NEW METHODOLOGY OF REPRESENTING THE POSITIONAL ERROR OF NON-POINT FEATURES IN GIS

H.X. Mao, W.Z. Shi

Department of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong-(Jessie.Mao, lswzshi)@polyu.edu.hk

KEY WORDS: Positional error, Non-point feature, Indicators, Shape similarity

ABSTRACT:

Traditionally, RMSE is used to represent the positional error of GIS features including point, linear and area features. Although RMSE can reflect the error of point features effectively and accurately, there are still some insufficiencies for the RMSE to describe the positional error of linear and area features. In this paper, the limitation of the RMSE for linear and area features has been depicted. Then some new indicators to depict the positional error of linear and area features has been introduced. These indicators include the parameters of area, perimeter, gravity, orientation, shifting, and shape similarity. First, these five parameters are configured and discussed respectively supposing only one factor influent the accuracy of the objects. Second, formulas as well as measurement methods are defined and described. Third, after individual error been measured and estimated, the cumulative error of the whole surveyed area can be integrated and computed. Based on the importance or other reasons, each error can be assigned with a certain weight when integrated them together. Consequently, we can achieve the objective of describing the positional error of GIS data, especially linear and area features, with a method different from traditional RMSE.

1. INTRODUCTION

RMSE is always used to describe the positional accuracy of features. No matter in the research field or industrial field, it is the common and acceptable way to measure the error and estimate the quality of features.

RMSE value is based on the description of points. For point, line, polygon or other ones with different shapes, the difference between the coordination of recorded value and that of universal value is used to define the value of RMSE (Bolstad et. al, 1990; Bogaert et. al., 2005). However, except points, such polyline and polygon, as we can not correctly define all the feature point location of its border, only the RMSE value to assess its quality would be not enough. In other words, an object can be described by its features, such as area, perimeter etc., if we can start with these parameters and define the appropriate measurements, and then we can find another better way to describe the positional uncertainty of different kind of GIS features.

In this study, we challenge the idea that RMSE is always used to describe the positional accuracy of feature, and proposed some other parameters to describe the problem more comprehensively and reasonably. The limitation of RMSE will be presented in section 2. In section 3, we will first discuss these five parameters respectively and then integrate them together with weighted based on their importance and contribution to the feature in section 4. Their fitness function will be defined one by one. Conclusion will be given in section 5.

2. LIMITATION OF RMSE

2.1 Description of RMSE

Based on ASPRS standards (ASPRS, 1990) or the NSSDA statistic, the root mean square error (RMSE) is defined by

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}[(x_{data,i} - x_{check,i})^2 + (y_{data,i} - y_{check,i})^2]}{n}} \qquad (1)$$

where $x_{data,i}$ and $y_{data,i}$ are the coordinates of the i^{th} check point in the dataset; $x_{check,i}$ and $y_{check,i}$ are the coordinates of the i^{th} check point in the independent source of higher accuracy; n is the number of check points tested; and i is an integer ranging from 1 to n.

2.2 Limitation of RMSE

The equation (1) is appropriate for the calculation of the RMSE of the point individuals. Similarly, for example, as the corners of a building are always sampled as check points and used to derive the accuracy measures. Compared the squared differences between the dataset coordinate values and the corresponding coordinate in the reference dataset, the RMSE values of the building corners are thought as the index of the building accuracy. Obviously, this indicator can not address issues concerned with shape similarity.

Suppose the point A, B, C and D are the four corners of a building in a reference map with higher accuracy, and their horizontal coordinate value can be denoted as (x_i, y_i), with i range from 1 to 4. The point A', B', C' and D' are the four corresponding corners of a surveyed building which is the representative of the building in the reference map. Suppose the point A'', B'', C'' and D'' are another surveyed building. The building ABCD, building A'B'C'D' and building A''B"C''D'' are illustrated in the Figure 1.

