

A New Anticorrelation-based Spectral Clustering Formulation

Julia Dietlmeier¹, Ovidiu Ghita¹, and Paul F. Whelan¹

Centre for Image Processing and Analysis, Dublin City University, Dublin, Ireland
julia.dietlmeier2@mail.dcu.ie, ghitao@eeng.dcu.ie, paul.whelan@dcu.ie

Abstract. This paper introduces the Spectral Clustering Equivalence (SCE) algorithm which is intended to be an alternative to spectral clustering (SC) with the objective to improve both speed and quality of segmentation. Instead of solving for the spectral decomposition of a similarity matrix as in SC, SCE converts the similarity matrix to a column-centered dissimilarity matrix and searches for a pair of the most anticorrelated columns. The orthogonal complement to these columns is then used to create an output feature vector (analogous to eigenvectors obtained via SC), which is used to partition the data into discrete clusters. We demonstrate the performance of SCE on a number of artificial and real datasets by comparing its classification and image segmentation results with those returned by kernel-PCA and Normalized Cuts algorithm. The column-wise processing allows the applicability of SCE to Very Large Scale problems and asymmetric datasets.

Keywords: Spectral clustering; Image segmentation; Dimensionality reduction; Latent variables

1 Introduction

Recent years have witnessed an enormous increase in research and applications devoted to spectral clustering (SC) where the problem of grouping is reformulated in an induced feature space. This attention comes not undeserved for several reasons. Firstly, SC can be referred to as a fully *unsupervised* classification method [13]. Secondly, SC excels in discovering hidden and secondary relationships [20], managing non-convex cluster shapes, non-metric data [14] and noise reduction [11] in a well defined and theoretically sound framework. Finally, the segmentation and grouping based on eigenvectors is able to return the perceptual organization features present in an image [16, 5, 6]. Conceptually, SC belongs to the domain of manifold learning methods aimed at the unsupervised extraction of a low-dimensional representation [19]. The term *spectral* therein refers to a broad family of clustering methods that make use of the eigenvectors of some normalized similarity matrix [15]. Different SC algorithms formalize the grouping problem in different ways and differ widely in the retained number and ranking of eigenvectors and matrix normalization steps [8, 7]. By far, the most popular

application to image segmentation is the *normalized cuts* (Ncut) algorithm presented by Shi and Malik in 1997 as the first application of SC to computer vision and image analysis domain [4].

Despite its merits, SC has also its limitations associated with the computational complexity of spectral decomposition [19, 6] and the problem of discretization of continuous eigenvectors [16]. Particularly, pixel classification tasks for large indefinite (possibly asymmetric) and fully dense similarity matrices form a considerable computational bottleneck for SC. Given an image with N pixels, the size of the similarity matrix increases to $N^2 \times N^2$ and the decomposition-based implementations of SC become quickly infeasible. This is a well-known fact which has been continually emphasized over the past decade [19, 6]. Approaches dealing with this paradigm range from exploiting the sparsity, subsampling of an image or similarity matrix to the low-rank approximation methods such as Nyström algorithm [21, 23]. We infer that this is still an open problem from the most recent work by Chen et al. in [25] where the team of researchers presents a parallel HPC implementation of SC. The other recent work by Tung et al. in [24] approaches the scalability problem of SC by using a combination of blockwise processing and stochastic ensemble consensus.

In this contribution we do not seek numeric or platform-related solutions to speed-up SC but rather a method which questions the optimality of eigenvectors. We therefore raise a question if the important aspects of the data can be represented through a less expensive alternative. This consideration opens the door to a much broader range of techniques in statistical machine learning and dimensionality reduction which forms the basis for any SC implementation. Therein, the family of learning methods encompasses but is not limited to principal component analysis (PCA), kernel-PCA, Linear Discriminant Analysis (LDA), other generative, discriminative, latent variable [18] and Independent Component Analysis (ICA) methods [11].

In this paper, we assume a low-dimensional manifold and search for a computationally less expensive alternative to eigenvector-based analysis. We next highlight the idea behind our algorithm that given a separable dataset [18], the core information related to the leading eigenvectors is contained in the columns of a kernel matrix. We call our algorithm Spectral Clustering Equivalence (SCE) and in the next sections we show its connection with kernel-PCA (Sections 2,3), *Ncut* (Section 4) and the Ising model [3] (Section 5).

