# An Algorithm Testbed for the Biometrics Grid*

Anlong Ming and Huadong Ma

Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,
School of Computer Sci. and Tech., Beijing Univ. of Posts and Telecommunications,
Beijing 100876, China
anthonyming@gmail.com, mhd@bupt.edu.cn
http://bklab.cs.bupt.cn/

**Abstract.** In this paper, we propose a novel application on grid, the biometrics grid, to promote the development of both biometrics technology and grid computing. The biometrics grid aims to overcome/resolve some main problems of existing biometric technology using grid computing. The most important service provided by the biometrics grid is an algorithm testbed for biometrics researchers on single biometric or multimodal biometrics. We give a case of two respective biometrics recognition processes in voiceprint and face on grid to show that it is feasible in deploying different biometrics applications on a testbed for performance evaluation.

**Keywords:** Grid computing, Biometrics, Face, Voiceprint.

## 1 Introduction

Biometrics usually refers to identifying an individual based on his or her distinguishing characteristics. The premise is that a biometrica measurable physical characteristic or behavioral traitis a more reliable indicator of identity than legacy systems such as passwords and PINs. Physiological biometrics is based on data derived from direct measurement of a body part (i.e., fingerprints, face, retina, iris), while behavioral biometrics is based on measurements and data derived from a human action [1] (i.e., gait and signature). Recent global terrorism is pushing the need for secure, fast, and non-intrusive identification of people as a primary goal for homeland security. As commonly accepted, biometrics seems to be the first candidate to efficiently satisfy these requirements. For example, from October 2004, the United States have controlled the accesses to/from country borders by biometric passports [2, 3].

Biometrics technology not only need advanced biometric technology interfaces but also the ability to deal with security and privacy issues. The integration of

* The work is supported by the National Natural Science Foundation of China (90612013), the National High Technology Research and Development Program of China under Grant No.2006AA01Z304, the Specialized Research Fund for the Doctoral Program of Higher Education (20050013010) and the NCET Program of MOE, China.

C. Cérin and K.-C. Li (Eds.): GPC 2007, LNCS 4459, pp. 447–458, 2007.

biometrics with access control mechanisms and information security is another area of growing interest. The challenge to the research community is to develop integrated solutions that address the entire problems from sensors and data acquisition, to biometric data analysis and systems design. Biometrics technology suffers problems in its way of research and applications:

**Multimodal biomsetrics and information fusion.** The performance of a biometric system is not reliable. This problem can be alleviated by installing multiple sensors that capture different biometric traits. Such systems, known as multimodal biometric systems, are expected to be more reliable due to the presence of multiple pieces of evidence. Use of multiple biometric indicators for identifying individuals has been shown to increase the accuracy and population coverage, while decreasing vulnerability to spoofing [4].Multimodal biometric systems are able to meet the stringent performance requirements imposed by various applications. Moreover, it will be extremely difficult for an intruder to violate the integrity of a system requiring multiple biometric indicators. However, an integration scheme is required to fuse the information churned out by the individual modalities. The key to multimodal biometrics is the fusion of various biometric modality data at the feature extraction, matching score, or decision levels [5].

**Duplicated works and cooperation in diverse fields.** Currently, most biometrics technology researches in offered production are either actually intra-organizational or operated by application domains, such as FaceVACS-SDK produced by Cognitec. It is wasteful with duplicated efforts in building test databases as well as difficulty in providing uniform performance standards. For example, face recognition researchers have spent great efforts in building face databases(i.e., FERET, PIE, BANCA, CAS-PEAL, AR) while these databases are not easily accessed by others. Furthermore, from a technical viewpoint, biometrics spans various technologies, such as fingerprint and face recognition, mathematics and statistics, performance evaluation, integration and system design, integrity, and last but not least, privacy and security. Therefore, there is a need for scientists and practitioners from the diverse fields of computing, sensor technologies, law enforcement and social sciences to exchange ideas research challenges and results.