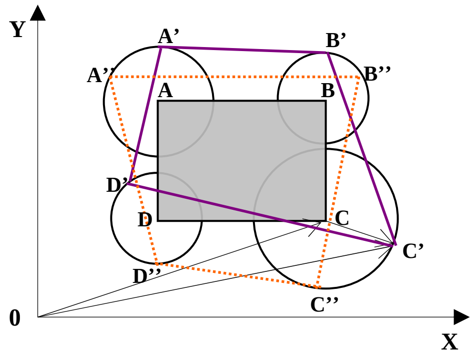

Figure 1. Different Shape of Area Features with same RMSE

Suppose the building A'B'C'D' as one of our object building which will be compared with the building ABCD. When we select the building corners as the check points, we can calculate the horizontal accuracy for the building based on the equation (1). We first need to calculate the difference between the corresponding corner, such as C with C' in Figure 1. Then we summarize all the difference of the corner to calculate the total difference between the two building object. Actually, when we derive the difference between corner C and C' is to measure the length of the vector CC'. However, if we draw a circle with the point C as the center of the circle and with length of the vector CC' as the radius, there are a lot of points on the circle which have the same distance to the point C with the point C', such as the point C''. That is so say, when we use the RMSE value to represent the accuracy of one corner of the building, although the RMSE value is identical, the corner of the building is not unique. That is to say, if there is a building with corner points A'', B'', C'' and D'' corresponding to the point A', B', C' and D' respectively. Based on the equation (1), building A'B'C'D' has the same RMSE with building A''B''C''D''. Consequently, RMSE has limitation to express the buildings' positional accuracy.

For example, in Figure 1, we calculate the RMSE of one surveyed building A'B'C'D', we have

$$RMSE_{A'B'C'D'} = \sqrt{\frac{\sum_{i=1}^{4}[(x_1 - x_1')^2 + (y_1 - y_1')^2]}{4}} \tag{2}$$

Similarly, we calculated the RMSE of the other surveyed building A''B''C''D'', we have

$$RMSE_{A''B''C''D''} = \sqrt{\frac{\sum_{i=1}^{4}[(x_1 - x_1'')^2 + (y_1 - y_1'')^2]}{4}} \tag{3}$$

Obviously, as each pair of corner points are on the circle with the corresponding reference building corner as the center, the distance respective but the same distance as the radius. In other word, as each components of buildings' RMSE are the same, the final RMSE value of two buildings are same. In another word, if the number of the surveyed building corners is the same, we can only get the same results based on the equation (1). Consequently, even though the locations of the surveyed building are different, the accuracy assessment results are same based on such error index.

3. PROPOSED INDICATORS

The corresponding building with some distinction in location, its RMSE value can possibly be no differences. In section 2.2, we have proved the RMSE value of the building corner can not comprehensively describe the horizontal accuracy, so we had better find some other error index as an auxiliary.

The most important limitation of the RMSE index is it can not measure the similarity of the shapes of two buildings. We anticipate the surveyed building polygon can resemble the true reference building polygon as closely as possible. Practically, the corners of the building, the perimeter of the building and the area of the building will show various degrees of dissimilarity in our survey. In the basis of these considerations, we proposed some new indicators to depict the positional accuracy of a building in the following pages.

3.1 Orientation

The indicator of the orientation means the shift of the building angle. Suppose the square area with white color is the reference building, and area with gray color is the surveyed building (Figure 2). There are angle shift in orientation between the two homonymous buildings. In order to define the orientation shift of two buildings, we first have to define the main orientations for a building. Second, we can calculate the shift angles between the corresponding main orientations and obtain the total angle shift of the building.

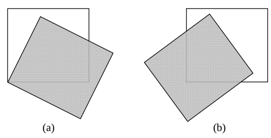

Figure 2. Orientation Shift of a building

Definition of the Main Orientation of a Building: Buildings on a topographic map always have no fixed shape, which means each building has its own outline and are different from the others. We can draw a minimum enclosing ellipse for each building in Figure 3Figure in order to establish the main orientation for the building.

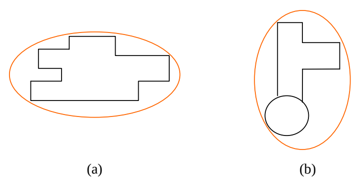

Figure 3. Minimum enclosing ellipse for a building

Normally, in order to establish the main orientation of a building, there are two steps. First, we have to find the gravity of a building. Second, we will draw a minimum enclosing ellipse for a building with the gravity as the center of the ellipse. Then, we look the direction of the long axis of the ellipse as the main orientation of the building.

In some cases, to find the gravity of a building with complex outline is not an easy work to do. Consequently, we will use the minimum enclosing rectangle of the building to instead of the enclosing ellipse as in Figure 4. As we use the minimum enclosing rectangle, we need not to find the gravity for the building, which reducing the great trouble and saving a lot of computation work and making the problem turn easy to solve. Similarly, we look the direction of the length of the rectangle as the main orientation of a building.