The major contribution of our paper is the reformulation of the standard spectral clustering through the construction of uncorrelated, orthogonal and centered components without the use of eigenvectors.

2 Development of Our Method

In multidimensional scaling (MDS), the spectral decomposition is carried out on the inner product matrix (Gram) \mathbf{G} in the feature space with the main emphasis to preserve the inner-point distances [9]. Let us denote a similarity matrix by \mathbf{S} and a dissimilarity matrix by \mathbf{A} (see Table 1). Either \mathbf{A} or \mathbf{S} can be viewed as a

dot product matrix \mathbf{G} in some feature space according to Schölkopf and Smola [27] or transformed to \mathbf{G} with kernelization and normalization [9] according to a particular SC formulation and application [6].

Table 1. SCE notations.

\mathbf{S}	similarity matrix
\mathbf{A}	kernel matrix (dissimilarity matrix)
\mathbf{C}	column-centered kernel matrix
\mathbf{R}	correlation matrix computed on the columns of \mathbf{C}
ρ_{min}	minimal signed Pearson correlation coefficient, $\rho \equiv \cos \angle c_1, c_2$
$\mathbf{c}_{1,2}$	a pair of most <i>anticorrelated</i> columns of \mathbf{C} , $\angle c_1, c_2 \equiv \theta_1 \rightarrow \pi$
$\mathbf{t}_{1,2}$	orthogonal complement to $\mathbf{c}_{1,2}$, $\theta_2 \rightarrow \pi/2$
\Rightarrow	next following operation
\mathbf{M}	mixing matrix $\mathbf{M} = \mathbf{C}^z$
z	equivalence coefficient
Φ	centered and uncorrelated SCE feature vectors
X	eigenvector-based feature vectors
$\tilde{\mathbf{G}}$	embedded (psd) Gram matrix [9]
λ_{min}	minimal eigenvalue (additive shift constant) [9]

We assume that \mathbf{A} is a generally indefinite (possibly asymmetric) matrix and interpret it as a multidimensional space spanned by its columns. Further, we center the columns of \mathbf{A} , call the new matrix \mathbf{C} and consider a 2-class data partitioning problem. In order to answer the question about which columns in \mathbf{C} carry more information about the binary class labels, we proceed with the analysis of linear dependencies present in \mathbf{C} . From the related works on linear dependency analysis, Srebro and Jaakkola in [17], for example, also seek to identify a low-dimensional subspace that captures the dependent and the "important" aspects of the data, and separate them from independent variations.

Thus, a natural way to conduct dependency analysis is to analyze correlations between different variables and the first step, prior to applying correlation analysis, is the centering of variables. Contrary to the formation of \mathbf{G} in kernel-PCA which involves double-centering [9], our normalization of \mathbf{A} in order to obtain \mathbf{C} does not involve row centering. Next, we define the correlation between two centered columns, \mathbf{c}_1 and \mathbf{c}_2 , according to the formula of Pearson product-moment correlation coefficient [1]:

$$\rho_{1,2} = \sum_{i=1}^N [\mathbf{c}_{1,i} \mathbf{c}_{2,i}] / \sqrt{\sum_{i=1}^N \mathbf{c}_{1,i}^2 \sum_{i=1}^N \mathbf{c}_{2,i}^2} . \quad (1)$$

A strong negative correlation provides a suitable measure of discrimination according, for example, to [2] and also indicates that the decrease in one variable is controlled by the increase in the second variable. In regards to natural images, it is reasonable to view foreground and background as the two most distinct and thus most anticorrelated image structures. We therefore are interested in the lower bound of $\rho \in [-1, 1]$ and define a pair of observations (columns of \mathbf{C}) with a strong negative correlation $\rho_{ij} \rightarrow -1$ dissociation patterns.

