Zoo Graph: a New Visualisation for Biometric System Evaluation

Romain Giot, Romain Bourqui, Mohamad El-Abed

Zoo Graph:
a New Visualisation for Biometric System Evaluation

Romain Giot*, Romain Bourqui* and Mohamad El-Abed†
*Labri, Univ. Bordeaux, France
†Rafik Hariri University, Lebanon

***Abstract*—Biometric authentication systems suffer from several performance limitations. Many performance metrics exist to assess the overall performance of such systems. However, these metrics provide a quantitative assessment in terms of errors without explaining the reasons behind the set of users who significantly contributed for these errors. Towards contributing to solve this problem, we present a novel method (named Zoo Graph) to visualize the performance of a biometric system as a graph thanks to a database of recognition scores. Our approach is an improvement of the Zoo Plot and emphasizes on the relations between the individuals of the database and allows interactive manipulations to track these relations and understand why the biometric authentication method reacts this way. This graph provides researchers with an additional visual assessment tool that would identify problematic users. Such information would allow researchers to update their developed authentication algorithms to reduce those errors.**

***Index Terms*—node link diagram, interaction, biometrics**

1. Introduction

Biometric authentication [1] systems are being used in many applications such as biometric passports. The biometric modalities are mainly classified into two types of families: the morphological modalities (such as iris recognition, face recognition, fingerprint recognition, etc.) where individuals are recognized thanks to a physical property and the behavioral modalities (keystroke dynamics, signature, gait, etc.) where individuals are recognized thanks to the way they do actions.

Biometric systems provide better security and easiness of use compared to traditional authentication methods (such as passwords or token presentation). Despite this, they contain several drawbacks decreasing their widespread of use in our daily applications. One of such drawbacks is related to performance as the matching of two biometric information provides a similarity score which cannot reach 100% of recognition due to many reasons. Among these reasons, we can list: the implementation of biometric authentication systems relies on machine learning and pattern recognition systems that are prone for errors, the biometric data can evolve over time, the biometric sensor can be noisy, the biometric modality is not universal, etc. For this reason, many researchers focused on developing performance evaluation frameworks to evaluate the overall performance of biometric systems [2]. This is done by developing dedicated biometric databases and performance metrics to be used to evaluate and compare new authentication algorithms. Such comparison is required to clarify the benefit of a new authentication algorithm vis-à-vis existing ones. The existing performance metrics are currently used to evaluate and compare biometric systems. With these metrics, researchers are able to sort the authentication algorithms based on their performance.

A common visualization used by the biometric community to show the efficiency of biometric authentication systems is the ROC curve (*Receiver Operating Characteristic*). Such curve allows first to compare the performance of several systems, and second to identify one or several settings that provide a good trade-off between these error rates. To obtain individual details, scatterplots are also widely used to represent some of the common evaluation metrics. In the biometric community, such scatterplots are the *Zoo Plot* [3] and are able to emphasize the biometric menagerie [4]. However such a visualization only provides a high level view on the system and cannot show the reasons (the relationship between individuals) that could explain low or high performances. Low performance of a system can be due to many reasons, *e.g* a unique individual can impersonate a lot of other individuals or on the contrary many individuals can impersonate very few other individuals.

In this work, we present a new visual evaluation method, named *Zoo Graph*, which helps to understand how the system performs. The set of users is represented as a graph whose topology depends on the performance of the biometric authentication system. The operator can navigate within the graph and track the problematic users by using various interaction tools. Once the problematic users are detected, he can use his own knowledge to understand why the users are problematics and how to modify the biometric authentication system to overcome this problem (if possible). So far to our knowledge, this is the first work where we search to exhibit the relations between the individuals of a biometric dataset.

The paper is organised as follows. Section 2 presents some generalities on biometrics and evaluation, section 3 describes our proposal, section 4 shows the experimental protocol, section 5 gives the results and section 6 concludes this work.

