The Consensus Clustering as a Contribution to Parental Recognition Problem Based on Hand Biometrics

Lúcia Sousa

Higher School of Technology and Management of Polytechnic Institute of Viseu, Portugal.

Abstract— *The clustering analysis is a subject that has been interesting researchers from several areas, such as health (medical diagnosis, clustering of proteins and genes), marketing (market analysis and image segmentation), information management (clustering of web pages). The clustering algorithms are usually applied in Data Mining, allowing the identification of natural groups for a given data set. The use of different clustering methods for the same data set can produce different groups. So, several studies have been led to validate the resulting clusters. There has been an increasing interest on how to determine a consensus clustering that combines the different individual clusterings, reflecting the main structure in clusters inherent to each of them, as a perspective to get a higher quality clustering.*

As several techniques of consensus clustering have been researched, the present work focuses on problem of finding the best partition in the consensus clustering. We analyze the most referred techniques in literature, the consensus clustering techniques with different mechanisms to achieve the consensus, i.e.; Voting mechanisms; Co-association matrix; Mutual Information and hyper-graphs; and a multi-objective consensus clustering existing on literature. In this paper we discuss these approaches and a comparative study is presented, that considers a set of experiments using two-dimensional synthetic data sets with different characteristics, as number of clusters, their cardinality, shape, homogeneity and separability, and a real-world data set based on hands' biometrics shape, in context of people's parental recognition. With this data we intend to investigate the ability of the consensus clustering algorithms in correctly cluster a child and her/his parents. This has an enormous business potential leading to a great economic value, since that with this technology a website can match data, as hands' photographs, and say if A and B are related somehow.

We conclude that, in some cases, the multi-objective technique proved to outperform the other techniques, and unlike the other techniques, is little influenced by poor clustering even in situations like noise introduction and clusters with different homogeneity or overlapped. Furthermore, shows that can capture the performance of the best base clustering and still outperform it. Regarding to real data, no technique was capable of identifying a person's mother/father. However, the research of distances between hands from a person and its father, mother, siblings, can retrieve the probability of that person being his/her familiar. This doesn't enable the identification of relatives but instead, decreases the size of database for seeking the matches.

Keywords— ***Consensus clustering, Hand biometrics, Hand geometry recognition, Hierarchical clustering algorithm, Validation.***

I. INTRODUCTION

The clustering methods are a powerful tool in data analysis because they allow, often without any previous information of the data structure, identifying natural clusters. Due to this, data clustering algorithms have applications in several areas, from medicine to marketing, or from image processing to taxonomy. In clustering analysis, given a set of elements, is intended to identify clusters of these elements such as members of the same cluster have very similar patterns, and elements of different clusters have very different patterns. The attainment of these clusters depends on the clustering method used, that is, on choices inherent to the methods, such as clustering criteria, number of clusters and initial conditions. The use of different clustering methods for a given data set, or the use of the same method but with different initializations (various parameters associated with the method) can produce different clustering, placing thereby the problem of choosing one of these clustering, or determining a consensus clustering which is the combination of different clustering obtained (the base clustering), in order to obtain a better clustering. A better clustering usually is understood as a more stable, robust and consistent clustering, intending to represent a data set in a natural structure in clusters.

Different consensus clustering techniques have been proposed which may result to several different consensus clustering for the same base clustering. In this work, we propose to analyze the performance of the different consensus clustering techniques comparing the clustering obtained and the known truthful clustering. Among some matching indexes suggested in the literature, we applied the Adjusted Rand Index because it is one of the most popular quantifying the proportion of pairs in agreement of two clustering. The consensus clustering techniques for analysis are: i) the traditional: the *Voting K-Means algorithm* [1], the EAC-*Evidence Accumulation Clustering* [2], another based on Mutual Information and hyper graphs [3], [4]; ii) a multi-objective approach, MOCLE- *Multi-objective Clustering Ensemble* [14]. The base clustering is obtained from hierarchical clustering algorithms, namely, *Single-Linkage*, *Complete-Linkage*, *Average-Linkage* and the *Ward* method. We evaluate these approaches with an empirical study using artificial and real-world data sets, being the artificial data with different characteristics regarding number of clusters, cardinality, cohesion and separability. Furthermore, the multidimensional real-world data set is achieved through biometrics of hands images which are related to people recognition. This work is intended to investigate the possibility of parental recognition using biometrics taken from hand images.

Several researches have been developed on the hand biometrics recognition area. Different biometric techniques have emerged, as techniques based on hand shape, hand geometrics and on palm print. The techniques based on hand shape have shown great accuracy since they are capable to achieve almost 100% of people's recognition, as the works [18], [19]. Likewise, in our recent work we achieve 100% of recognition for some considerable samples of hand set [21]. The ability to identify a person by his hand image can be helpful, for instance, for parental relationships identification. There are situations where it's necessary to identify whether a person is another person's child, for example, in the case of children that went missed. Although one can use a genetic test to identify parenting of a child, the photography of the hand is fast, cheap, no need for a technique and can be used remotely to query an online database.