**Large scale biometric database.** The population in a database can significantly affect performance [6]. In a system with a large scale database, the ordinary recognition processes perform poorly: with the increase of the database scale, the identification rates of most algorithms may decline rapidly; meanwhile, querying in a large scale database may be quite time-consuming. So how to deal with a large scale database has been a difficult problem faced by researchers on biometrics technology in recent years. Su Guangda et al presented a face recognition system framework constructed on the client-server architecture [7]. A distributed and parallel architecture is introduced to this system (see Fig. 1 (a)). The clients and servers are connected by 1000$MB$ networking switch. Although this system has gained good performance: querying one face image in 2,560,000 faces costs only 1.094s and the identification rate is above 85% in most cases, it is limited in accessing and extending due to its C/S framework.

A grid [8, 9] is a high-performance hardware and software infrastructure providing scalable, dependable and secure access to the distributed resources. Grid systems are the gathering of distributed and heterogeneous resources (CPU, disk, network, etc.). Unlike distributed computing and cluster computing, the individual resources in grid computing maintain administrative autonomy and allow system heterogeneity; this aspect of grid computing guarantees scalability and vigor. Therefore, the grids resources must adhere to agreed-upon standards to remain open and scalable. They are promising infrastructures for executing large scale applications and to provide computational power to everyone. In order to promote both biometrics technology and grid computing, we combine biometrics applications with grid computing to give a novel grid application - the biometrics grid (BMG).

The remainder of this paper is organized as follows: related work is presented in Section 2, design issues of system are described in Section 3. The BMG-specific testbed is discussed in Section 4. Finally, we give a case study in Section 5. We conclude our work in Section 6.

## 2 Related Work

Biometric systems have been defined by the U.S. National Institute of Standards and Technology (NIST) [10, 11] as systems exploiting "automated methods of recognizing a person based on physiological or behavioral characteristics" (biometric identifiers, also called features). Biometric systems are being used to verify identities and restrict access to buildings, computer networks, and other secure sites [12]. A biometric system is essentially a pattern-recognition system. Such a system involves three aspects: data acquisition and preprocessing, data representation, and decision-making. Biometric systems are traditionally used for three different applications [13]: *physical access control* for the protection against unauthorized person to access to places or rooms, *logical access control* for the protection of networks and computers, and *time and attendance control*. Due to have been designed for only traditional biometrics applications, biometric systems can't used to solve the problems mentioned in Section 1.

However, the proposed BMG is more than a biometrics system. Considering *multimodal biometrics*, *duplicated works*, *cooperation in diverse fields*, *information fusion* and *Large scale biometric database*, BMG provides an algorithm testbed for the research on single biometric or multimodal biometrics. The testbed enables researchers mainly focus their energy on algorithm design and programming.

Also, BMG can conquer disadvantages of C/S framework because in the heterogeneous grid environments, we can hide the heterogeneity of computational resources and networks by providing Globus Toolkit Services and can implement the distributed parallel computing of a large scale problem by taking full advantage of Internet resources. According to the applied demand, grid MPI parallel program is offered for specialized applications.

# 3 Design Issues of System

## 3.1 Concepts

BMG is designed to develop integrated solutions that address the entire problems from sensors and data acquisition, to biometric data analysis and system design. BMG aims to

1. Provide a testbed for the research on biometrics algorithms. The testbed enables researchers mostly or only pay their attention on algorithm design and programming. BMG would test modules designed by researchers on uniform databases.