2. Biometrics Generalities

A biometric system is composed of two main modules: the *enrolment* module and the *verification* module. The enrolment module serves to compute a *template* for each user thanks to a *gallery* of biometric samples. The verification module serves to verify if the *identity* claimed by the *claimant* is the good one. The *query* (which is a biometric sample) of the claimant is compared to the template of the claimed user. This comparison produces a score which is then compared to a threshold in order to take the decision to accept or reject the claimant. To evaluate a biometric authentication system it is necessary to dispose of a gallery (set of labeled samples) to enrol the users and a probe (disjoint set of labeled sample) to compute the comparison scores.

The biometric recognition field disposes of various evaluation tools [5]. We can count the following error metrics: (i) Failure-to-enroll rate (FTE): proportion of the user population for whom the biometric system fails to capture or extract usable information from the biometric sample. (ii) Failure-to-acquire rate (FTA): proportion of verification or identification attempts for which a biometric system is unable to capture a sample or locate an image or signal of sufficient quality. (iii) False-match-rate (FMR): the rate for incorrect positive matches by the matching algorithm for single template comparison attempts. (iv) False-non-match rate (FNMR): the rate for incorrect negative matches by the matching algorithm for single template comparison attempts. (v) False rejection rate (FRR): proportion of authentic users that are incorrectly denied. If a verification transaction consists of a single attempt, the false reject rate would be given by $FRR(\tau) = FTA + FNMR(\tau) * (1 - FTA)$. (vi) False acceptation rate (FAR): proportion of impostors that are accepted by the biometric system. If a verification transaction consists of a single attempt, the false accept rate would be given by $FAR(\tau) = FMR(\tau) * (1 - FTA)$. (vii) Equal Error Rate (EER): this error rate corresponds to an operational point where FAR and FRR are equal (compromise between FAR and FRR). It is widely used to evaluate and to compare biometric authentication systems. The more the EER approaches 0%, the better the performance of the target system.

In addition to these error metrics, some visualisations exist: (viii) Receiver operating characteristic curve (ROC)[6]: plot of the rate of FMR as well as FAR (*i.e.*, accepted impostor attempts) on the x-axis against the corresponding rate of FNMR as well as FRR (*i.e.*, rejected genuine attempts) on the y-axis plotted parametrically as a function of the decision threshold. (ix) The *Zoo Plot* [3] which displays all the individuals of the dataset in a scatter plot where their coordinates correspond to the mean genuine and mean impostor scores. One advantage of this representation is the ability to quickly see various categories [4] of users depending on their performance on the system. In the rest of the paper, we consider that FMR (resp. FNMR) is equal to FAR (resp. FRR).

Among all these standard evaluation metrics and graphs, only the *Zoo Plot* allows to get information for all the individuals of the tested database ; the other metrics or visualisations only give an information on the global system. The *Zoo Plot* is a simple and efficient method to display the individuals of the tested dataset on a 2d plan [3] (figure 2a). Each individual is presented by a dot on the plan with its X position corresponds to the mean value of its intra scores (the queries are compared to the template of the same user) while the Y position corresponds to the mean value of its inter scores (the queries are compared to the template of different users). The 25^{th} maximum and minimum percentiles of these mean scores are presented. We can see the spread of the scores, but we lack several information such as the relationship of each individual to the performance of the system (we globally see who are the best and worst users, but we cannot know how they interact together).

Due to non a uniform distributions of scores and to the use of linear scales for axis, another limitation of *Zoo Plot*s is that most of the individuals are usually displayed in a narrow region of the *Zoo Plot* (see figure 2a).

3. *Zoo Graph*: A New Representation of Biometric Performance

Because of the limitations of the *Zoo Plot*, we have designed its evolution, the *Zoo Graph* (figure 2d), which is based on a node link representation instead of a scatter plot representation. The rest of this section describes the employed methodology (figure 1).

3.1. Graph Data Model

The *Zoo Graph* displays a graph $G = (V, E)$ which is constructed in several steps. The first step consists in constructing a comparison graph $GC = (VC, EC)$ from a dataset of scores composed of tuples $t_i = (s_i^{probe}, s_i^{gallery}, s_i^{comparison}, s_i^{score})$ interpreted as follows: s_i^{score} is the comparison score for the $s_i^{comparison}$ th comparison between the user s_i^{probe} (who owns the query) and the user $s_i^{gallery}$ (who owns the template).

