Constrained affinity matrix for spectral clustering: A basic semi-supervised extension

C. Castro, D. H. Peluffo, C. G. Castellanos

***Abstract*—Category 2. Spectral clustering has represented a good alternative in digital signal processing and pattern recognition, however decisions concerning the affinity functions among data is still an issue. In this work its presented an extended version of a traditional multiclass spectral clustering method which employs prior information about the classified data into the affinity matrixes aiming to maintain the background relation that might be lost in the traditional manner, that is using a scaled exponential affinity matrix constrained by weighting the data according to some prior knowledge and via k-way normalized cuts clustering, results in a semi-supervised methodology of traditional spectral clustering. Test was performed over toy data classification and image segmentation and evaluated with and unsupervised performance measures (group coherence, fisher criteria and silhouette).**

***Index Terms*—Affinity matrix, kernel methods, prior information, semi-supervised analysis, spectral clustering.**

I. INTRODUCTION

SPECTRAL clustering, which is an unsupervised method of data analysis (that is does no require any prior data), is a discriminative method based upon the graphs theory and has as initialization parameters affinity matrixes and the number of groups for classification[1]. In [2] some methods for estimating the number of groups from small subsets of data based on the spectral information provided from the data itself and from affinity matrixes were explored, here a constrained affinity matrix is proposed for a spectral clustering in order to enhance the performance by ensuring some affinity coherence in the clustering process.

In graph theory, affinity matrixes have the property and capability of represent the relationship grade among nodes since they are constructed as positive semi-definite symmetric matrixes which values are can be define by several functions including random, trivial (inner products row-wise from the data matrix); in [3] is presented a weighted affinity matrix, based on a vector α such that $\mathbf{W}_\alpha = \mathbf{X}\,\mathrm{diag}(\alpha)\,\mathbf{X}^\top$ and the vector is obtained trough an iterative process called $Q-\alpha$.

Yet as the distance stands as a similitud measure, also have been proposed affinity matrixes that includes distances among data but that are smoothed by exponential functions as in [4] where a global scale parameter in introduced into an exponential affinity matrix. A more robust and enhanced version of the previously mentioned was introduced by [5] using a local scale parameter instead of a global one.

Cristian Castro-Hoyos: ccastroh89@gmail.com, Master student, Universidad Nacional de Colombia, Sede Manizales.

Diego Hernán Peluffo-Ordoñez: dhpeluffoo@hotmail.com, Ph.D. student, Universidad Nacional de Colombia, Sede Manizales.

César Germán Castellanos-Domínguez: cgcastellanosd@unal.edu.co, Ph.D., Professor Department of electrical engineering, electronic engineering and computer sciences of Universidad Nacional de Colombia, Sede Manizales.

Previous mentioned affinity matrixes uses weights as a parameter of smoothing the affinity and maintaining the internal structure relation of the data, this is then the same concept that uses Weighted Principal Component Analysis (WPCA) where the covariance matrix is weighted either in the samples or in the features matrixes as in [6], where a Weighted kernel PCA method is used for a multi-way spectral clustering algorithm with out-of-sample extension.

In general, spectral clustering uses the information provided by such affinity matrix to classify the data according to their relations, nonetheless the criteria for which the affinity matrix considers data to be similar or not is not always the same even when data has the same nature, and thus the clustering performance might decrease. Yet as an alternative you might introduce prior knowledge to the clustering procedure to ensure some stability, that is, a semi-supervised clustering. Here, its proposed an alternative solution to this by introducing a constrained affinity matrix, the affinity is define by a scaled exponential matrix, which is constrained by assigning a great value to those elements that prior to clustering are known to belong to a given class or cluster. For evaluation of the proposed affinity matrix, clustering is performed for toy-data classification and image segmentation based on pixel-by-pixel clustering, via different clustering algorithms, first as reference k-means is performed, second for the evaluation of the affinity matrix, we test the semisupervised affinity matrix via the evaluation of two different spectral clustering algorithms, a multiclass spectral clustering proposed by [7] and a multiway spectral clustering [6]. Performance is evaluated in terms of unsupervised measures as groups coherence[1], fisher criteria and silhouette.

II. THEORETICAL FRAMEWORK

A. Preliminaries on Spectral Analysis and Graph theory

Signal analysis and data clustering based upon spectral information provided by the internal distribution and structure of the observations is a common discriminative technique that does not require any prior information such as structure supposition, partitions distributions etc. Instead is based in statistical global criteria that estimates the probability that two observations belong to the same class. This criteria or analysis can be describe from the graph theory where such probability is define as the affinity between the nodes in the graph, that is, the weight value of the link that connects them.

