Study of Temporal Variability in On-Line Signature Verification

Nicolas Ragot[1] *Julie Fortune*[1,2] *Nicole Vincent*[3] *Hubert Cardot*[1]

[1]Laboratoire d'Informatique, Université François Rabelais de Tours,
64 av. Jean Portalis, 37200, Tours, France
{nicolas.ragot, julie.fortune, hubert.cardot}@univ-tours.fr
[2]Atos Worldline, 19 rue de la Valle Maillard, 41000 Blois, France
[3]Laboratoire SIP / CRIP5, Université René Descartes,
45 rue des Saints-Pères, 75270 Paris Cedex 06, France
nicole.vincent@math-info.univ-paris5.fr

Abstract

In the field of on-line signature verification, much attention has been paid on classifiers and features used to elaborate authentication systems able to deal with this modality and its specificities such as forgeries and signers variability. Nevertheless, this variability has not really been studied through time as it is, for example, in the field of on-line handwriting recognition. Indeed, most of the databases used were acquired during one or two sessions but there is no database that could describe the variability of the signers through several months. This work presents preliminary conclusions on such database. It seems that there are no trends at all concerning the evolution of signatures. But this study also reveals that the signer variability seems much higher than the one observed in previous databases. A direct conclusion that could be drawn is that performances announced until know could be overestimated considering a real-life exploitation. This should be nuanced considering the experiment size (small sample size and duration) but what seems obvious is that there is a need for further studies and for new on-line signature databases taking into account the temporal variability of signatures.

Keywords: Biometrics, on-line signature verification, temporal variability.

1. Introduction

In the field of biometric authentication systems, on-line handwritten signature has been widely studied [2][3][6][11][13] in the past few years. Indeed this modality is a part of our habits, it is non intrusive, and it models behavioral characteristics of the human being that can be neither stolen nor lost. Until know, most attention has been paid on: the features or classifiers to use for this specific task of authentication with this specific modality [2][3][6][10][11][12]; on the way to fuse several modalities (face, voice, palmprint, fingerprint, signatures, etc.) [5]; on the way to improve the reliability and the security of the systems, making them able to deal with skilled forgeries [5][9].

Unfortunately, little attention has been paid about one important specificity of handwritten signature: its temporal evolution (or variability). Whereas this aspect is know well studied for on-line handwriting recognition to adapt a recognition system to a specific user and to the way its handwriting could vary through time [1][4], we have not yet seen similar studies for the signatures. One of the main reasons is that there exist only few databases for the evaluation of on-line signature authentication systems and that such databases are difficult to create under real working conditions, for privacy and logistic reasons. It is then likely that databases that could represent the temporal evolution of the signature for the same users during several months are much more difficult to create. However, similarly to handwriting, it is reasonable to think that a signature could evolve through time or at least vary, depending on the context in which the authentication is performed ("humor" of the signer (boring, speed, angry, numbers of signatures performed just before), stability of the acquiring material, etc.). Consequently, there is a need for new databases based on this temporal aspect, and for results on such databases. For example, there are two well known databases in the community: MCYT [7] and SVC [13]. In the former (baseline corpus), the acquisition was performed by sets of 5 signatures, genuine and skilled alternatively. In the latter, the acquisition was performed during two sessions (10 signatures each time) with at "least one week" between them. For both of them, the temporal evolution of signatures is not representative enough and could not be studied.

In this paper, we try to investigate this problematic by characterizing the evolution of signatures and evaluating the impact of such phenomenon on the performances of authentication systems. To do this, we have at first created a preliminary database with signature acquisitions that

were done regularly since 10 months by the same users. This paper describes in the section 2 this database. Next, section 3 gives some first conclusions obtained by statistics results. Finally, section 4 show preliminary results on the impact of the temporal aspect on the performances of a DTW authentication system [10][11][12].