The number of comparison scores of the database is the number of tuples. Each node $vc_u \in VC$ corresponds to an individual of the dataset ($set\left(s_\bullet^{probe} \cup s_\bullet^{gallery}\right)$). There is an edge $\{vc_{u1}, vc_{u2}\} \in EC$ if there is at least one tuple t_j where $s_j^{probe} = vc_{u2}$ and $s_j^{gallery} = vc_{u1}$. For each edge $e \in EC$ are associated the following mappings: $scores : EC \rightarrow \mathbb{R}^*$ gives the list of possible scores for the current edge; $bioScore : EC \rightarrow \mathbb{R}$ gives the mean value of $scores(e)$.

For each node $v \in VC$ are associated the following mappings: $bioScoreSelf : VC \rightarrow \mathbb{R}$ gives the mean value of $scores(\{v, v\})$ (the comparison scores when the genuine user is the probe user); $bioScoreIn : VC \rightarrow \mathbb{R}$ gives the mean value of $\cup_{u \in in\_neighbors(v)} scores(\{u, v\})$ (the comparison scores when other users try to impersonate the current user); $bioScoreOut : VC \rightarrow \mathbb{R}$ gives the mean value of $\cup_{o \in out\_neighbour(v)} scores(\{v, o\})$ (the comparison scores when user try to impersonate the other users).

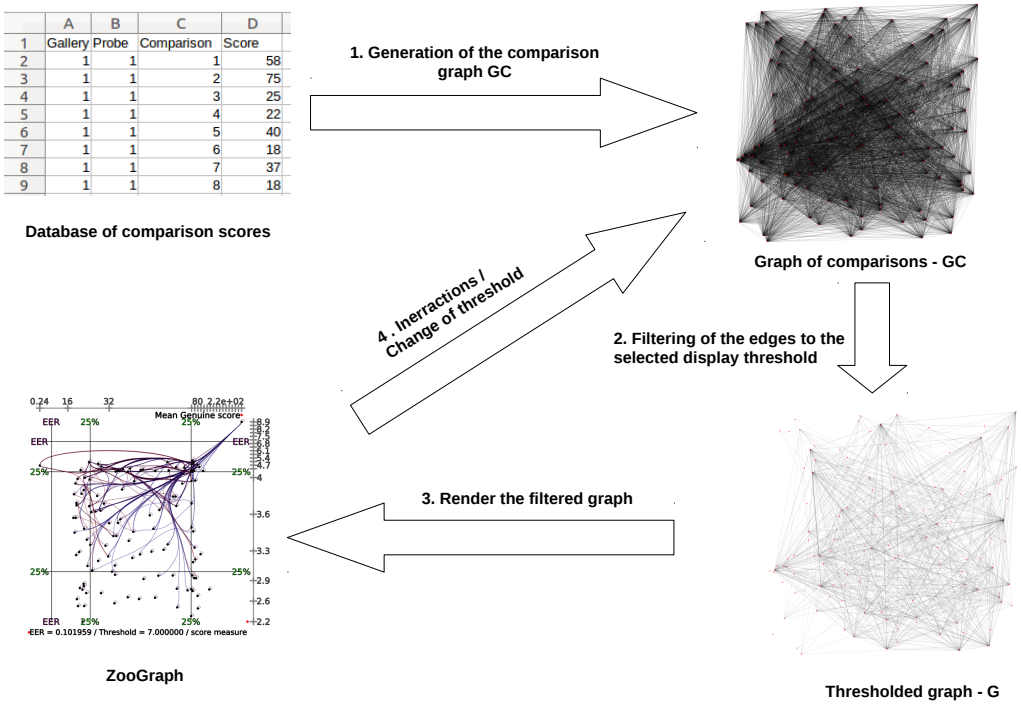


Figure 1. Summary of the workflow to generate the *Zoo Graph* from the dataset of scores. A comparison graph (CG) is generated from the database of scores. It is then converted to a thresholded graph (G) thanks to a decision threshold τ. Biometric information and topology of G is used to draw it on a 2d plan.

Finally, the thresholded graph G is constructed from GC by filtering (i.e., removing) all the edges which are below an authentication threshold τ. It means there is an oriented edge between two nodes of G if the averaged recognition score between the source node and the target node is higher or equal to τ (i.e. the source node is recognised as being the target node).

