AUTOMATIC MAP NAME PLACEMENT FOR 1:250000 TOPOGRAPHICAL MAPS

Fan Hong Zhang zuxun Du Daosheng Zhang jianqing
The State key Lab. Of WHU,129 Luoyu Road,Wuhan,China,430079
Tel: 86-027-87881292, Fax:86-027-87643969,fh@hp01.wtusm.edu.cn

KEY WORDS: Hopfield neural network, Energy function, Map name placement

ABSTRACT

The placement of map name has been a difficult problem in the filed of cartography. The latest research showed that the kernel problem of map name placement is a NP-hard problem. The traditional approaches as search in status space of the problem and expert system approach are not able to solve this problem well because they both take the sequential searching strategy. One way to solve this problem is to try some way to decrease the complexity of searching in order to find the near-optimal position in the user-tolerable time.

This paper concentrated on the problem of automatic map name placement for 1:250000 topographical maps. After some guidelines for map name placement were given, a hopfield neural network algorithm was proposed to find the best label position for point feature. The experimental results proved that this algorithm has good permanence and high speed. For name placement for linear feature such as streams and roads, an improved exhaustive searching strategy was proposed to find the most optimal label position. The experimental results on streams map showed its permanence and speed is also satisfactory. For polygon features, map names should be placed on the sketch line inside a polygon. An algorithm to utilize mathematical morphology was proposed to find sketch lines.

In this paper, the corresponding methods and experiment results would be presented and analyzed in detail.

1 PREFACE

Map name is an important component of the map, whether labels in place or not plays an important role on the readability and usefulness of a map. Thus far, the placement of cartographic names has been a manual process which takes plenty of time and energy. An automatic or even semi-automatic way of map name placement will increase the speed of map name placement greatly.

More recently, many different algorithms have been developed to aid this map name placement process. Among them are the priorities suggested by Yoeli[Yoeli ,1972] for the placement of names for point feature. The integer programming is designed for the placement of names for point and linear feature[Zoraster, 1986], an expert system approach for dense-map name placement[Doerschler, 1989]. Yet the automation of name placement has proved difficult to be implemented successfully. Because an automated name-placement system must select names from scale-independent databases and place them in a manner acceptably similar to that of a trained cartographer. Latest research showed that the kernel problem of map name placement is a NP-hard problem[Marks,1991], The traditional approaches like search in status space of the problem and expert system approach are not able to solve this problem well because they both take the sequential searching strategy, a way out to solve this problem is to try some way to decrease the complexity or improve the efficiency of problem in order to find the near-optimal position in the user-tolerable time .

A hopfield neural network algorithm was proposed in this paper to be applied to add name to the map for the point feature. The hopfield neural network algorithm took advantage of parallel searching instead of sequential searching. The experimental results proved that this algorithm had good permanence and high speed. For name placement for linear features as streams and roads, an exhaustive searching strategy was proposed to find the most optimal label positions. An encoding scheme based on binary bit was adopted to improve its permanence. For area features such as water bodies and political or statistical areas, map names are often placed on the sketch line inside a polygon. In our paper, an algorithm to utilize mathematical morphology to find sketch line was applied. It was proved to have better performance and could handle more complicated situations than the method based on scan-line midpoint.

The proposed algorithm have been tested and verified on three different 1:250000 topographical maps.

In this paper, the corresponding methods and experiment results would be presented and analyzed in detail.

2 NAME PLACEMENT GUIDELINES

As usual, different map placement guidelines are required for different kinds of features in terms of point ,line and polygon. However, the following principles will be general.

(1) “refer to” principle : label should refer unambiguously to its intended referent feature.

(2) “avoiding off” principle : label name should not overprint features of same category ,other labels and other important features.

(3) “accustomed to” principle : the character position、character order etc should accord with the reading custom of the readers.

Referring to the above general principles and the map-making regulations of 1:250000 topographic of the People’s Republic of China, a practical guideline for point, line and polygon features can be given as follow.

For point feature such as residents. the guidelines was given as follow.

(1) One of the eight ranked positions surrounding a feature symbol is used to place a point feature label. In order of priority, these positions are 1) to the right of the symbol; 2) above; 3) to the left; 4) below; 5) above and to the right; 6)above and to the left; 7) below and to the left; 8) below and to the right, showed in Figure 1.Position with higher priority are preferred over positions with lower priority.