   In biometrics algorithms test, such efforts are wasteful, with duplicated work in building test databases as well as difficulty in providing uniform performance standards. A basic requirement is for tools that allow data managers to make licensed and uniform "person" data available to the BMG community. These tools include the means to create searchable databases of persons, provide catalogs of the data that locate a given piece of data on an archival system or online storage, and make catalogs and data accessible via the Web. Prior to the advent of the grid, these capabilities did not exist, so potential users of the model data had to contact the data managers personally and begin the laborious process of retrieving the data they wanted.
2. Create a virtual collaborative environment linking distributed centers, users, models, and data to simplify both the resource management task, by making it easier for resource managers to make resource available to others, and the resource access task, by making biometrics data as easy to access.
3. Support mature biometric applications with different QoS demands including applications with large scale databases or applications of multi-modal biometrics can be solved by grid computing. However, BMG does not guarantee that biometrics applications are meeting with the QoS goals, when defining QoS more broadly than the bandwidth and capacity.
4. Develop a specialized grid workflow for multimedia computing and data mining in biometrics applications on BMG.

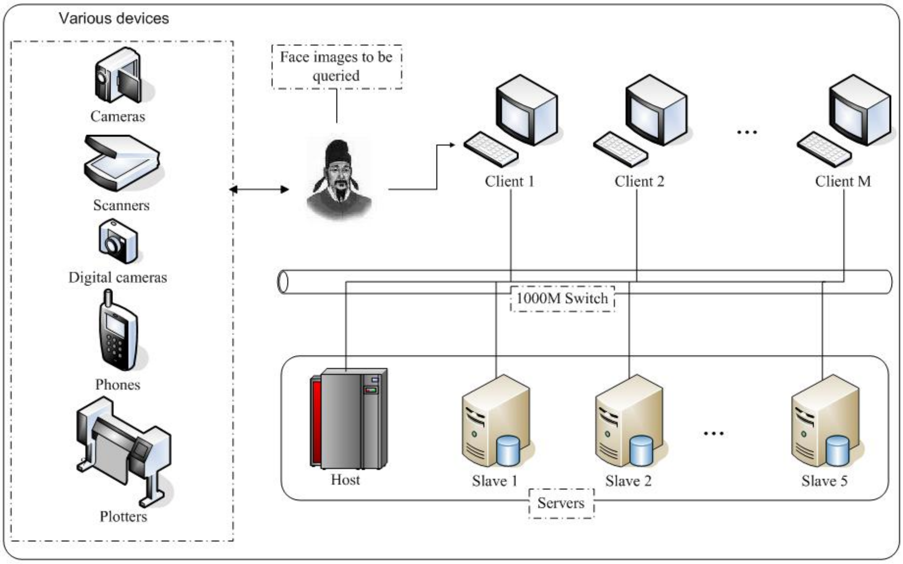
In this paper, we only discuss one of the BMG issues, the algorithm testbed.

## 3.2 A Framework of the Biometrics Grid

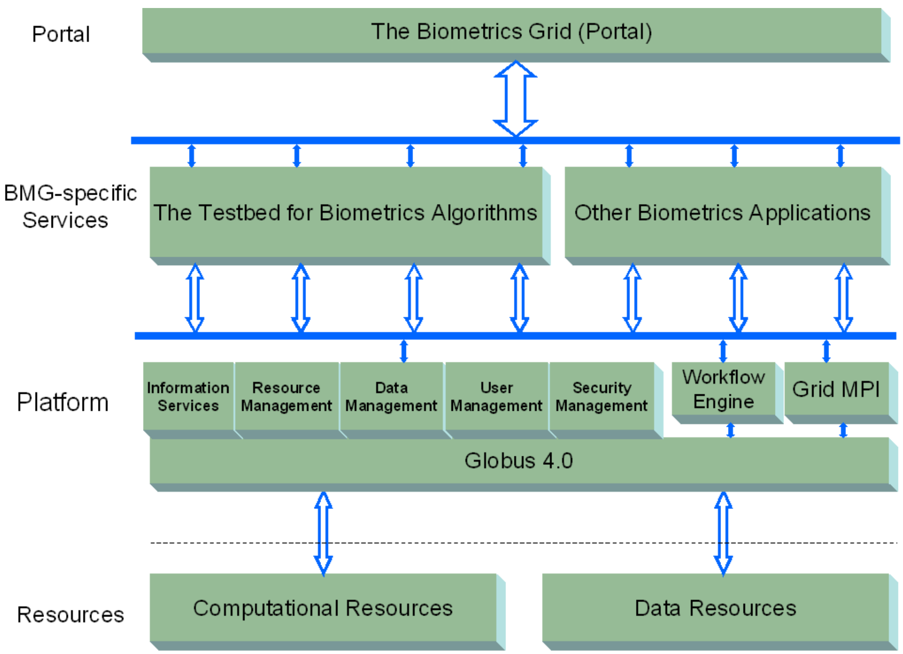
We present a description of the BMG framework in Fig. 1 (b). BMG is divided into four layers:

**Resources.** These are the basic resources on which BMG is constructed including computational resources and data resources.

**Platform.** This provides remote, authenticated access to shared BMG resources. All these components are based on the Globus Toolkit and the WS-Resource




(a)




(b)

**Fig. 1.** a) An example of C/S framework to support large scale database. (b) The BMG framework schematic showing four layers.

Framework (WSRF) which enables the discovery of, introspection on, and interaction with stateful resources in the standard and interoperable ways.

**BMG-specific services.** The testbed is the most important of these biometrics applications. The testbed enables researchers intend to focus their energy on algorithm designing and programming. All biometrics applications are wrapped to Web Services specified by WSRF and deployed into Web Services container.

**Portal.** Web portal control and render the user interface-interaction. BMG creates a virtual collaborative environment which provides advantages to urge co-operations in diverse fields. Generally, portal let you take multiple Web pages, automatically produce controls to link between them, and let subsets of them be displayed on a single Web page. All biometric applications are wrapped to Web Services (each with user-facing ports) are aggregated for the user into a single client environment. We assume that all data and information presented to users originates from a Web Service, called a content provider. This content could come from a simulation, data repository, or stream from an instrument. Each Web Service has resource- or service-facing ports that communicate with other services [14]. However, we are more concerned with the user-facing ports, which produce content for users and accept input from client devices.

### 3.3 The BMG Workflow for the Algorithm Testbed

The BMG workflow is simply defined as a set of Grid resources and services, a quality expectation defined by the user(s) and a workflow model acting on them.

The BMG workflow pays more attention to multimedia computation and data mining in biometrics applications on BMG, the BMG-MPI parallel programming interfaces are offered for the BMG testbed to run algorithm jobs. Its design sustains and integrates closely with parallel processing from the bottom, so it can be applied in different applications.

Further, the BMG workflow has strong self-adaptability to effectively overcome the dynamic variation during the operating process of a biometrics algorithm, and the BMG workflow engine can also perform dynamic resource discovery and allocation, dynamically collects the status of nodes of BMG by MDS modules in Globus.

## 4 A Testbed for Biometric Algorithms

To solve the problems of *duplicated works*, *multimodal biometrics*, and *information fusion*, BMG provides the testbed for the research on single biometric or multimodal biometrics to enable researchers intend to only focus their energy on algorithm designing and programming. For example, researchers' works are limited to design the modules of feature extracting, feature matching, information fusion, etc. Researchers code these modules according to the testbed interface description and then submit their works to BMG.