This paper is organized with the following structure: Section 2 is devoted to the hierarchical clustering, consensus clustering algorithms and validation indexes, of interest for this work. Section 3 addresses real-world biometric application, namely patterns of recognition by hand shape biometrics. It follows the work methodology developed to analyze the performance of the consensus clustering techniques in Section 4. Results and conclusions are provided in Section 5 and 6, respectively.

II. HIERARCHICAL CLUSTERING, CONSENSUS CLUSTERING AND VALIDATION

2.1 Hierarchical clustering algorithms

A hierarchical clustering algorithm produces a hierarchy of partitions, represented in a dendrogram. The agglomerative hierarchical algorithm considers, at first, each element of the data set as a cluster, and then, successively aggregate pairs of clusters until all clusters are combined into a single cluster containing all the elements. These algorithms are the most often used so, this work addresses these algorithms, and henceforth we refer only to these algorithms.

The methodology of sequentially aggregate clusters is based on proximities or similarities matrix containing the distance between clusters. According to this proximity matrix and in accordance with the shortest distance, the clusters are aggregated forming a new cluster. Then, with a new cluster, distances are recalculated and hence obtaining a new proximity matrix. The process goes on, and ends with all elements in the same cluster. The representation of this process can be seen on a dendrogram, which is a hierarchy of partitions. At each level of the dendrogram, there is defined a partition with a specific number of clusters. In our studies, being known the structure in clusters of the data sets and consequently the number of clusters we fix a level obtaining a partition or clustering.

Different definitions of distance between clusters lead to different aggregation methods, and then different hierarchies, hence different clusterings for the same data set. For the same aggregation method, existing different definitions of distance between elements, can conduct also to different clustering. In our studies we considered the Euclidian distance between elements, and the following aggregation methods: *Single-Linkage* (SL), *Complete-Linkage* (CL), *Average-Linkage* (AL) and *Ward* (W). Having different clustering for the same data set, the consensus clustering is a contribution towards the resolution of this problem.

2.2 Main methodologies of consensus clustering processing

The various consensus clustering techniques consist of two principal steps: Generation, defining how to produce the set of individual clusterings to combine; Consensus Function, describing how to combine them to find the consensus clustering. In the Generation step, several clusterings to combine must have certain diversity between them, providing more information in the processing of consensus [7]. On second step, the Consensus Function combines these individual clusterings to obtain the consensus clustering. The Consensus Function is so the main step, and can be based for instance, on Voting, Co-association Matrix, Graph and Hyper graph Partitioning, Information Theory, Finite Mixture Models, Genetic Algorithms. Additionally, some consensus functions are based on more than one of these approaches [11]. The methodologies of consensus clustering can be classified as traditional or multi-objective. While traditional aims to get a consensus clustering, the multi-objective can find more than one consensus clustering.

2.2.1 Methods based on the traditional approach

The important contributions based on the traditional approach are the works of, Fred [1], Fred and Jain [2] and Strehl and Ghosh [3], [4]. Also are perhaps, the most referred in the literature.

In [1], the Consensus Function is based on Voting and Co-association Matrix. Is countered the number of times that pair of elements is in the same cluster in the different clustering. This number is set on a matrix, the co-association matrix. The consensus clustering is formed putting in the same cluster, pair of elements having a co-association value higher than 0.5 (the threshold pre-defined). In [2] the EAC (Evidence Accumulation Clustering) technique, the co-association matrix, as in [1], is represented by a graph. The weak links between nodes are cut by a threshold called “highest lifetime”, which corresponds to the minimum weight in the edges. This is analogous to cut the dendrogram produced by SL algorithm, at a level, where the lifetime, the range of threshold, is obtained by the distance between two consecutive levels on the dendrogram, where for each level is delivered a clustering with k clusters, and the range with the highest value is selected as the consensus clustering [11]. In [3], [4], the consensus clustering is achieved by an optimization problem, the maximization of the Consensus Function and hyper graphs representations. The Consensus Function is based on Mutual Information [10]. There are three proposed algorithms based on hyper graph representation and partitioning algorithms: CSPA - Cluster-*based Similarity Partitioning Algorithm*; HGPA - *Hyper Graph Partitioning Algorithm* and MCLA - *Meta-Clustering Algorithm*. The partitioning algorithms used are METIS [8] and HMETIS [9]. The result of each one of these algorithms is a consensus clustering. The final consensus clustering is one of these clusterings that shares most information with the others.

