SVC2004: First International Signature Verification Competition

Dit-Yan Yeung[1], Hong Chang[1], Yimin Xiong[1], Susan George[2], Ramanujan Kashi[3], Takashi Matsumoto[4], and Gerhard Rigoll[5]

[1] Hong Kong University of Science and Technology, Hong Kong
[2] University of South Australia, Australia
[3] Avaya Labs Research, USA
[4] Waseda University, Japan
[5] Munich University of Technology, Germany

Abstract. Handwritten signature is the most widely accepted biometric for identity verification. To facilitate objective evaluation and comparison of algorithms in the field of automatic handwritten signature verification, we organized the First International Signature Verification Competition (SVC2004) recently as a step towards establishing common benchmark databases and benchmarking rules. For each of the two tasks of the competition, a signature database involving 100 sets of signature data was created, with 20 genuine signatures and 20 skilled forgeries for each set. Eventually, 13 teams competed for Task 1 and eight teams competed for Task 2. When evaluated on data with skilled forgeries, the best team for Task 1 gives an equal error rate (EER) of 2.84% and that for Task 2 gives an EER of 2.89%. We believe that SVC2004 has successfully achieved its goals and the experience gained from SVC2004 will be very useful to similar activities in the future.

1 Introduction

Handwritten signature verification is the process of confirming the identity of a user based on the handwritten signature of the user as a form of behavioral biometrics [1–3]. Automatic handwritten signature verification is not a new problem. Many early research attempts were reviewed in the survey papers [4, 5]. The primary advantage that signature verification has over other types of biometric technologies is that handwritten signature is already the most widely accepted biometric for identity verification in daily use. The long history of trust over signature verification means that people are very willing to accept a signature-based biometric authentication system.

However, there has not been any major international effort that aims at comparing different signature verification methods systematically. As common benchmark databases and benchmarking rules are often used by researchers in such areas as information retrieval and natural language processing, researchers in biometrics increasingly see the need for such benchmarks for comparative studies. For example, fingerprint verification competitions (FVC2000 and FVC2002) have been organized to attract participants from both academia and industry to compare their algorithms objectively. As inspired by these efforts, we organized the First International Signature Verification Competition (SVC2004) recently.

The objective of SVC2004 is to allow researchers and practitioners to compare the performance of different signature verification systems systematically based on common benchmark databases and benchmarking rules. Since on-line handwritten signatures collected via a digitizing tablet or some other pen-based input device can provide very useful dynamic features such as writing speed, pen orientation and pressure in addition to static shape information, only online handwritten signature verification was included in this competition.

We made it clear to all participants from the very beginning that this event should not be considered as an official certification exercise, since the databases used in the competition were only acquired in laboratory rather than real environments. Moreover, the performance of a system can vary significantly with how forgeries are provided. Furthermore, handwritten signature databases are highly language dependent. Nevertheless, it is hoped that through this exercise, researchers and practitioners could identify areas where possible improvements to their algorithms could be made.

2 Participants

The Call for Participation announcement was released on 30 April 2003. By the registration deadline (30 November 2003), 33 teams (27 from academia and six from industry) had registered for the competition showing their intention to participate in either one or both tasks of the competition. Of the 33 teams registered, 16 teams eventually submitted their programs for Task 1 while 13 teams for Task 2 by the submission deadline (31 December 2003). Some teams participated in both tasks. One team submitted a program that requires a licensed software to run it. Eventually this team withdrew. So we ended up having a total of 15 teams for Task 1 and 12 teams for Task 2. All are academic teams from nine different countries (Australia, China, France, Germany, Korea, Singapore, Spain, Turkey, and United States). Table 1 shows all the participating teams, with nine decided to remain anonymous after the results were announced. Team 19 submitted three separate programs for each task based on different algorithms. To distinguish between them when reporting the results, we use 19a, 19b and 19c as their Team IDs.