(2) Point label should not overprint the previously placed label and any point feature.

(3) Point label should not overprint the important linear feature of the same color such as railways and major roads etc. While overlap is unavoidable, efforts should be made to decrease overlap.

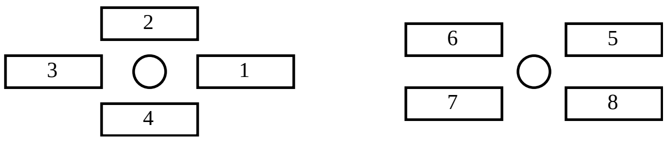


Figure 1 Ranked label positions for point feature

For name placement for linear feature such as streams and roads, the following guidelines were used:

(1) Name should be placed right-reading and follows the curvature of the feature.

(2) A long linear feature should be divided into several short segments being independently placed with one name.

(3) The written direction of name should be consistent with the direction of linear feature.

(4) The distance between label and linear feature had better be 1-2mm.

(5) The distance between characters of label had better be the size of 3-5 characters.

(6) Name is not allowed to overprint the linear feature;

(7) Name is not allowed to overprint previously placed label and other important features.

For polygon feature such as water bodies, political or statistical regions, double-line rivers and lake with small area, the following guidelines were used.

(1) For big regions such as areal lakes, political or statistical areas, map names are often placed on the sketch line inside a polygon.

(2) For a long stripe regions such as double-line rivers, too narrow to hold the label name, be treated as a linear feature, label name be placed outside the regions and along the long side of outline of the regions.

(3) For small regions such as lakes with small area and area reservoirs, too small to hold the label name, be treated as a point feature , label names be placed outside the regions and around the point feature.

(4) For scattered area residents, label names be placed outside the regions and around the outlines of regions.

3 THE SOLUTION TO AUTOMATED NAME PLACEMENT FOR RESIDENT FEATURES

Point features involves residents, flag points and elevation points etc., among which residents are representative for their largest number and highest density. According to the guideline of name placement for point feature mentioned above , the following strategies were taken to automate name placement for resident of topographical map.

The total solutions contain the following 3 steps:

step 1: coarsely choose the candidate position and determine the optimal level for every position.

Firstly, 8 available position for every resident are chosen, showed in Figure 1, through evaluation of "readability" and "belong to" conditions of the 8 positions, a weight factor is given for every possible position and detailed method is depicted as follow.
The weight of the candidate position is the function of the primary weight and the secondary weight, that is : weight = the primary weight * the secondary weight
Specify a basic weight from 1.0 to 0.93 corresponding to the priority from high to low.
Use a rectangle area to represent the area that a label name occupies and judge if the area will overlap the important feature such as railway, major road and minor road. Different secondary weight factor is assigned accordingly. For example, if the label rectangle of one feature overlaps with a railway or a major road , according to the extent of overlap, its secondary weight should be assigned to 0.1, 0.2, 0.3 or 0.4 respectively; In order to control overlap with the minor road , the secondary weight of 0.51,0.52,0.53 should be assigned for different extent of overlap.
Judge whether the label and the referent feature appears on the same side of the boundary and whether the label overlap with the boundary, To handle this case, an appropriate secondary factor should be designated.
Step 2: construct a hopfield network, set the initial value for it and make it run. If necessary, make the network run many times, observe its convergence, select the best results and record it.
Step 3: after taking the previous steps, local optimal processing is made so as to resolve the remaining conflicts between the residents labels. Few remained conflicts that have not yet been resolved in the end, will be adjusted and revised manually.
The core of the whole algorithm is step 2, that is to use hopfield network to find the best local or global label position for every resident on topographical map.

