Quadratic Problem Formulation with Linear Constraints for Normalized Cut Clustering

D.H. Peluffo-Ordóñez[1], C. Castro-Hoyos[2],
Carlos D. Acosta-Medina[2], and Germán Castellanos-Domínguez[2]

[1] Univerisdad Cooperativa de Colombia – Pasto, Calle 18 No. 47 – 150, Pasto, Colombia
diego.peluffo@campusucc.edu.co
[2] Universidad Nacional de Colombia – Manizales, Km 7 vía al Magdalena, Manizales, Colombia
{cdacostam,cgcastellanosd}@unal.edu.co

Abstract. This work describes a novel quadratic formulation for solving the normalized cuts-based clustering problem as an alternative to spectral clustering approaches. Such formulation is done by establishing simple and suitable constraints, which are further relaxed in order to write a quadratic functional with linear constraints. As a meaningful result of this work, we accomplish a deterministic solution instead of using a heuristic search. Our method reaches comparable performance against conventional spectral methods, but spending significantly lower processing time.

1 Introduction

Graph-based clustering is a suitable alternative to classical clustering techniques, mainly in case of problems involving highly non-linearly separable clusters. Within a graph framework, data points to be clustered can be seen as nodes of a weighted and non-directed graph, where the weights are given by a pairwise relationship called similarity or affinity. Being non-supervised, graph-based clustering approaches have no need to training the model, but initial parameters such as the number of clusters and a similarity matrix are required. Algebraically, data can be divided into disjoint clusters by the well-known normalized cuts clustering (NCC) criterion, which consists of either minimizing the cuts between clusters (min-cuts) [1] or maximizing the within-cluster similarity [2]. NCC has successfully been utilized for several applications [3–5]. Mostly, NCC approaches have been aimed to solve the graph partitioning problem by imposing orthonormality constraints, and then, leading to an eigenvector-based solution. For instance, there are spectral approaches aimed at either optimizing a minimum cuts formulation [6], or formulating a quadratic problem to determine clustering indicators as done in the multi-cluster spectral clustering (MCSC) proposed in [2]. By taking into account that any orthonormal rotation of the span of eigenvectors is also a solution to NCC, MCSC heuristically searches for the best rotation. Other spectral approaches resort to using kernels to yield k-means generalizations [1] and variations [7]. Generally, these methods are known as kernel k-means and very often require an additional clustering procedure being accomplished by a conventional k-means. Furthermore, more elaborated approaches involving kernel models within support vector machines formulations have also been introduced [8, 9].

E. Bayro-Corrochano and E. Hancock (Eds.): CIARP 2014, LNCS 8827, pp. 408–415, 2014.

Nonetheless, despite their applicability and performance [10,11], spectral approaches are prohibitive for some applications due to the high computational cost that often involves the calculation of eigenvectors, e.g. when dealing with large data. Some studies have concerned about to get alternatives for solving the normalized cuts clustering without using eigenvectors: In [12], from the mathematic equivalence between weighted kernel k-means objective and weighted graph clustering objective, a multilevel algorithm is proposed. Other approaches are focused on quadratic problem formulations with linear constraints. For instance, a fast algorithm for NCC is introduced in [13], where linear constraints are imposed based on labeled samples. Previous approach suffers from having the need for supervised information. With the attempt to avoid using labeled samples and/or pre-clustering stages, [14] proposes an alternative solution to the problem presented in [2] by determining required rotation matrix via a deterministic solution of an equation matrix, reducing then the processing time meanwhile a comparable performance is reached. Also, other approaches rely on heuristic searches within the weighted graph [5].