3 Signature Databases

3.1 Database Design

SVC2004 consists of two separate signature verification tasks using two different signature databases. The signature data for the first task contain coordinate information only, but the signature data for the second task also contain additional information including pen orientation and pressure. The first task is suitable for

Table 1. SVC2004 participating teams

Team ID	Institution	Country	Member(s)	Task(s)
3		Australia	V. Chandran	1 & 2
4	*anonymous*			1 & 2
6	Sabanci University	Turkey	Alisher Kholmatov Berrin Yanikoglu	1 & 2
8	*anonymous*			2
9	*anonymous*			1 & 2
12	*anonymous*			1
14	*anonymous*			1 & 2
15	*anonymous*			1
16	*anonymous*			1
17	*anonymous*			1 & 2
18	*anonymous*			1 & 2
19	Biometrics Research Laboratory, Universidad Politecnica de Madrid	Spain	Julian Fierrez-Aguilar Javier Ortega-Garcia	1 & 2
24	Fraunhofer, Institut Sichere Telekooperation	Germany	Miroslav Skrbek	1
26	State University of New York at Buffalo	USA	Aihua Xu Sargur N. Srihari	1
29	Institut National des Télécommunications	France	Bao Ly Van Sonia Garcia-Salicetti Bernadette Dorizzi	2

on-line signature verification on small pen-based input devices such as personal digital assistants (PDA) and the second task on digitizing tablets.

Each database has 100 sets of signature data. Each set contains 20 genuine signatures from one signature contributor and 20 skilled forgeries from at least four other contributors. Unlike physiological biometrics, the use of skilled forgeries for evaluation is very crucial to behavioral biometrics such as handwritten signature. Of the 100 sets of signature data, only the first 40 sets were released (on 25 October 2003) to participants for developing and evaluating their systems before submission (by 31 December 2003). While the first 40 sets for the two tasks are totally different, the other 60 sets (not released to participants) are the same except that the pen orientation and pressure attributes are missing in the signature data for Task 1. Although both genuine signatures and skilled forgeries were made available to participants, user enrollment during system evaluation accepted only five genuine signatures from each user, although multiple sets of five genuine signatures each were used in multiple runs. Skilled forgeries were not used during the enrollment process. They were only used in the matching process for system performance evaluation. Evaluation of signature verification performance for each user was only started after all users had been enrolled. Therefore, participants could make use of genuine signatures from other users to improve the verification accuracy for a user if they so wished.

3.2 Data Collection

Each data contributor was asked to contribute 20 genuine signatures. For privacy reasons, the contributors were advised not to use their real signatures in daily use. Instead, they were suggested to design a new signature and to practice the writing of it sufficiently so that it remained relatively consistent over different signature instances, just like real signatures. Contributors were also reminded that consistency should not be limited to spatial consistency in the signature shape but should also include temporal consistency of the dynamic features.

In the first session, each contributor contributed 10 genuine signatures. Contributors were advised to write naturally on the digitizing tablet (WACOM Intuos tablet) as if they were enrolling themselves to a real signature verification system. They were also suggested to practice thoroughly before the actual data collection started. Moreover, contributors were provided the option of not accepting a signature instance if they were not satisfied with it. In the second session, which was normally at least one week after the first one, each contributor came again to contribute another 10 genuine signatures.

The skilled forgeries for each data contributor were provided by at least four other contributors in the following way. Using a software viewer, a contributor could see the genuine signatures that he or she tried to forge. The viewer could replay the writing sequence of the signatures on the computer screen. Contributors were also advised to practice the skilled forgeries for a few times until they were confident to proceed to the actual data collection.

The signatures are mostly in either English or Chinese. Although most of the data contributors are Chinese, many of them actually use English signatures frequently in daily applications.

3.3 Signature Files

Each signature is stored in a separate text file. The naming convention of the files is `UxSy`, where `x` is the user ID and `y` is the signature ID. Genuine signatures correspond to `y` values from 1 to 20 and skilled forgeries from 21 to 40. However, random re-numbering was performed during the evaluation process to avoid the class information from being revealed by the file names.