Now, having a definition of affinity parting from a graph, we can expect that given a set of observations or pixels, those which belongs to the same class or subset of $\mathbb{G}$ have a higher affinity among them and lower to those in different classes. In the next sections we show two different affinity matrices, the locally scaled exponential affinity matrix[5] and the one proposed here, the constraint affinity matrix where the restriction of high affinity for known observations is expected to enhance the clustering process since some regularity is ensure in the probabilistic information of the data.

B. Exponential Affinity Matrix

Let $\mathbf{X}$ be an $n \times p$ matrix, where n is the number of observations and p the number of features, we define the exponential affinity matrix $\hat{\mathbf{W}}$ as,

$$\hat{w}_{ij} = \exp\left(\frac{-\,\mathrm{d}^2\left(\mathbf{x}_i,\mathbf{x}_j\right)}{\sigma_i\sigma_j}\right) \quad \text{with } \sigma_i = \mathrm{d}\left(\mathbf{x}_i,\mathbf{x}_N\right) \tag{1}$$

where in Equation 1 we have that, $\mathbf{x}_i$ is each data vector, $\mathrm{d}\left(\mathbf{x}_i,\mathbf{x}_j\right)$ is the Euclidean distance between vectors and σ_i is the local scale parameter, with $\mathbf{x}_N$ being the $N-$th neighbor of $\mathbf{x}_i$.

Now, let us define the diagonal normalization matrix $\mathbf{D}$ as, $d_{ii} = \sum_{j=1}^{n} \hat{w}_{ij}$ and construct the normalized affinity matrix $\mathbf{L}$

$$\mathbf{L} = \mathbf{D}^{-1/2}\hat{\mathbf{W}}\mathbf{D}^{-1/2} \tag{2}$$

This matrix is then the used in the clustering process.

C. Constraint Affinity Matrix

As mentioned before, in a given set of observations described by a graph $\mathbb{G}$, is expected that those which belong to the same class to have higher affinity among them a lower with respect to the rest. Moreover, based on the Perron-Frobenius theorem and the eigen-decomposition of irreducibly positive matrixes, the space generated by the eigen-vector is directly related with the quality of a clustering process [8].

Hence, where introducing some prior information into the affinity distribution of a given set, the eigen-space generated is expected to have some regularity and coherence within the clusters. This constraint, induces then a semi-supervised procedure into the clustering process, but is expected to enhance the performance.

Thereby we can define a constraint affinity matrix $\tilde{\mathbf{W}}$ as

$$\tilde{w}_{ij} = \begin{cases} \exp\left(\frac{-\,\mathrm{d}^2\left(\mathbf{x}_i,\mathbf{x}_j\right)}{\sigma_i\sigma_j}\right) & \text{if } ij \notin \mathbf{I} \\ \Upsilon & \text{if } ij \in \mathbf{I} \end{cases} \tag{3}$$

where in Equation 3 we have that $\mathbf{I}$ is a $m \times 2$ constraint matrix, with m being twice the number of known links connections row-wise and Υ a positive constant such as $\Upsilon \in \mathbb{N}$, i.e. given the graph depicted in Figure 1, where the red and blue nodes denotes two different classes and the orange links are known to belong to a given class, the constraint matrix $\mathbf{I}$ is,

$$\mathbf{I} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \\ 5 & 6 \\ 6 & 5 \\ 4 & 8 \\ 8 & 4 \end{bmatrix}$$

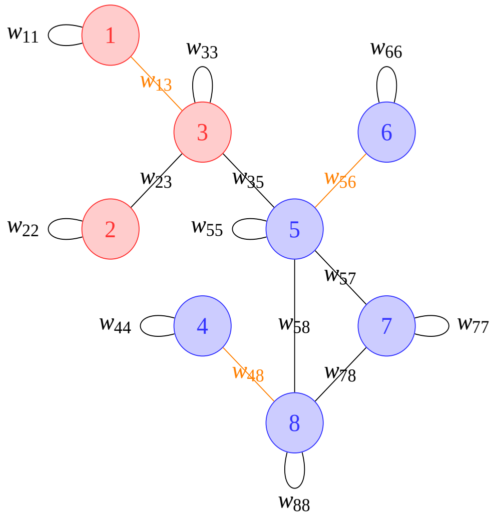


Fig. 1: Undirected graph ($w_{ij} = w_{ji}$), with known links(orange links).