2. On-line signature database with temporal variability

To create a database that reflects the temporal evolution of signatures, we asked 20 persons to sign during a regular meeting that occurs nearly every two weeks in early afternoon. First acquisition sessions started 10 months ago. The device used is always the same: a classical TabletPC from which we get x, y positions of the pen, the pressure (binary value: up or down) and the time. All acquisitions are performed inside a window of a fixed size. During one acquisition session, the protocol is always the same:

- **Training step:** during this stage, the user can train to sign on the device, i.e. there is no recording of the signal. Indeed, signers are not always at ease with this kind of device and the variability phenomenon we want to study must not be influenced by the fact that the user is getting accustomed during the acquisition session. Consequently, this step should limit the variability of the signatures but mainly during each session. The inter-session variability (temporal phenomenon we are interested in) should remain unchanged;
- **Enrollment step:** the signer provides 5 valid signatures. If there is a problem during the acquisition, the user can cancel the current signature to do it again;
- **Test step:** the signer performs 5 valid signatures in the same conditions as the enrollment stage. At the end, a verification is performed: if one valid signature from the test step is roughly different from the ones acquired during the corresponding enrollment step, the signer must start again from the enrollment step. See section 4.1 for the criterions used for this verification.

3. Statistical Analysis of the database

After 10 months of acquisitions, we have at most 17 acquisition sessions for the most regular signers and only 14 persons had performed 10 acquisition sessions or more. In the following we only consider these persons since the results for the others should not be significant. There are 14 acquisition sessions on average per person with 10 signatures each time. We are of course aware that this database contains only few people and few sessions. It is why the acquisition protocol is still running. We are also working to create another database with new people and with the same experimental protocol. Nevertheless, as we will see in following sections, we can yet draw interesting conclusions using this first database.

For each signature, we operate a classical preprocessing: rotation along the principal inertia axis, centering and scale normalization [11]. Then we compute the three following features: total length of the signatures (sum of distances between each point), total duration (elapsed time between first and last point) and average local speed (mean of the velocity in each point). Next, for each acquisition session (10 signatures), we computed the means and variances of each feature. The Figures 3, 4 show the temporal evolution of these features along the different sessions for several signers, plotting the mean value for each session plus/minus one standard deviation. We have also computed the ratio between the total variance (for all sessions) and the average variance during a session for all these features (see Table 1). Finally, we give the correlation coefficients between the total length and the average local speed and between the total duration and the total length. From these results, the following conclusions can be drawn.

3.1. Total duration of signatures

The persons that sign very quickly (u0, u7, u9 and to some extent u4, u8, u12, u14, u15, u16) have a stable signature (considering total duration) for both intra-session and inter-session variability. Persons that sign slowly (u19, u13) generally produce less stable signatures considering both intra-session and inter-session variability. Two signers (u5, u11) were relatively stable through all sessions except for two of them for which the variability is higher. This could probably explained by some kind of perturbations (the signers had to repeat the whole acquisition process because the test verification failed for example) but we have no way to verify.

3.2. Total length of signatures

This feature appears clearly to be more variable over the sessions but also during a single session. Users u7 and u9, who performed short signatures, are quite stable. This is not the case for u11, u12 and u16 who, even with short signatures also, bear witness to a relative high variability between the several sessions. u8 and u18, with longer signatures, have also a high variability, especially through the different sessions. More generally, we cannot see a stable temporal evolution through time, except for the signer u19 (and to some extent u0) who tends to sign bigger and bigger.

3.3. Average local speed of signatures

This feature has a variability similar as the one of total length but this time, the stable signatures are the slower ones (u5, u11, u12, u13, u16 and u19). The less stable signers (generally the speeder) are u0, u7, u8, u9 and u18. Again, there is no significant temporal evolution through the different sessions.

3.4. Correlation between length and duration (see Table 1)

For these two features, the correlation goes from -0.06 to 0.64. The average is 0.41. This is a week correlation, less than what we could expect. For example, u4, u8, u14, and u18 have a length that varies in a significant way comparatively to other signers whereas the total duration remains constant and stable. For u13, both length and duration are variable but not in the same way. The correlation becomes stronger when the variability is low (see u7, u9, u16), except for u19 for whom length and duration are increasing through sessions.

Table 1. Correlation between total length and total duration - corr(l,d) -, between total length and average local speed – corr(l,s) – and ratio between total variance and mean variance of the sessions for length – r(l) -, duration – r(d) - and speed – r(s) -.

