RESEARCH of a COMPUTER SCIENCE PROFESSOR

BY BROOKE HADLEY

Dr. Qi Li

THE RESEARCH CONDUCTED BY DR. QI LI, AN ASSISTANT PROFESSOR OF COMPUTER SCIENCE AT WESTERN KENTUCKY UNIVERSITY, EMBODIES WKU'S VISION OF BECOMING A LEADING AMERICAN UNIVERSITY WITH INTERNATIONAL REACH. DR. LI'S RESEARCH IN THE FIELD OF FEATURE SELECTION AND EXTRACTION WAS RECOGNIZED INTERNATIONALLY AS AN ESSENTIAL SCIENCE INDICATOR IN APRIL 2007 BY THOMPSON SCIENTIFIC, AN ORGANIZATION WHICH ANALYZES RESEARCH PUBLISHED IN OVER 11,000 JOURNALS FROM AROUND THE WORLD TO IDENTIFY THE MOST INNOVATIVE SCIENTIFIC RESEARCH.

Dr. Li began his academic career in China, where he earned a BS in math. In 2000 he moved to the United States, where he obtained a Master's degree as well as a Doctorate in computer science. His previous research has included a comparative study on content-based music genre classification, a survey on wavelet applications in data mining, and two-dimensional linear discriminant analysis.

Attracted to WKU because of "expectations that faculty members are to have a good balance between teaching and research, and good internal funding opportunities," Dr. Li has made Western his home for the past three years.

One of Dr. Qi Li's fields of research involves work with facial recognition systems, a complex computer application that has the ability to automatically identify or verify a person's identity using an image.

One of his fields of research involves work with facial recognition systems, a complex computer application that has the ability to automatically identify or verify a person's identity using an image. The facial recognition software compares selected facial features from a digital or video source in a high-dimensional, or 3-D format, (from a security camera for example) to a known image of a person, such as a passport, visa, or driver's license picture which is in 2-D, or low-dimensional format. Pattern recognition is then used to determine identity based on similarities between the high-dimensional and low-dimensional images. Typically, large amounts of high-dimensional data are generated. The data collected then must go through a process known as data mining to sort the information so that only the specific patterns, or facial features, such as the distance between the eyes, are incorporated in a search of the existing database of known images. However, as the data mining process occurs, the software must be able to recognize features and classify the information received into specific data despite variations in facial appearance due to factors such as light quality, facial size, and distance from an image source (i.e., security camera) through a process known as linear discriminant analysis.

Dr. Li explains, "Previous linear discriminant analysis methods are based on singular value decomposition, or generalized singular value decomposition, which are computationally expensive. Also, many engineering applications involve high-dimensional data that challenges an engineer's hardware configurations, such as CPU, memory, and network. The high computational

By applying QR decomposition, the linear discriminant analysis method we proposed first maximizes the between-class distance, and then minimizes the within-class distance. This process is known as dimension reduction. It is this two-step procedure that leads to the scalability and low computational cost."

Currently, Dr. Li has focused his research efforts on embryonic

pixel-to-pixel correspondence so that the comparison between embryonic images are biologically meaningful (see picture 2)."

Have you ever used digital photography software to edit pictures before printing? Image standardization, although much more difficult a task, is similar. Look again at the two images. The images are actually the same, but appear different because image two has been standardized (or "photo shopped"). Like cropping a picture to remove unwanted background scenery so the primary focus of the image is on the intended subject, Dr. Li has focused the image on the gene intended for study, thus removing any unnecessary and distracting background. He has also sharpened the image to make details more apparent.

"A complete computational system contains the following steps: image standardization, stage classification, clustering, and querying. My current focus"

cost of these previous approaches becomes the bottleneck for research and implementation. Therefore, it is highly desirable to obtain a compact representation for high-dimensional data with minimal computational overhead."

In 2002, Dr. Li began work on a method to reduce the costs involved in the data mining and linear discriminant analysis process while maintaining the reliability of the software and improving efficiency. Through a collaborative effort with Dr. Jeping Ye of Arizona State University, a long time friend and former colleague, Dr. Li began using a method of linear discriminant analysis called QR decomposition, which he explains is "scalable and has a much lower computational cost than previous methods with comparable classification accuracy.

imaging technology, specifically, image standardization. He explains, "Embryonic images contain spatial-temporal gene information, which can help genetic biologists discover gene expression patterns. With the rapidly increasing amount of publicly available embryonic image data, computational methodology was recently introduced to assist genetic biologists to examine an unknown gene expression pattern. A complete computational system contains the following steps: image standardization, stage classification, clustering, and querying. My current focus," continued Li, "is on image standardization. A raw embryonic image may contain the following imaging variations: orientation, size, partial embryos, etc (see picture 1). Image standardization aims to remove these variations to build

Unlike consumer versions of photo editing software which allow users to manipulate images with just a few clicks of the mouse, image standardization used in the bioinformatics field is actually a series of operations which contributes to the overall quality of a microscopic image displayed on a computer monitor. The operations related to image standardization include sample preparation, formation of the image via microscope, digitization of the image through digital cameras, compression of the image, and transmission of the image onto a monitor display. Standardized images allow researchers to compare change over time and identify specific cells or other biologic information (i.e., genes), thus promoting human knowledge of the most basic intricacies of life and conditions that effect the quality of life. The knowledge gained from such research will result in tremendous gains in the areas of genetic, disease, and developmental research. ■
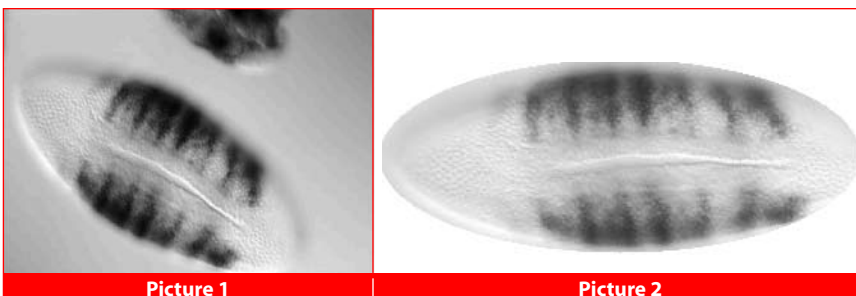

Picture 1 | **Picture 2**

Image standardization aims to remove these variations to build pixel-to-pixel correspondence so that the comparison between embryonic images is biologically meaningful