Accurate Calculation Approach for Auto-estimation on Anisotropic and Sporadic Small Area

Huang Kai
School of software engineering
Shanghai Jiao Tong University
Shanghai, China
e-mail: huangkaikiki@sjtu.edu.cn

Qi Zhengwei
School of software engineering
Shanghai Jiao Tong University
Shanghai, China
e-mail: qizhwei@sjtu.edu.cn

***Abstract*—This paper proposes a statistical estimation approach for the sporadic especially antitropic shapes of the source picture. In the field of GIS, the application such as the estimation of the green coverage rate in established districts will need the aid of the area estimation approach. Series of statistical estimation approaches have been put forward for different geographical regions, such as mountainous, open, and forested areas. However, no effective method has been found to function well on sporadic and antitropic regions. In this condition, the error brought out by the border of the sporadic and antitropic regions cannot be neglected, or the accumulation of these errors will come to be considerable. We propose an approach which will bring out no error in the calculation process. Our approach includes the two parts-shape dividing and area integral. The time cost of the shape dividing algorism is lower than the existing algorism in our condition. The adaptation of the integral approach based on the result of the shape dividing algorism can guarantee the accuracy of the statistical estimation result.**

Keywords- SAE(small area estimation), Domain integral, triple integral convex vertex, Barycenter, miss rate

I. INTRODUCTION

Applications of Geographic Information Systems(GIS) such as geographic survey, weather forecast, water source exploration, aerial sowing and artificial precipitation have become extremely prevalent.

Various kinds of maps or pictures such as USGS Topographic maps, aerial photographs, remotely sensed images can easily be achieved by radars, satellites or some kinds sensors with the development of wireless sensor network. However, the crucial problem is the analysis of the Data but not the achievement of it.

Although the auto-dispose and auto-analysis can save human work, however, the accuracy of the result and the cost of the algorism remains a problem which impedes the adaptation of it. The main frustrating difficulty is the error brought out by the calculation process when pictures or maps are combined with sporadic and antitroptic shapes. The error bring out by the irregular boundary can be accumulated and will influence the following mathematical dispose process.

Our approach is designed for the specific condition that the target figure or picture for calculation is full of sporadic and antitroptic shapes and its error rate and time cost lower than other approaches.

The main idea of our approach is to adapt our algorism to divide the sporadic and antitroptic shape into triangular area, and then calculate the area integral on these small triangular fragments for specific purposes.

II. RELATED WORKS

The GIS system has been adapted into various of conditions such as the loss estimation in a earthquake, percentage of forest cover estimation, the extent of desertification, Snow Covered Area Estimation[1-3,5]. Some research works has been done in the area of medical diagnosis or organic tissue auto-analysis[4]. some diagnosis like the area assess of the necrotic tissue well greatly depend on the area estimation approach to give the best treatment. In these application scenes, the statistical estimation approach we proposed in this paper can be adapted which will greatly improve the performance and the accuracy, because there is so many sporadic and antitroptic shapes in these source picture or figure in these scenes.

The whole process of auto-estimation includes the achievement of the source data, the target shape recognition, the basic data analysis, the calculation process, and the statistical analysis for specific application. Our work is mainly related with "the basic data analysis, the calculation process, and the statistical analysis for specific application".

The state-of-art machine learning methods has been adapted in the process of target recognition such as the neural network[17], the support vector machine which are so widely used in pattern match[6]. After the target area is distinguished, the source data must be disposed before statistical analysis. Our work is mainly related with the work of the period.

There have been so many achievements in the estimation work of this kind. So many works are just based on the comparison of the source picture one and two. The TTK is one case of this kind. It includes two steps. At first the source picture is disposed through the canopy compensation[22-23]. Then the pixel wise linear interpolation is performed to compare the difference of the picture[24]. Then estimation result will be achieved. Someone has modified the approach into 2 order (conic) and 3 order (cubic) bivariate polynomials to model the area values in the areagram[25]. It is an implement of 2D interpolation. There have been some works on the squared district estimation[26]. Some other statistical approaches also been analyzed such as the MDLC,MLC,MLCN,MLC&PPC algorisms[19]. All these approaches will be with large error rate in the condition of pictures with sporadic and antitroptic shapes, on the contrary our approach get good performance.