In order to construct the orthogonal and uncorrelated kernel-PCA estimates we first draw on the idea of canonical spaces [28]. Let \mathcal{X} and \mathcal{Y} be the two unknown subspaces spanned by the columns of \mathbf{C} . The largest canonical angle between \mathcal{X} and \mathcal{Y} is defined as $\theta(\mathcal{X}, \mathcal{Y}) = \max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} \angle(x, y)$ [28]. We know that the correlation between centered variables is equivalent to the cosine of the angle between these variables [1]. The cosine of the largest (canonical) angle $\theta \rightarrow \pi$ can therefore be interpreted as the minimal signed Pearson correlation coefficient between the columns of \mathbf{C} , $\rho_{min} \rightarrow -1$ [1]. According to the cosine

input		output
image \Rightarrow features $\Rightarrow \mathbf{S} \Rightarrow \mathbf{A} \Rightarrow$	Dimensionality $\Rightarrow X_{1,2}$ (SC) Reduction $\Phi_{1,2}$ (SCE)	set of discrete clusters

$$\mathbf{A} \Rightarrow \mathbf{G} = \mathbf{U}\mathbf{A}\mathbf{U}^T \Rightarrow \lambda_{min} \Rightarrow \tilde{\mathbf{G}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{U}}^T \Rightarrow X = \tilde{\mathbf{U}}\sqrt{\tilde{\mathbf{\Lambda}}} \Rightarrow X_{1,2} . \quad (2)$$

$$\mathbf{A} \Rightarrow \mathbf{C} \Rightarrow \mathbf{R} \Rightarrow \rho_{min} \Rightarrow \mathbf{c}_{1,2} \Rightarrow \perp \Rightarrow \mathbf{t}_{1,2} \Rightarrow \mathbf{M}\mathbf{t}_{1,2} \Rightarrow \text{decorrelate} \Rightarrow \Phi_{1,2} . \quad (3)$$

with 2×2 PCA
and center

Fig. 1. Algorithm description and the comparison of SC (2) and SCE (3). Both methods take a similarity matrix \mathbf{S} as an input and produce a pair of orthogonal and uncorrelated feature vectors. SC is based on the eigen-decomposition of the pseudo-Gram matrix \mathbf{G} . Conversely, SCE is a decomposition-free method which is based on the dependency analysis of the column-centered kernel matrix \mathbf{C} . Because SCE works on the columns of \mathbf{C} , it can generally be applied on asymmetric, non-PSD and rectangular datasets. The convention to write a pipeline of equations with " \Rightarrow " has been adopted from [9].

definition, π is the maximum possible angle corresponding to $\cos(\theta) = \rho_{min} = -1$ [1]. Thus, ρ_{min} not only defines a pair of mostly anticorrelated columns \mathbf{c}_1 and \mathbf{c}_2 but also provides the link with the first canonical angle θ_1 . According to Stewart in [28], the number of canonical angles in the case of $\dim(\mathcal{X}) < \dim(\mathcal{Y})$ is equal to $\dim(\mathcal{X})$, which in our case $\dim(\mathcal{X}) = 2$. This fact allows the construction of the second SCE-based feature component by employing the orthogonality

constraint to obtain θ_2 . For this purpose for $\forall k \in [1, N^2]$ we seek an orthogonal complement vector \mathbf{c}_k to $\mathbf{c}_{1,2}$, and in this process we discard the least orthogonal pair of vectors:

$$\mathbf{t}_{1,2} = \begin{cases} \mathbf{c}_{1,k}, & \text{if } |\angle \mathbf{c}_1, \mathbf{c}_k - \pi/2| < |\angle \mathbf{c}_2, \mathbf{c}_k - \pi/2| \\ \mathbf{c}_{2,k}, & \text{otherwise} . \end{cases} \quad (4)$$

Further, we multiply $\mathbf{t}_{1,2}$ by \mathbf{C}^z ("C to the power of z") to maximize linear dependency and subsequently decorrelate (and center) with PCA on the computed 2×2 covariance matrix (refer to Table 1 for notations) in order to obtain the pair $\Phi_{1,2}$ of orthogonal, uncorrelated and centered SCE feature vectors. In this formulation, the SCE approximation of kernel-PCA feature vectors $\mathbf{X}_{1,2}$ is controlled by the coefficient z . After decorrelation and centering and similarly to the ranking of the PCA components [11], our selection of the leading feature vector Φ_1 is based on the maximum variance principle such as $\sigma^2(\Phi_1) > \sigma^2(\Phi_2)$. The product $\mathbf{C}^z \mathbf{t}$ can also be viewed as a multivariate polynomial regression model with \mathbf{t} being a vector of $N^2 \times 2$ regression coefficients. The algorithm description and the progression from \mathbf{c} to Φ is outlined in Fig. 1.