2.2.2 Methods based on multi-objective approach

The most common clustering techniques use only an objective function which allows obtaining a single structure, limiting the knowledge that can be extracted from the data. The techniques, by which the clusterings are obtained by multi-objective optimization, have the intention to overcome this limitation hence it handles simultaneously with more than one objective function, called multi-objective clustering algorithms. One of the main multi-objective clustering algorithms is MOCK - *Multi Objective Clustering with automatic K-determination* [12], which is able to find structures in clustering with multi criteria and also determine the number of clusters which, many times is difficult, as the structure of the data may be unknown. The MOCK finds different structures in clustering with different number of clusters. By the clustering in the optimal set of Pareto decides, through an evolutionary multi-objective algorithm, PESA II – *Pareto Envelope based Selection Algorithm* [13] and two objective functions, which are compactness and connectivity of the clusters to minimize on the optimization process. The compactness is measured by the variance intra-cluster, and connectivity reflects the degree in which elements of different clusters are placed in the same cluster in the new clustering.

Considering that multi-objective clustering algorithms can find many solutions, usually lead to more difficult analysis by domain experts. Thus, the consensus clustering multi-objective allows to give an answer to this problem. The multi-objective consensus clustering, MOCLE - *Multi- Objective Clustering Ensemble* [14], is related to both, multi-objective algorithms and consensus clustering techniques in the optimization process [23], [24]. The MOCLE applies an evolution process to individual clusterings and pairs of the resulting clustering are combined iteratively by a consensus clustering technique to optimize the criteria, resulting in a consensus clustering. The MOCLE uses individual clusterings as the initial population in an evolutionary algorithm based on Pareto, the genetic algorithm NSGA-II - *Non-dominated Sorting Genetic Algorithm* [22], which only uses crossover operator of the individuals. In combination of clustering pairs, it uses a graph representation and the MCLA algorithm. The graph is partitioned into k parts, by METIS partition algorithm, being k the number of clusters of the resulting clustering of this combination and is randomly chosen within the range of the number of clusters of the two combined clustering. The optimization criterion is defined by objective functions in the genetic algorithm, considering a semi-supervised context, consisting of usage of knowledge of a partition in clusters of the data set. The objective functions are three, compactness, connectivity and "information gain" which is based on the concept of entropy of Information Theory [10].

About the optimal set in the front of Pareto based on multi-objective clustering scenario, differences in the assignment of only one element to a different cluster in two partitions can result in different values of the measures optimized. Hence, this can result in several very similar partitions in the approximation of Pareto's front obtained. By this fact, MOCLE generates a concise set of solutions that are representative of the Pareto's front hence finding high quality partitions according to their clustering criteria [25].

2.3 Validating clustering

The current procedure is to calculate, for each clustering obtained, the value of validity indexes, for instance, by the analysis of cohesion or homogeneity and separability of their clusters, allowing to compare them. The various criteria of partition validation, in accordance with the strategy adopted, can be classified in external, relative or internal [6].

Indexes of internal validation evaluate a partition, for instance, by the separability and homogeneity of the clusters. Indexes of external validation evaluate a partition comparing it with the reference partition, usually by knowing the "real" partition, as the index *Adjusted Rand* [5]. Indexes of relative validation compare two partitions, many times applying the same indexes as in external criteria.

Considering the hierarchical base clustering and the consensus clustering, we propose to evaluate the accuracy of these partitions by external criteria, comparing the partitions obtained and the known partition through ARI index.

In this study, situations in which the partition found by the multi-objective procedure is close enough to the structure in clusters which are underlying the data, particularly for the simulated data sets, are illustrated.

III. HAND'S BIOMETRICS FOR RECOGNITION

It has been long known that people's hands differ in their size and shape, and that these differences can be used to distinguish one person from another. Many geometric characteristics of the hand can be measured and used to distinguish identity. For example, measurements such as length width, area, perimeter and thickness, are unique to an individual. Also, combinations of these measurements, such as the ratio of length to width, can be used with good effect as identity discriminator. Recognition systems based on hand biometry are one of the oldest recognition systems, and have been developed in recent decades. Researches in the field of biometrics found that the human hand contains features that can be used for personal identification, as geometry and shape [15]. A biometric system of hand's geometry recognition extracts the most relevant hand features, and with these creates the identity of the person [16]. Some properties of these systems are medium cost, as it needs a platform and a resolution camera, use low-computational cost algorithms, leading to fast results and very easy and attractive to users. An important issue is that the information is collected automatically which could bring ethical problems, but the automatic systems can only proceed to the identification after the person authorizes.