3.2. Node Positioning and Visual Encoding

While the *Zoo Plot* is displayed as a scatterplot, the *Zoo Graph* is displayed as a node-link diagram whose layout comes from G topology and its mappings.

In the *Zoo Plot*, the space taken by the interval of scores is not proportional to the number of nodes it represents. For instance, in figure 2a, a majority of the individuals are displayed within a narrow region that occupied about 5% of the visualization. While such representation makes sense as it shows the distribution of values, overplotting many individuals complicates the identification of single individual as well as its neighbors. For this reason we have chosen to apply a non linear transformation to the coordinates of each individual in order to let the 25% minimum and maximum percentile take 25% of the space and the rest to take 50% of the space.

In order to reduce the clutter due to the drawing of the edges, we use the edge bundling technique explained in this paper [7]. Various information have been encoded in the visual attributes of the graphs, they are presented in table 1.

3.3. Additional Visual Information

As with the *Zoo Plot*, we display the first and last 25 percentiles as well as the ticks to give an idea of the spread of the score values. Remember that a non linear transformation has been applied to our score space. In order to emphasize this aspect, the ticks are computed by linearly selecting T values between the maximum and minimum value of the scores (in X or Y axis) and projecting them on the non linear space. Thus a distribution of scores which is dense in a small interval of the original space will be represented by few ticks, while a distribution of scores which is sparse in a large interval will be represented by several ticks.

As the *Zoo Graph* is threshold dependant, the value of the threshold is represented with a vertical and horizontal axis in order to delimit the zones which correspond to acceptance patterns and the zones which correspond to rejection patterns.

3.4. Interaction tools

To analyse a static image of a *Zoo Graph* may be not sufficient because for a configuration threshold τ which has a high False Match Rate, the number of edges to draw is

Table 1. LIST OF THE INFORMATION AVAILABLE IN THE *Zoo Graph* AND COMPARISON WITH THE *Zoo Plot*.

Information	**Encoding of proposal** (*Zoo Graph*)	**Encoding of baseline** (*Zoo Plot*)
Mean genuine score (bioScoreSelf)	Node X position	Node X position
Mean impersonate score (bioScoreOut)	Node Y position	Node Y position
Mean impersonated score (bioScoreIn)	Node color (relative mapping)	-
Number of time individual is impersonated (in degree)	Node size	-
Mean comparison score between 2 individuals (bioScore)	Edge color (relative mapping)	-
Relationship for τ	Edges between individuals	-
Exploration	Interactive analysis	-

large and clutters the visualization. It is then necessary to use an interactive representation for such cases.

Remember that *Zoo Graph* is a directed graph where an edge represents an individual source which is often able to be recognised as an individual destination. To understand the interactions of a specific individual with the others, we can emphasize the directed neighborhood of its representing node [8]. By inspecting individual nodes and their neighborhood (either in or out), the expert can identify for any problematic individual, the individuals who can impersonates him and on the contrary the individuals he can be recognized as.

The performance of a biometric system depends on the selected threshold τ; it is also interesting to vary the value of this decision threshold. Our visualization technique can support fast filtering of edges, indeed the rendering step of our workflow runs in linear time (see figure 1). The user can therefore change this threshold on the fly and analyse the appearance/disappearance of edges according to this value.

4. Experimental Protocol

We have used several datasets from the literature to evaluate *Zoo Graph* (see table 2). We have generated the biometric scores ourself for the datasets which contain only to raw biometric data.

Zoo Graph is threshold dependant. For this reason, we have chosen to use for τ the threshold which corresponds to the EER (i.e. to similar FNMR and FMR). To reduce ticks label overplot, we have chosen to use $T = 10$. The non linear mapping of the space hardly allows to use automatic methods [15] to set the number and position of the ticks.

Among the different evaluation methodologies, we want to compete with the *Zoo Plot*. For this reason, it is the baseline method we use.

Zoo Graph has been prototyped as a set of plugins applications which intensively rely on the Tulip software [16]. It is written in Python and C++14.