3 Results of Comparison to kernel-PCA

In this section we compare SCE-based classification to the kernel-PCA-based result and consider an experiment with an asymmetric dataset. We created an interlocked spirals dataset, shown in Fig. 2, which is considered to be a challenging benchmark for spectral clustering [22]. To alleviate this challenging clustering

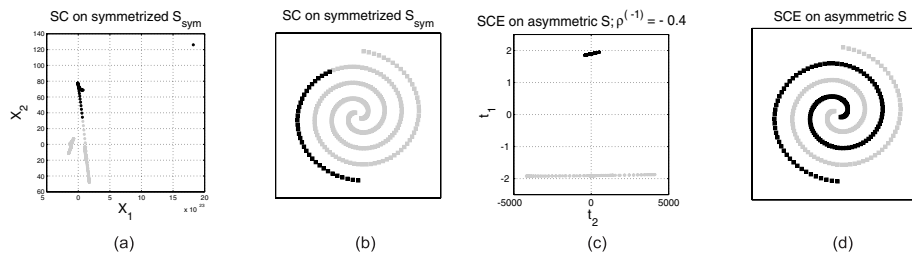


Fig. 2. Segmentation of asymmetric data with binary k-means in SC (a) and SCE (c) constructed feature spaces. An interlocked two spirals dataset is considered as a challenging benchmark for SC [22]. As can be seen in (c), SCE results in a better inter-cluster separation in the feature space. As the symmetry condition constitutes one of the four metric axioms [12], this example also tests the non-metric invariance of SCE.

problem, Chang and Yeung present in [22] a robust path-based spectral clustering algorithm with the use of a Gaussian kernel. The main objective of our experiment is to demonstrate the asymmetric and non-metric invariance of SCE

and we designed the following asymmetric similarity measure:

$$\begin{aligned} s_{ij} &= b_1 \cdot (dx_i - dx_j), \\ s_{ji} &= b_2 \cdot (dy_i - dy_j) . \end{aligned} \quad (5)$$

In the experiment outlined in Fig. 2 each spiral consists of 151 data points and is generated according to the equation of Archimedean spiral. Two separately computed coordinate vectors have been further concatenated to form x and y vectors. In (5) we are using the first derivatives dx and dy of the raw coordinates x and y . The model parameters b_1 and b_2 control the degree of asymmetry and with $b_1 = 20$ and $b_2 = 2$ we obtain a highly asymmetric \mathbf{S} . Due to the symmetric formulation of SC we decompose \mathbf{S} into its symmetric and skew-symmetric parts $\mathbf{S} = \mathbf{S}_{sym} + \mathbf{S}_{skew}$ according to [10]. Because of its symmetric formulation, SC disregards \mathbf{S}_{skew} and diagonalizes only \mathbf{S}_{sym} .

Given a high degree of asymmetry, SC fails to correctly identify the two separate spirals as illustrated in Fig. 2(b). Conversely, SCE fully utilizes the information in the asymmetric component of \mathbf{S} to achieve the correct separation (Fig. 2(d)) and results in a better projection and thus higher inter-cluster separability than kernel-PCA. The application of SCE to image segmentation and its relation to $Ncut$ will be investigated in the next section.

4 Connection of SCE with Normalized Cuts

$Ncut$ is the graph-theoretic formulation of SC with the objective to minimize a normalized measure of disassociation [7]. $Ncut$ operates on the 2nd generalized eigenvector of a normalized weight matrix \mathbf{W} where the normalization procedure has the purpose to penalize large image segments. $Ncut$ then computes the diagonal matrix \mathbf{D} containing the sum of all edges and solves for the eigenvectors of $\mathbf{N} = \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}}$ with $\mathbf{N}(i, j) = \mathbf{W}(i, j) / \sqrt{\mathbf{D}(i, i)} \sqrt{\mathbf{D}(j, j)}$. The second smallest generalized eigenvector λ_2 of \mathbf{W} is a componentwise ratio of the second and first largest eigenvectors of \mathbf{N} [6]. We are interested if our Φ -based approximation can provide computational savings over the $Ncut$ algorithm while maintaining the same image partition.