Since the 70's, several authors aim measuring hand characteristics and capture some features for persons' identification. From there, have been increasing the contributions, which developed many systems and different sets of hand features were identified. Those features are, for instance, length and width of the fingers, thickness, area, perimeter and height of the palm hand and finger deviations. Hand geometry recognition systems comprise several steps, such as images acquisition, pre-processing the images, detection and measurement of the feature points, features extraction, including the construction of the data base, and lastly the recognition [21]. Different systems have different forms of processing, at least in one of the steps above. Many authors apply in their studies biometrics as palm print, hand gesture and hand shape. The algorithm available at [17] of a recognition system based on the shape and silhouette of the hand, consist of three steps. The first refers to image processing, where each hand image undergoes a process of normalization of its contour which encompasses segmentation, localization of extremities, ring artifact removal and registration of fingers and wrist. In this step, the hands images are converted into pixels. The second is the feature extraction, where is applied the Independent Component Analysis (ICA) on hands images. The third step is the recognition. Two different architectures were proposed for the recognition. Due to the high dimensionality of the pixels of an image, there is a reduction stage prior applying the PCA (Principal Component Analysis). In our experiments, the features are extracted from those algorithms. We apply the algorithms at the first and second steps, not the recognition phase. As in a recent paper [21] was shown, by these features and the hierarchical clustering algorithms, it is possible to identify people with great percentage of correct identification. Now, we investigate if it's possible to identify peoples parents based on this.

IV. EXPERIMENTAL DESIGN

This work proceeds to computational implementation in Matlab of the clustering algorithms, individuals and the consensus, as well as the clustering validation index which was referred. For obtaining the individual clusterings we consider the hierarchical clustering algorithms *Single Linkage* (SL), *Complete Linkage* (CL), *Average linkage* (AL) and *Ward* method. Cutting on a determined level of the hierarchy, one gets a partition (according to the known partition). As traditional consensus clustering, we employ the *Voting K-Means* *algorithm* (Tec.1), the EAC (Tec.2) and the ones based on Mutual Information and hyper graphs (Tec.3). Regarding to the multi-objective technique, we use the version of MOCLE (Tec.4), available at the server laboratory of the Intelligent and Distributed System of the Federal University of São Carlos, Brazil. Regarding this technique and unlike the traditional ones, the resulting consensus clustering can be more than one. Despite this, in the results, we show the one with greater ARI value. We proceed to a series of experiments for performance analysis and comparison of these different approaches. The evaluation of the clustering obtained is performed using the *Adjusted Rand Index* (ARI) comparing them with the known partition.

In the set of experiments performed are considering simulated data, taking different situations concerning dimensions of the data sets, number of clusters and number of elements that constitute the different clusters, their internal cohesion and separability. We also consider data sets with added noise and data sets with overlapped clusters. Regarding the real-world data set, it is related with a parental recognition by the system based on hands shape mentioned above. We aspire to investigate by consensus clustering, whether it is possible to find the parents of a child through the picture of the right hand. This matter has applications for example, in identifying parents of people who were lost at an early age, during natural calamities and wars. Also, if we want to know who is the father (and the mother) of a person, one cannot perform genetic testing to all people, as it would be very expensive among many other restrictions. If the hand images constitute a database of the potentials reducing the probable parents to a much smaller number, we are saving money. That's the idea, reducing the size of demand without the need to identify exactly who is the father or mother.

A description of each data set is given below.

Simulated data sets

In Fig. 1 to Fig. 5 are represented the 2-dimensional simulated data sets used in our experiments and in Table 1 are the details of those data. The data sets are with random data (according to their partition into clusters) and Normal distribution. Some of them are data sets used by others papers. There are five data sets assigned, D1-4g, D2-3g, D2-3gr10 (data sets D2-3g, with 10% noise), D4-10g [20] and D4-10gSS [20] (data set D4-10g, without overlapped clusters).

Table 1: Details of the simulated data sets. Data generated by Normal distribution, $N(\mu\,,\sigma^2)$ where μ is the mean and σ^2 is the variance. D is the dimensionality, C the number of clusters, Ni the number of data elements for cluster i, OC and AN means overlapped clusters and add noise, respectively. The data noise are generated by Uniform distribution U(a,b) where (a,b) is the support interval.

Name	**D**	**C**	**Ni**	**Source**	**OC**	**AN**
D1-4g	2	4	15×35×35×35	C1: $N((0.5,0)\,,(0.05,0.05))$, C2: $N((-1,4)\,,(0.2,0.2))$ C3: $N((2,0)\,,(0.2,0.2))$, C4: $N((2,3.5)\,,(0.2,0.2))$	No	No
D2-3g	2	3	3×50	C1: $N((-1,0)\,,(0.25,0.25))$, C2: $N((1.5,2.5)\,,(0.25,0.25))$ C3: $N((8.5,10)\,,(2.25,2.25))$	No	No
D2-3gr10	2	3	50×56×59	C1: $N((-1,0)\,,(0.25,0.25))$, C2: $N((1.5,2.5)\,,(0.25,0.25))$, U(3,4) C3: $N((8.5,10)\,,(1.5,2.25))$, U(6,7)	No	Yes
D4-10g	2	10	25×5 50×5	Ci: $N(([0,50],[0,50])\,,(\,[0.1,0.3],[0.1,0.3]))$ i=1,..10.	Yes	No
D4-10gSS	2	10	25×5 50×5	Ci: $N(([0,50],[0,50])\,,(\,[0.1,0.3],[0.1,0.3]))$ i=1,…,10. For each 2 clusters, d(c_k, c_l)>3($\sigma_k+\sigma_l$) where c_k and c_l are the center points respectively [20].	No	No