The evaluation of the application has been done by an expert in evaluation of biometric systems who visualised static pictures of the datasets shown with *Zoo Graph* and *Zoo Plot* and manipulated the interactive representation.

5. Results

With our method, the best biometric system is represented by a graph where all nodes have a self-loop (everybody is able to be authenticated) and where there exist no other edges (nobody is able to be recognised as someone else). An example is available in Figure 2f.

Figure 2 compares some results of our proposal against the baseline for datasets. The non linear projection of the score values on the screen space allows to reduce node overlaps ; if we analyse the results of the AR *Zoo Graph*, the top right individual of Figure 2d corresponds to an individual hidden by the legend of Figure 2a). The threshold line (EER) is clearly visible and emphasizes the top left individual which is unable to recognize itself (it is the only node with no self loops). There are a majority of nodes with no edges (except their self loop). Similar information is obtained from the ENSIB *Zoo Graph* (Figure 2e). This validates the interest in using the non linear projection.

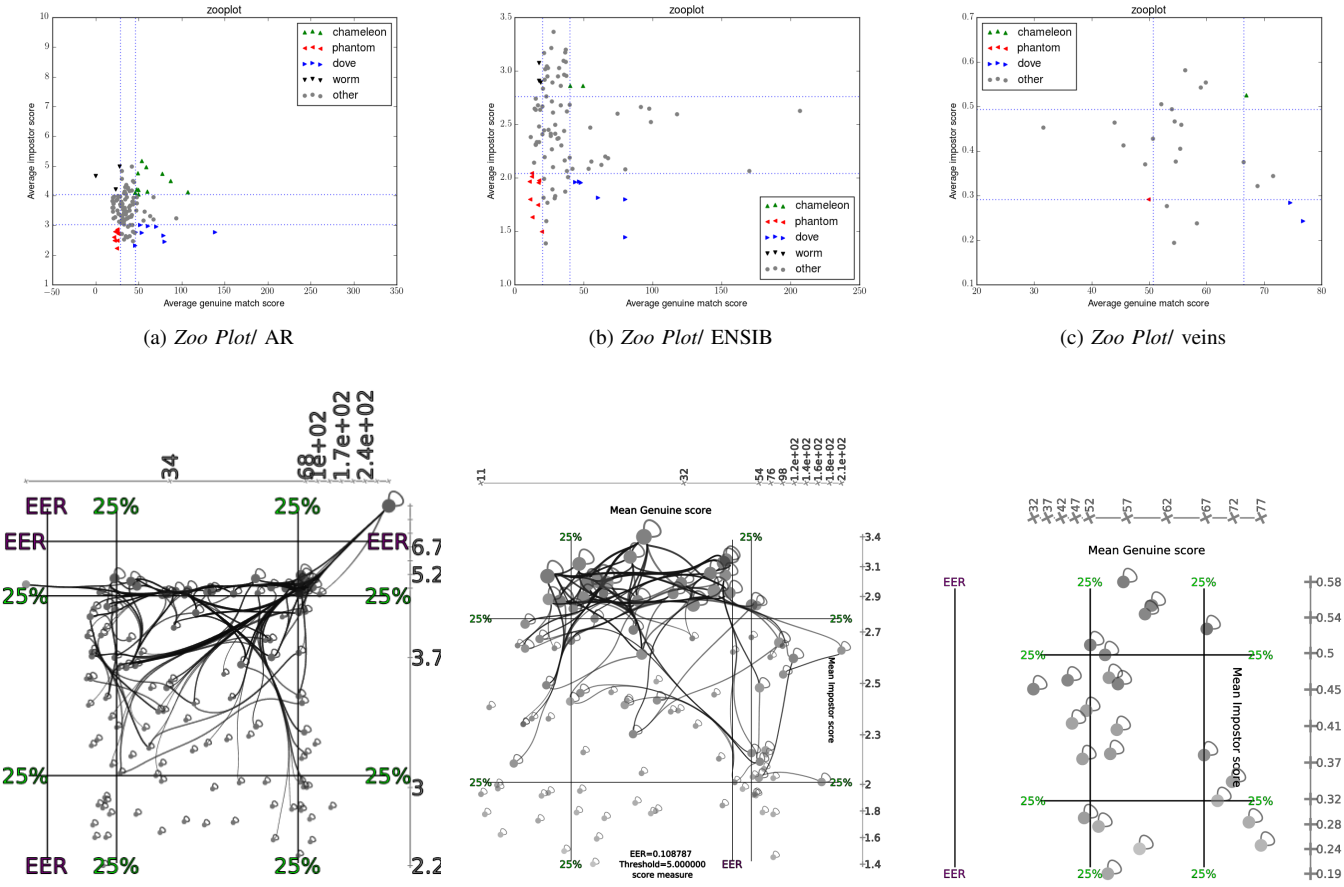
The edges of the *Zoo Graph* clearly provide additional information. We quickly see the users able in average to authenticate to the system (i.e., the nodes with a self loop) of those unable in average to authenticate to the system (i.e., the nodes without a self loop). In systems where the FMR is limited (i.e., no more than 10%) the number of edges (not self-loop) is moderated and we can see which parts of the individuals is responsible of those errors (for AR, Figure 2d, and ENSIB, Figure 2e, these nodes are more on the upper side of the picture). This information is not available with the *Zoo Plot*. The power of similarity between two users is represented by the darkness of the edge ; we can see that this power is lower for the nodes more distant of the upper side of the pictures. This validates the interest in using the edges to show the relations between the individuals.

