

图件扫描数字化的误差分析

许捍卫

李纪人

(河海大学土木工程学院 南京 210098) (水利部遥感中心 北京 100761)

摘 要 从图件扫描数字化误差理论分析出发,利用定向点和采样点的自动对中和多种人工采点方法的实验数据,对误差原因及分布规律进行了验证,得出了扫描数字化精度与手扶跟踪数字化精度相比具有一定的优势,对提高大比例尺图件数字化的精度和速度具有借鉴意义。

关键词 图件扫描,数字化,定向点,采样点,精度分析

中图分类号 TP751

当今地理信息系统(以下简称 GIS)市场迅速扩展,而在 GIS 建立中最多的工作就是数字数据的获取。如果获取的数据精度不能满足要求,那么对系统的打击是致命的,所有的数据将成为一堆垃圾,系统将处于瘫痪状态。

GIS 中数字数据的来源主要是对纸质地图的数字化,根据数字化工作原理的不同可分为手扶跟踪数字化和扫描数字化。由于技术原因扫描数字化还未能广泛使用,对此讨论的文章也较少。当数据采集量较大时需要扫描数字化,因此有必要对它的精度情况作一深入的了解,以利于在利用扫描数字化作业时采取适当措施,提高扫描数字化的精度。

1 数据采集过程中的误差来源

1.1 地图固有的误差

除了控制点和碎部点野外测量存在误差外,地图至少还具有下列主要误差:控制点展绘误差,其大小一般为 0.10~0.20 mm;绘图误差,其范围为 0.10~0.18 mm;绘图材料的变形误差,对于不同的绘图材料,具有不同的变形误差,为减少图纸变形,可采用一些变形率小($<0.02\%$)的图纸,如聚脂薄膜等。

1.2 扫描数字化误差

除地图固有误差外,扫描数字化引起的误差还与要素对象、软件处理技术和扫描仪三者有关。

1.2.1 要素对象

要素本身的宽度、复杂程度、粘连以及图面的整洁和清晰程度都对扫描数字化误差有一定的影响。例如,图面不整洁产生了噪声,易引起软件误判;线条的粘连,结合处易出现较大的误差;线条发虚,会得到多个实体等等。

1.2.2 软件处理技术

在扫描数字化过程中,三个主要参数:分辨率、阈值(灰度值或对比度值)和滤波值的确定将对扫描图的质量产生重大的影响,而图像处理、几何校正和矢量化等后处理技术,其功能的强弱、模型的优化将直接影响扫描数字化的精度。如采用合适的校正模型(仿射变换、双线性变换、多项式变换等)定向点的自动对中、采样点的自动对中等,都将有效的提高扫描数字化的精度。

1.2.3 扫描仪

在扫描过程中,由于使用 CCD 扫描仪,会产生一些误差,主要包含有:扫描仪的分辨率;光学误差;电信号传输过程中造成的辐射误差;沿导轨扫描过程中,由于机械运动的速度不均或其它原因所造成的直线性误

差^[1-3]线阵方向与扫描方向不垂直所引起的 CCD 线阵的直线性误差 ;外界因素影响产生的误差 ;随机性误差等等。

在实际操作过程中 ,很难对这些扫描误差一一进行变形改正 ,只能综合考虑它们的影响 ,综合校正。

2 图像坐标向大地坐标数据的转换

扫描后的地图 ,其所形成的是图像数据 ,而我們所需要的结果是大地坐标 .从图像坐标 (x_T, y_T) 转换到大地坐标 (X_D, Y_D) ,一般可采用仿射变换 ,即

$$X_D = a_1 + a_2x_T + a_3y_T ; Y_D = b_1 + b_2x_T + b_3y_T \tag{1}$$

其中 $a_1 \sim b_3$ 为仿射变换转换参数 ,共有 6 个 ,只需三个控制点就可唯一确定转换关系 .当控制点数目大于 3 个时 ,可用最小二乘法确定 ,这有利于提高转换参数的精度 .转换参数一经确定 ,就对所有采样点适用 .因此 ,转换参数的求取精度将直接影响所有采样点的精度 .