In each signature file, the signature is represented as a sequence of points. The first line stores a single integer which is the total number of points in the signature. Each of the following lines corresponds to one point characterized by features listed in the following order (the last three features are missing in the signature files for the first task): x-coordinate, y-coordinate, time stamp, button status, azimuth, altitude, and pressure.

4 Performance Evaluation

4.1 Testing Protocol

Both tasks used the same code submission scheme. For each task, each team was required to submit two executable files, one for performing enrollment and the other for matching. Executable files were for the Windows platform and could run in command-line mode without any graphical user interface.

The testing protocol is as follows. Each program was evaluated on two signature databases. The first database, which was released to the participants, consists of genuine signatures and skilled forgeries for 40 users. The second database consists of similar signature data for 60 users. This set was not released to the participants. For each user from either database, 10 trials were run based on 10 different random subsets of five genuine signatures each from files S1-S10 for enrollment. After each enrollment trial, the program was evaluated on 10 genuine signatures (S11-S20), 20 skilled forgeries (S21-S40), and 20 random forgeries selected randomly from genuine signatures of 20 other users. Whenever randomness was involved, the same random sets were used for all teams.

For each signature tested, a program is expected to report a similarity score, between 0 and 1, which indicates the similarity between the signature and the corresponding template. The larger the value is, the more likely the signature tested will be accepted as a genuine signature. Based on these similarity scores, we computed false rejection rates (FRR) and false acceptance rates (FAR) for different threshold values. Equal error rates (ERR) and Receiver Operating Characteristics (ROC) curves were then obtained separately for skilled forgeries and random forgeries.

4.2 Results

The programs of some teams encountered problems during the evaluation process. In particular, they failed to report similarity scores for some input signatures. For fairness of comparison, EER statistics and ROC curves are not reported for these programs. Besides reporting the average EER over all users and all 10 trials for each team, we also report the standard deviation (SD) and maximum EER values.

Tables 2 and 3 show the EER results for both tasks evaluated on signature data from 60 users not released to participants. Figures 1 and 2 show the corresponding ROC curves for the evaluation with skilled forgeries. The results of some teams (Teams 3 and 9 for Task 1 and Teams 3, 9 and 29 for Task 2) are not included in the tables since their programs failed to report similarity scores for some signatures. For both tasks, Team 6 from the Sabanci University of Turkey gives the lowest average EER values when tested with skilled forgeries. Due to page limit, some results are not included in this paper. Readers are referred to `http://www.cs.ust.hk/svc2004/results.html` for more details.

5 Discussions

We have noticed that the EER values tend to have relatively large variations as can be seen from the SD values. While behavioral biometrics generally have larger intra-class variations than physiological biometrics, we speculate that this is at least partially attributed to the way in which the signature databases were created for SVC2004. Specifically, the signatures are not the real signatures of the data contributors. Although they were asked to practice thoroughly before signature collection, larger variations than expected were still expected.

Table 2. EER statistics for Task 1 (60 users)

Team ID	10 genuine signatures + 20 skilled forgeries			10 genuine signatures + 20 random forgeries		
	Average	SD	Maximum	Average	SD	Maximum
6	**2.84%**	**5.64%**	**30.00%**	2.79%	5.89%	50.00%
24	4.37%	6.52%	25.00%	1.85%	2.97%	15.00%
26	5.79%	10.30%	52.63%	5.11%	9.06%	50.00%
19b	5.88%	9.21%	50.00%	2.12%	3.29%	15.00%
19c	6.05%	9.39%	50.00%	2.13%	3.29%	15.00%
15	6.22%	9.38%	50.00%	2.04%	3.16%	15.00%
19a	6.88%	9.54%	50.00%	2.18%	3.54%	22.50%
14	8.77%	12.24%	57.14%	2.93%	5.91%	40.00%
18	11.81%	12.90%	50.00%	4.39%	6.08%	40.00%
17	11.85%	12.07%	70.00%	3.83%	5.66%	40.00%
16	13.53%	12.99%	70.00%	3.47%	6.90%	52.63%
4	16.22%	13.49%	66.67%	6.89%	9.20%	48.57%
12	28.89%	15.95%	80.00%	12.47%	10.29%	55.00%