The problem like this is called the SAE(small area estimation) problem. However the mainly finished research work is about the statistical analysis of the predisposed data[21]. Our approach is mainly about how to calculate the accurate integral value on the sporadic and antitroptic shapes in the picture which includes the predispose of the source data and the accurate statistical calculation stage.

Someone has adapted the approach of dividing the graph into convexes and then divides the convex into triangulations[7]. Some other researches has been done in the block shaped picture.

Some research work has been done in convex decomposition even though there are holes in the graph[8]. Of course, the exact convex decomposition problem is NP-hard.

However our approach is, with no doubt, better than the others in the picture full of sporadic and antitroptic shapes combined with the area integral approach we adapt with less time cost and low error rate.

III. TRIANGULATION ALGORISM

A. Original algorism

There have been many approaches in this area triangular-dividing problem. They are widely used in the field of image processing [9].

The most widely used method is first decomposing the concave shape into small convex ones which don't overlap with each other. And then decompose the convex shape into triangular through the dynamic programming algorism [10-12]. Although it also works well with picture full of the sporadic and antitroptic shapes, the cost is too high for a GIS system whose main function is to finish the auto-analysis and data mining of massive data which may be achieved through the satellite photo taking or some newly developed techniques like sensor networks.

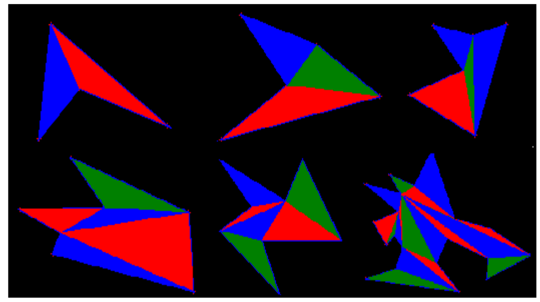

Figure 1. demo graph

Definition:[13]

1. **convex vertex**: draw a line between one point's left neighbor and right neighbor point to form a new shape, if this point is not in the boundary of the new shape, the point is called the convex vertex
2. **n**: The n donates the number of vertexes of one shape.
3. **N** : The N donates the number of convex vertex of one shape.

Our algorism is search through all the vertexes of one shape. If one vertex is a convex vertex with its two neighbor vertexes, then this triangular area defined by the convex vertex and its two neighbor points is removed with the remaining vertexes unchanged. The program flow diagram is displayed below.

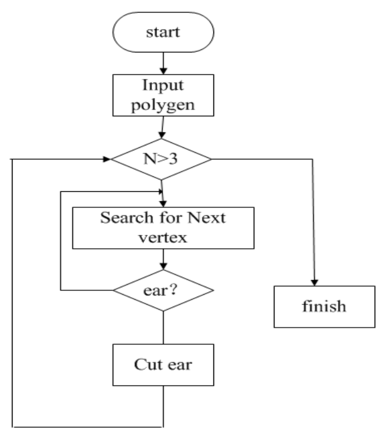

Figure 2. algorism diagram

```
//**********************************
The pseudo code is listed below:
//**********************************
While (finish==false)
{
        While(pointNotFount)
       {
             Ponit=getnextvertex();
```

```
            If(IsConvexVertex(Ponit))
                  pointNotFount=false;
            else
                  pointNotFount=true;
}
If(Ponit!=NULL)
    CutConvexVertex(Ponit);

If(LeftPointCount==3)
     Finish=true;
}
//*********************************
```

The original algorism has great limitation. In it we just search for the next convex vertex in the clock-wise order. The possibility of find a non-convex vertex is so high which greatly influence the performance of our algorism. So we proposed one improved algorism.

B. Improved algorism

Definition:

1. **barycenter**[13]: The point in a system of bodies or an extended body at which the mass of the system may be considered to be concentrated and at which external forces may be considered to be applied. It also is called as barycenter centroid.