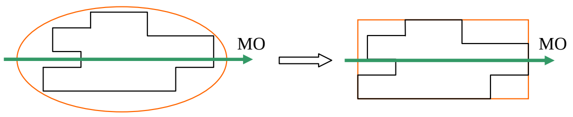

Figure 4. Enclosing rectangle instead of enclosing ellipse

However, if the enclosing ellipse of a building turns into enclosing circle, it can be looked as a special case. However it will have no influence on the drawing of the enclosing square for the building. Although there will be a great difference when deciding the main direction for a square, we can establish the main orientation with the help of the gravity and one identical corner of a building.

In this special case, we first draw the enclosing square for a building. As the width and the length of the square is the same, we have two alternative directions of the square for main direction of the building. However, based on the identical corner point of a reference building and a surveyed building, the main orientation of the building can be decided.

Commonly, we first draw the minimum enclosing square for a building as in Figure 5, and then find two identical building corner points. For example, A is one of the building corners of a reference map, and A' is also one of the building corners of a surveyed building, which is the corresponding identical corner point with A. Meanwhile, O and O' are the gravity of the reference building and surveyed building respectively. We name the direction which is paralleled with the side of the enclosing square as the main orientation of a building, and with the corner A and A' as the reference points respectively.

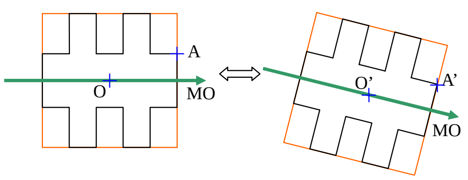

Figure 5. Orientation of the enclosing-square of a reference building

The minimum enclosing rectangle and enclosing square are used to decide the main orientation of a building. Further more, we need to calculate the orientation shift of a building in the basis of the main direction of the buildings.

Computation of Orientation Shift of a Building: If the main orientation of a reference building is denoted as R_A, and the main orientation of the corresponding surveyed building is denoted as $R_{A'}$. A measure of the orientation rotation of the surveyed building is proposed, based on the Euclidean between the main direction of a reference building and the corresponding surveyed building.

$$OR_{AA'} = \sqrt{(R_A - R_{A'})^2} \tag{4}$$

3.2 Positional Accuracy

Position accuracy is a most important indicator of the map, which is really indispensable for many different map users. In Figure 6, we use point O to denote the gravity of the reference building and the point O' to denote the gravity of the surveyed building. From the offset of the two corresponding gravity, we can obviously judge that there is some positional offset between these two corresponding buildings. Commonly, the different positional changes of most corresponding building corners will accompany with the change of the gravity.

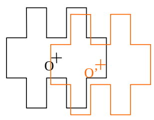

Figure 6. The Positional offset of a building

In many cases, the position accuracy of a building is represented and calculated by the position accuracy of the center point of a building. This method can obtain the position offset of a building to some extent, and it can bring the great convenience for the computation. However, it can not always reflect the true shift of a building in some cases, such as in Figure 7.

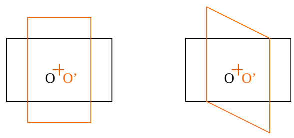

Figure 7. No Position offset between a reference building and surveyed building

No matter the shape of the building is changed or not, only if the center of the reference building and the surveyed building are the same, there will be no position offset. Actually, for some cases listed in Figure 7, we can find there is really some position error between the surveyed data and the reference data. In another word, it does not always work if we use the position accuracy of the gravity to represent the accuracy of a building.

In this study, we prefer to identical building corners rather than gravity to be used to calculate the position accuracy of a building. There are some reasons for this selection as follows:

- Gravity is a special identical point of a building. There is one and only one gravity of a reference building and surveyed building respectively.
- Gravity offset can not reflect the positional change of a building in some cases. The details are introduced and described in Figure 7.
- Gravity of a building with irregular and complex shape can not easy to be find and calculated.

Extraction of the Identical Building Corner: When we use the coordination changes of some identical building corners, we first need to find which two buildings are the same building in reference map and surveyed map. Only after we can distinguish two identical buildings, we can match the building corners. In topographic maps, we always look the corners of a building as our points with identical name and use them to calculate the positional accuracy of a building.

The work flow for the software to judge the building corners with identical name is drafted in Figure 8.