If we analyse the size of a node, we quickly access to the individuals which are problematic and attract to much other individuals. For example, in AR, Figure 2d, the worst individual is in on the top right of the picture, whereas in ENSIB, Figure 2e, there are more individuals with a bad behavior (in the upper part of the picture and even in the middle of the picture). This validates also the need to map the size of a node to the attractiveness of its user.

Despite these good points, the *Zoo Graph* presents some limitations. The method can fail when there are too many recognition issues. For instance, in Figure 3, the number of edges is too large resulting in a cluttered visualization. In this case, it is probably better to use the *Zoo Plot* or to interactively navigate within the graph with the neighborhood interaction tool.

Table 2. DIFFERENCES BETWEEN THE DATASETS USED TO EVALUATE THE PROPOSAL.

Dataset	Type	Modality	Methodology	# users	# scores	# sample/user	EER
AR [9]	Score	Face	SIFT based matching	120	26	360000	10.19%
ENSIB [10]	Score	Face	SIFT based matching	100	40	390000	10.88%
FC94 [11]	Score	Face	SIFT based matching	152	20	438976	0.29%
FVC [12]	Score	Fingerprint	SIFT based matching	100	8	70000	10.27%
veins [13]	Score	Vein	SIFT based matching	24	30	16704	0.0%
OU-ISIR BSS3 [14]	Distance	Gait (accelerometer)	Distance between 2 signals	736	variable	10175181	14.88%



(a) *Zoo Plot*/ AR (b) *Zoo Plot*/ ENSIB (c) *Zoo Plot*/ veins

(d) *Zoo Graph*/ AR (e) *Zoo Graph*/ ENSIB (f) *Zoo Graph*/ veins

Figure 2. Illustration of the difference between the baseline method (*Zoo Plot*) and the proposal (*Zoo Graph*) for 3 biometric databases.

6. Conclusion

Biometric authentication systems are prone for errors. Therefore, the performance evaluation of such systems is important for deployment in our daily life applications. Existing evaluation methodologies are able to provide the global performance of the system or to visualize the performance of the individuals of the dataset but not the reason of the bad performance.

We have presented in this paper a novel method to visualize the performance of a biometric system (named *Zoo Graph*) as a graph thanks to a database of recognition scores which is an improvement of a well know method (*Zoo Plot*). Our approach emphasizes on the relations between the individuals of the database and allows interactive manipulation to track these relations and understand why the biometric authentication method reacts this way. The result shows that most of the time this new approach is interesting with static images, and the limits of the static visualisation can be bypassed by interacting with the graph.