4 THE ALGORITHM OF HOPFIELD NEURAL NETWORK TO FIND THE OPTIMAL LABEL POSITIONS FOR RESIDENT FEATURES

Hopfield neural network is a one-level of feedback network , Let N1, N2, …Nn stand for the n neural unit, Wij stands for the connection weight from Ni to Nj. If we use W to represent the connection strength between the n nodes, Hopfield is symmetric, then:

$$W_{ij} = W_{ji} \qquad i, j \in \{1, 2, 3 \ldots n\}$$

For the continuing feedback network, when the network is working, the relation between the input and the output can be represented in terms of the following status equation , among which, g(。) is a continuing monotony ascending function with up limit, among which sigmond or hyperbolic tangent function most commonly used, Ui represents the input of the neural unit i , Vi represents the output of the neural unit I, Ii represents the bias of the neural network namely the stimulus coming from the external world.

$$C_i \frac{du_i}{dt} = -\frac{u_i}{t} + \sum_j W_{ji} V_j + I_i$$

$$V_i = g(U_i)$$

It has been proved that hopfield network is a non-linear motion system. The hopfield neural network owns the feature of convergence of the energy, namely, the energy function of network will converged to a minimum point of network energy as the network runs a certain times in an iterative way. If we can express the goal function of the optimal combinatorial problem into a energy function, mapping the variant of the problem to the status of neural network, while the energy function is converged to a minimum point, the best solution for the problem will be gained accordingly. So hopfield neural network often be successfully used to solve the optimal combinatorial problem.
The problem of finding the best label position for resident feature can be regarded as a combinatorial optimal problem. A hypothesized map contains m residents, and every resident has n candidate points(such as, n = 8, 16…). There are m×n candidate points , and these residents will be listed as a matrix in which n candidate label positions of one resident will be lined as one row, altogether there are m row * n column, as showed in table 1. Considering this table as a matrix of m rows * n columns, m stands for the number of the residents, n stands for the candidate label position for every resident. Corresponding one candidate label position to one neural unit a hopfield neural network that comprises m ×n neural unit can be constructed.
In order to define the energy function , we describe the problem as the the sum of constraint condition and the optimal goals as follows.

	Candidate label position							
Resident point	1#	2#	3#	4#	5#	6#	7#	8#
1#	1	0	0	0	0	0	0	0
2#	0	1	0	0	0	0	0	0
3#	0	0	1	0	0	0	0	0
...	...							
...	...							
2514#	1	0	0	0	0	0	0	0
2515#	0	1	0	0	0	0	0	0

Table 1 Candidate label position for resident

Constraint condition: every resident can choose only one label position.
The most optimal goal: the number of overlapping between two label rectangle area is the smallest.
According to the above constraint condition and the most optimal goal, we can give out the energy function of the network as follows.

$$E = \frac{B}{2}\sum_i\left(\left(\sum_j V_{ij}\right)-1\right)^2 + \frac{A}{2}\sum_i\sum_j\sum_k\sum_l D(i,j,k,l)V_{ij}V_{kl}$$

Of the above equation, the first item represents the constraint condition, and only one label position is allowed for one resident feature. When this condition is met, the first item will be equal to 0. The second item is optimal goal, D(i,j,k,l) is specified as follow:

$$D(i,j,k,l) = 1 \quad when V_{ij,}V_{kl} \quad overlap \quad each \quad over$$
$$0 \quad when V_{ij,}V_{kl} \quad do \quad not \quad overlap \quad each \quad other$$

Thus the second item is the multiplied number of the overlap between every two label rectangles, if the choice of label position is the optimal, E_2 can reach the minimum value, and if the choice of label position is approximately optimal, E_2 can reach the smaller value.
By comparing the energy function with the standard energy function, the connecting weight between unit i and unit j will be determined as:

$$T_{ij,kl} = -AD(i,j,k,l) - B\delta_{ij,kl} \quad \ldots\ldots(1)$$

$$\delta_{ij,kl} = \begin{cases} 1 & \text{当}ij=kl\text{时} \\ 0 & ij\neq kl\text{时} \end{cases}$$

$$I_{ij} = B \quad \ldots\ldots(2)$$

Because $W_{ij,kl} \propto T_{ij,kl}$,let$W_{ij,kl} = T_{ij,kl}$, then the running equation will be derived as follow:

$$\frac{du_{ij}}{dt} = \frac{-u_{ij}}{\tau_i} + \sum_{j\neq i}\left(AD(i,j,k,l) - B\delta_{ij,kl}\right)V_{ij} + B \quad \ldots\ldots(3)$$