2. **centripetal distance** : centripetal distance donates the distance from the point to the barycentor.

3. **miss rate**: it donate the ratio of the condition that finding a non-convex vertex to a convex vertex in searching for the next candidates.

In this improved algorism, we rank the vertex points of the graph by its centripetal distance from high to low. Then we browse all the vertexes to find the convex vertex. If one is found, then the triangular area defined by the convex vertex and its two neighbor points is removed with the remaining vertexes unchanged like the original algorism. As the vertex points in the out-skirt area is more likely to be the convex vertexes, so the performance of the improved algorism can be better. Through comparison, we found the miss rate sharply declined after the modification of the algorism. More detail is with the experiment section.

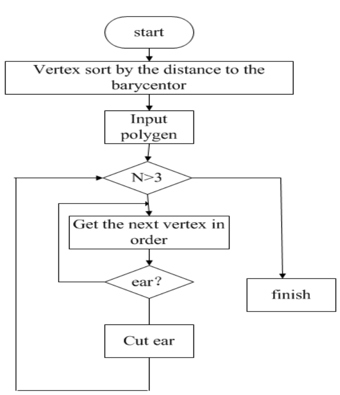

Figure 3. improved algorism diagram

C. Algorism time cost analysis

At first, a comparison between the widely adapted approach and our algorism in the time cost is listed below.

TABLE I. TIME COST COMPARISON

	Time cost		
Convex shape Cut	$O(n+N^3)$		
triangulation	$O(n)$		
all	$O(n+N^3)+O(n)$		
	worst	normal	best
Original triangulation	$O(N^2)$	$O(n)$	$O(n)$
Improved triangulation	$O(N^2)$	$O(n)$	$O(n)$

The traditional algorism contains two parts: 1. convex decomposition, 2. triangulation.

In the phase of convex decomposition, the concave graph is divided into convex shapes, and the time cost is $O(n+N^3)$ [10]. In the phase two, they adapt the dynamic programming approach to divide the convex shape into triangular shapes which only with the time cost of $O(n)$ [11].

Definition:

1. **N/n**: the percentage of concave vertex in the graph. N and n has been defined above.

The main time cost of our algorism is spent on the searching for the convex vertex point. The cost of other part of the algorism is constant level. The performance of the algorism is greatly related with the shape of the graph. It also relates with the N/n rate of the graph. The algorism can have the best performance if it can always find the convex vertex. And the algorism will have the worst performance if it always should browse through the vertex list and find the convex vertex at last.

The best condition time cost is:

$$\sum_{c=1}^{n} 1 \quad \text{n donates the vertex number}$$

The worst condition time cost is:

$$\sum_{c=1}^{n} n - c \quad \text{n donates the vertex number}$$

The average condition should be calculated:

We define p as the possibility that the algorism will find a concave vertex, so the (1-p) is the possibility for the convex vertex..

The possibility of finding the convex vertex through t steps search is [14]:

$$P_t = \left(\frac{N}{n}\right)^{t-1} \left(1 - \frac{N}{n}\right)$$

So the expected value of search step count should be:

$$\sum_{t=1}^{n} P_t * t$$

The Table Ⅰ shows the changes of the expect value according to the n and $\frac{N}{n}$ value.

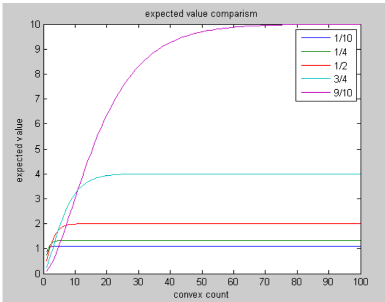

Figure 4. expected step count value comparison

Through the figure, you can learn that the higher the $\frac{N}{n}$ rate, the higher the expect value will be. Nevertheless the expected value all converges to a constant value with $n \to \infty$. On this fact we can know the time cost of our algorism is $O(n)$ rather than $O(n^2)$.