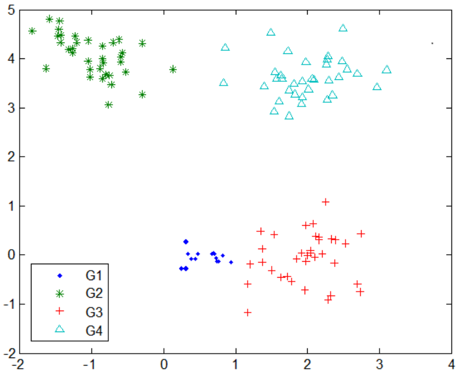

Fig.1: Representation of data set D1-4g.

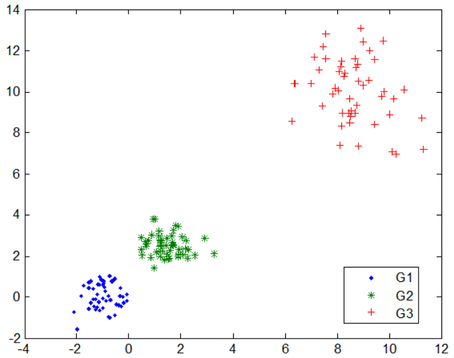

Fig. 2: Representation of data set D2-3g.

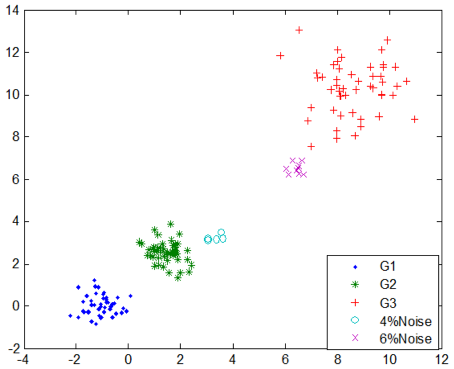

Fig. 3: Representation of data set D2-3gr10.

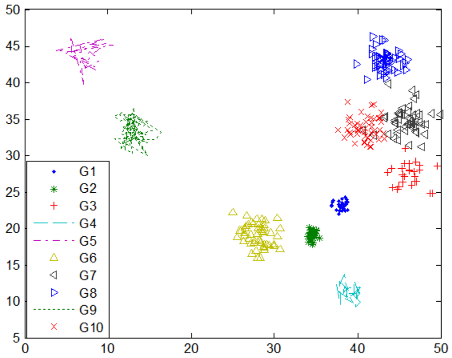

Fig. 4: Representation of data set D4-10g.

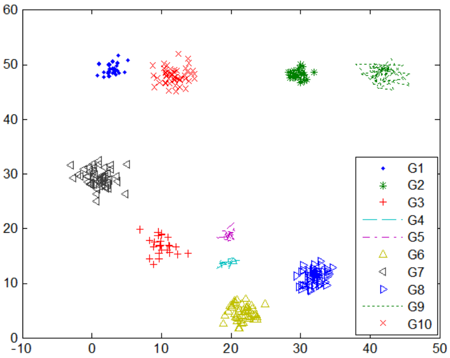

Fig. 5: Representation of data set D4-10gSS.

Real-world data set

In order to carry this research, the experiments are performed over the hand images. Our hand images database consists of right hand images (palm and fingers) from 187 persons, parents and children, and 3 hand images from each person. 271 features per image were extracted using the algorithms, available at the Bosphorus Hand Database [17]. Those features are based on the hand shape silhouette.

This hand images database was created to develop a parental biometric recognition system, and the unique constraint is the hand must be placed ober a black background. All images were acquired through normal mobile phone, in different situations of luminosity and proximity, or scanner. These images initially saved as JPG images, were converted to bitmap images with 382×525 bits definition, 588Kb, with color image resolution. The features of these images were extracted by the algorithms described in [17]. These algorithms started by normalize the hand shape, procedure that extracts the hand region from the background, detects and localizes the hand extremities. The data set to clustering by the hierarchical and consensus clustering algorithms are created by the features extracted using Independent Component Analysis, which consist on coefficients statistically independents of sources pixels of hand images. The experiences based on normalized hand images were performed on families formed by fathers, mothers and children. Our goal is to take a persons' photos (with 3 photos) and see if someone else corresponds to it in terms of father, mother or sibling. This will happen if they are placed together in the same cluster. So, we analyze different databases considering: 1) fathers and children (F); 2) mothers and children (M); 3) siblings (S) and 4) the family, i.e., parents and children (P). Some of these images are illustrated in Fig. 6.