In this work, we present a novel quadratic formulation for solving the NCC problem. Such formulation is done by establishing simple and suitable constraints, which are further relaxed in order to write a quadratic functional with linear constraints. Then, problem can be readily solved by standard quadratic programming procedures. The performance of our method is assessed regarding image segmentation accuracy and processing time. For experiments, images extracted from the free access Berkeley Segmentation Data Set [15] are considered. A comparative is done regarding three baseline methods kernel k-means (KKM) [7], min cuts (Min-cuts) [6] and multi-cluster spectral clustering (MCSC) [2]. All the methods are performed with a given number of clusters and a similarity matrix is set as the scaled exponential similarity matrix described in [7]. Our approach reaches comparable performance against conventional spectral methods, spending significantly lower processing time. It is noteworthy to mention that instead of a heuristic one, a deterministic solution is accomplished.

The outline of this paper is as follows: Section 2 introduces notation and describes the NCC matrix formulation. Section 3 explains the proposed approach starting from a simplified formulation to reach an improved formulation. Experimental results are shown in section 4. Finally, section 5 presents concluding remarks and future work.

2 Normalized Cuts Based-Clustering (NCC)

Let us define a weighted non-directed graph as $\mathbb{G} = (\mathbb{V}, \mathbb{E}, \boldsymbol{\Omega})$, where $\mathbb{V} = [N] = \{1, \cdots, N\}$ are the nodes being N the number of nodes, $\mathbb{E}$ is the set of edges, and matrix $\boldsymbol{\Omega} \in \mathbb{R}^{N \times N}$ is the similarity matrix. Since $\mathbb{G}$ is a non-directed graph, matrix $\boldsymbol{\Omega} = [\Omega_{ij}]$ should be selected as a positive semi-definite symmetric matrix. The aim of NCC is to split $\mathbb{V}$ into K disjoint sub-graphs or groups, i.e., $\mathbb{V} = \cup_{k=1}^{K} \mathbb{V}_k$ where $\mathbb{V}_l \cap \mathbb{V}_k = \emptyset, \forall l \neq k$. Let us denote the data set to be clustered as $\boldsymbol{X} \in \mathbb{R}^{N \times d} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^\top$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ is the i-th sample associated with node i. In matrix terms, NCC is aimed at determining a binary indicator matrix $\boldsymbol{M} \in \{0,1\}^{N \times K}$, such that $\boldsymbol{M} = (\boldsymbol{m}^{(1)}, \ldots, \boldsymbol{m}^{(K)})$, where $\boldsymbol{m}^{(k)} \in$ denotes the k-th column vector formed by data point membership regarding cluster k. Each entry ik of matrix $\boldsymbol{M}$, being in

turn the i-th entry of vector $\boldsymbol{m}^{(k)}$, is defined as $m_{ik} = m_i^{(k)} = 1$ if $i \in \mathbb{V}_k$ otherwise, with $k \in [K], [K] = \{1, \ldots, K\}$, and $i \in \mathbb{V}$. To guarantee that nodes belongs to a single sub-graph, the condition $\sum_{k=1}^{K} m_i^{(k)} = (m_i^{(1)}, \ldots, m_i^{(K)})^\top \mathbf{1}_K = 1,\ i \in [N]\ \therefore \boldsymbol{M}\mathbf{1}_K = \mathbf{1}_N$, must be fulfilled, where $\mathbf{1}_n$ stands for a n-dimensional all ones vector. Given the duality of weighted graph problems [2], sought clusters can be determined by maximizing within-cluster similarity (associations) and minimizing between-cluster similarity (cuts). In [2], it is demonstrated that once normalized associations and cuts formulations are equivalent. Hence, we opt by the maximization setting that has a cost function in the form:

$$\sum_{\substack{i,\, j \,\in\, \mathbb{V}_k, \\ k \,\in\, [K]}} \Omega_{ij} = \ \mathrm{tr}(\boldsymbol{M}^\top \boldsymbol{\Omega} \boldsymbol{M}) = \sum_{k=1}^{K} \boldsymbol{m}^{(k)\top} \boldsymbol{\Omega} \boldsymbol{m}^{(k)}. \tag{1}$$