D. Multiclass Spectral Clustering

Multiclass spectral clustering proposed by [7] its based on the solution of the objective function

$$\begin{aligned} \text{maximize}\left\{\varepsilon\left(\mathbf{M}\right)\right\} &= \frac{1}{k}\sum_{l=1}^{k}\frac{\mathbf{M}_l^{\mathsf{T}}\mathbf{W}\mathbf{M}_l}{\mathbf{M}_l^{\mathsf{T}}\mathbf{D}\mathbf{M}_l} \\ \text{s.t.}\,\mathbf{M} \in \{0,1\}^{n\times k} & \qquad \mathbf{M}\mathbf{1}_k = \mathbf{1}_n, \end{aligned} \tag{4}$$

which is called the Normalized Cuts Problem (NCPM, where M denotes that the optimization problem is respect to the partition matrix $\mathbf{M}$). And $\mathbf{W}$ is the affinity matrix, $\mathbf{D}$ is the degree matrix

$$\mathbf{M} = [\mathbf{m}_l] \quad l = 1,\ldots,k \tag{5}$$
$$\mathbf{c} = \mathbf{W}\mathbf{1}_n \tag{6}$$
$$\mathbf{D} = \mathrm{Diag}\left(\mathbf{c}\right) \tag{7}$$

where in equation 5, $\mathbf{M}$ is a binary partition matrix that represents the membership of a given observation to a partition, that is, $\mathbf{m}_l$ is a binary column formed as $m_{il} = \langle i \in \mathbb{V}_l \rangle \;\; i \in \mathbb{V}$, being $\langle \cdot \rangle$ a binary operator that is 1 if true and 0 otherwise and $\mathbb{V}_l$ is the l-th partition of the graph set. Matrix $\mathbf{D}$ is the degree matrix.

However, this spectral analysis obtains global-optimal solutions in the continuos eigenspace generated by the partition and degree matrix, and hence in contained in a high dimensional space and has no restrictions and hast therefor the need of getting a discrete solution with an approximate discrete optimal as it is detailed in [7].

III. Experimental Setup

Evaluation of the proposed semi-supervised affinity is made in comparison to common unsupervised clustering methods under the analysis of toy-data and image segmentation. Toy-data is numerically generated with different patterns and number of groups to demonstrate stability of the affinity proposition; 15 different toy-data two dimensional patterns are used for clustering, having different number of classes being the lowest 2 and 31 the highest.

On the other hand multiclass spectral clustering has been successfully applied to image segmentation from a pixel by pixel clustering so the Berkeley database for image segmentation is used [9].

To asses clustering we compare against two unsupervised methods: *k*-means and multiclass spectral clustering with scaled exponential affinity matrix. The performance of those is the compared against the performance of the multiclass spectral clustering method with a semi-supervised affinity matrix.

A. Database

1) Synthetic Data: Synthetic data used is a set of different multiclass datasets, with multiple observation, features and classes. This data is not linearly easily separable and is often used for testing and tuning of clustering methods. Table I shows the summary of the synthetic data

TABLE I: Synthetic Database Description

Dataset	Observations	Features	Classes	Cite
Cluster2Circle1	238	2	3	
HappyFace	266	2	3	
Bullseye3	299	2	3	
Clusters2Noise	300	2	3	
Cluster3	303	2	3	
Bullseye2	500	2	2	
Lines4	512	2	4	
Gaussians4	1000	2	4	[5]
Clusters5Noise	622	2	5	
Spiral	312	2	3	
PathBased	300	2	3	[10]
Flame	240	2	2	[11]
Jain	373	2	2	[12]
Compound	622	2	5	[13]
Aggregation	788	2	7	[14]
D31	3100	2	31	[15]

2) Berkeley Segmentation Dataset: This database is composed of 300 481×321 or 321×481 pixels jpeg images. Database is separated into *train* with 200 images and *test* with the rest 100. Since segmentation is based in a pixel by pixel unsupervised and semisupervised clustering, each pixel has to be considered an observation, hence affinity matrixes are too big for numerical computation so images are resized into a one eight of their original size, that is from 481×321 to 61×41 and from 321×481 to 41×61. Now each pixel is characterized into color spaces forms **rgb, hsv, ycc, lab and luv**, so that for color RGB images we have 3 features per color per pixel for a subtotal of 15 features, finally the spatial x, y location of the observation is added. Hence the image pixel feature matrix size would be 2501×17.