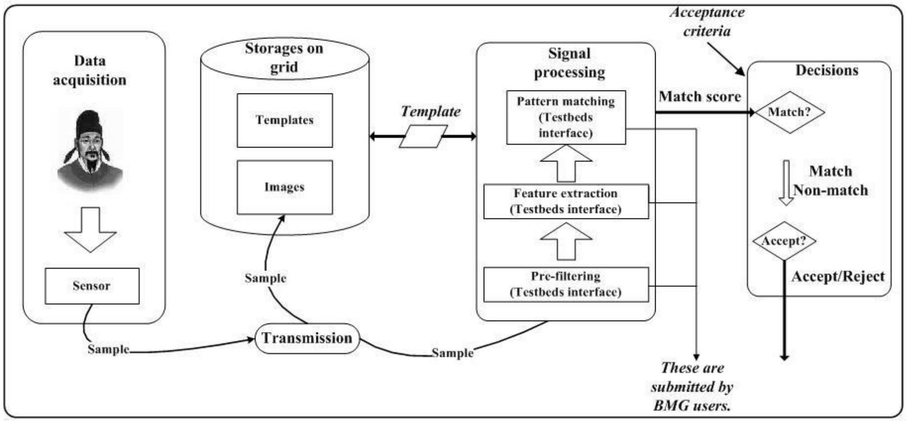



**Fig. 2.** Structure of a general single biometric system

### 4.1 Single Biometric Test

A biometric system has a general structure [15]. First of all, a sensor acquires a sample of the user presented to the biometric system (i.e., fingerprint, face, iris images). As defined in [15], a sample is a biometric measure presented by the user and captured by the data collection subsystem as an image or signal. The sample can be transmitted, eventually exploiting compression/decompression techniques. BMG stores the complete sample data in the storage unit. BMG uses and stores only a mathematical representation of the information extracted from the presented sample by the signal processing module that will be used to construct or compare against enrolment templates: the biometric feature. If the extracted feature is stored (enrolled) into BMG, a template for future identification or verification (matching) is added. BMG has a measure of the similarity between features derived from a presented sample and a stored template. The measure produces a typical index called matching score. Hence, a match/nonmatch decision may be made according to whether this score exceeds a decision threshold or not. The term transaction refers to an attempt by a user to validate a claim of identity or nonidentity by consecutively submitting one or more samples, as allowed by the system decision policy [16]. Lastly, a transmission process is implemented to transmit the collected data to the signal processing section. The signal-processing module represents the core of the system and is generally composed by sub-modules implementing preprocessing functions (i.e., image filtering and enhancement), the feature extraction, and the matching between two features.

BMG deploys this general single biometric system on the testbed. Of course, some definitions should be firstly done such as feature extracting interface, feature matching interface, pre-filtering interface. All these definitions together are defined as part of the testbed interface description. For example, a simple feature extracting interface can be defined as $c$ executable file (e.g. *FeatureExt.exe*) with a parameter (e.g. a file name of a sample), *FeatureExt.exe* can be invoked by command line mode as follows:

$$FeatureExt.exe \quad a\_sample\_file\_name$$

When a user of BMG submits *featureExt.exe* to the BMG Web portal, *featureExt.exe* itself will be wrapped into a Web Service specified by WSRF. Then BMG can provide this service as a part of the testbed using BMG components (e.g. GRAM).

### 4.2 Multimodal Biometrics Test

Multimodal biometrics fusion that is possible when combining multiple biometric systems:

① Fusion at the feature extraction level, where features extracted using multiple sensors are concatenated.
② Fusion at the confidence level, where matching scores reported by multiple matchers are combined [15,16].
③ Fusion at the abstract level, where the accept/reject decisions of multiple systems are consolidated [17].

