Title	染色体画像の検索システムの研究
Author(s)	川口, 昌宏
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/651
Rights	
Description	Supervisor:佐藤 賢二, 知識科学研究科, 修士

A Database System for the Management and Information Retrieval of Chromosome Image Data

Akihiro Kawaguchi

School of Knowledge Science,
Japan Advanced Institute of Science and Technology

March 2000

Keywords: chromosome image, FISH, karyotype analysis, Snakes, feature extraction, information retrieval, ODBC

Usually, in many laboratories of biology, researchers have been storing, retrieving, and managing their experimental data with manual operation. Because they tend to keep their knowledge on the experiments covert in order to maintain their potential advantages against their competitors.

However, in many laboratories of biology which are usually doing FISH and GISH experiments, it is important to store, retrieve, and manage large amount of chromosome image data easily. FISH and GISH are the techniques to catch a gene inside a cell visually. A sign has specific genetic arrangement concretely with a fluorescent pigment, and the physical position and distribution of the gene are visualized by fluorescent makers on the chromosome. These experiments are done without discrimination between animals, plants and all kind of eucaryotes. So several thousand sheets of image data accumulated in a year even if they are the laboratories of 2-3 personnel, because it can be done in such small number and in short time. For this reason, conservative way of data management, that is, technicians keep their experimental data in their own PCs, will ruin for thousands or millions of handy reused image data. Furthermore, version management of image data is also needed in such laboratories since karyotype analysis means a series of image processing from raw image data taken by cooled CCD camera to final image data in which all the chromosomes are cropped, directed vertically, and sorted in the order of their size. To assist such a chromosome image data management, we are developing a database system for store, retrieval, and karyotype analysis

---

of them.

The system consists of 3 stages of image data management. The first is image synthesis, The second is chromosome image extraction, and the last is karyotype analysis. A fluorescence microscope and cooled CCD camera provide R, G and B plane gray-scale image data. The image synthesis stage combines 3 gray-scale image data into Full-Color image data. This stage has noise filter and place reviser to get clear image. In the following stage, chromosome images are extracted from original image by “Snakes” that is major method of region extraction using active contour model. There is implemented karyotype analysis and integrated the function into our system on the third stage. Karyotype analysis is one of the major methods in chromosome image analysis that is the serialization of sequence of the extracted chromosome image.

In older to provide easy operation with a PC, this system consists of Windows client and backend DBMS. The client is designed to connect backend DBMS via ODBC (Open Database Connectivity). It can be freely chosen from any DBMSs (e.g. ORACLE, Sybase, etc.) when there are appropriate drivers available. Currently, PostgreSQL is used in our system. In the information retrieval, the system uses two search methods. One is a function of DBMS. DBMS search responds to user's request by SQL, that uses search key about date, species, experimental methods, and so on. The client side search function enables retrieval of similar images. The client performs feature extraction from colors and shapes of each chromosome in a image. Since these two methods are integrated in the user interface, a user does not have to be aware of the difference of retrieval methods. In the analysis of FISH and GISH experimental data, it is usually needed to reprocess the same raw image data for analyzing them from various viewpoints. So the system should be able to keep parent-child relationships in the transformation process of image data for allowing easy navigation and comparison of each experimental data. Using this relationship as a search key, a user retrieves a stack of images derived from the same raw image data.

The system was coded by C++ language and tested on Microsoft Windows OS. There are the following advantages as comparing with existent systems.

- It is adapted to the operation of the chromosome image that is obtained from FISH and so on.
- It is done from injection, reference and processing of the image to the accumulation from

the viewpoint of assembly line.

- Flexible system composition is taken by the corresponding with ODBC.
- Image retrieval system function is done, and a resemblance chromosome can be looked for intuitively.
- Fast turn-around time and high scalability.
- The lowest initial cost (because it is free software!).

Additional implementations such as strengthening of the karyotype analysis function and the image retrieval system function will lead to further improvements of efficiency of the chromosome analysis, and it will reduce the burden of the researchers.

Related research

[1] A. Kawaguchi and K. Satou, A Database System for the Management and Karyotype Analysis of Chromosome Image Data, The Tenth Workshop on Genome Informatics, Tokyo, 1999.