如图 1 所示 , a, b, c, d 为图像定向点 ,其图像坐标可用一定方法量测 (如自动对中、人工屏幕采点等) 而得 ,而其大地坐标可从图纸格网上求得 ,这样 ,仿射变换公式可用矩阵表示为

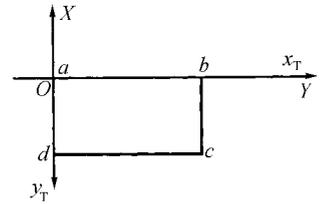


图 1 图像坐标系与大地坐标系的关系
Fig.1 Image coordinate system and gauss coordinate system

$$X = F \cdot A \quad X = \begin{bmatrix} X_{D1} \\ X_{D2} \\ X_{D3} \end{bmatrix} \quad F = \begin{bmatrix} 1 & x_{T1} & y_{T1} \\ 1 & x_{T2} & y_{T2} \\ 1 & x_{T3} & y_{T3} \end{bmatrix} \quad A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

$$Y = F \cdot B \quad Y = \begin{bmatrix} Y_{D1} \\ Y_{D2} \\ Y_{D3} \end{bmatrix} \quad F = \begin{bmatrix} 1 & x_{T1} & y_{T1} \\ 1 & x_{T2} & y_{T2} \\ 1 & x_{T3} & y_{T3} \end{bmatrix} \quad B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

从而 ,可求得 $A = F^{-1} \cdot X \quad B = F^{-1} \cdot Y$ (2)

这样 ,转换参数 A, B 就唯一确定了 ,图像坐标与大地坐标也唯一的一一对应。

3 定向点对大地坐标点位精度影响分析

为不失一般性 ,又便于精度分析 ,如图 1 所示 ,假设 :

$$\begin{aligned} x_{Ta} = y_{Ta} = 0 & \quad x_{Tb} = L_{Tx} & \quad y_{Tb} = 0 & \quad x_{Tc} = L_{Tx} & \quad y_{Tc} = L_{Ty} \\ X_{Da} = Y_{Da} = 0 & \quad X_{Db} = 0 & \quad Y_{Db} = L_{Dy} & \quad X_{Dc} = -L_{Dx} & \quad Y_{Dc} = L_{Dy} \end{aligned}$$

这样 $F = \begin{bmatrix} 1 & 0 & 0 \\ 1 & L_{Tx} & 0 \\ 1 & L_{Tx} & L_{Ty} \end{bmatrix} \quad X = \begin{bmatrix} 0 \\ 0 \\ -L_{Dx} \end{bmatrix} \quad Y = \begin{bmatrix} 0 \\ L_{Dy} \\ L_{Dy} \end{bmatrix}$ (3)

把式 (3) 代入式 (2) ,可求得 $A = \begin{bmatrix} 0 \\ 0 \\ -\frac{L_{Dx}}{L_{Ty}} \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ \frac{L_{Dy}}{L_{Tx}} \\ 0 \end{bmatrix}$ (4)

对仿射变换式 (1) 进行微分 ,有

$$dX_D = da_1 + da_2x_T + da_3y_T + a_2dx_T + a_3dy_T \tag{5}$$

在求转换参数时 , X_D, Y_D 为大地控制点 ,一般假定为无误差 ,因此 $dX_D = 0$.把 a 点图像坐标 (x_{Ta}, y_{Ta}) 及式 (4) 代入式 (5) ,得

$$da_1 - \frac{L_{Dx}}{L_{Ty}} \cdot dy_{Ta} = 0$$

即 $da_1 = \frac{L_{Dx}}{L_{Ty}} \cdot dy_{Ta}$

同理把 b, c 两点图像坐标及式(4)分别代入式(5)经整理后有

$$da_2 = \frac{L_{Dx}}{L_{Tx} \cdot L_{Ty}} \cdot (dy_{Tb} - dy_{Ta}) \quad da_3 = \frac{L_{Dx}}{L_{Ty}^2} \cdot (dy_{Tc} - dy_{Tb})$$

另外有 $dY_D = db_1 + db_2x_T + db_3y_T + b_2dx_T + b_3dy_T$ 把 a, b, c 点坐标及式(4)代入经整理后得

$$db_1 = -\frac{L_{Dy}}{L_{Tx}} \cdot dx_{Ta} \quad db_2 = \frac{L_{Dy}}{L_{Tx}^2} \cdot (dx_{Tb} - dx_{Ta}) \quad db_3 = \frac{L_{Dy}}{L_{Tx} \cdot L_{Ty}} \cdot (dx_{Tc} - dx_{Tb})$$