Sign.	corr(l,d)	corr(l,s)	r(l)	r(d)	r(s)
u0	0.29	0.84	3.6	3.4	5.2
u4	0.46	0.65	3.5	1.8	3.2
u5	0.47	0.75	1.6	1.4	2.0
u7	0.61	0.30	1.7	3.3	3.3
u8	0.35	0.90	3.8	3.0	5.3
u9	0.56	0.64	2.2	2.5	2.7
u11	0.23	0.79	3.7	1.2	6.3
u12	0.46	0.61	3.3	3.2	4.0
u13	0.37	0.72	9.7	2.8	4.6
u14	0.51	0.75	3.4	1.7	4.8
u15	-0.06	0.69	3.0	2.5	2.5
u16	0.65	0.69	2.0	1.7	1.7
u18	0.29	0.83	3.5	1.8	4.6
u19	0.62	0.75	7.3	2.1	4.4
Mean	0.41	0.71	3.7	2.3	3.9

3.5. Correlation between length and average local speed (see Table 1)

This time, the correlation is much stronger: it goes from 0.30 up to 0.89 with 0.71 on average. This correlation is particularly strong for u0, u8 and u18. For the others, the correlation remains strong in general but it must be discussed knowing the fact that the average local speed is low. For one signer (u7), the correlation is very week which could be explained by the fact that the signature is at the same time small and very quickly performed.

3.6. First conclusions

We can conclude from this statistical analysis that there is not temporal evolution of the signatures (no trends at all) but that there is a temporal variability that could be quite important for several signers. This variability is much more important for the total length and the average local speed of the signatures, whereas total duration seems to be more stable (see r(l), r(d) and r(s) in Table 1). We can also conclude that there is not necessarily a correlation between the length and the duration, which is a little bit surprising, whereas length and average local speed are much more correlated.

Another important point is the fact that inter-session variability is very high. Indeed the ratio between the total variability and the average variability during a session is always positive and is, in average more than 2 for all features (see Table 1). This means that if the parameters of the system are learnt from only one acquisition session, they could become incorrect (and the system less accurate) for other sessions and thus for real-life authentication processes. This phenomenon is clearly visible on the Figures 1 and 2 showing the average values for length and duration at each session and the same values for all signatures respectively. For example, on Figure 1, we can see that for u8, for a given session (noted here u8_1), the average length of his signature is 1648. During this session, the length varies from 1489 to 1712. During another session (noted u8_2), the average length is 1037 and it varies from 990 to 1206. If we consider now u15 and its session for the one the average length is the higher (1006; we note it u15_1), the length varies from 893 between 1089, which overlap u8_2 but not u8_1. Consequently, if the authentication system and its parameters (learning) are determined using u8_1 and u15_1, it is quite possible that an u8 signature with a length similar to those produced during u8_2 become more similar to u15 signatures and thus to be rejected.

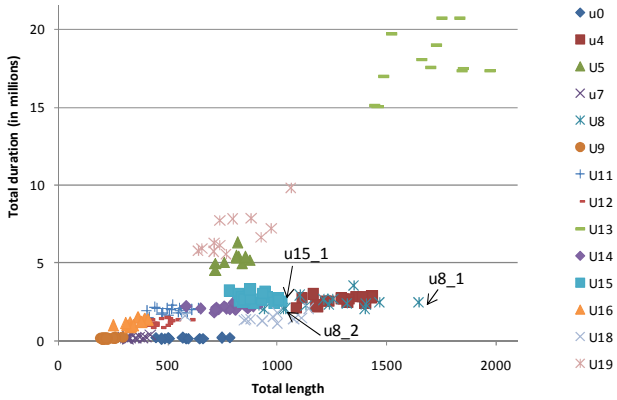

Figure 1. Average length and duration of signatures for each signer and each sessions.

Of course, this phenomenon described here for few global features (we can see the same thing for average speed), could be compensated -or increased- by other features. But it is reasonable to think that temporal variability of signatures could have an important impact on the performances of authentication systems and then we can say that present results of authentication systems

provide biased results. Thus, the entire evaluation protocol should be revised and, at the same time, the databases used for these performance evaluations, so that they take into account the temporal variability aspect of the signatures more significantly.

Finally, this study also corroborates previous works that have shown how to evaluate the stability of features (see [3] e.g.). Our results seems to show that, to improve authentication systems, we should adapt the feature set to the signer to authenticate, taking into account the variability of these features trough time. In general, a feature is more stable for a signer if it has small values for this feature.