To make feasible solutions lies into a unit hyper-sphere, previous cost function can be length normalized by using the degree matrix $\boldsymbol{D} \in \mathbb{R}^{N \times N}$, which is a diagonal matrix such that its entries are given by $d_{ii} = \sum_{j \in [N]} \Omega_{ij}$. As well, $\boldsymbol{D} = \ \mathrm{Diag}(\boldsymbol{\Omega}\mathbf{1}_N)$, being $\mathrm{Diag}(\cdot)$ a diagonal matrix formed by its argument vector. In this sense, the K-way normalized cuts-based clustering can be expressed as:

$$\max_{\boldsymbol{m}^{(k)}} f(\boldsymbol{m}^{(k)}) = \frac{1}{K} \frac{\sum\limits_{k=1}^{K} \boldsymbol{m}^{(k)\top} \boldsymbol{\Omega} \boldsymbol{m}^{(k)}}{\sum\limits_{k=1}^{K} \boldsymbol{m}^{(k)\top} \boldsymbol{D} \boldsymbol{m}^{(k)}},\ \text{s. t.} \quad \boldsymbol{M} \in \{0,1\}^{N \times K},\ \boldsymbol{M}\mathbf{1}_K = \mathbf{1}_N. \tag{2}$$

3 Proposed Approach to Solve NCC Problem

The solution proposed here aims to pose quadratic problem with sufficient linear constraints. First, we rewrite the cost function f as a quadratic form by eliminating the denominator. To do so, notice that term $\sum_{k=1}^{K} \boldsymbol{m}^{(k)\top} \boldsymbol{D} \boldsymbol{m}^{(k)}$ equals to $||\boldsymbol{\Omega}||_{L_1}$ when ensuring $\boldsymbol{M}$ be binary by satisfying $\boldsymbol{M}\mathbf{1}_K = \mathbf{1}_N$. Furthermore, condition $\boldsymbol{m}^{(k)\top}\mathbf{1}_N \geq 1$ is necessary to ensure each cluster has assigned at least one node. As well, previous constraint is upper bounded by $N - K + 1$ to avoid that all the nodes will be assigned to a unique cluster. For further algorithmic procedures, we summarize these constraints as $0 < \boldsymbol{m}^{(k)\top}\mathbf{1}_N < N - K$. In addition, the squared norm $||\boldsymbol{m}^{(k)}||_{L_2}^2 = \boldsymbol{m}^{(k)\top}\boldsymbol{m}^{(k)}$ should match the number of nodes belonging to cluster k (N_k). Thereupon, by recalling optimization problem shown in in equation (2), we can pose a new problem given by

$$\begin{aligned} &\max_{\boldsymbol{m}^{(k)}} \sum_{k=1}^{K} \boldsymbol{m}^{(k)\top} \boldsymbol{\Omega} \boldsymbol{m}^{(k)} \\ &\text{s. t.} \quad \boldsymbol{m}^{(k)} \succ 0, 0 < \boldsymbol{m}^{(k)\top}\mathbf{1}_N < N - K,\ \forall k \in \{1, \ldots, K\}, \\ &\qquad \sum_{k=1}^{K} \boldsymbol{m}^{(k)\top} \boldsymbol{D} \boldsymbol{m}^{(k)} = ||\boldsymbol{\Omega}||_{L_1}, \sum_{k=1}^{K} \boldsymbol{m}^{(k)\top}\boldsymbol{m}^{(k)} = N_1 + \cdots + N_k = N. \end{aligned} \tag{3}$$

3.1 Simplified Approach

By relaxing the quadratic constraints of the problem given by equation (3), cost function can be written as follows:

$$f(\boldsymbol{m}^{(k)}|\lambda,\mu)=\sum_{k=1}^{K}\left(\boldsymbol{m}^{(k)\top}\boldsymbol{\Omega}\boldsymbol{m}^{(k)}+\lambda\boldsymbol{m}^{(k)\top}\boldsymbol{D}\boldsymbol{m}^{(k)}+\mu\boldsymbol{m}^{(k)\top}\boldsymbol{m}^{(k)}\right)-\lambda||\boldsymbol{\Omega}||_{L_1}-\mu N,$$

where λ and μ are regularization parameters.