We have identified several improvements which are left for future work. First, even if we have evaluated the proposal on a database of several hundreds of users, it could be interesting to evaluate it on even larger datasets. Finally, *Zoo Graph* is constructed thanks to a set of scores without any knowledge of the biometric samples used ; it is interesting to

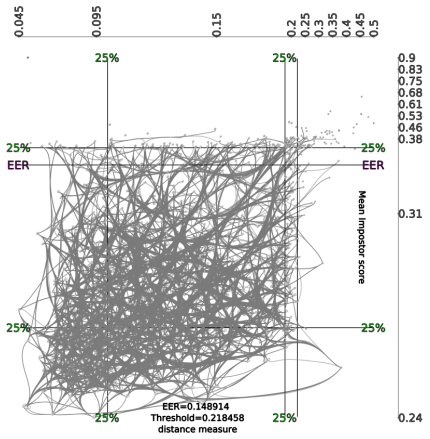


Figure 3. Cluttered visualization due to a large number of edges. Attention in this case the biometric measure is a distance, not a score, so the worst individuals are below the EER.

track the biometric samples which produced these biometric scores in order to allow additional interactions with the ability to show these samples on demand and ease the comprehension of the dataset (i.e., the gallery of each user or the pairs of the edges).

Acknowledgments

We want to thanks the LaBRI for the financial support of this work.

References

[1] A. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*. Springer Science & Business Media, 2007.

[2] R. Giot, B. Dorizzi, and C. Rosenberger, "A review on the public benchmark databases for static keystroke dynamics," *ELSEVIER International journal on Computer & Security*, vol. 55, pp. 46–61, 2015.

[3] N. Yager and T. Dunstone, "The biometric menagerie," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 2, pp. 220–230, 2010.

[4] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation," DTIC Document, Tech. Rep., 1998.

[5] ISO/IEC 19795-1, *Information technology – biometric performance testing and reporting – Part 1: Principles and framework*, International Organization for Standardization tt Std., 2006.

[6] J. R. Matey, G. W. Quinn, P. Grother, E. Tabassi, C. Watson, and J. L. Wayman, "Modest proposals for improving biometric recognition papers," in *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*. IEEE, 2015, pp. 1–7.

[7] A. Lambert, R. Bourqui, and D. Auber, "Winding roads: Routing edges into bundles," in *Computer Graphics Forum*, vol. 29, no. 3. Wiley Online Library, 2010, pp. 853–862.

[8] T. Moscovich, F. Chevalier, N. Henry, E. Pietriga, and J.-D. Fekete, "Topology-Aware Navigation in Large Networks," in *SIGCHI conference on Human Factors in computing systems*, A. Press, Ed. Boston, États-Unis: ACM, 2009, pp. 2319–2328. [Online]. Available: http://hal.inria.fr/inria-00373679

[9] A. Martinez and R. Benavente, "The AR face database," *CVC Tech. Report*, 1998.

[10] B. Hemery, C. Rosenberger, and H. Laurent, "The ENSIB database : a benchmark for face recognition," in *International Symposium on Signal Processing and its Applications (ISSPA), special session "Performance Evaluation and Benchmarking of Image and Video Processing"*, 2007.

[11] U. of Essex, "Faces94 database, face recognition data," 1994. [Online]. Available: http://cswww.essex.ac.uk/mv/allfaces/faces94.html

[12] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "Fvc2002: Second fingerprint verification competition," in *International Conference on Pattern Recognition (ICPR'02)*, vol. 3, 2002, pp. 811 – 814.

[13] P.-O. Ladoux, C. Rosenberger, and B. Dorizzi, "Palm vein verification system based on sift matching," in *the 3rd IAPR/IEEE International Conference on Biometrics (ICB'09)*, 2009, pp. 1290–1298.

[14] N. T. Trung, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "Performance evaluation of gait recognition using the largest inertial sensor-based gait database," in *Biometrics (ICB), 2012 5th IAPR International Conference on*. IEEE, 2012, pp. 360–366.

[15] J. Talbot, S. Lin, and P. Hanrahan, "An extension of wilkinson's algorithm for positioning tick labels on axes," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 16, no. 6, pp. 1036–1043, 2010.

[16] D. Auber, D. Archambault, R. Bourqui, M. Delest, J. Dubois, B. Pinaud, A. Lambert, P. Mary, M. Mathiaut, and G. Melancon, "Tulip III," in *Encyclopedia of Social Network Analysis and Mining*. Springer, 2014, pp. 2216–2240.