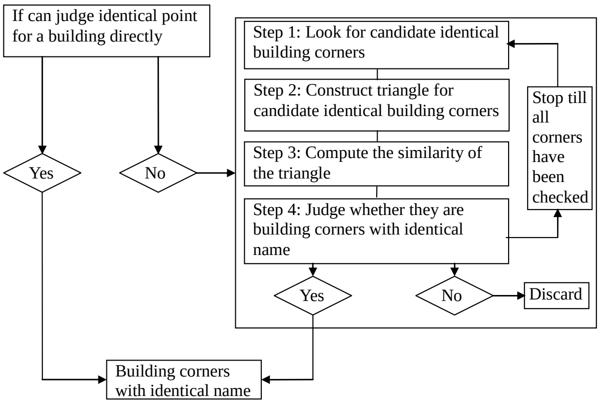

Figure 8. Flow of automatic checking building corners with identical name

Computation of Position Error: If we have a building A in a sample map, which has the building corners (A_1,…A_m), and a building B in a reference map, which has the building corners ($A_{1'}$,…$A_{m'}$), A measure of the position offset of the surveyed building is proposed, based on the Euclidean between.

$$PA_{AA'} = \sqrt{D^2(P_1, Q_1)^2 + ... + D^2(P_m, Q_m)} \tag{5}$$

3.3 Perimeter

The coordinate value of the building corner and the length of all the building sides in a topographic map are recorded in the database, it’s easy for us to obtain the perimeter for a building in sample map and in reference map respectively.

If there is a building A in the sampled map and a corresponding building A’ in the reference map, we can calculate the perimeter difference between the two buildings:

$$PD_{AA'} = \left| Perimeter_A - Perimerter_{A'} \right| \tag{6}$$

3.4 Area

Different buildings always have different outline and shape, however, the area of the geometric graph can be compute by a certain formula. The most important thing is to find the building with identical name. Actually, this information can be obtained from the attribute table corresponding to the building feature. After comparing with the area difference of the building in surveyed map and in reference map, we can get the third error index for a building to be evaluated.

If there is a building A in the sampled map and a corresponding building A’ in the reference map, we can calculate the area difference between the two buildings:

$$AD_{AA'} = \left| Area_A - Area_{A'} \right| \tag{7}$$

3.5 Shape Similarity

The situation of the shape change of a building during the production of a topographic map always happens. In this paper, we have used the area and perimeter difference as the indicators of a building accuracy to evaluate the appearance change of a building. However, these two indicators can neither reflect the shape of a building individually nor unite to depict the shape of the building commonly. Consequently, we use another indicator – Circle to measure the shape change of a building.

Definition of the Circle: Circle of a geometric graph is denoted as the square of the perimeter to the area. P is the perimeter, and A is the area of a certain geometric graph, then C is the circle, we have

$$C = P^2 / A \tag{8}$$

Similarity of Two Buildings: Although the circle can help us to judge the shape change of a certain geometric graph, it can not give a shape similarity index for the feature shape change from one to another. In order to depict the similarity of two shapes, we consider the use and extension of Trigon Onetric Function Distance model to measure the degree of similarity between the buildings in sample map and those in reference map.

The two points x and y in the Trigon Onetric Function Distance model can be regarded as circles of two buildings to be matched. To use and extend this model to describe the degree of similarity between two buildings, we can have some improvement the model as follows:

$$Similarity_{A_i A'_i} = Sin\left[\arctan\left(\left|\frac{1}{C_i - C'_i}\right|\right)\right],\ (i = 1...m) \tag{9}$$

Here, A is the building in a sampled map with the circle C_i, and A’ is the corresponding building in a reference map with the circle C'_i, and the i means the Number of all the buildings in a sampled map, which is from 1 to m.

From the formula (9), we can see if $C_i = C'_i\ (i = 1...m)$, which means the circle value of the buildings are the equal. Meanwhile, we can get the value of the degree of similarity is 1, which means the buildings are completely similar. Usually, as there are shape change between two corresponding buildings, and C_i is not always equal to $C'_i\ (i = 1...m)$, but the closer this two values, the more similar between two building shapes. So, on the basis of formula (9), we can get the value of the degree of the similarity, which range from 0 to 1 for each building pair.

4. INTEGRATED INDICATOR

We can now provide the error expression for a single building from above five aspects, namely orientation rotation, position accuracy, area difference, perimeter difference and circle difference. To evaluate the total accuracy of all the buildings in a sampled map comprehensively, we need to summarize these five measurements of error of all the buildings respectively. And then, we can get the overall accuracy evaluation of the map based on five error indicators.