Letting matrix $\boldsymbol{P}_{\lambda,\mu}\in\mathbb{R}^{N\times N}$ be $\boldsymbol{P}_{\lambda,\mu}=\boldsymbol{\Omega}+\lambda\boldsymbol{D}+\mu\boldsymbol{I}_N$, and eliminating constant terms, the following simplified quadratic functional is accomplished $f(\boldsymbol{m}^{(k)})=\sum_{k=1}^{K}\boldsymbol{m}^{(k)\top}\boldsymbol{P}_{\lambda,\mu}\boldsymbol{m}^{(k)}$. Now, in order to pose a pure quadratic problem, we recall membership matrix $\boldsymbol{M}$ and write the vectorization of problem as:

$$\begin{aligned}&\max_{\boldsymbol{M}}\ \ f(\boldsymbol{M})=\operatorname{vec}(\boldsymbol{M})^{\top}(\boldsymbol{P}\otimes\boldsymbol{I}_K)\operatorname{vec}(\boldsymbol{M})\\&\text{s. t.}\quad \operatorname{vec}(\boldsymbol{M})\succ 0,\quad 0\prec\left(\mathbf{1}_N^{\top}\otimes\boldsymbol{I}_K\right)\ \operatorname{vec}(\boldsymbol{M})\prec N-K,\end{aligned}\tag{4}$$

where $\operatorname{vec}(\boldsymbol{M})$ is a NK-dimensional vector representing the vectorization of $\boldsymbol{M}$ such that $\operatorname{vec}(\boldsymbol{M})=(\boldsymbol{m}^{(k)\top},\ldots,\boldsymbol{m}^{(k)\top})^{\top}$, and $\otimes$ stands for Kronecker product. Since constraints are linear and functional is quadratic, the dual version of previous problem can be solved by a conventional quadratic programming algorithm.

3.2 A Better Approach: Ensuring Orthogonality

Yet, in spite of involving the condition $\sum_{k=1}^{K}\boldsymbol{m}^{(k)\top}\boldsymbol{m}^{(k)}=\operatorname{tr}(\boldsymbol{M}^{\top}\boldsymbol{M})=N$, optimization problem yields infinite solutions being local optima, and then an extra binaryzing procedure is required as done in [14]. To tackle this matter, we suggest to analyze the out-diagonal elements of the quadratic form $\boldsymbol{M}^{\top}\boldsymbol{M}$. Since $\boldsymbol{M}$ should be binary, $\boldsymbol{M}^{\top}\boldsymbol{M}=\operatorname{Diag}(N_1,\ldots,N_k)$, which means column vectors $\boldsymbol{m}^{(k)}$ are orthogonal. Since condition $\operatorname{vec}(\boldsymbol{M})>0$ was beforehand ensured, such orthogonality can be guaranteed by imposing the constraint: $\sum_{k\neq s}\boldsymbol{m}^{(k)\top}\boldsymbol{m}^{(s)}=0$. To incorporate such constraint into the original problem statement (see equation (3)), we introduce a Kronecker's delta $\delta_{sk}=1$ if $s=k$ and 0 otherwise, such that modified functional becomes

$$\begin{aligned}f(\boldsymbol{m}^{(k)}|\lambda,\mu,\gamma)=&\sum_{k=1}^{K}\sum_{s=1}^{K}\boldsymbol{m}^{(k)\mathsf{T}}\delta_{sk}\left(\boldsymbol{\Omega}-\lambda\boldsymbol{D}-\mu\boldsymbol{I}_N\right)\boldsymbol{m}^{(s)}\\&-\gamma\sum_{k=1}^{K}\sum_{s=1}^{K}\left(1-\delta_{sk}\right)\boldsymbol{m}^{(k)\mathsf{T}}\boldsymbol{m}^{(s)}+\lambda\left\|\boldsymbol{\Omega}\right\|_{L_1}+\mu N,\end{aligned}$$