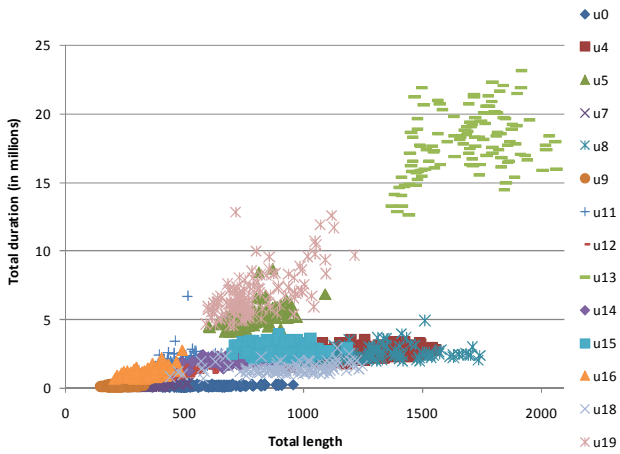

Figure 2. Length and duration of all signatures of each signer.

4. Performance evaluation

In this part, we try to evaluate the impact of the temporal variability of the signatures during the authentication process. This gives complementary results to the previous ones. Indeed, we use a DTW classifier (function-based approach [8]) that performs local comparisons, contrary to the previous feature-based.

4.1. Authentication prototype

To verify if a signer is or not who he claims to be, we used a *Coarse to Fine* approach previously developed [10][11][12] and experimented on several databases.

The *Coarse* step ensures to get rid of the more distant signatures. It is based on supposed stable features [3][11][12]: total length and total duration of signatures. To accept a signature, both features must have a value between min and max values of the signatures of reference multiplied by a coefficient.[1]

If the signature is accepted, the *Fine* step proceeds. It is based on a DTW comparison using spatial distance between points and modified to operate normalization depending on the number of points in the signature [12]. To accept a signature at this step, the DTW score must be less than a threshold specified during the training stage: it corresponds to the threshold that give the EER on the training database (see section 4.2).

4.2. Experimental protocol

To have a balanced dataset, we used the previous database but limited to the 10 first acquisition sessions. Thus, there are 14 signers with 10x10 signatures for each. To evaluate the accuracy of our prototype, we performed tests using the leave-one-out method.

For the learning, 13 signers are used to determine the rejection threshold: for each signer, the 5 first signatures are used as the signatures of reference (enrollment) and the others (the 5 unused from the first session and the 9x10 from the others sessions) plus all the signatures of other signers (12 signers x 10 sessions x 10 signatures) are used to evaluate the FAR and FRR. The mean FAR and FRR could be determined for the 13 signers and the threshold is defined when we obtain the EER.

For the test phase, the threshold defined previously is kept and we operate the test with the 14th signer (that was not used during learning) as a new person to authenticate: we use its 5 first signatures as references (enrollment) and then we use the remaining signatures of this signer (5+ 9x10 from the others sessions) and all the others of the database (13 signers x 10 sessions x 10 signatures) to evaluate the FAR and FRR.

The entire procedure (learning and testing) is repeated such as every signer is used once in the test phase. The result of the evaluation is the mean FAR and the mean FRR obtained during these 14 testing phases.

4.3. Results (Table 2)

Table 2. Results of the authentication prototype on several databases.

Database	EER (%)
Temporal	6.19
Static[2]	1.8
SVC [13]	1.94
MCYT [7]	3.5

The mean FAR and mean FRR obtained using the leave-one-out protocol previously defined are of 6.28% and 6.09% respectively. The corresponding EER is the

[1]The coefficients used here are the same than the one used for previous experiments on other databases

[2]This database contains 800 signatures performed by 40 people (20 signatures each). The material, protocol and signers are similar to those used in our temporal database.

average: 6.19%. This is to compare with the results obtained with the same authentication prototype on other databases that are more complex considering the number of signers (see Table 2). The EER on our temporal database is quite high for such small database without skilled forgeries and we could expect this is mainly because of the temporal variability of signatures. Another argument highlighting this thesis is the result obtained in the same conditions but using only the first session of each signer from our temporal database (14 signers x 10 signatures). In these conditions, the temporal aspect is completely removed and the EER decrease significantly to 2.15%.