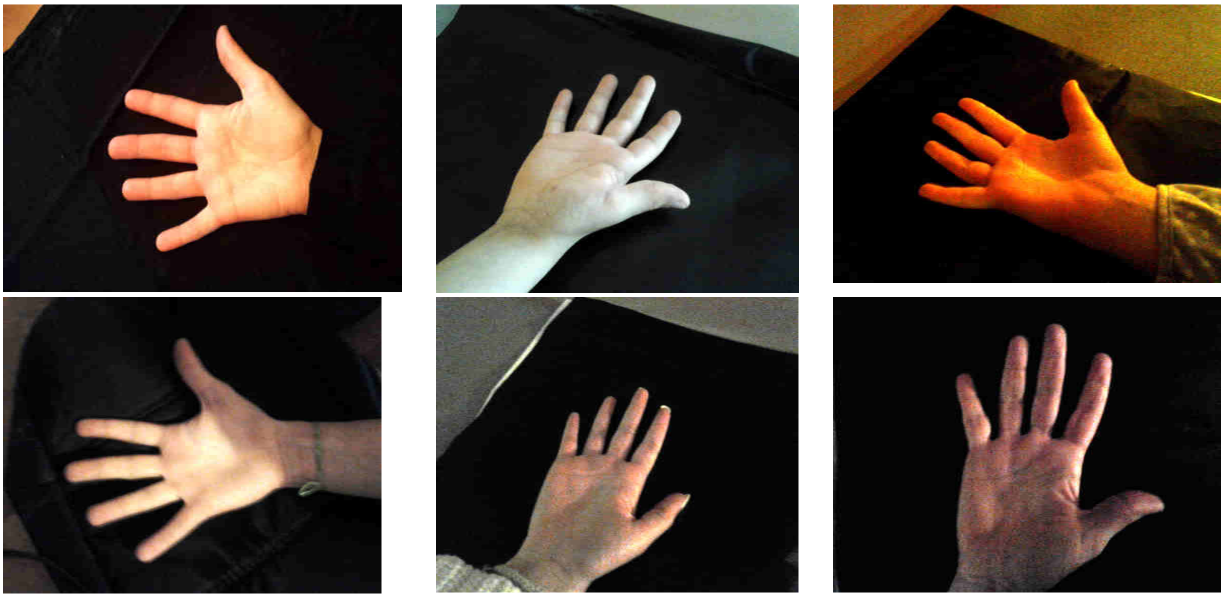

Fig. 6: Six examples of hand images of six different persons in our database.

V. RESULTS AND DISCUSSION

The results obtained in the experiments are presented on Tables 2 - 4. First, focusing on simulated data sets and observing the results of the ARI values on Table 2, in accordance with the partitions having clusters with different cardinalities, even having close clusters (D1-4g and D4-10gSS) or overlapped (D4-10g), most of the individual hierarchical clusterings show approximately the same and a good performance, as well as the consensus clustering techniques. Now, for the partitions with clusters having big difference between them regarding to the homogeneity and close clusters (D3-2g), or with data noise added (D2-3gr10), the individual hierarchical clusterings show different performance between them, as also the consensus clustering techniques. In fact, some individual hierarchical clusterings, as some consensus clustering technique, present relatively better performance than others. In general, the performances of the consensus clustering techniques are affected by the performance of the individual clusterings. One can say that the performance of the traditional techniques is in accordance with the performance of most of individual clusterings. On the other hand, the multi-objective technique seems to be influenced by the clustering with good performance and not by most of them, as the examples D2-3g and D4-10g. Moreover, it can outperform the individual clusterings (see D4-10gSS).

Noting Table 3, for the real-world data sets, derived from databases of hand images of parents and their children, the ARI values of consensus clustering techniques do not reveal a great performance. But, despite that, in some cases regarding to the database and the consensus clustering technique, the ARI values are close or equal to 0.30, it reveals that, it’s possible to exist some agreement or association between the parents and children. With the purpose of finding out this association, we can think about finding the distances between a child and his/her father, or mother or siblings, by the features extracted from their hands. Because, as the clusters are formed by the distances and the ARI values suggest that the parents and children are not so close, we want try to know how far is a child from his/her parents. This is another analysis of this framework. The procedure is: we get a person's photo and calculate the distances to all others photos. Having, 3 hand images for each person, and calculating the distances between each two people, we have 9 distances. Our statistic is the distance between each two persons, using the minimum of these 9 distances. Analyzing the distribution of these distances for all the people allows us to verify, for instance, if A has his/her father, mother or sibling among 10% of the closest persons. According to the probability of a child have his/her father, mother or sibling among 10% of the closest persons, if it's for instance 95%, then the search for the parent of a child can be reduced for the 10% of the closest people in the database.