B. Methods and Algorithms

C. Performance Evaluation

In order to asses the clustering performance, we employed three measures namely. Firstly, cluster coherence that quantifies the clustering quality in terms of the cluster affinity associations. The second one is the well known fisher criterion index. Finally, the mean and standard deviation of silhouette of each data point is also considered.

Algorithm 1 Evaluation of proposed constraint for clustering

1: Initialization: For a dataset matrix $\mathbf{X}$, choose a k number of classes, a clustering algorithm, either *k-means* or *Multiclass Spectral Clustering*.
2: **if** k-means **then**
3: Do k-means on X matrix and go to step 8.
4: **else if** Multiclass Spectral Clustering **then**
5: Determine if procedure is unsupervised or semisupervised, and compute the respective affinity matrix as in section II-B or II-C.
6: Realize the clustering procedure with the respective affinity matrix and go to step 8.
7: **end if**
8: Compute performance measures: Fisher Criteria, Groups Coherence and Silhouette to evaluate clustering.

TABLE II: Applied performance measures

Measure	Description
Fisher Criterion	Indicates an adequate clustering when the greater is its value.
J	$J = \frac{\sum_{j=1}^{k} \mathbf{q}_k - \hat{\mathbf{q}}}{\mathrm{tr}(\Sigma\Sigma_j)}$ where $\mathbf{q}_j$ is the mean of i-th cluster, $\hat{\mathbf{q}}$ is the mean of whole data $\mathbf{X}$ and Σ_j is the covariance matrix associated to cluster j.
Cluster Coherence	It is ranged into $[-1, 1]$ and close to 1 when well clustering.
$\varepsilon_{\mathbf{M}}$	$\varepsilon_{\mathbf{M}} = \frac{1}{k} \sum_{l=1}^{k} \frac{\mathbf{M}_l^\top \blacksquare \mathbf{M}_l}{\mathbf{M}_l^\top \mathbf{D} \mathbf{M}_l}$
Silhouette	It ranges from -1 to 1, being 1 when clustering adequately.
S	$s_i = \frac{\min(\mathbf{b}_i - a_i)}{\max(a_i, min(\mathbf{b}))}$ where a_i is the average distance from the i-th point to the other points in its cluster, $\mathbf{b}_i = (b_1^i, \ldots, b_k^i)$ and b_j^i is the average distance from the i-th point to points from cluster j.

IV. Results and Discussion

Results shown that including prior information into the affinity structure of the data may influence into the spectral analysis, moreover the value of the Υ constant in the semisupervised analysis has a major influence into the performance of the clustering process.

Hence it is necessary to generate affinity matrixes that can take this semisupervised approach into a better account, since the nature of the exponential affinity matrix based upon an euclidean distance limits the performance.

A. Toydata

Tables III, IV and V show the perfomance results obtained for the toydata, it can be seen that spectral approaches to hard clustering problem allow to enhance the performance, however the semisupervised approach does not provide an stable value for the constan Υ which might be given the exponential nature of the matrix.

Nonetheless the results of the semisupervised approach might be useful for providing a tradeoff performance when dealing with several performance measures, since the behavior has exposed that the semisupervised approach, more specifically the value of the constant Υ influence all the used measures since affects directly both the output clustering labels and the affinity used for analysis.

TABLE III: K-means clustering performance for toydata.

Dataset	k	μ_S	σ_S	J	ε
Aggregation	**8**	0.6313	0.2356	29.5832	0.9427
Bullseye2	**3**	0.5670	0.2312	3.2782	0.9816
Bullseye3	**4**	0.4649	.2761	4.0039	0.9577
Clusters2Circle1	**4**	0.8855	0.1930	13.4637	0.9324
Clusters2Noise	**4**	0.7022	0.2097	6.8112	0.9070
Clusters3	**4**	0.7469	0.2278	31.8204	0.9581
Clusters5Noise	**7**	0.7822	0.2152	13.3284	0.9210
Compound	**6**	0.6062	0.2534	32.7492	0.8971
D31	**32**	0.6813	0.2358	164.9440	0.8580
Flame	**3**	0.5772	0.2479	25.5294	0.9252
Gaussians4	**5**	0.7384	0.1965	21.9565	0.9706
HappyFace	**3**	0.7142	0.2242	10.9893	0.9833
Jain	**3**	0.6591	0.2523	12.6131	0.9802
Lines4	**5**	0.6224	0.2725	10.6681	0.9788
Pathbased	**4**	0.6287	0.2262	7.8660	0.9476
Spiral	**4**	0.5243	0.2211	5.0509	0.9408

TABLE IV: Multiclass Spectral Clustering performances with regular affinity matrix.