where γ is a parameter to regularize the introduced orthogonality condition. Again, eliminating constant terms and arranging, we have

$$\begin{aligned}&\sum_{k=1}^{K}\sum_{s=1}^{K}\boldsymbol{m}^{(k)\mathsf{T}}\left(\delta_{sk}\left(\boldsymbol{\Omega}-\lambda\boldsymbol{D}-(\mu-\gamma)\boldsymbol{I}_N\right)-\gamma\boldsymbol{I}_N\right)\boldsymbol{m}^{(s)}\\&=\sum_{k=1}^{K}\boldsymbol{m}^{(k)\top}\boldsymbol{P}_{\lambda,\mu}\boldsymbol{m}^{(k)}-\gamma\sum_{k=1}^{K}\sum_{s=1}^{K}\boldsymbol{m}^{(k)\mathsf{T}}\boldsymbol{m}^{(s)}.\end{aligned}\tag{5}$$

Then, we can rewrite the problem in terms of matrix $\boldsymbol{M}$ as follows

$$f(\boldsymbol{M}|\gamma) = \text{ tr}\left(\boldsymbol{M}^\top \boldsymbol{P}_{\lambda,\mu}\boldsymbol{M}\right) - \frac{\gamma}{(K-1)!}\sum_{\ell=1}^{L} \text{tr}(\boldsymbol{Z}_\pi^{(\ell)}\boldsymbol{M}^\top\boldsymbol{M}),$$

where π is a permutation of K elements given by $\pi : \{1,\ldots,K\} \to \{1,\ldots,K\}$, $\boldsymbol{Z}_\pi^{(\ell)}$ is the $\ell-$th $K \times K$ dimensional permutation matrix and $L = K!$ is the total number of permutations. Finally, by defining the auxiliary matrix $\boldsymbol{H} \in \mathbb{R}^{NK\times NK}$ as

$$\boldsymbol{H} = (\boldsymbol{I}_K \otimes \boldsymbol{P}_{\lambda,\mu}) - \frac{\gamma}{(K-1)!}\sum_{\ell=1}^{L}\left(\boldsymbol{Z}_\pi^{(\ell)} \otimes \boldsymbol{I}_N\right),$$

it is possible to vectorize $f(\boldsymbol{M}|\gamma)$ as

$$\begin{aligned} &\max_{\boldsymbol{M}} \;\; f(\boldsymbol{M}) = \text{ vec}\,(\boldsymbol{M})^\intercal\, \boldsymbol{H}\, \text{vec}\,(\boldsymbol{M}) \\ &\text{s. t.} \quad \text{vec}(\boldsymbol{M}) \succ 0, \quad 0 \prec \left(\mathbf{1}_N^\top \otimes \boldsymbol{I}_K\right)\, \text{vec}\,(\boldsymbol{M}) \prec N - K. \end{aligned} \tag{6}$$

Again, functional $f(\boldsymbol{M})$ becomes quadratic regarding the vectorization of $\boldsymbol{M}$. Previous problem can be rewritten as minimizing its dual function given by

$$g(\boldsymbol{M}) = \text{ vec}\,(\boldsymbol{M})^\top\,(\boldsymbol{I}_{NK} - \boldsymbol{H})\, \text{vec}\,(\boldsymbol{M}) = \text{ vec}\,(\boldsymbol{M})^\top \text{vec}\,(\boldsymbol{M}) - f(\boldsymbol{M}). \tag{7}$$

Notice that function g naturally ensures maximal variance of matrix $\boldsymbol{M}$. Under the same linear constraints as (6) and because matrix $\boldsymbol{I}_{NK} - \boldsymbol{H}$ is positive semi-definite matrix, then the problem of minimizing g is convex. Therefore, it can be solved by means of standard quadratic programming algorithms.