In respect to the distances between people in the database, we search to fulfil the sentence: “Running the hand images of a person by the database where M is, there is the probability P of be identified among p closest persons.” We consider, M= {father, mother, sibling, at least one of these familiar} and p= {10%, 25%, 50%}. The probabilities P are on Table 4. According to these results, we can state that running hand images of a person by the database where are the father, mother and a brother, there is 95% probability of at least one of the family be in the half of those closest. This does not allow the identification of one of the relatives but can restrict the search space for half, for instance, in a genetic test.

Table 2: For each simulated data set, the ARI values of the, A- individual clusterings; B- consensus clustering techniques.

A

Data sets	Algorithm	ARI
D1-4g	SL	0.8143
	CL	0.9823
	AL	0.9823
	W	0.9823
D2-3g	SL	0.5584
	CL	0.4448
	AL	0.5584
	W	1
D2-3gr10	SL	0.3500
	CL	0.7937
	AL	0.3500
	W	0.7937
D4-10g	SL	0.7681
	CL	0.9518
	AL	0.9402
	W	0.9402
D4-10gSS	SL	0.9945
	CL	0.9946
	AL	0.9946
	W	0.9946

B

Data sets	Technique	ARI
D1-4g	Tec.1	0.9823
	Tec.2	0.9823
	Tec.3	0.9823
	Tec.4	0.9823
D2-3g	Tec.1	0.5584
	Tec.2	0.5681
	Tec.3	0.5681
	Tec.4	**1**
D2-3gr10	Tec.1	0.7274
	Tec.2	0.7274
	Tec.3	0.7937
	Tec.4	0.7937
D4-10g	Tec.1	0.9402
	Tec.2	0.9377
	Tec.3	0.9402
	Tec.4	**0.9518**
D4-10gSS	Tec.1	0.9946
	Tec.2	0.9946
	Tec.3	0.9946
	Tec.4	**1**

Table 3: ARI values of the clustering according to the database and the consensus clustering technique.

Database	Technique	ARI
F	Tec.1	0.1571
	Tec.2	0.3032
	Tec.3	0.2280
	Tec.4	0.2463
M	Tec.1	0.2030
	Tec.2	0.2901
	Tec.3	0.2460
	Tec.4	0.2299
S	Tec.1	0.2711
	Tec.2	0.2875
	Tec.3	0.2414
	Tec.4	0.2915
P	Tec.1	0.1283
	Tec.2	0.2165
	Tec.3	0.1762
	Tec.4	0.1883

Table 4: The entries are probabilities of M be among p of closest persons of a child.

M\p	10%	25%	50%
Father	29,3%	53,3%	79%
Mother	40%	57,5%	78%
Sibling	52,2%	76%	94%
A familiar	64,2%	85%	95%

VI. CONCLUSIONS

In this paper we proposed to analyze the performance of some of the approaches most referred in literature of consensus clustering. The traditional and multi-objective consensus clustering techniques have different consensus functions applied to individual hierarchical clusterings.

The experiments were performed by simulated and real-world data sets. For the simulated data sets, we can conclude that the traditional consensus clustering algorithms are more susceptible to the existence of individual clustering of worse quality than MOCLE algorithm.

Regarding the real data sets, these are derived from hand images of parents and children. Each one of these images provides 271 features extracted from shape of the hands. This work was intended to identify the father or mother of a person by consensus clustering techniques. All the techniques presented approximately the same performance, and not a good one. Although the performances of these consensus clustering have not been good, meaning that, by the ARI values, parents and their children are not close enough to be placed in the same cluster, the ARI values allowed to conclude that the consensus clustering and the real clustering are not in total disagreement, i.e., there are some proximity between parents and their children. So, in another analysis, the obtaining of the distances between all the people enables to conclude that, regarding a person that has the father, mother and/or a sibling in the database, there is a great probability of at least one of them be in 50% of the closest persons. This is a good result that, although does not enable to identify the parents of a child, instead, allows reducing the domain of research substantially.

As final remarks we must refer that most of the hand images were taken with a mobile phone and at different conditions, as luminosity, which may be loss making. We believe that, being the collection of images made by scanner, the results by the consensus clusterings will be predictable better. Thus, this is reserved for a future work.

REFERENCES

[1] A. Fred, Finding consistent clusters in data partitions, in J. Kittler and F. Roli, editors, Multiple Classifier Systems, volume LNCS 2096. Springer, 2001, pp. 309-318.

