

Run length coding and efficient compression of hexagonal raster data based on Gosper curve

Rui Xin ^a, Tinghua Ai ^{a, *}

^a School of Resource and Environment Sciences, Wuhan University, xinrui@whu.edu.cn, tinghuaai@whu.edu.cn

* Corresponding author

Keywords: hexagonal grids, grid encoding, Gosper curve, data compression

Abstract:

Compared with regular quadrilateral grid, regular hexagonal grid is isotropy and has higher cell compactness and sampling density. This gives regular hexagonal grid advantages in visual display, spatial analysis, and many other aspects. However, the studies of raster data mainly focus on regular quadrilateral grid, and various encoding methods are also focused on it. The researches on hexagonal raster data are relatively insufficient.

In this paper, encoding and compression for regular hexagonal grid are studied. By introducing Gosper curve which has good spatial aggregation and takes into account the morphological structure of regular hexagonal grid, the bidirectional correlation between Gosper curve and regular hexagonal grid is established. Then, a new encoding framework is built to determine the Gosper coding of each grid unit. The lossless compression is completed by performing run-length coding on adjacent coding sets in the target region.

The Gosper fragments in different colour that are cut by the target region in Figure 1 correspond to the adjacent coding sets. These fragments are mapped in order to a one-dimensional colour band and length ratio. The red segments represent the adjacent coding sets in the target region and the white segments represent the non-self adjacent coding sets. The existence of a large number of short white segments blocks the connection of the red segments, resulting in a great quantity of adjacent coding sets in the target region. At the same time, there are many short red segments in the colour band. One short adjacent coding set occupies the same coding amount as long adjacent coding set. All of these factors affect the encoding efficiency but also provide breakthroughs for the reduction in coding amount. The specific implementation of lossy compression is as follows: following the Gosper curve, the attribution of some grid units are changed under certain threshold t to achieve region fusion which can reduce the number of encoding objects. Based on this, the run length coding is carried out again to reduce the coding amount. The colour bands after lossy compression in Figure 2 show that the larger the threshold t , the less the red segments.

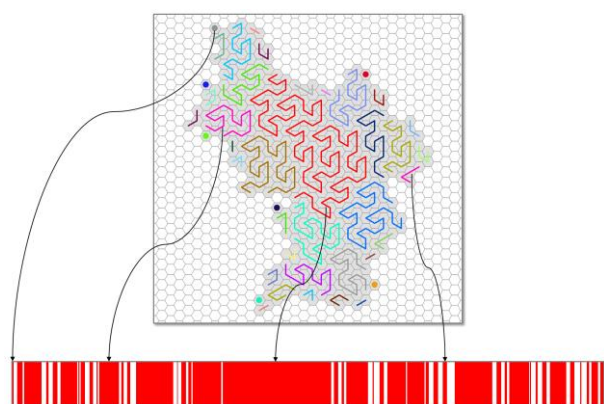


Figure 1. Correspondence between the Gosper curve fragments and the colour band.



Figure 2. Colour bands after lossy compression in different thresholds.

Encoding and compression experiments were carried out on regular hexagonal raster data at multiple resolutions. The apothems of the hexagons in each resolution were 100、 $50 \times \sqrt{2}$ 、50、 $\frac{50}{\sqrt{2}}$ (unit: meter). For lossless compression experiment, the result in Table 1 shows that as the resolution increases, the compression rate decreases and the compression effect becomes better and better. The compression rates of the individual elements under different resolutions were also analysed. As shown in Figure 3, the horizontal axis represents the elements arranged in ascending order by area from left to right. The vertical axis represents the compression rate. It shows that with the growth of the element area, the compression rate presents a decreasing trend. We set three fusion thresholds ($t=1, 2, 3$) to carry out lossy compression experiment at 4 resolutions. The results in Table 2 show that as the accuracy requirement is continuously reduced (i.e., a larger t value is selected), the fusion rate is increased. The lossy compression can even complete the fusion of most adjacent coding sets, and further reduce the coding amount.

Resolution	Compression rate
1	24.95%
2	18.09%
3	12.97%
4	9.25%

Table 1. Overall lossless compression results.

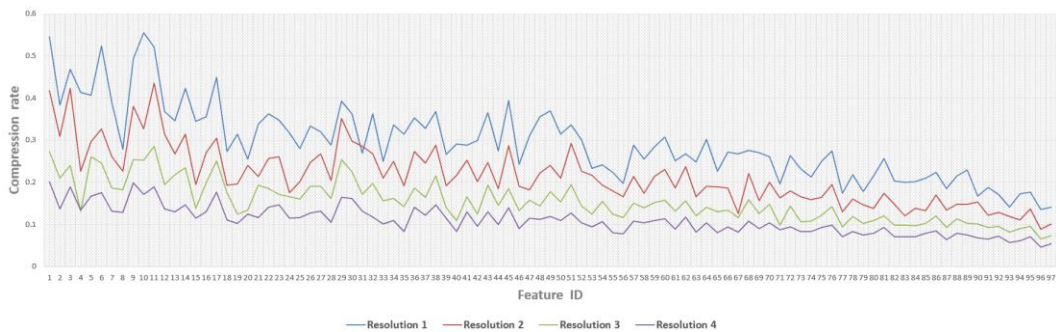


Figure 3. Line diagrams of the element compression rates.

Resolution	Number of adjacent coding sets before fusion	Fusion threshold	Number of adjacent coding sets after fusion	Fusion rate
1	2226	$t=1$	1303	41.46%
		$t=2$	890	60.02%
		$t=3$	719	67.70%
2	3228	$t=1$	1891	41.42%
		$t=2$	1327	58.89%
		$t=3$	1090	66.23%
3	4630	$t=1$	2692	41.86%
		$t=2$	1954	57.80%
		$t=3$	1577	65.94%
4	6664	$t=1$	4051	39.21%
		$t=2$	2843	57.34%
		$t=3$	2378	64.32%

Table 2. Experimental results of lossy compression.

In this paper, a new encoding method is proposed by introducing Gosper curve into regular hexagonal grid, and a new type of compression method is constructed to realize lossless compression and lossy compression. In the future, we will study how to choose the appropriate fusion threshold for different resolutions and precision requirements.