathered and visualized in one image. Figure 6 left shows pixels forming segments which are greater than twenty pixels for $b = 0.002$. Right image contains pixels from segments which have more than ten pixels for $b = 20$. Figure 6 is performed for the set of variables no. 1, Figure 7 is prepared for the set of variables no. 2 and Figure 8 is arranged for the set of variables no. 3 where parameter value b is chosen from explored range 0.002-20, and size of segment is from range 20-80, in each case exact values are settled in descriptions under the figures.
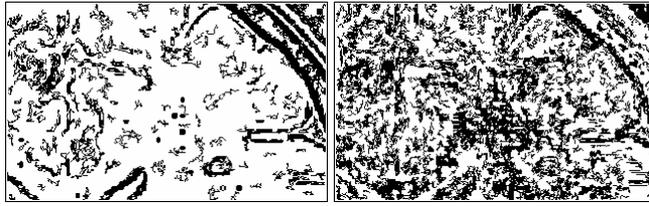
Fig.6 Images showing pixels belonging to segments greater than 20 pixels, $b = 0.002$ and
to segments greater than 10 pixels, $b = 20$, the set of variables no. 1

Fig.7 Images showing pixels belonging to segments greater than 60 pixels, $b = 0.002$ and
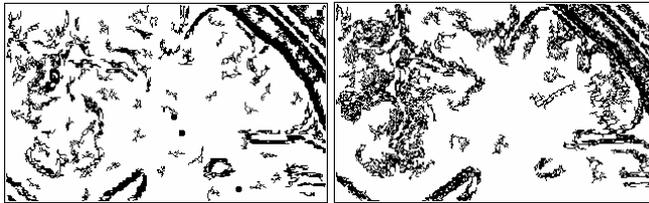to segments greater than 20 pixels, $b = 20$, the set of variables no. 2

Fig.8 Images showing pixels belonging to segments greater than 30 pixels, $b = 0.02$ and
to segments greater than 80 pixels, $b = 0.2$, the set of variables no. 3

4. CONCLUSIONS

Pixels in the image are characterized by features described by variables. Values of variables form data matrix which is processed by procedure GCCA. GCCA orders a sequence of pixels and groups together similar pixels into subsets. These subsets are visualized in subimages. Some of subimages contain distinct structures which are derived from the image. Other contain dispersed pixels and at the moment there is no advantage of these pixels.

Current inquiry results in a large amount of subimages, which will be larger when the size of image sequence increases from twenty to for instance fifty. Both questions solves segments detection performed in each subimage separately. Submages containing larger structures are cleared from single pixels or small segments. Images with dispersed pixels are detected and marked as unabled to further processing.

Experiments with different sets of variables shows that there exists easy way to improve appearance of image structures in the sequences of subimages. Next activity will concern influence of a size of the image on image decomposition and establishing suitable number of subimages. Important task is to prepare specialized implementation of GCCA which would have additional functions adapted to image processing.

BIBLIOGRAPHY

[1] GRZEGOREK M., Homogeneity of pixel neighborhoods in gray level images investigated by the Grade Correspondence Analysis, in: Kurzyński M., Puchała E., Woźniak M., Żołnierek A., Computer Recognition Systems 2, pp 76-83, Springer, series Advances in Soft Computing, Berlin-Heidelberg, 2007.

[2] GRZEGOREK M., Towards an exploration of GCA ordered pixels, in: Kurzyński M., Puchała E., Woźniak M., Żołnierek A., Computer Recognition Systems 2, pp 156-163, Springer, series Advances in Soft Computing, Berlin-Heidelberg, 2007.

[3] http://gradestat.ipipan.waw.pl/

[4] KOWALCZYK T., PLESZCZYŃSKA E., RULAND F., Grade Models and Methods for Data Analysis, With Applications for the Analysis of Data Populations, Series: Studies in Fuzziness and Soft Computing, Vol. 151, Springer Verlag Berlin Heidelberg New York, 2004.

[5] KSIĄŻYK J., MATYJA O., PLESZCZYŃSKA E., WIECH M., Analiza danych medycznych i demograficznych przy użyciu programu GradeStat. Instytut Podstaw Informatyki Polskiej Akademii Nauk, Instytut "Pomnik – Centrum Zdrowia Dziecka", Warszawa, 2005.

[6] PLESZCZYŃSKA E., Application of grade methods to medical data: new examples, Biocybernetics and Biomedical Engineering, Vol. 27, No. 3, pp. 77-93, 2007.