Dataset	k	μ_S	σ_S	J	ε
Aggregation	**8**	0.6539	0.2789	39.2189	0.9775
Bullseye2	**3**	0.1847	0.2986	1.1506	0.9995
Bullseye3	**4**	−0.0350	0.5909	0.4049	0.9925
Clusters2Circle1	**4**	0.6519	0.6396	2.9398	0.9806
Clusters2Noise	**4**	0.2606	0.7384	0.7177	0.9773
Clusters3	**4**	0.7493	0.2691	19.7965	0.9682
Clusters5Noise	**7**	0.6258	0.6971	3.0167	0.9870
Compound	**6**	0.5437	0.3150	38.2338	0.9132
D31	**32**	0.7593	0.1907	186.0452	0.9240
Flame	**3**	0.5681	0.2774	25.6426	0.9494
Gaussians4	**5**	0.7850	0.1881	22.9379	0.9862
HappyFace	**3**	0.6286	0.4194	9.3491	0.9998
Jain	**3**	0.5985	0.3357	9.4346	0.9946
Lines4	**5**	0.5549	0.4156	5.3322	0.9931
Pathbased	**4**	0.6396	0.2924	7.4782	0.9687
Spiral	**4**	−0.2249	0.5599	0.6750	0.9821

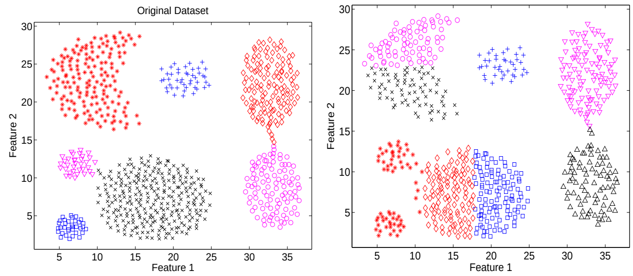
Figures 2 to 3 shows some clustering results and the original clustering.

Notice the in the case of k-means clustering and multiclass spectral clustering, the number of clusters is one more of the original, this was done to provide a certain error margen in the experiments and visualiza how the classes distribution could change.

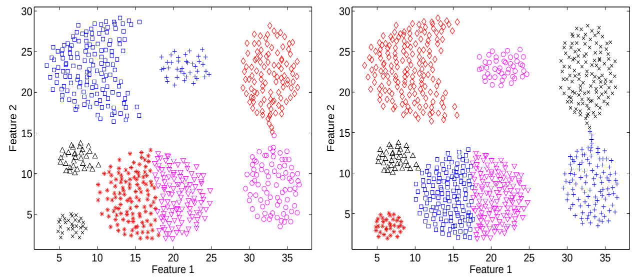
It can be seeing that even when the semisupervised analysis does not enhances the performances in a great measure, it gives a certain structure to the clustering, and the clustering error can be attribute to the extra class restriction, moreover that class is usually adjacent to another and can be corrected with a hierarchical clustering procedure.

TABLE V: Multiclass Spectral Clustering performances with constraint affinity matrix..

Dataset	k	Υ	μ_S	σ_S	J	ε
Aggregation	**8**	**0.00**	0.6508	0.2825	38.8230	0.9773
Bullseye2	**3**	**0.00**	0.1931	0.2699	0.5590	0.9997
Bullseye3	**4**	**1.00**	−0.1761	0.5455	6.9584	0.9318
Clusters2Circle1	**4**	**0.00**	0.6684	0.6277	2.9568	0.9847
Clusters2Noise	**4**	**0.00**	0.4821	0.7385	1.3175	0.9731
Clusters3	**4**	**0.01**	0.7674	0.2559	20.5451	0.9497
Clusters5Noise	**7**	**0.00**	0.6179	0.6982	3.9521	0.9856
Compound	**6**	**0.01**	0.5413	0.3033	38.5231	0.9202
D31	**32**	**0.00**	0.7577	0.1986	188.5219	0.9270
Flame	**3**	**0.01**	0.5762	0.2649	25.3342	0.9459
Gaussians4	**5**	**0.01**	0.7549	0.2276	25.6871	0.9472
HappyFace	**3**	**0.00**	0.6377	0.3986	9.5429	0.9993
Jain	**3**	**0.00**	0.5861	0.3481	9.5113	0.9932
Lines4	**5**	**0.01**	0.3910	0.5198	9.3212	0.9628
Pathbased	**4**	**0.01**	0.6325	0.2702	7.9252	0.9513
Spiral	**4**	**1.00**	−0.1963	0.4985	4.9115	0.9341