4 Results and Discussion

To assess the performance of our method, we employ five images extracted from the free access Berkeley Segmentation Data Set [15]. Images are characterized by RGB color space and the xy position of each pixel. Due to memory usage restrictions, images are resized at 20% of the original size. To solve the quadratic problem of proposed method (See Equation (7), primal active-set algorithm QP is used, which is initialized by setting the vectorization of matrix satisfying the Hölder inequality described in [16]. To do so, graph Laplacian effect is involved by setting $\lambda = 1$ as well as to simplify the problem according to Equation (5) we eliminate a free parameter doing $\mu = \gamma$. Also, experimentally the regularization parameter γ is established as $0.4 < \gamma < 0.6$ (about 50%). A comparatives is done regarding three baseline methods kernel k-means (KKM) [7], min cuts (Min-cuts) [6] and multi-cluster spectral clustering (MCSC) [2]. All the methods are performed with a given number of clusters K set as shown in shown in Fig. 1. The similarity matrix is set as the scaled exponential similarity matrix described in [7]:

$\Omega_{ij} = \exp(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2_{L_2}/(\sigma_i\sigma_j))$ where $\sigma_i = ||\boldsymbol{x}_i - \boldsymbol{x}_i(N_{ngh})||^2_{L_2}$ and $\boldsymbol{x}_i(N_{ngh})$ is the N_{ngh}-th nearest neighbor to $\boldsymbol{x}_i$. To fairly compare the methods, we standardize the results by setting the same initial parameters (number of clusters K and similarity matrix) for all cases. Experiments were done using MatLab Version 7.12.0.635 (R2011a) on a computer with RAM 8Gb, and processor Intel(R) Xeon(R) CPU X5660 2.8GHz. Image segmentation results are shown in Fig. 1. The segmentation performance is quantified by a supervised index noted as Probabilistic Rand Index (PR) explained in [17], such that $PR \in [0, 1]$, being 1 when regions are properly segmented. Proposed approach achieves in most cases a comparable result regarding MCSC but spending lower processing time. Table 1 depicts the processing time employed by each method. Processing times are given as a proportion of the highest ones. Clustering algorithms are iterated 10 times. The highest average processing times for all considered images are $T_p^{(1)} = 112.87s$, $T_p^{(2)} = 65.93s$, $T_p^{(3)} = 66.43s$, $T_p^{(4)} = 66.56s$ and $T_p^{(5)} = 66.75s$.

Table 1. Average clustering processing time ratio along 10 iterations. Mean and standard values are shown.

Method	**Image**				
	113044	118035	12003	181091	24004
Proposed method	$0.5414T_p^{(1)}$ ± 0.0072	$0.5605T_p^{(2)}$ ± 0.0064	$0.5612T_p^{(3)}$ ± 0.0089	$0.5431T_p^{(4)}$ ± 0.0094	$0.5945T_p^{(5)}$ ± 0.0079
KKM	$0.8538T_p^{(1)}\pm$ ± 0.0092	$0.9339T_p^{(2)}\pm$ ± 0.0085	$0.9244T_p^{(3)}\pm$ ± 0.0091	$0.9327T_p^{(4)}\pm$ ± 0.0125	$0.9342T_p^{(5)}\pm$ ± 0.0179
min-cuts	$0.7663T_p^{(1)}\pm$ ± 0.0091	$0.8951T_p^{(2)}\pm$ ± 0.0088	$0.7787T_p^{(3)}\pm$ ± 0.0083	$0.7412T_p^{(4)}\pm$ ± 0.0095	$0.8253T_p^{(5)}\pm$ ± 0.0078
MCSC	$T_p^{(1)}\pm$ ± 0.0081	$T_p^{(2)}\pm$ ± 0.0072	$T_p^{(3)}\pm$ ± 0.0083	$T_p^{(4)}\pm$ ± 0.0089	$T_p^{(5)}\pm$ ± 0.0075

Note that our method spends the least processing time, reducing at around 50% the reference processing time TP in all cases. Index PR compares the resultant segmentation with multiple manually labeled ground-truth images through a variability function regarding each pair of pixels in the ground-truth set. Then, we can say that, in terms of PR values, our approach achieves a suitable trade-off between performance and computational cost in contrast with the other considered methods.