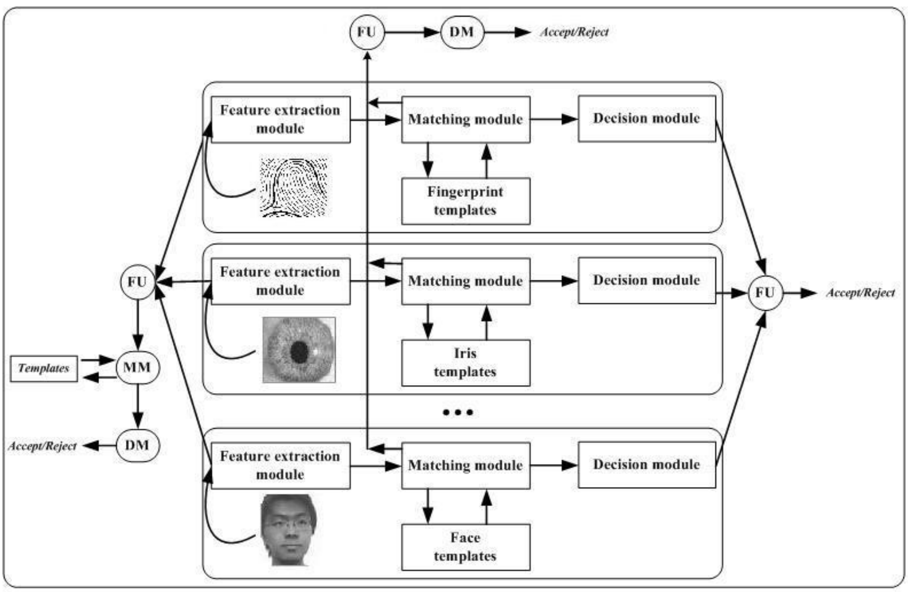



**Fig. 3.** Structure of a general multimodal biometrics system showing the three levels of fusion; FU: Fusion Module, MM: Matching Module, DM: Decision Module. FU, MM, DM are programed and submitted by BMG users.

Fusion in the context of biometrics can take the following forms: ① Single biometric multiple classifier fusion, where multiple classifier on a single biometric indicator are combined [18]. ② Single biometric multiple matcher fusion, where scores generated by multiple matching strategies (on the same representation) are combined [19]. ③ Multiple biometric fusion, where multiple biometrics are utilized [20, 21, 22]. An important aspect that has to be dealt with is the normalization of the scores obtained from the different domain experts [23]. Normalization typically involves mapping the scores obtained from multiple domains into a common framework before combining them. This could be viewed as a two-step process in which the distributions of scores for each domain is first estimated using robust statistical techniques and these distributions are then scaled or translated into a common domain.

Also, BMG deploys this general multimodal biometric system on the testbed just like that mentioned in single biometric.

## 5 A Case Study

### 5.1 The Environment

In the case, we carry out two respective biometrics recognition processes for voiceprint and face on grid. The voiceprint recognition approach we used is described in [24]. The face recognition approaches we used are listed as: the line based face recognition algorithm [25], the improved line based face recognition algorithm [26], PCA and PCA+LDA [27]. Our development OS is Linux Fedora Core 4, and the development toolkit is Globus Tookit 4.0, Web server platform is Apache Tomcat 5.0, DBMS is MySQL 5.0, the development languages are HTML, JSP, Servlet, Java Bean, Java class and XML.

In voiceprint recognition, 24 samples from 44 persons are collected. The first 20 samples are put in the training set, and 4 samples left are made as the test set.

In face recognition, we use a face database established by ourselves to evaluate the performance of our algorithm. Pictures of 35 persons are taken by a standard camera (6 pictures per person) under different illumination intensity (weak, medium and strong). We select 3 views of each person for training, and the other 3 views (in weak, medium, and strong illumination intensity respectively) is used to test.

### 5.2 Two Biometrics Recognition Processes

We define 3 simple interfaces, which are executable files of *c* language in Linux platform, to run two respective biometrics recognition processes in speech and face.

– Interface 1 *Training.exe*, an executable file for biometrics data training, can be invoked as follows:

$$Training.exe\ \ samples$$

– Interface 2 *FeatureExt.exe*, an executable file for extracting feature vectors using training results, can be invoked as follows:

$$FeatureExt.exe \quad a\_samples$$

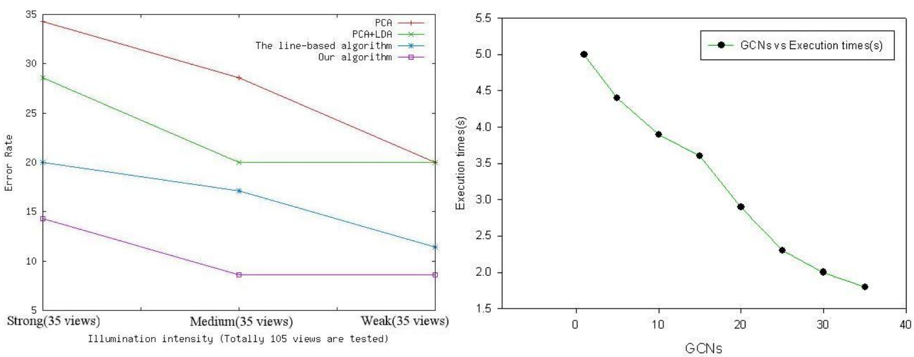
– Interface 3 *FeatureMat.exe*, an executable file for matching feature vectors between two samples, can be invoked as follows:

$$FeatureMat.exe \quad sample\_1 \quad sample\_2$$

We program each recognition method and build $3 \times 2$ exe files respectively, then these files would be submitted to grid by GRAM Server and RSL (XML file) for recognition.

In voiceprint recognition, one is selected among 44 persons and tagged as *imposter*, and 43 persons left are seen as *client*. Every person can enter at his own status, *imposter* tries to enter at other 43 persons' status and repeats 20 times. Then we get $44 \times 43 \times 20$ verification results. In our voiceprint recognition job, the FAR (False Accept Rate) is 0.092%, FRR (False Reject Rate) is 2.27%.

In face recognition, we have tested four face recognition methods on the same, but individually processed, face database. The performance of different algorithms in face recognition is shown in Fig. 4 (a). Moreover, as illustrated in Fig. 4 (b), the average execution times of the improved line based face recognition algorithm can be shorten by increasing the number of grid computation nodes.




(a) (b)

**Fig. 4.** (a) The performance of different algorithms on error rate (Totally 105 pictures are tested). (b) The relation between the execution times and the numbers of grid computation nodes (GCNs) of the improved line based face recognition algorithm.

### 5.3 Analysis

According to results of the case study above, we can conclude that:

– It's feasible to deploy biometrics applications on BMG.
– The algorithm testbed of BMG can provide uniform interfaces to different algorithms belonged to different types of biometrics.

– The algorithm testbed of BMG can provide uniform interfaces to different algorithms belonged to a same type of biometrics.
– BMG can meet with some QoS demand (i.e. execution times) by using methods such as increasing computation nodes.

## 6 Conclusions

We propose a concept of BMG to simplify both the resource management task, by making it easier for resource managers to make resource available to others, and the resource access task, by making biometrics data as easy to access as Web pages via a Web browser. BMG would test modules designed by users based on uniform database, modules would be wrapped to Web Services based on WSRF and deployed into Web Services container. We give a case study about two respective biometrics recognition processes in voiceprint and face deployed to grid. The results show that it is feasible in deploying not only algorithms belonged to different types of biometrics (i.e., face recognition, voiceprint recognition) but also different algorithms belonged to a same type of biometrics (such as face recognition) on grid to provide services using grid computing. Also, the time-consuming algorithms can be shortened by grid computing.

However, there exist great difficulties in building BMG nowadays. For example, it is not an easy case to build an uniform biometrics database because there are fears of an invasion of privacy. The advent of BMG should be under the legal guidelines of governments. With the development of grid computing, the technical scheme of BMG will also be improved.