Table 3. EER statistics for Task 2 (60 users)

Team ID	10 genuine signatures + 20 skilled forgeries			10 genuine signatures + 20 random forgeries		
	Average	SD	Maximum	Average	SD	Maximum
6	**2.89%**	**5.69%**	**30.00%**	2.51%	5.66%	50.00%
19b	5.01%	9.06%	50.00%	1.77%	2.92%	10.00%
19c	5.13%	8.98%	51.00%	1.79%	2.93%	10.00%
19a	5.91%	9.42%	50.00%	1.70%	2.86%	10.00%
14	8.02%	10.87%	54.05%	5.19%	8.57%	52.63%
18	11.54%	12.21%	50.00%	4.89%	6.65%	45.00%
17	12.51%	13.01%	70.00%	3.47%	5.53%	30.00%
4	16.34%	14.00%	61.90%	6.17%	9.24%	50.00%

We have also noticed that the results for Task 1 are generally slightly better than those of Task 2. This seems to imply that additional dynamic information including pen orientation and pressure is not useful and can lead to impaired performance. While conflicting results have been seen in the literature, we believe this is again due to the way of collecting our signature data, as discussed above. The invariance of pen orientation and pressure is likely to be less than that of other dynamic information used for Task 1.

From these findings, we are further convinced that establishing benchmark databases that faithfully reflect the nature of signature data found in real-world applications is of great importance to the research community. We hope SVC2004 can facilitate collaborative efforts in establishing such databases before long.

More performance criteria may be considered in the future. While this competition considers only accuracy as measured by EER, it would be useful, particularly from the application perspective, to include other criteria such as running time. Moreover, we may also allow a program to reject a signature during the enrollment and/or testing phase.

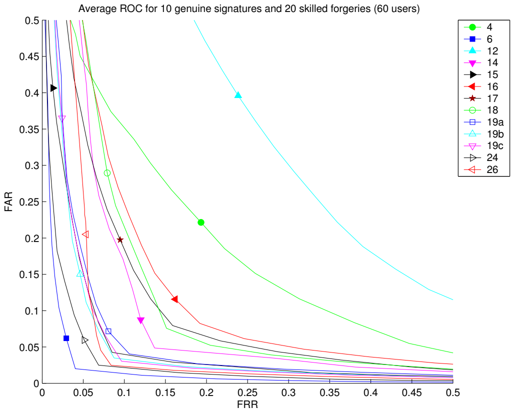

Fig. 1. ROC curves for Task 1 (60 users)

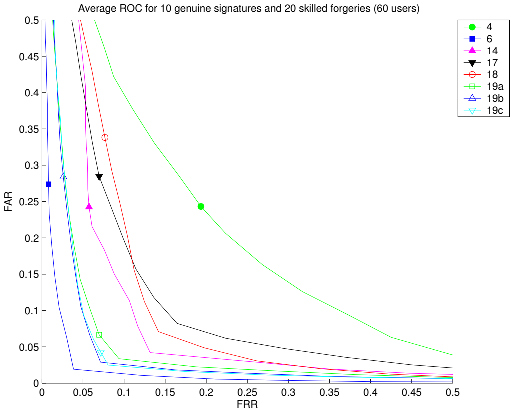

Fig. 2. ROC curves for Task 2 (60 users)

References

1. V.S. Nalwa. Automatic on-line signature verification. *Proceedings of the IEEE*, 85(2):215–239, 1997.
2. A. Jain, R. Bolle, and S. Pankanti. *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers, Boston, MA, USA, 1999.
3. A.K. Jain, F.D. Griess, and S.D. Connell. On-line signature verification. *Pattern Recognition*, 35(12):2963–2972, 2002.
4. R. Plamondon and G. Lorette. Automatic signature verification and writer identification – the state of the art. *Pattern Recognition*, 22(2):107–131, 1989.
5. F. Leclerc and R. Plamondon. Automatic signature verification: the state of the art – 1989–1993. *International Journal of Pattern Recognition and Artificial Intelligence*, 8(3):643–660, 1994.