Additional remarks: By ensuring a proper initialization as that one used here, considered quadratic programming algorithm since each iteration can be performed in $O(N^2)$. Meanwhile, MCSC in $O(N^3)$. By other hand, KKM approaches involves a maximum complexity of $O(N^3 + N^{dK+1} \log N)$ due to the calculation of eigenvectors and iterate KM procedure.

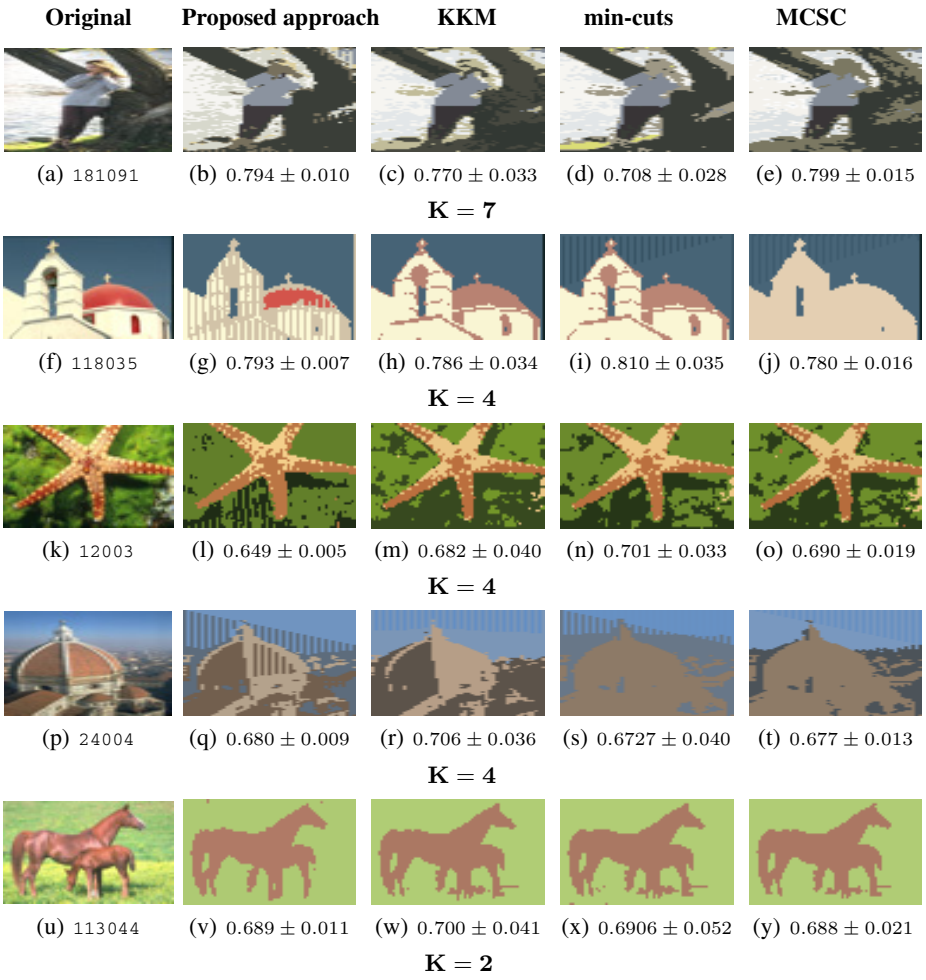

Fig. 1. Performance of analyzed clustering methods on image segmentation. PR index is used as a performance measure. The values of mean and standard deviation are calculated after 10 iterations.

