UDC 004.75

ПАРАЛЛЕЛЬНОЕ МАСШТАБИРОВАНИЕ ИЗОБРАЖЕНИЙ В РЕАЛЬНОМ ВРЕМЕНИ НА ОСНОВЕ ТЕХНОЛОГИИ MAPREDUCE HADOOP

Темирбекова Ж.Е., Тюлепбердинова Г.А., Черикбаева Л.Ш., Адилжанова С.А., Толеугазы Б.

Казахстан, г. Алматы, Казахский Национальный университет им. аль-Фараби

zhanerke_3089@mail.ru

Статья посвящена обработке цифровых изображений, в частности задаче повышения разрешающей способности изображений. Рассматриваются методы масштабирования изображений, а также параллельное реализация масштабирование изображений в реальном времени на основе технологии MapReduce Hadoop.

Ключевые слова: интерполяция, бикубическая интерполяция, алгоритм масштабирования, платформа Hadoop, технология MapReduce

PARALLEL SCALE IMAGES IN REAL-TIME BASED ON THE TECHNOLOGY MAPREDUCE HADOOP

Temirbekova Zh.E., Tyulepberdinova G. A. Cherikbaeva L.Sh, Adilzhanova S.A., Toleugazy B.

Article is devoted to digital image processing, in particular the objective of improving the resolution of the images. The methods of image scaling and parallel implementation scale images in real-time based on the technology MapReduce Hadoop.

Key words: interpolation, bicubic interpolation algorithm upscaling, platform Hadoop, MapReduce technology.

Digital image processing finds wide application in almost all areas of industry. Often its use allows to reach a new technological level. Here are the most complex issues related to automatic extraction of information from image and its interpretation, which is the basis for decision-making in the production process control.

The computer image is represented by a finite number of discrete points. Digital Image Scaling associated with bringing an array of information in accordance with the resolution and size of the illustrations. Depending on the purpose of the amount of information you need to either increase or decrease. Simple multiplication or reduction is a simple task for the computer systems. However, the possibility of a significant distortion of the geometry of small parts and the appearance of false patterns on textures. To these losses were minimal in the transformation, it is necessary to use interpolation algorithms. But these methods require to process a large amount of computer time [1].

In order to make the image more with the same resolution, it is necessary to add new pixels therein. The task is to calculate the color of new pixels to be added between the existing ones.

Conventional interpolation algorithms can be divided into two categories: adaptive and non-adaptive. Adaptive methods vary depending on the object interpolation (sharp edges, smooth texture), whereas non-adaptive methods treat all pixels equally [2-3].

Adaptive algorithms include many commercial algorithms licensed programs such as Qimage, PhotoZoom Pro, Genuine Fractals and others. Many of them use different versions of their algorithms (based on analysis of the pixel) when they detect the presence of the border, in order to minimize various defects interpolation in areas where they are most visible.

Non-adaptive algorithms include nearest neighbor, bilinear, bicubic, spline function cardinal sine, Lanczos method. Depending on the complexity, they are used from 0 to 256 (or more) contiguous pixels to be interpolated. The more adjacent pixels are included, the more the image quality will be, but this is achieved at the expense of significant increase processing time. These algorithms can be used for scanning, and to scale the image.

nearest neighbor method. This is the most common of all interpolation algorithms, which requires the least processing time, since one pixel includes only - the closest to the interpolated point. In the method of the nearest neighbor new points are as near-site. The image is heavily pixelated.

The bilinear interpolation. In computational mathematics bilinear interpolation is called extension of linear interpolation for functions of two variables. The key idea is to conduct a conventional linear interpolation first in one direction and then in the other.

Bilinear interpolation considers 2×2 famous square pixels surrounding the unknown. As an interpolated value using a weighted average of these four pixels. The resulting image looks much smoother than the result of the method of the nearest neighbor. With a significant increase in digital images there is a strong image pixelation.

Bicubic interpolation in computational mathematics is an extension of the cubic interpolation in case of a function of two variables, the values of which are given on a regular two-dimensional grid. The surface resulting from bicubic interpolation, is smooth, in contrast to the surfaces resulting from bilinear interpolation or nearest neighbor interpolation. Bicubic interpolation is often used in image processing, giving better picture quality as compared with bilinear interpolation. In the case of bi-cubic interpolation value of the function at the desired point is calculated in terms of its value in the 16 neighboring points.

Bicubic produces much sharper images than the previous two methods, and may be the optimal ratio of processing time and quality of the output.

Significant results can be achieved by parallelizing the computational process. Since successively received pixel blocks are processed completely independent from each other, their parallelization is not complicated, thus it gives considerable acceleration scale [4]. In this paper image scaling algorithm Bicubic interpolation was chosen, which was developed for the parallel embodiment.

The algorithm is divided into two phases, which are determined by the scope of MapReduce technology. Experimental results demonstrate that the proposed algorithm has the scalability and significantly reduces the computation time by increasing the number of nodes.

The algorithm is implemented on Hadoop platform [5], which is a framework of open source, designed to create and run distributed applications that process large amounts of data. Hadoop has the ability to automatically parallelize a sequential program and has its own distributed file Hadoop Distributed File System (HDFS) [6] to store data on the internal drives of servers.

Configuring Hadoop infrastructure carried out as follows. operating system Linux Ubuntu 11.10 has been established, then copied hadoop archive in the prospectus / usr / local, which was extracted, were identified access rights (hadoop-master $ sudochmod 777 - Rhadoop), have been set up configuration files, registered IP addresses and names of all the nodes distributed system, and then start Hadoop (Fig. 1) was produced.

Figure 1. Configure the Hadoop infrastructure

The task consists of 3 classes: Main, Parallel, Image Resizer. Main is the main class. Parallelization of image is performed in Parall class. In class ImageResizer we find the scaling factors for width and length. Conventional 3x3 square aperture loops through all the pixels of the future images. At each step, we turn across the scaling factors to the desired pixel in the original image. 9 Multiply values rounding the vicinity of the current source pixel bicubic factor summarize all this, and paste it in the destination pixel. Smoothing is carried out through the use of bicubic factor.

GUI in Eclipse was developed for visualization of the results. Fig. 2 shows an image to clustering, and Fig. 3 image after clustering. Implementation of clustering tasks were distributed to 6 compute nodes.

Figure 3 - Example for clustering

Figure 4 - Image after clustering

Thus, presented in this paper MapReduce technology can be successfully used for scaling the image with the result will be a significant gain in productivity.

Literature

1. Маркелов К.С. Модель повышения информативности цифровых изображений на базе метода суперразрешения. – М.: ФГБОУ ВПО МГТУ им. Н.Э. Баумана. 2013.
2. George Y. Lu1, David W. Wong, “An adaptive inverse-distance weighting spatial interpolation technique”, Computers & Geosciences, Vol. 34, No.9, 2008, pp.1044-1055.
3. Jonathan A. Greenberg, Carlos Rueda, Erin L. Hestir, etc, “Least cost distance analysis for spatial interpolation”, Computers & Geosciences, Vol. 37, No. 2, 2011, pp. 272-276.
4. Егоров И.В., Внуков А.А. Разработка высокопроизводительного параллельного алгоритма масштабирования изображений. – М.: МИЭМ НИУ ВШЭ, 2012.
5. Venner J. Pro Hadoop. – New York: Apress. 2009. – 442 p.
6. Чак Лэм. Hadoop in work. – М.: DMK Press, 2012. – 424 p.