IV. TRIANGLE AREA INTEGRAL APPROACH

The purpose of dividing the concave shape into triangular is to adapt the effective area integral approach which will bring no additional error into the result. Out approach will get the exact integral value of the triangular area.

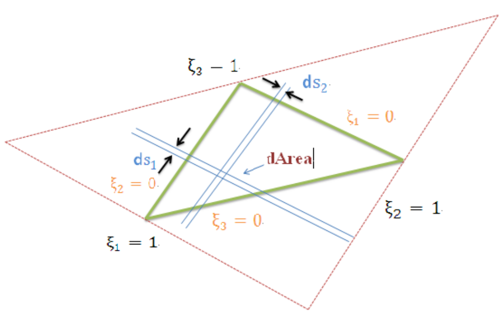

Figure 5. coordination transform demo figure

We transform the canonical coordination to the homogeneous coordination..

$$dArea = ds * \cos(\alpha) * dv$$
$$dArea = h_1 * d\xi_1 * \cos(\alpha) * h_2 * d\xi_2$$
$$dArea = 2 * Area * d\xi_1 * d\xi_2$$

Through the transformation, the double integration are transformed into the $d\xi_1 * d\xi_2$.

The integral variables meets the limitation[16]:

$$\xi_1 + \xi_2 + \xi_3 = 1$$

Then the area integral should be transformed into the double integral of $d\xi_1 * d\xi_2$.

$$I = \int_{Area} \xi_1^a \xi_2^b \xi_3^c \, dArea$$

$$I = 2A * \int_0^1 \int_0^1 \xi_1^a \xi_2^b (1 - \xi_1^a - \xi_2^b) \, d\xi_1^a d\xi_2^b$$

Then variable substitution is performed on the basis of equations below:

$$\xi_1 = e^{\xi_2}(1 - \xi_2) \Rightarrow d\xi_1 = (1 - \xi_1) * de^{\xi_2}$$

Also the integral upper limit and lower limit are transformed on the equations below:

If ξ_1=0 then $\xi_1 = e^{\xi_2}(1 - \xi_2) = 1$
If ξ_1=1 then $\xi_1 = e^{\xi_2}(1 - \xi_2) = 0$

the integral formula after transformation is :

$$I = 2A * \int_0^1 \int_1^0 e^{\xi_2} * (1 - \xi_2)^a * (\xi_2)^b * (1 - e^{\xi_2} * (1 - \xi_2) - \xi_2)^c * (1 - \xi_2) \, de^{\xi_2} d\xi_2$$

$$I = 2A * \int_0^1 \int_1^0 e^{\xi_2} * (1 - \xi_2)^a * (\xi_2)^b * (-(1 - e^{\xi_2}) * ((1 - e)^{\xi_2})^c) * (1 - \xi_2) \, de^{\xi_2} d\xi_2$$

$$I = 2A * \int_0^1 \int_1^0 [e^{\xi_2 * a} * \left(1 - e^{\xi_2}\right)^c][\left((\xi_2)^b * \left(e^{\xi_2}\right)^b\right)^c * (1 - \xi_2)^{(a+c+1)} * de^{\xi_2} d\xi_2]$$

Through it we can get the integral formula. Use it the integral result will be easily achieved. The integral result of this approach through the transformation is very accurate. A further comparison with the common integral approach through the definition of area integral will be illustrated in the next section.

V. EXPERIMENT TEST

A. Triangulation

The experiment is mainly about the improvement between the original algorism and the improved algorism. The miss rate is the main issue of the algorism. The miss rate means the times that the algorism search for a convex vertex but only with a concave one be found. In Figure 6, The red line donates the original algorism performance in the concave graph condition. The blue line demonstrate the improved algorism. The other 2 donate the two algorisms for the convex graph. The source experiment result is listed in the Table Ⅱ and Table Ⅲ.

Theory: a convex graph will remain a convex graph after the removal of a convex triangle.

Due to the theory mentioned above, the miss rate remains zero when the two algorisms are adapted in a convex graph.