5 Conclusions and Future Work

Despite being a suitable alternative, spectral solutions can be prohibitive for problems involving large data In this work, we introduced a novel alternative to solve the NCC problem. From the conventional formulation, we lead a quadratic functional with suitable linear constraints. Proposed work reaches comparable results but reducing significantly the processing time. The noteworthy benefit of our method is that a deterministic solution is reached.

As a future work, different quadratic models should be explored. As well, sparsity and positivity properties are to be studied for ensuring global optima convergence and reducing complexity.

References

1. Dhillon, I.S., Guan, Y., Kulis, B.: Kernel k-means, spectral clustering and normalized cuts (2004)
2. Stella, Y., Jianbo, S.: Multiclass spectral clustering. In: ICCV 2003: Proceedings of the Ninth IEEE International Conference on Computer Vision, p. 313. IEEE Computer Society, Washington, DC (2003)
3. Tung, F., Wong, A., Clausi, D.A.: Enabling scalable spectral clustering for image segmentation. Pattern Recognition 43(12), 4069–4076 (2010)
4. Joulin, A., Bach, F., Ponce, J.: Discriminative clustering for image co-segmentation. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1943–1950. IEEE (2010)
5. Peluffo-Ordonez, D.H., Castro-Ospina, A.E., Chavez-Chamorro, D., Acosta-Medina, C.D., Castellanos-Dominguez, G.: Normalized cuts clustering with prior knowledge and a pre-clustering stage
6. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 888–905 (2000)
7. Zelnik-manor, L., Perona, P.: Self-tuning spectral clustering. In: Advances in Neural Information Processing Systems, vol. 17, pp. 1601–1608. MIT Press (2004)
8. Alzate, C., Suykens, J.A.K.: Multiway spectral clustering with out-of-sample extensions through weighted kernel pca. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(2), 335–347 (2010)
9. Peluffo-Ordonez, D.H., Alzate, C., Suykens, J.A.K., Castellanos-Dominguez, G.: Optimal data projection for kernel spectral clustering
10. Molina-Giraldo, S., Álvarez-Meza, A., Peluffo-Ordoñez, D., Castellanos-Domínguez, G.: Image segmentation based on multi-kernel learning and feature relevance analysis. In: Advances in Artificial Intelligence–IBERAMIA 2012, pp. 501–510 (2012)
11. Peluffo-Ordóñez, D.H., García-Vega, S., Álvarez-Meza, A.M., Castellanos-Domínguez, C.G.: Kernel spectral clustering for dynamic data. In: Ruiz-Shulcloper, J., Sanniti di Baja, G. (eds.) CIARP 2013, Part I. LNCS, vol. 8258, pp. 238–245. Springer, Heidelberg (2013)
12. Dhillon, I.S., Guan, Y., Kulis, B.: Weighted graph cuts without eigenvectors a multilevel approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(11), 1944–1957 (2007)
13. Xu, L., Li, W., Schuurmans, D.: Fast normalized cut with linear constraints. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 2866–2873. IEEE (2009)
14. Peluffo, D., Acosta, C.D., Castellanos, G.: An improved multi-class spectral clustering based on normalized cuts. In: XVII Iberoamerican Congress on Pattern Recognition CIARP (2012)
15. Martin, D., Fowlkes, C., Tal, D., Malik, J.: 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, pp. 416–423 (July 2001)
16. Pate, T.H.: Extending the hölder type inequality of blakley and roy to non-symmetric non-square matrices. Trans. Amer. Math. Soc 364, 4267–4281 (2012)
17. Unnikrishnan, R., Pantofaru, C., Hebert, M.: Toward objective evaluation of image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6), 929–944 (2007)