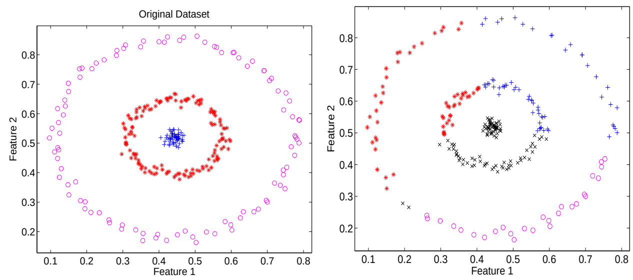


(a) Original Aggregation Dataset. (b) K-means Clustering.

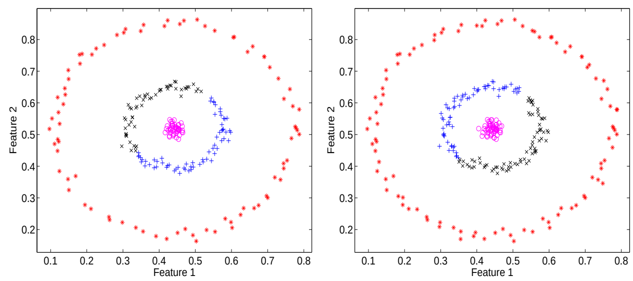


(c) Multiclass Spectral Clustering with regular affinity matrix. (d) Multiclass Spectral Clustering with constraint affinity matrix.

Fig. 2: Results for Aggregation Dataset.



(a) Original Bullseye 3 cores, Dataset. (b) K-means Clustering.

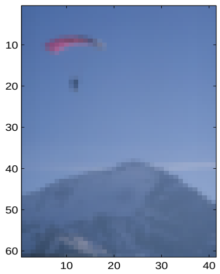


(c) Multiclass Spectral Clustering with regular affinity matrix. (d) Multiclass Spectral Clustering with constraint affinity matrix.

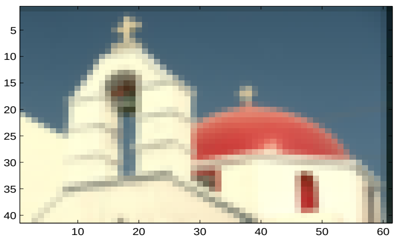
Fig. 3: Results for Bulleye3 Dataset.

B. Images

In general, we can appreciate that taking advantages of the original labels of data clustering performance can be enhanced. We set to be constant value Υ the affinity value between data points priori established to belong into the same cluster in accordance the labels. Nonetheless, from experimental results we can note that setting a constant value for connected data

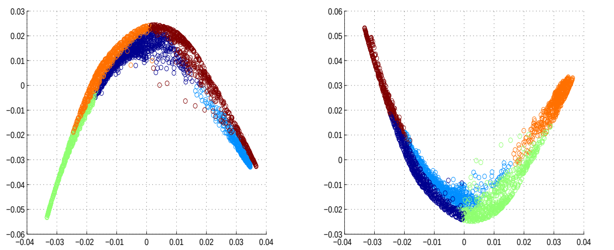


(a) Image 60079 resized to 1/8 of the original size.



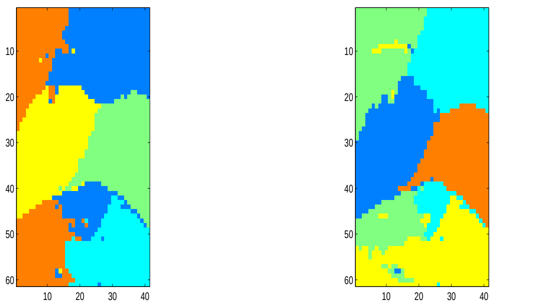
(b) Image 118035 resized to 1/8 of the original size.

Fig. 4: Resized Images.