5. Conclusion and perspectives

In this paper we worked on a new on-line signature database that reflect the temporal evolution of signatures. The acquisition process started 10 months ago and we have up to 17 acquisition sessions, regularly performed, by 14 signers. This small database is still in creation to improve its significance both for sample size and duration. Nevertheless, we can still draw several conclusions. Firstly, it seems that there is not evolution of signatures (no trends) considering duration, length and average speed. However, the variability between sessions is very important (much more than the intra-session variability). This makes us think that the performances determined on classical databases are overestimated since they do not deal with the real variability of signatures over a long time. This hypothesis is reinforced by experimental results with a DTW classifier that performs well on previous databases (EER less than 3.5%) but not as well on our temporal database with 14 signers only (EER of 6.19%). This is only preliminary studies that should be continued but we think that there is a need for new databases with temporal variability of signatures and that authentication systems should deal with this variability to be as accurate as they seems to be on classical databases with one or two acquisition sessions only.

References

[1] S. D. Connell, A. K. Jain, “Writer Adaptation for Online Handwriting Recognition”, IEEE Trans. on PAMI, vol. 24, Issue 3, pp. 329 – 346, 2002.

[2] G. Dimauro, S. Impedovo, R. Modugno, G. Pirlo, L. Sarcinella, “Analysis of Stability in Hand-Written Dynamic Signatures,” Proc. of IWFHR’02, pp. 259-263, 2002. A. K. Jain, F. D. Griess and S. D. Cornell, "On-line Signature Verification," Pattern Recognition, vol. 35, no. 12, pp. 2963-2972, 2002.

[3] H. Lei, V. Govindaraju, “A comparative study on the consistency of features in on-line signature verification”, Pattern Recognition Letters, vol. 26, pp. 2483–2489, 2005.

[4] H. Mouchère, E. Anquetil, N. Ragot, “Writer Style Adaptation in On-line Handwriting Recognizers by a Fuzzy Mechanism Approach: The ADAPT Method”, Int. Journal of Pattern Recognition and Artificial Intelligence, 21(1):99-116, 2007.

[5] A. M. Namboodiri, S. Saini, X. Lu and A. K. Jain,“Skilled Forgery Detection in On-Line Signatures: A Multimodal Approach”, LNCS, vol.3072, pp. 505-511, 2004.

[6] L. Nanni, A. Lumini, “A novel local on-line signature verification system”, Pattern Recognition Letters, vol. 29, pp. 559-568, 2008.

[7] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero and Q. –I. Moro, "MCYT baseline corpus: a bimodal biometric database", IEEE Proc.-Vis. Image Signal Process., Vol. 150, No. 6, 2003.

[8] R. Plamondon and G. Lorette, “Automatic Signature Verification and Writer Identification – The State of the Art,” Pattern Recognition, vol. 22, no.2, pp. 107-131, 1989.

[9] T.H. Rhee, S.J. Cho, J.H. Kim, “On-line signature verification using model-guided segmentation and discriminative feature selection for skilled forgeries”, ICDAR’01, pp. 645-649, 2001.

[10] M. Wirotius, J.-Y. Ramel, N. Vincent, "Selection of Points for On-Line Signature Comparison," International Workshop On Frontiers in Hanwriting Recognition (IWFHR), Tokyo (Japon), pp. 503-508. 2004.

[11] M. Wirotius, J.-Y. Ramel, N. Vincent, “Contribution of global temporal information for authentification by on-line handwritten signatures”, In International Graphonomics Society (IGS), 2005.

[12] M. Wirotius, J.-Y. Ramel and N. Vincent, “Distance and matching for authentification by on-line signature”, In AutoID, 2005.

[13] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto and G. Rigoll, "SVC2004: First International Verification Competition," ICBA’04, pp. 16-22. 2004.

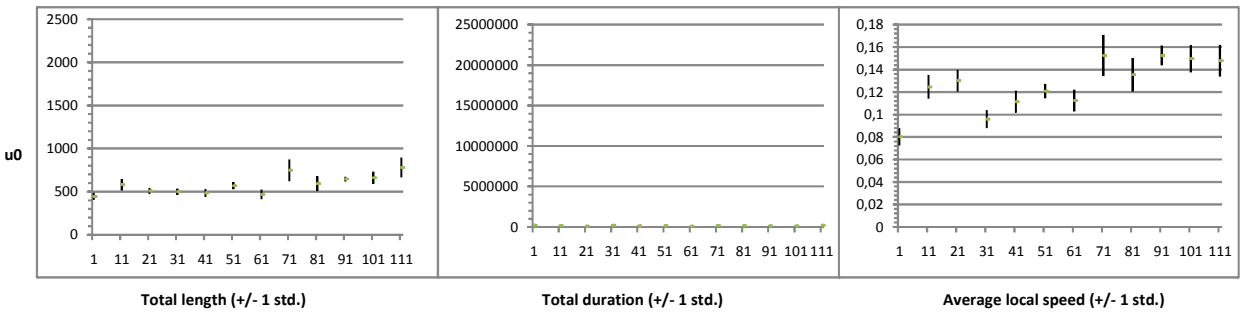

Figure 3. Variability of global features through different sessions.