## References

1. R. Bolle, S. Pankanti, and A. K. Jain: Guest editorial, IEEE Computer (Special Issue on Biometrics), vol. 33, No. 2, (2000) 46-49
2. S. Waterman: Biometric borders coming, Times, Washington (2003)
3. General Accounting Office (GAO).: Technology assessment: Using biometrics for border security, GAO-03-174, Washington, (2002)
4. A.K. Jain, R. Bolle, et al.: Biometrics: Personal Identification in Networked Society, Kluwer Academic (1999)
5. D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar: Handbook of Fingerprint Recognition, Springer (2003)
6. PJouathou Phillips, Patrick Grother, et al.: Face recognitiou vendor test 2002: Evaluatiou report, Audio- and Video-Based Person Authentication (AVBPA) (2003)
7. Kai Meng, Guangda Su, et al.: A High Performence Face Recognition System Based on A Huge Face Database, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou (2005)
8. Satoshi Matsuoka, et al.: Japanese computational grid research project: NAREGI, Digital Object Identifier , Vol. 93, Issue 3, (2005) 522–533
9. David Bernholdt, et al.: The Earth System Grid: Supporting the Next Generation of Climate Modeling Research, Digital Object Identifier, Vol. 93, Issue 3 (2005) 485–495

10. M. Gamassi, Massimo Lazzaroni, et al.: Quality Assessment of Biometric Systems: A Comprehensive Perspective Based on Accuracy and Performance Measurement, IEEE Transactions on Instrumentation Measurement, Vol. 54, No. 4, (2005)
11. R. Bolle, S. Pankanti, and A. K. Jain: Guest editorial, IEEE Computer (Special Issue on Biometrics), vol. 33, No. 2, (2000) 46-49
12. J. D. M. Ashbourn: Biometrics: Advanced Identify VerificationThe Complete Guide, Springer-Verlag, Berlin (2000)
13. R. Norton, The evolving biometric marketplace to 2006, Biometric Technology Today, vol. 10, No. 9, (2002) 7C8
14. Geoffrey Fox: Grid computing environments, Digital Object Identifier, Vol.5(2) (2003): 68-72
15. A. J. Mansfield, J. L. Wayman: Best practices in testing and reporting performance of biometric devices, National Physical Lab., Center for Mathematics and Scientific Computing (2002)
16. Arun Ross, Anil Jain, Jian-Zhong Qian: Information Fusion in Biometrics, 3rd International Conference on Audio- and Video-Based Person Authentication (AVBPA), Sweden (2001) 354-359
17. Y. Zuev and S. Ivanon: The voting as a way to increase the decision reliability, in Foundations of Information/Decision Fusion with Applications to Engineering Problems, Washington (1996) 206–210
18. R. Cappelli, D. Maio, and D. Maltoni, Combining fingerprint classifiers, in First International Workshop on Multiple Classifier Systems, (2000) 351-361
19. A. K. Jain, S. Prabhakar, and S. Chen: Combining multiple matchers for a high security fingerprint verification system, Pattern Recognition Letters, vol. 20 (1999) 1371 - 1379
20. J. Kittler, M. Hatef, R. P. Duin, and J. G. Matas: On combining classifiers, IEEE Transactions on PAMI (1998) 226-239
21. E. Bigun, J. Bigun, B. Duc, and S. Fischer: Expert conciliation for multi-modal person authentication systems using bayesian statistics, in First International Conference on AVBPA, Crans-Montana (1997) 291-300
22. S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz: Fusion of face and speech data for person identity verification, Research Paper IDIAP-RR 99-03, Switzerland (1999)
23. R. Brunelli and D. Falavigna: Person identification using multiple cues, IEEE Transactions on PAMI, vol. 12, (1995) 955-966
24. Liu Y: Research on identity verification system based Institute of Automation, Chinese Academy of Sciences on voiceprint and semanteme[Mnater dissertation], Beijing, china, (2002)
25. O.de Vel and S.Aeberhard.: Line-based face recognition under varying pose, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21 (1999) 1081-1088
26. Anlong Ming, Huadong Ma: An improved Approach to the line-based face recognition.pdf, Proceedings of the 2006 IEEE International Conference on Multimedia and Exposition (ICME), Toronto, 2006
27. Zhao WY, Chellappa R, Phillips PJ, Rosenfeld A: Face recognition: A literature survey, ACM Computing Surveys, Vol. 35 (2003) 399-458