(a) Unsupervised analysis eigenvectors. (b) Semisupervised analysis eigenvectors.

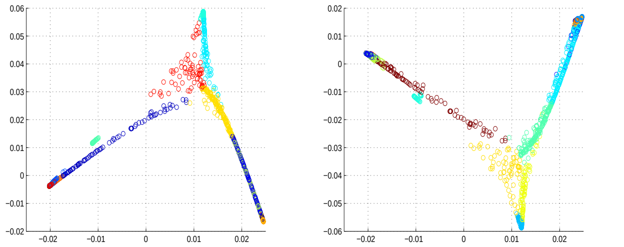
Fig. 5: Eigenvector Plot for unsupervised and semisupervised analysis of image 4(a).



(a) Regular affinity matrix. (b) Constraint affinity matrix.

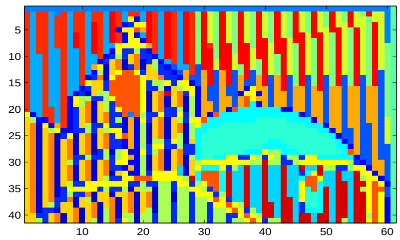
Fig. 6: Segmentation of image 4(a) with multiclass spectral clustering.

points is not a stable approach. It works well when some compactness level is guaranteed. This can be attributed to either selection of parameter Υ is not an arbitrary task and

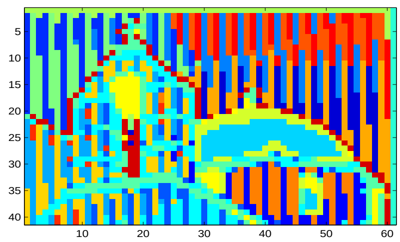


(a) Unsupervised analysis eigenvectors. (b) Semisupervised analysis eigenvectors.

Fig. 7: Eigenvector Plot for unsupervised and semisupervised analysis of image 4(b).



(a) Regular affinity matrix.



(b) Constraint affinity matrix.

Fig. 8: Segmentation of image 4(b) with multiclass spectral clustering.

semi-supervised extensions should not only take into account information given by original data but also that given by eigenvectors.

As mentioned in the toydata results section, a hierarchical procedure can be added to the semisupervised analysis to enhance the performance of the segmentation, this is visually achievable in the segmentation images, since adjacency of the clusters is regular. Moreover if considering the resolution loss

TABLE VI: Multiclass Spectral Clustering with Regular Affinity Matrix

Image	k	μ_S	σ_S	J	ε_m
12003	8	0.6070	0.3187	7.9393	0.7008
60079	5	0.6062	0.2925	7.0013	0.7184
37073	13	0.5358	0.3133	5.4993	0.5722
113044	7	0.6339	0.3187	7.7124	0.6478
145086	25	0.3849	0.3100	3.9569	0.2735
118035	22	0.4810	0.3499	6.1655	0.3967

TABLE VII: Multiclass Spectral Clustering with Constraint Affinity Matrix

Image	k	Υ	μ_S	σ_S
12003	**8**	2.5	0.6926	0.3319
60079	**5**	2.5	0.6706	0.2795
37073	**13**	10	0.5631	0.2786
113044	**7**	2.5	0.6690	0.3029
145086	**25**	100	0.5569	0.2692
118035	**22**	5	0.5273	0.3123
Image	k	Υ	J	
12003	**8**	100	10.1774	
60079	**5**	25	11.3640	
37073	**13**	10	6.2830	
113044	**7**	5	8.5933	
145086	**25**	100	5.8939	
118035	**22**	0	7.4803	
Image	k	Υ	ε_m	
12003	**8**	0	0.7058	
60079	**5**	0	0.7184	
37073	**13**	0.01	0.5840	
113044	**7**	0	0.6513	
145086	**25**	0.00001	0.5225	
118035	**22**	0	0.4843	

for the resizing transformation, performance can be penalized.

V. Conclusions and Future Work

Semi-supervised approaches take place under the premise that by adding priori true information about the original data, i.e., labels to set in advance the cluster membership of some data points, clustering performance can be significantly enhanced. In this work, we presented a basic semi-supervised extension for basic multi class spectral clustering by setting a constant value for connected data points. From experimental results, method showed to work well when some compactness and separability level is guaranteed. This fact can be attributed to parameter selection for constrained affinity matrix must be done taking into consideration the original features space representation but eigenspace information as well.