求得的 $da_1 \sim db_3$ 就是定向点误差引起的转换参数的误差,当采样时,由定向点引起的大地点位误差为

$$dX_D = da_1 + da_2x_T + da_3y_T \quad dY_D = db_1 + db_2x_T + db_3y_T \quad (6)$$

假设地图比例尺为 $1:M$ 扫描分辨率为 DPI (每英寸多少点,其中 $1 \text{ inch} = 0.0254 \text{ m}$) 则有

$$L_{Ty} = \frac{L_{Dx}}{M} \cdot DPI/0.0254 \quad \text{即} \frac{L_{Dx}}{L_{Ty}} = \frac{L_{Dy}}{L_{Tx}} = 0.0254 \cdot \frac{M}{DPI} \quad (7)$$

把转换参数误差 $da_1 \sim db_3$ 及式(7)代入式(6)经整理后得

$$\left. \begin{aligned} dX_D &= 0.0254 \cdot \frac{M}{DPI} \cdot \left[\left(1 - \frac{x_T}{L_{Tx}}\right) \cdot dy_{Ta} + \left(\frac{x_T}{L_{Tx}} - \frac{y_T}{L_{Ty}}\right) \cdot dy_{Tb} + \frac{y_T}{L_{Ty}} \cdot dy_{Tc} \right] \\ dY_D &= -0.0254 \cdot \frac{M}{DPI} \cdot \left[\left(1 - \frac{x_T}{L_{Tx}}\right) \cdot dx_{Ta} + \left(\frac{x_T}{L_{Tx}} - \frac{y_T}{L_{Ty}}\right) \cdot dx_{Tb} + \frac{y_T}{L_{Ty}} \cdot dx_{Tc} \right] \end{aligned} \right\} \quad (8)$$

从式(8)可以看出定向点误差对扫描数据采样的影响 (a)定向点 x 方向误差仅影响大地坐标 Y 向精度,定向点 y 方向误差仅影响大地坐标 X 向精度。(b)扫描分辨率越高,则精度越好,但数据存储量成幂次增加。

当 $M = 1000, DPI = 250, m_a = m_b = m_c = \pm 1$ 个像素时,在最不利条件下,其 $m_{Dx} = m_{Dy} = \pm 0.2 \text{ m}$;在最有利条件下,其 $m_{Dx} = m_{Dy} = \pm 0.1 \text{ m}$ 。这已能满足 GIS 数据精度的要求。

4 采样点误差对数字化结果的影响

由仿射变换式(1)微分得

$$dX_D = a_2dx_T + a_3dy_T = -\frac{L_{Dx}}{L_{Ty}} \cdot dy_T \quad (9)$$

把式(6)代入式(9)得

$$m_{X_D} = \pm 0.0254 \cdot \frac{M}{DPI} \cdot m_{y_T} \quad m_{Y_D} = \pm 0.0254 \cdot \frac{M}{DPI} \cdot m_{x_T} \quad (10)$$

从式(10)可以看出 (a)采样点图像 x 方向误差仅影响大地坐标 Y 向精度,采样点 y 方向误差仅影响大地坐标 X 向精度。(b)其大地坐标误差与采样点在图像中的具体位置无关。(c)扫描分辨率越高,则大地坐标精度也越高,但数据存储量成幂次增加。

当 $M = 1000, DPI = 250, m_{x_T} = m_{y_T} = \pm 1$ 个像素时, $m_{Dx} = m_{Dy} = \pm 0.1 \text{ m}$ 。这当然能满足 GIS 数据精度的要求。

另外,在扫描分辨率 DPI 增大的情况下,并不仅仅是数据存储量成幂次迅速增加,它还将导致后处理时间也成幂次迅速增加,使后处理变得困难重重。因此, DPI 值应较小一些,同时采用一定的技术来提高定向点,采样点的数据精度,从而满足大地坐标的数据精度要求。

5 定向点和采样点数据的自动对中方法

从上面的误差分析可以知道,定向点的量测精度的提高,有利于提高大地坐标的精度。对于定向点,一般采用自动对中算法。对于一般几何图形,找到其几何中心点。如图2所示,在屏幕上拾取一点 P ,以 P 点为中心,上下左右各移动 n 个像素,作垂线,与像素实体相交得 $1 \sim 8$ 点,对应点取中点得 A, B, C, D ,直线 AB 与直线 CD 的交点 E 即为自动对中的定向点坐标,通过自动对中的定向点,其精度可以达到子像素级。