$$V_{ij} = g(u_{ij}) \quad \ldots\ldots(4)$$

while an appropriate parameter A,B and an appropriate constant of the time τ_i were specified, after both connection weights of the network and the initial status of the network were set up in the light of the above guideline, running the network in an iterative way, after a certain times of iterations, network approached to a steady status, namely, a minimum energy status. According the status, optimal label positions were obtained.
The 3 tested maps comprise 3251, 1651 and 2734 residents features respectively. Making use of the above algorithm to find the label position for the residents, after 10 times of iteration of the neural network, the conflict number was reduced to 196, 97 and 123. After the local optimization and adjustment, all conflicts were resolved, the running time were respectively 19, 11 and 15 minutes. The result was satisfactory. Figure 2 shows the result of adding label to a resident map with this algorithm .
The experimental results showed that the neural network method can be used to solve the combinatorial optimal problem including finding the best label position. Because the network can converge swiftly,

therefore complex problem with “combinatorial explosion” danger can be converted to a simple problem, which will largely improve the efficiency of algorithm.

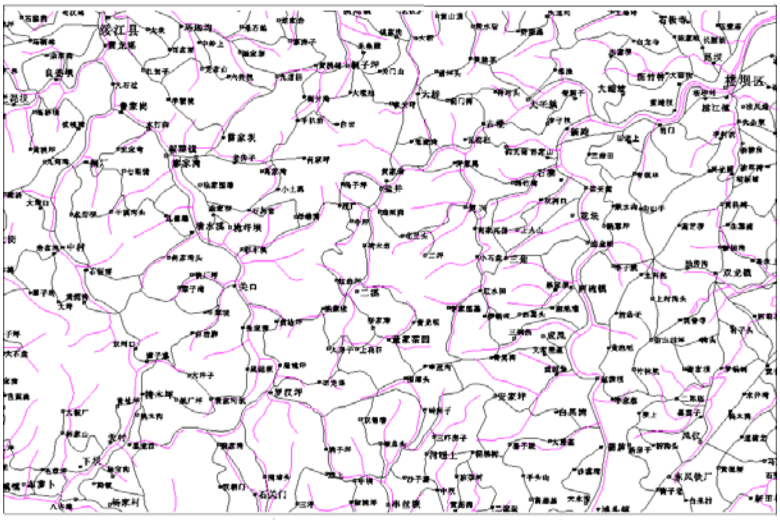


Figure 2 A part of the result of map name placement for a resident map.

5 THE SOLUTION TO AUTOMATED NAME PLACEMENT FOR LINEAR FEATURES

For name placement for linear feature as streams and roads, an exhaustive searching strategy was proposed to find the optimal label position. In implementing the algorithm, according to the guideline for linear features above given, the following technical step were taken: (1)derive the coordinates of the parallel line based on the coordinate points of linear features and place name on the parallel line; (2)divide the long linear features such as streams into several short segments and place one name for each segment; (3) take the exhaustive searching strategy to find out all possible label positions on each segment and assign it an value to represent the priority. (4) find the optimal label position for each segment. Our algorithm will be introduced in more detail as follow.

(1) find the parallel line for linear feature

Linear features are made up of a set of link-up straight segments and the algorithm of finding the parallel for linear features is very complex. In our algorithm, parallel was only used for providing a reference position for map name placement therefore it was no necessary to find very fine parallel.

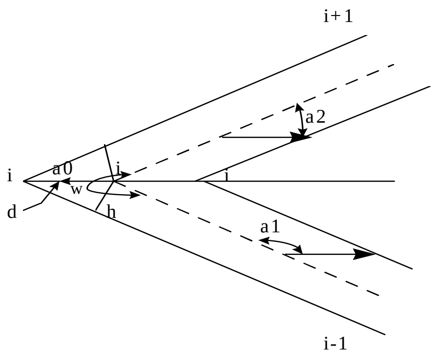


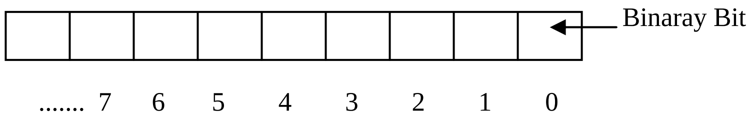
Figure 3 calculating the corresponding vertex of parallel for every vertex based on angle bisector
(dashed line represents the linear feature, solid lines represent the obtained parallel lines)