From the TableⅡ, Table Ⅲ and figure 6,you can easily find that the performance of improved algorism is better that the original one, although there is some fluctuation in the performance.

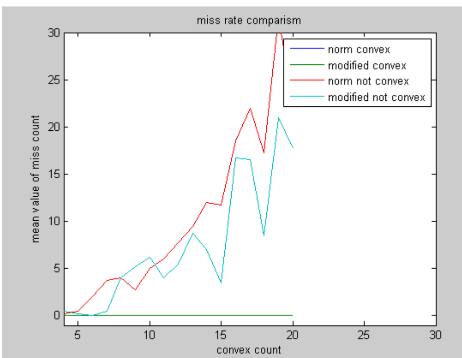

Figure 6. miss rate comparison

TABLE II. MISS RATE OF NORMAL ALGORISM

Vertex count	4	5	6	7	8	9	10	11	12
Test 1	0	1	3	1	3	2	6	5	7
Test 2	1	1	0	2	3	1	4	6	8
Test 3	0	0	4	8	8	7	5	5	9
Test 4	0	0	1	4	2	1	5	8	7
Average	0.25	0.5	2	3.75	4	2.75	5	6	7.75

Vertex count	13	14	15	16	17	18	19	20
Test 1	16	8	13	22	16	13	27	20
Test 2	3	2	8	19	15	17	12	12
Test 3	10	12	19	16	45	29	28	20
Test 4	9	26	7	17	12	10	62	39
Average	9.5	12	11.75	18.5	22	17.25	32	22.75

TABLE III. MISS RATE OF IMPROVED ALGORISM

Vertex count	4	5	6	7	8	9	10	11	12
Test 1	1	0	0	0	4	2	3	8	3
Test 2	1	1	0	1	7	1	10	6	0
Test 3	0	0	0	1	1	13	0	0	13
Test 4	0	0	0	0	4	5	12	2	6
Average	0.5	0.25	0	0.5	4	5.25	6.25	4	5.5

Vertex count	13	14	15	16	17	18	19	20
Test 1	4	8	5	19	17	6	34	9
Test 2	5	1	0	2	28	17	14	13
Test 3	25	8	6	14	0	1	19	45
Test 4	1	11	3	32	21	10	27	4
Average	8.75	7	3.5	16.75	16.5	8.5	21	17.75

Due to this reason, our algorism will not account any miss in a convex graph. Our algorism will get the best performance O(n) in this condition. It can be easily learned from the Figure 6.

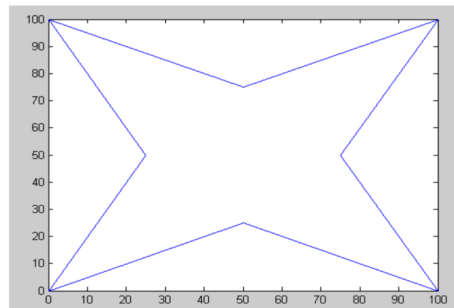

Figure 7. the graph to be tested

B. Integral on triangulation

Our approach has very low error rate. Here we compare it with the traditional integral approach through the definition of the double integral. The result is shown in Figure 8. The test graph is shown in Figure 7. The traditional approach divided the plain contain the graph into dells, And then check whether the cell overlap the polygon area(contains some point in the polygon). If so add the area of the cell to the whole area estimation value. The Table Ⅳ listed the experiment result in estimating the area of the figure displayed in Figure 7. The Table Ⅳ listed the experiment result in estimating the integral value of the area displayed in Figure 7.