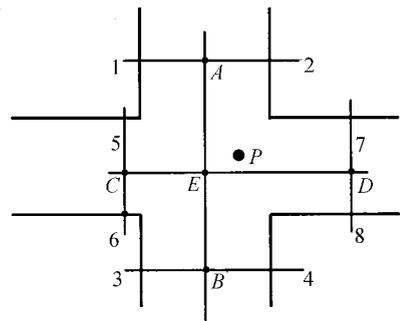


图2 自动对中示意

Fig.2 Automatical alignment of orientation points

对于具体的采样点,就全自动细化矢量化来说,要经过一系列的技术处理.如点状符号,需要符号识别;线面状符号,需要细化处理.这些技术已超出本文内容,欲了解,可参考文献[4 5].但就半自动的扫描数字化作业及人工扫描屏幕拾取来说,可采用自动对中算法,提高采样点的精度.

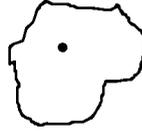


图 3 点状实体
Fig.3 Object of points

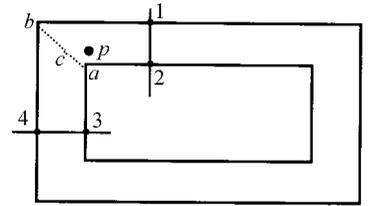


图 4 线面状实体
Fig.4 Object of polylines

点状实体如图 3,可以以所有边界像素点的重心坐标作为采样点的坐标,即

$$x_T = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad y_T = \frac{1}{n} \cdot \sum_{i=1}^n y_i$$

对于线面状实体的点,如图 4 所示的矩形,在屏幕上拾取点为 P,以 P 点为中心,上下左右各移动 n 个像素,作垂线,与像素实体相交得 1~4 点,以 1 A 追踪得一角点 b,以 2 3 两点追踪得一角点 a(可利用直线方程与夹角关系判断求得),ab 连线的中点 c 即为所要求的拾取点.

6 地图扫描数据精度实验分析

为了对扫描数字化结果的精度进行验证,作者利用南京市某 1:500 数字化地籍图一幅(地籍图的数字化坐标已知),用喷墨绘图仪(Nova jet2)按 1:500 绘制,经扫描仪(Micro Tek)按分辨率 300DPI 扫描成图像文件,如图 5 所示.分别以定向点的自动对中方式和人工屏幕采集求得仿射转换参数,进而利用作者自编的程序,由不同的人员进行自动采集、人工采集 1 及人工采集 2,利用 MapInfo 系统人工采集求得各图像点的大地坐标,与地籍图的数字化坐标进行比较,得出一些结论.

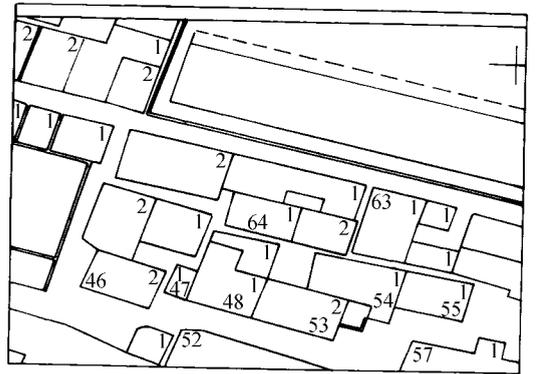


图 5 1:500 局部扫描地图

6.1 控制点精度计算

当控制点数目大于 3 个时,可用最小二乘法求得仿射转换参数,此时可求得控制点转换精度.设大地点实际坐标为 X_D, Y_D , 图像坐标经仿射转换后的大地计算坐标为 X_D^1, Y_D^1 , 则:

$$\Delta_X = X_D - X_D^1 \quad \Delta_Y = Y_D - Y_D^1$$

$$m_X = \sqrt{\frac{[\Delta_X \cdot \Delta_X]}{n - 3}} \quad m_Y = \sqrt{\frac{[\Delta_Y \cdot \Delta_Y]}{n - 3}} \quad m = \sqrt{m_X^2 + m_Y^2}$$

试验对利用 4 个图像点经仿射转换后,得两组大地坐标中误差及图纸精度(见表 1).