[2] A. Fred and A. Jain, Combining Multiple Clusterings Using Evidence Accumulation, IEEE Trans Pattern Analysis and Machine Intelligence 27(6), 2005, pp. 835-850.

[3] A. Strehl and J. Ghosh, Cluster Ensembles - A Knowledge Reuse Framework for Combining Partitionings, in Proc. Conference on Artificial Intelligence. Edmonton, 2002, pp. 93-98.

[4] A. Strehl and J. Ghosh, Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions, Journal of Machine Learning Research, 3, 2002, pp. 583-617.

[5] L. Hubert and P. Arabie, Comparing Partitions, Journal of Classification, 2, 1985, pp. 193-218.

[6] A. K. Jain and R. C. Dubes, Algorithms for clustering data, Ed. Prentice Hall, Inc, 1988.

[7] S. T. Hadjitodorov, L. I. Kuncheva and L. P. Todorova, Moderate diversity for better cluster ensembles, Information Fusion, vol. 7(3), 2006, pp. 264-275.

[8] G. Karypis and V. Kumar, A fast and high quality multilevel scheme for partitioning irregular graphs, SIAM Journal of Scientific Computing, 20 (1), 2004, pp. 359-392.

[9] G. Karypis, R. Aggarwal, V. Kumar and S. Shekhar, Multilevel hypergraph patitioning: Applications in VLSI domain, in Proc. of the Design and Automation Conference, 1997.

[10] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd edn., Wiley Series in Telecommunications and Signal Processing (Wiley-Interscience), 2006.

[11] S. Vega-Pons and J. Ruiz-Shulcloper, A Survey of Clustering Ensemble Algorithms, Internation Journal of Pattern Recognition and Artificial Intelligence, vol. 25 (3), 2011, pp. 337-372.

[12] J. Handl and J. Knowles, Multiobjective clustering with automatic determination of the number of clusters, Technical Report TR-COMPSYSBIO-2004-02, MIST, Manchester, 2004.

[13] D. Corne, N. Jerram, J. Knowles and M. Oates, PESA-II: Region-based Selection in Evolutionary Multiobjetive Optimization, in Proc. of the Genetic and Evolutionary Computation Conference (GECCO-2001),San Francisco, California, USA, Morgan Kaufmann, 2001, 283-290.

[14] K. Facelli, Um framework para análise de agrupamento baseada na combinação multi-objetivo de algoritmos de agrupamento, in Tese apresentada ao ICMC-USP para obtenção do título de Doutor em Ciências de Computação e Matemática Computacional. USP, 2007.

[15] T. Fotak, P. Koruga and M. Baca, Trends in hand geometry biometrics, Central European Conference on Information and Intelligent Systems, 2012, pp. 319-493.

[16] H. Matos, Reconhecimento biométrico baseado na geometria da mão, Mestrado em Engenharia Electrotécnica e Computadores, Major Automação, 2011.

[17] Bosphorus Hand Database. <http://bosphorus .ee.boun.edu.tr/hand/home.aspx>, accessed on 02/03/2016.

[18] E. Konukoglu, E. Yoruk, J. Darbon and B. Sankur, Shape-based hand recognition, IEEE Trans. on Image Processing, 15(7), 2006, pp. 1803-1815.

[19] E. Yoruk, H. Dutagaci and B. Sankur, Hand Biometry, Image And Vision Computing, 24(5), 2006, pp.483-497.

[20] I. Gurrutxaga, I. Albisua, O. Arbelaitz, J. Martin, J. Muguerza, J. Perez and I. Perona, SEP/COP: An efficient method to find the best partition in the hierarchical clustering based on a new cluster validity index, Pattern Recognition, 43, 2010, pp. 3364-3373.

[21] L. Sousa, J. Gama, The application of hierarchical clustering algorithms for recognition using biometrics of the hand. International Journal of Advanced Engineering Research and Science, IJAERS Journal, vol. 1, Issue 7, ISSN: 2349-6495, Dec-2014.

[22] K. Deb, A. Pratap, S. Agarwal and T. Meyrivan, A Fast and Estilist Multiobjective Genetic Algorithm: NSGA-II, IEE Transion on Evolutionary Computation, 6(2), 2002, pp. 182-197.

[23] R. Liu, Y. Liu and Y. Li, An Improved Method for Multi-objective Clustering Ensemble Algorithm, WCCI 2012 IEEE, 2012, pp. 10-15.

[24] K. Faceli, A. Carvalho and M. de Souto, Multi-objective clustering ensemble for gene expression data analysis, Neurocomputing 72, 2009, pp. 2753-2774.

[25] K. Faceli, A. C. P. L. F. Carvalho and M. C. P. Souto, Multi-objective Clustering Ensemble, International Journal of Hybrid Intelligent Systems - Hybridization of Intelligent Systems, Volume 4 Issue 3, 2007, pp. 145-156.