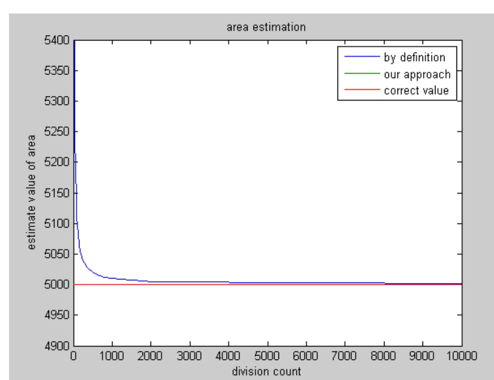

Figure 8. area estimation comparism

TABLE IV. AREA ESTIMATION THROUGH DEFINITION

Cell count	10	20	40	50	70	90	100
Estimated	5800	5450	5237.5	5197	5138	5108	5098
Cell count	120	150	180	200	250	300	350
Estimated	5081.94	5065.78	5054.94	5049.5	5039.68	5033.11	5028.41
Cell count	400	500	600	800	1000	1200	180
Estimated	5024.88	5019.92	5016.61	5012.47	5009.98	5008.32	5005.55
Cell count	2000	10000					
Estimated	5004.99	5001					

The integral test of function “$X^2 + Y^2$” is done by these two approach. The result is shown in Figure 8. Of the traditional approach, the only difference between the area estimation and the integral estimation is to calculate the function value with one point in the cell which overlaps the polygon, and add the multiple result of cell area and the function value into the final result.

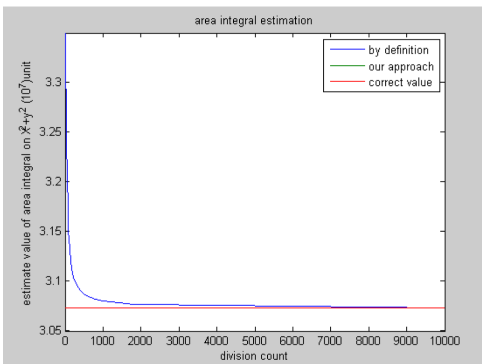

Figure 9. function integral comparison

TABLE V. INTEGRAL ESTIMATION THROUGH DEFINITION

Cell count	10	20	40	50	70	90	100
Estimated	3.534	3.36875	3.23789	3.20767	3.17142	3.1505	3.14305
Cell count	120	150	180	200	250	300	350
Estimated	3.13175	3.12029	3.11257	3.10868	3.10164	3.09691	3.09352
Cell count	400	500	600	800	1000	1200	180
Estimated	3.09097	3.08739	3.08499	3.08199	3.08018	3.07897	3.07696
Cell count	2000	10000					
Estimated	3.07656	3.07365					

The green line is just overlapped with the red one, so you cannot clearly recognize it in the Figure 8 and 9. The accuracy will be proved through the increase of the cell count, however our approach will immediately get the accurate value of it.

VI. CONCLUSION

This paper has presented our approach which combined with the graph decomposition and an accurate area integral approach. Our work is a basis for the future statistical analysis or graphic data mining. The paper has focused on the time cost of our triangulation algorism. And the low error rate of our integral approach. Our work can be summarized as a triangulation decomposition approach and an integral approach on the triangulation area. Our approach has been proved to be a better one than the others in the condition of the sporadic and antitroptic structures in the figure.

In the field of SAE, a lots work remains to be explored. Our work is just a pioneer one. The nowadays approach still cannot dispose the condition with curves in the figure, however curve is very common in the GIS applications. The further work will be focused on this field to estimate the area of an area surrounded by the curve lines.

VII. ACKNOWLEDGEMENT

This work was supported by the National Nature Science Foundation of China under Grant No 60773093. The National Grand Fundamental Research 973 Program of China under Grant No .2007CB316506, the 863 Research and Development Program of China under Grunt No. 2006AA01Z169.

VIII. REFERENCES

[1] Luojus, K., Pulliainen, J., Hallikainen, M “Accuracy Assessment for HUT Snow Covered Area Estimation Method ” .