The algorithm of finding the simple (coarse) parallel line for linear feature included the following steps. firstly, the linear feature was generalized by douglass-Peuker algorithm and then simplified by removing some small acute angle vertex. then, for a given distance d, a algorithm based on angle bisector was applied to calculate the corresponding vertex of parallel for every vertex of original line, the principle for the algorithm see figure 3. Finally, every vertex was connected in proper order to obtain the parallel. the resultant parallel maybe had some small problems, further processing such as removing the self-intersection was applied to obtain optimal parallel. This algorithm had a high speed and result could meet the requirement of afterwards processing

(2) detection of conflict and overlap

There are two key factors to affect a candidate label position's validation and priority: label's conflict and overlap conditions. Here conflict refers to a label overprinting another label and overlap refers to a label overprinting map features. While searching the map for the optimal position, conflict and overlap for every candidate label position will be detected and analyzed. A valid label position is defined as one without conflict. The optimal position is defined as one that owns the highest priority. Detection of conflict and overlap based on vector data is a time-consuming procedure. For example, detection of a label's conflict need compare this label with all other labels in the map. A detection strategies of conflict and overlap based on raster data format was adopted in our research: firstly, any feature was stored as a raster map in advance, where each cell had a code to represent if there are features through it. With these raster map, detection of conflict and overlap just need check the cells within the scope of rectangle outline of a label on corresponding map, therefore the speed of detection would be increased greatly.

However, raster data occupied a large volume space, an encoding scheme based binary bit was applied to get a more compact storage, the encoding scheme see figure 4. Compared with an encoding schema using a binary byte, in the best conditions, the new encoding scheme saved about seven-eighth of storage space. Meantime, by means of bit arithmetic the operation of retrieving the feature information and label information from the bit raster map was the same flexible and handy.



Encoded bit stand for:
0 bit: existed label
1 bit: resident
2 bit: railway
3 bit: major road
4 bit: minor road
5 bit: linear water bodies
6 bit: area water bodies
7 bit: terrain contour

•••••••

Figure 4 binary bit encoding scheme

(3）improved exhaustive searching to find out the optimal placement

The goal of map name placement is searching the validate label positions and find out the optimal one. For a linear feature whose name include n Chinese character, there are two point different from for the point name placement. One is a label position need find n positions for n characters, the another is a valid candidate label position requires all n positions of n characters have no conflict. The searching strategies was introduced as follow:

(a)Find a valid label position: regarding each character of a label as a rectangle, the proper order of placing the character is always from the middle to the two ends, for example, if map name contains odd characters, the middle character was firstly placed, then to the next two characters near the middle one, then to the next two characters near the last two....…;If the name contains even characters, then the middle two characters are placed firstly, then to the next two characters near the middle two, and then to the next two characters near the last two.

(b) Search all valid label position: for a given distance Dist and offset Off (Dist standing for the distance between parallel and the linear feature, Off standing for the offset of the middle point of label to the midpoint of the parallel lines), from on the both left/right or up/down side of the linear feature, search out n group of (for example n = 2) valid label positions. Change two parameters Dist and Off, repeat searching procedure to find more valid positions.

(c)Find the optimal position: While a validate candidate position was searched out, the procedure of detection of overlap was started. A structure called ValidAnno was used to record the status of this candidate position such as overlap information and the priority. Finally, all validate label position were compared with each other, among all the m*n groups of valid candidate positions, the optimal position with the highest priority was found out and output as a result.

(4)Experimental results for linear features

In the study of map name placement for linear feature of 1:250000 topographical map, we experimented the above algorithm on 3 sheets of streams maps, because the number for linear feature was relatively small, although we took the exhaustive searching, after applying the raster and detection strategies to improve it, the efficiency was still satisfactory. In our experiment, the 3 maps include 296、542，350 rivers separately, run time was about 1 minutes. The quality of results of label placement was also satisfactory(see Figure 5).

Figure 5 a part of the result of map name placement for a streams map.

6 THE SOLUTION TO AUTOMATED NAME PLACEMENT FOR POLYGON FEATURES

As well known, regions is a kind of complex geographic feature for map name placement, however, due to its small amount, it is not the emphasis of the research for 1:250000 topographical map. Because of the limited space of this paper, here we only give a simple introduction.

According to the above guideline, 4 cases should be handled separately. For case (b), (c) and (d), existed algorithm can be used to solve these problems. Here the research was focused on case (a):a category of region which is big enough to hold a map name. Typically, water bodies and political or statistical areas belong to this case. For this case, map names should be placed on the sketch line inside the polygon., therefore the key task is to find out the major sketch line for a polygon.