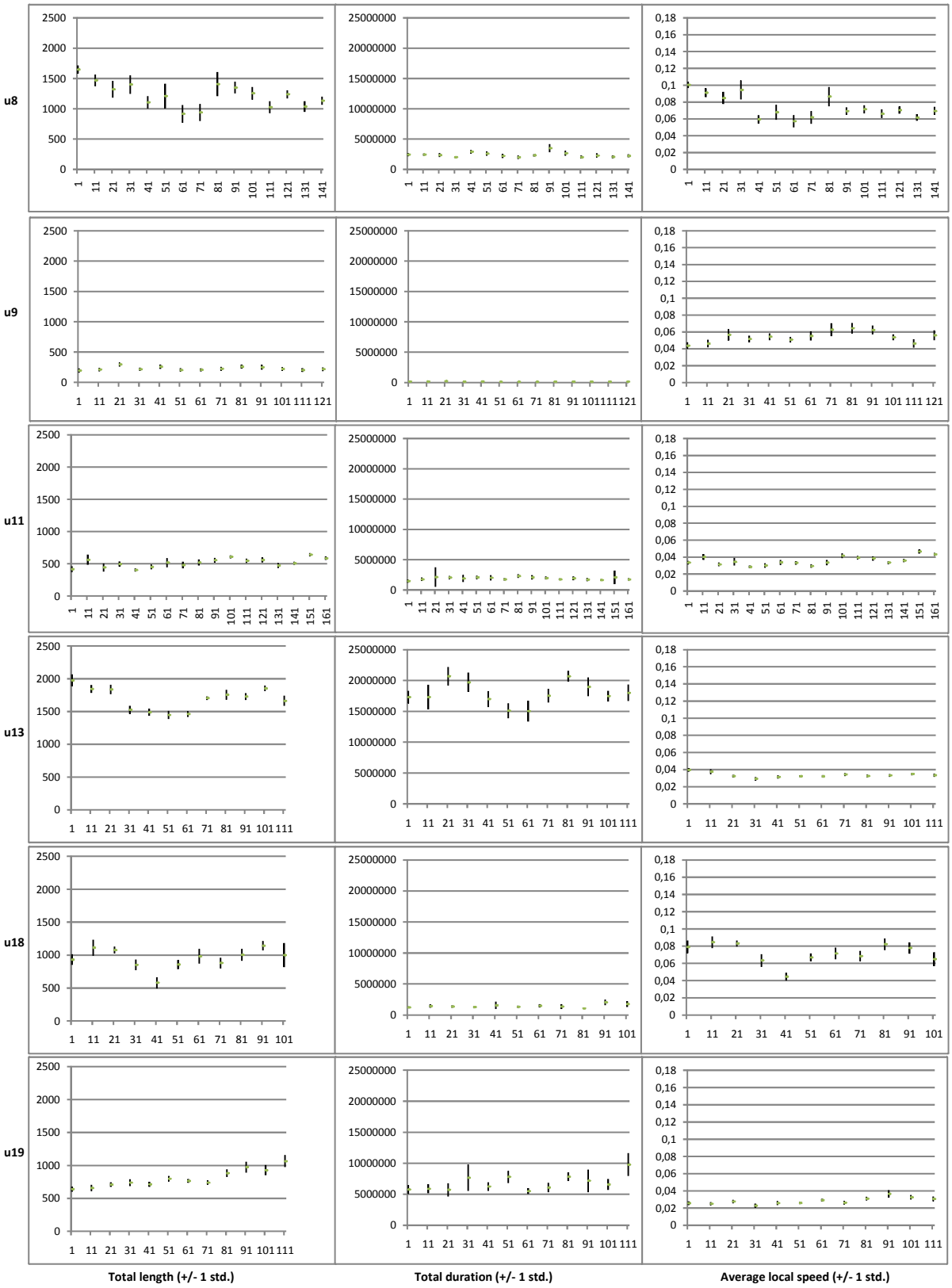

Figure 4. Variability of global features through different sessions.