[2] Kari P. Luojus, Student Member, IEEE, Jouni T. Pulliainen, Senior Member, IEEE, Sari J. Metsämäki, andMartti T. Hallikainen, Fellow. “Accuracy Assessment of SAR Data-BasedSnow-Covered Area Estimation Method, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 44, NO. 2, FEBRUARY 2006

[3] K. Luojus1,*, J. Pulliainen1, S. Metsämäki2 and M. Hallikainen1. “Snow Covered Area Estimation Using Satellite Radar Wide Swath Images”. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 45, NO. 4, APRIL 2007

[4] Pablo Arias, Alejandro Pini, Gonzalo Sanguinetti, Pablo Sprechmann, Pablo Cancela, Alicia Fernández, Alvaro Gómez, and Gregory Randall. “Ultrasound Image Segmentation With Shape Priors: Application to Automatic Cattle Rib-Eye Area Estimation”. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 16, NO. 6, JUNE 2007

[5] K. Luojus':", J. Pulliainen', s. Metsamaki', S.Anttila3 and M. Hallikainen1. “Assimilating Spacebome Radar and Ground-Based Weather Station Data for Operational Snow-Covered Area Estimation”.

[6] Ben Gorte and Alfred Stein “Bayesian Classification and Class Area Estimation of Satellite Images Using Stratification”. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 36, NO. 3, MAY 1998

[7] MARK KEIL “On the time bound for convex decomposition of simple polygons”. International Journal of Computational Geometry applications

[8] Jyh-Ming Lien Nancy M. Amato “Approximate Convex Decomposition of Polygons”

[9] Levente Kova´ cs, Member, IEEE, and Tama´s Szira´ nyi, Senior Member, IEEE “ Focus Area Extraction by Blind Deconvolution for Defining Regions of Interest IEEE” TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 29, NO. 6, JUNE 2007

[10] Chazelle B， Dobkin D "Optimal convex decompositions". Toussaint G． Ed Computational Geometry． Amsterdam : North·

[11] Holland． 1985． 63． 133.

[12] E Anagnostou, D Corneil. "Polynomial-time instances of the minimum weight triangulation problem" - Computational Geometry: Theory and Applications, 1993 - portal.acm.org.

[13] LI JianPing "dynamic programming for the best triangulaion decomposition"

[14] Mark de Berg, Marc van Kreveld "Computational Geometry:Algorithms and Applications" Second Edition

[15] Jay L.Devore " Probability and Statistics for Engineering and the Sciences Brooks/Cole Publishing Company ."

[16] McGraw-Hill Walter Rudin "Principles of Mathematical Analysis" ,Third Edition

[17] Carsten Carstensen "Finite Element Methods",Susanne C. Brenner

[18] Hugh G. Lewis and Mark S. Nixon "Empirical Models For Estimating Land Cover Areas From Remotely Sensed Imagery"

[19] WAYNE WEI-MING DAI "Hierarchical Placement and Floorplanning in BEAR"

[20] Parzi Zeaiean Firouzabadi "Performance Evaluation of Supervised Classification of Remotely Sensed Data For Crop Acreage Estimation"

[21] [20] A Sowter', G Lemoine' and H Van Leeuwen "SAR TECHNIQUES FOR RAPID AREA ESTIMATION OF AGRICULTURAL CROPS DURING AND SPRING WINTER"

[22] Danny Pfeffermann "Small Area Estimation – New Developments and Directions" appears in International Statistical Review, 2002

[23] J. Pulliainen, J. Koskinen, and M. Hallikainen, "Compensation of forest canopy effects in the estimation of snow covered area from SAR data",200l.Proc. IGARSS'01, Sydney, Australia, 9-13 July

[24] K. Luojus, J. Pulliainen, S. Metsamaki, and M. Hallikainen, "Snow Covered Area Estimation Using Satellite Radar Wide Swath Images",Proc. IGARSS'05, Seoul, Korea, 25-29 July 2005.

[25] J. Koskinen, J. Pulliainen, and M. Hallikainen, "The use ofERS-1 SAR data in snow melt monitoring", IEEE Trans. Geosci. Remote Sens. 35:601-610 1997.

[26] Milind S. Shah and Prem C. Pandey "Estimation of Vocal Tract Shape for VCV Syllables for a Speech Training Aid"

[27] WAYNE WEI-MING DAI "Hierarchical Placement and Floorplanning in BEAR"