Five Individual Indicators for All Buildings: If the buildings in a sampled map are denoted as (P_1,…P_m) , and the corresponding buildings in a reference map are denoted as (P_1',…P_m'), then the five total accuracy are listed as follows:

$$\text{Total Orientation of all buildings} = \sum_{i=1}^{m} \left| OR_{P_i P'_i} \right| \tag{10}$$

$$\text{Total Position Accuracy of all buildings} = \sum_{i=1}^{m} \left| PA_{P_i P'_i} \right| \tag{11}$$

$$\text{Total Area Difference of all buildings} = \sum_{i=1}^{m} \left| AD_{P_i P'_i} \right| \tag{12}$$

$$\text{Total Perimeter Difference of all buildings} = \sum_{i=1}^{m} \left| PD_{P_i P'_i} \right| \tag{13}$$

$$\text{Total Shape Difference of all buildings} = \sum_{i=1}^{m} \left| Similarity_{P_i P'_i} \right| \tag{14}$$

In the basis of above five formulas, we can calculate each individual error index for all buildings in a sampled map. Furthermore, we can fill in the following Table with these five total accuracy indicators.

Orientation Rotation	Position Accuracy	Area Difference	Perimeter Difference	Shape Similarity

Table 1. Five error indicators for the building feature of a topographic map

Integrated Indicator for All Buildings: Maybe one or several aspects of these indicators are used to evaluate the error of a topographic map. However till now, they are not as comprehensive as the methodology referred in this study. What's more, we can not only provide the error indicators individually, but also provide an integrated indicator.

In order to obtain the integrated error indicator for all the buildings, we can first construct a vector U= {U1, U2, U3, U4, U5}. Each dimension of the vector represent the total orientation rotation, the total position accuracy, the total perimeter difference, the total area difference and the total shape similarity of a building respectively. To define the integrated error index, we should first arrange the indicator based on its importance according to the users' requirement, and then give a weight for each indicator. Integrated Error Indicator can be evaluated by

$$IEI = a_1 \sum_{i=1}^{m} \left| Similarity_{P_i P'_i} \right| + a_2 \sum_{i=1}^{m} \left| PA_{P_i P'_i} \right| + a_3 \sum_{i=1}^{m} \left| AD_{P_i P'_i} \right| + a_4 \sum_{i=1}^{m} \left| PD_{P_i P'_i} \right| + a_5 \sum_{i=1}^{m} \left| Similarity_{P_i P'_i} \right| \tag{15}$$

Here, a_1, a_2, a_3, a_4, a_5 are the weight of the each error indicator of a building lying on the importance decided by map users or producers.

To emphasize, the integrated error index is defined based on the requirement, and it will vary from one to another with the different definition of each weight. Also, it's a relative value, because it only gives an overall impression for the map user the error of a topographic map. If users want to know the absolute error of a map, they should evaluate the accuracy on the basis of the five individual indicators.

4. CONCLUSION

The object of this paper is to develop an accuracy indicator for area objects in GIS. An area object can be described by its features, such as area, perimeter, gravity, shape, orientation. These features from a vector which represents the features of the area object as the following: {area, perimeter, gravity, shape, orientation} or X={x1, x2, x3, x4, x5}, and assume that the true values of the above features are U={u1, u2, u3, u4, u5}. The difference between X and U form can be used to measure the accuracy of the area feature.

In this paper, we will first discuss these five aspects respectively, research the accuracy indicate supposing only this aspect influent the accuracy of the area objects. Second, we define this accuracy indicate and describe it with a formula as well as measure the value with a certain method. Third, after we have finished measure these five accuracy indicators, we can estimate individually the cumulative error of the whole surveyed area in the basis of the result of the spatial sampling individuals. Last, we have had these five indicators to describe the accuracy of the topographic map from different aspects, however, we'd better integrate them into an integrated accuracy indicator, which can provide a whole and intuitionistic sense about the accuracy of a topographic map for data producers and users.

REFERENCES

American Society for Photogrammetry and Remote Sensing (ASPRS) Specifications and Standards Committee, 1990. ASPRS Accuracy Standards for Large-Scale Maps: Photogrammeric Engineering and Remote Sensing, v. 56, no. 7, pp. 1068-1070.

Bogaert, P., Delincé, J. and Kay, S., 2005. Assessing the error of polygonal area measurements: a general formulation with applications to agriculture. Measurement Science and Technology 16(5): 1170–1178.

Bolstad P.V., Gessler P., and Lillestand T.M., 1990, Positional uncertainty in manually digitized map data. International Journal of Geographical Information System, No. 5, pp: 159-168.

ACKNOWLEDGEMENTS

This work was supported by the funds from The Hong Kong Polytechnic University (Project No. G-YF24 and RGMG).