6.2 图件扫描数字化的精度计算

由于每一个采样点其地籍图的数字化坐标已知,采样后的图像坐标经仿射转换后得计算坐标,假设坐标表示与上一致,则有

$$m_X = \sqrt{\frac{[\Delta_X \cdot \Delta_X]}{n}} \quad m_Y = \sqrt{\frac{[\Delta_Y \cdot \Delta_Y]}{n}}$$

$$m = \sqrt{m_X^2 + m_Y^2}$$

在本次试验中,表 2 中的位移差为数字化后的图形坐标与原地籍图相对应的数字化坐标的差值,以 3 cm 为间隔进行统计.

表 1 控制点精度情况

Table 1 Accuracy of control points

点号	大地坐标/m		人工采集 1		自动对中	
	X	Y	X	Y	X	Y
1	6759.260	4901.580	213	81	213.3	81.2
2	6764.661	4809.010	334	2289	333.1	2289.0
3	6850.000	4800.000	2348	2505	2348.5	2505.0
4	6870.100	4883.808	2832	507	2831.8	506.2
控制点误差/mm	实际		34.5		13.2	
	图纸		0.069		0.026	

从表 1 2 可以看到：

a. 控制点的自动对中精度为 0.026 mm ,高于人工对点方式的精度 0.069 mm.

b. 运用人工对点方式采样点误差产生的原因 ,一是控制点对点误差的影响 ,二是屏幕采点人员的熟练程度.

c. 采用自动对中方法比人工对点的方法精度提高不显著 ,这是由于一是自动对中的算法还不是很完善 ,二是扫描图的质量不是很高 ,使自动对中算法难以克服某些因素的影响.

d. 采用自动对中算法比人工对点方式节省工作时间 ,利于减轻作业人员劳动强度.

e. 利用 MapInfo 来屏幕扫描数字化精度不高 ,其原因是屏幕图像拖拉作业时 ,图像会局部移动.

表 2 采样点精度情况统计

Table 2 Accuracy of sampling points

采样方法	采样 总点数	时间 /min	采样位移差点数						精度/mm	
			0~3 cm	3~6 cm	6~9 cm	9~12 cm	12~15 cm	15~18 cm	实际	图纸
人工采集 1	252	35	17	62	88	64	21	0	77.78	0.156
人工采集 2	251	30	43	84	83	38	3	0	59.47	0.119
自动对中	232	22	41	83	76	30	1	1	50.15	0.100
MapInfo 采集	249	30	23	83	77	47	10	3	68.03	0.136

7 结 论

a. 采用完善的自动对中算法 ,有利于提高扫描数字化的作业精度 ,同样有利于节省劳动时间、减轻劳动强度.

b. 与文献 4 的手扶跟踪数字化的绝对精度 0.240 mm 相比 ,表 2 中的扫描数字化精度具有明显优势 ,对于 1:500 ,1:1000 的大比例尺图件 ,采用 300DPI~400DPI 为合理分辨率 ,既能满足 GIS 数字制图的精度要求 ,又使图像存储空间不致很大.

参 考 文 献

- 1 姜崇洲,李鹰,黄伟.扫描技术在城市规划信息系统中的应用.城市规划,1998(1):55~58
- 2 黄幼才.GIS 空间误差分析和处理.武汉:中国地质大学出版社,1995:22~35
- 3 花向红,潘正风.基于大比例尺地形图的扫描数字化算法研究.城市勘测,1997(1):22~25
- 4 张桥平,许捍卫.图件数字化采集的精度分析.测绘工程,1997(3):61~64
- 5 许捍卫.GIS 数据采集与管理系统[学位论文].南京:河海大学,1998

Error Analysis of Map Scanning Digitalization

Xu Hanwei

(College of Civil Engineering, Hohai Univ., Nanjing 210098)

Li Jiren

(RS Application Center, Ministry of Water Resources, Beijing 100761)

Abstract On the basis of experimental data of orientation points and sampling points, a detailed theoretical analysis for errors of map scanning digitalization is made. The conclusion of scanning digitalization is more precise than that the tracking digitalization gives. The present study can help increase map scanning digitalization precision and speed.

Key words map scanning; digitalization; orientation points; sampling points; precision analysis