Generally, it is sufficient to find a sketch line based on scan-line midpoint, but it can’t work very effectively when the polygon becomes irregular. In our research, an algorithm to utilize mathematical morphology to find sketch line was proposed. It was proved to have better performance and could handle more complicated situation than the scan-line midpoint method. After sketch lines obtained, it was easy to place name along the sketch line in terms of the name placement algorithm developed for linear feature. The detailed algorithm to utilize mathematical morphology to find sketch lines see the reference (Wan youchuan, 1998).

7 CONCLUSION

In our research, the practical guidelines of automatic map placement for 1:250000 topographical map were firstly put forward, and some automatic map name placement algorithms were proposed and based on it an experimental map name placement system for 1:250000 topographical map was built. Besides, the flexible interactive editing functions were also provided in this system to edit labels and correct label errors after labels being automated placed. This research system experimented on 3 sheets of 1:250000 topographical map, which are near Chengdou city, which is famous for its condensed population in China. The results of experiment showed the quality of map name placement by the above algorithm approached or achieved to the global optimal solution and could compare with the results of map placement of manual work and meantime the speed and efficiency were far superior to manual work.

Of course, map name placement is a very complex problem, as we know, the requirement for name placement differed greatly while maps differ in scale, theme and use. We plan in the future to extend our algorithm suitable for more map scales, more thematic map categories and in the meantime to make our system more powerful so as to solve more problems and handle more cases in this field.

ACKNOWLEDGEMENTS

Thanks for the supporting from Natural Science Fund of P.R.China (No. 40001019) and Surveying Development Fund of the National Bureau of Surveying and Mapping of P.R.China(No.95014).

REFERENCES:

1) Hiersch,S.A.,1982,“An Algorithm for Automatic Name Placement Around Point Data.” The American Cartographer, vol 9., No. 1, pp 5-17
2) Ann,J. And Freeman.H., 1983, “A Program for Automatic Name Placement”, Proceedings of Auto-Carto VI,Vol. 2, pp. 444-453.
3) Lee R.Ebinger and M. Goulette,1990,“Noninteractive Automated Names Placement for the 1990 Decennial Census.”, Cartography and Geographic Information Systems, Vol. 17, No.1 , pp. 69-78
4) Doerschler,J.,and H. Freeman,1989,”An Expert System for Dense-Map Name Placement.” Proceedings of Auto-Carto 9,pp. 215-224
5) Steven Zoraster,1986, “Integer Programming Applied to the Map Label Placement Problem.”, Cartographica , Vol.23, No.3 , pp16-27
6) Cormely,R.G.,1985, “An LP Relaxation Procedure for Annotating Point Feature Using Interactive Graphics.” Proceedings of Auto-Carto 7,pp. 127-132
7) Dacid S. Johnson, Umit Basoglu.,1989,“The Use of Artificial Intelligence in The Automated Placementof Cartogrphic Names”, Proceedings of Auto-Carto 9
8) Freeman,H.,J.Ahn.,1984,“AUTOMAP:An Expert System for Automatic Map Name Placement” Proceedings of the First International Symposium on Spatial data Handling, pp 544-571
9) Imhof, E., 1975, “Positioning Names on Maps”,The American Cartographer, vol 2(2), pp 128-144，.
10) Marks J, Shiebar S.,1991 “The Computational Complexity of Cartographic Label Placement”, Technical Report, Harvard University
11) Yoeli,P.,1972, “The logic of Automated Map Lettering”,The cartographical Journal, vol. 9(2),pp.99-108
12) James E. Mower, 1993 ,“Automated Feature and Name Placement On Parallel Computers”, Cartography and Geographic Information Systems, Vol.20, No.2, pp. 69-82
13) Zoraster S, 1991, “Expert Systems and the Map Label Placement Problem.” Cartographica, vol. 28, no. 1, pp. 1-9
14) Zhang liming, 1995, “Neural Network Models and its Application”，The Press of Fudan Univ.
15) Fan Hong, Zhang zuxun, Du daosheng ,1998, “ Find the optimal position for point feature based on neural network model”，The journal of Wuhan technical univ.,Vol23(1)
(17) Wan youchuan, Fanhong,1998, “Map name placement for polygon features based on mathematics morphology”, The journal of computer and digital engineering, Vol26(2) ,pp11-16