As a future work, new semi-supervised extensions for spectral clustering are to be designed where formulation and parameter tunning are carried out in such way eigenvectors are separable.

References

[1] D. H. Peluffo-Ordoñez, "Estudio comparativo de métodos de agrupamiento no supervisado de latidos de señales ECG," Master's thesis, Universidad Nacional de Colombia Sede Manizales, Manizales, Colombia, Nov 2009.

[2] C. C. Hoyos, A. E. C. Ospina, D. H. P. O. nez, and C. G. C. Domínguez, "On the Groups Number Estimation for Unsupervised Clustering," in *XVI Simposio de Tratamiento de Señales, Imagénes y Visión Artificial - STSIVA 2011*, September 2011.

[3] L. Wolf and A. Shashua, "Feature selection for unsupervised and supervised inference: The emergence of sparsity in a weight-based approach," *J. Mach. Learn. Res.*, vol. 6, pp. 1855–1887, December 2005.

[4] P. Perona and W. Freeman, "A factorization approach to grouping," in *Computer Vision, ECCV 1998*, ser. Lecture Notes in Computer Science, H. Burkhardt and B. Neumann, Eds. Springer Berlin, Heidelberg, 1998, vol. 1406, pp. 655–670, 10.1007/BFb0055696.

[5] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," *Advances in Neural Information Processing Systems 17*, vol. 2, no. 1601-1608, pp. 1601–1608, 2004.

[6] C. Alzate and J. Suykens, "Multiway spectral clustering with out-of-sample extensions through weighted kernel PCA," *IEEE transactions on pattern analysis and machine intelligence*, pp. 335–347, 2010.

[7] Y. S. X. and S. Jianbo, "Multiclass spectral clustering," in *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2003, p. 313.

[8] B. T. Anh and D. D. X. Thanh, "A perron-frobenius theorem for positive quasipolynomial matrices associated with homogeneous difference equations," *Journal of applied mathematics*, 2007.

[9] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, vol. 2, July 2001, pp. 416–423.

[10] P. Chan, M. Schlag, and J. Zien, "Spectral k-way ratio-cut partitioning and clustering," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 13, no. 9, pp. 1088 –1096, Sep. 1994.

[11] L. Fu and E. Medico, "FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data," *BMC Bioinformatics*, vol. 8, no. 1, pp. 3+, Jan. 2007.

[12] A. Jain, "Data clustering: Userś dilemma," in *Machine Learning and Data Mining in Pattern Recognition*, ser. Lecture Notes in Computer Science, P. Perner, Ed. Springer Berlin / Heidelberg, 2007, vol. 4571, pp. 1–1, 10.1007/978-3-540-73499-4_1.

[13] C. T. Zahn, "Graph-theoretical methods for detecting and describing gestalt clusters," *IEEE Trans. Comput.*, vol. 20, no. 1, pp. 68–86, Jan. 1971.

[14] A. Gionis, H. Mannila, and P. Tsaparas, "Clustering aggregation," *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, Mar. 2007.

[15] C. J. Veenman, M. Reinders, and E. Backer, "A maximum variance cluster algorithm," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 24, pp. 1273–1280, 2002.

Cristian Castro-Hoyos was born in Bogota D.C., Cundinamarca, in 1989. He received his degree in electronic engineering from the Universidad Nacional de Colombia, Manizales, Colombia, in 2011. Currently, he is M.Eng. student in the same university. His main research interests are computer applied maths, pattern recognition and signal processing and their applications in biosignal analysis and telecommunication systems.

Diego Hernán Peluffo-Ordoñez was born in San Juan de Pasto, Nariño, in 1986. He received his degree in electronic engineering and the M.Eng. degree in industrial automation from the Universidad Nacional de Colombia, Manizales, Colombia, in 2008 and 2010, respectively. Currently, he is PhD student in the same university. His main research interests are applied maths and unsupervised learning and their applications in biosignals analysis.

César Germán Castellanos-Domínguez received his undergraduate degree in radiotechnical systems and his Ph.D. in processing devices and systems from the Moscow Technical University of Communications and Informatics, in 1985 and 1990 respectively. Currently, he is a Full Professor in the Department of Electrical, Electronic and Computer Engineering at Universidad Nacional de Colombia Sede Manizales. He is also the Leader of the Signal Processing and Recognition Group at the same university. His teaching and research interests include information and signal theory, digital signal processing and bioengineering.