For comparative purposes we have acquired the $Ncut$ demo software from [26] and used the supplied parameters for the calculation of the adjacency matrix based on intervening contour similarities. In order to compare the segmentation results, we take the returned \mathbf{N} matrix and compute Φ as outlined in Fig. 1. There are two aspects which are non-trivial in connection with $Ncut$: definition of a feature similarity and selection of a partitioning threshold which can take the values of 0, median or a point that minimizes $Ncut(A, B)$ [7]. In our SCE formulation we center the columns of \mathbf{N} before mixing with \mathbf{t} for a number of z iterations. We can also implement $\mathbf{N}^z \mathbf{t}$ mixing iteratively by centering only the mixed $N^2 \times 2$ components after each iteration. The successive centering results in the ideal non-parametric case, where we partition the graph according only to the signs ($A = \{\Phi_1 > 0\}, B = \{\Phi_1 \leq 0\}$). For our experiments we used a Dell Precision M6300 dual-core notebook with 2GB RAM and Matlab R2009a

environment. The results of an experiment on the full resolution 321×481 images from the Berkeley segmentation database are shown in Fig. 3.

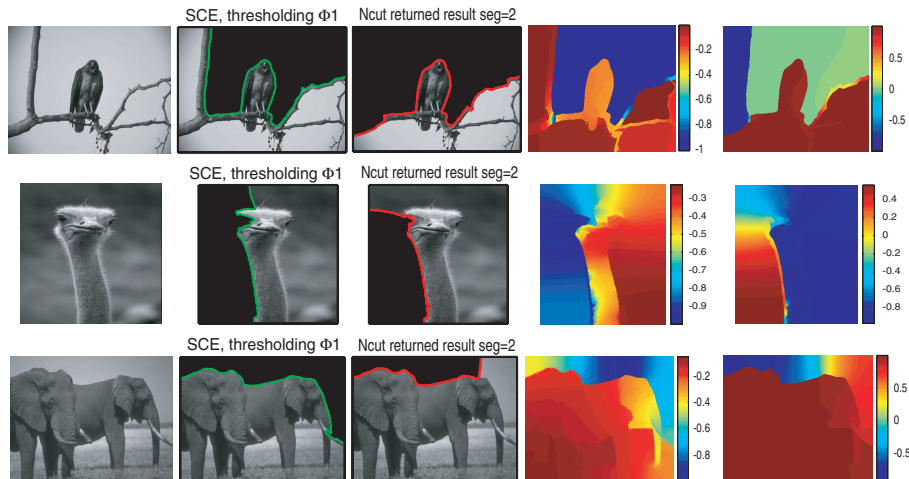


Fig. 3. Original 321×481 images from the Berkeley database are shown in the first column and SCE-based results in the second column. The last three columns show the *Ncut*-based result and the first and the second eigenvector returned by *Ncut*. This diagram is best viewed in color.

The returned binary *Ncut* partition (Fig. 3, third column) is given by the second computed eigenvector (fifth column). We observe the qualitative equivalence between SCE segmentation based on Φ_1 and the first eigenvector (fourth column) returned by *Ncut* algorithm which also outputs a very narrow-banded sparse matrix \mathbf{N} with $\approx 0.1\%$ non-zero elements. Due to the inherent sparsity advantage which has its roots in the definition of similarities [7] *Ncut* does not rely on the direct eigen-decomposition of \mathbf{N} . Instead, it uses the iterative Lanczos eigensolver [28] which, similarly to SCE, is also based on sparse matrix-vector multiplications. In our experiments with the returned sparse \mathbf{N} , the ρ_{min} is marginally low $\rho_{min} = -8.6025e - 004$ which explains the high number of SCE iterations ($z \approx 1e4$) needed to approximate the first *Ncut* eigenvector.

5 SCE Extension with Latent Variables

It is known that although the eigenvectors are efficient in capturing the perceptual organization features [16, 5, 6], binary *Ncut* solution does not guarantee the correct discrete image partitions [7]. Conversely, in our SCE formulation, the approximation to the first *Ncut* eigenvector is connected with the rotation and scaling of two hyperplanes, implemented through iterative \mathbf{N} by \mathbf{t} multiplications. Therefore, it is reasonable to assume that the foreground innovations

can be extracted ahead of the eigenvector equivalence condition and in a much shorter time. To approach these problems, we initially followed the idea of a greedy search and designed an optimization procedure detailed in Table 2.

Table 2. Ising-based SCE extension.

```

function [ $\Omega, A, B$ ] = sccut( $\mathbf{N}, \mathbf{t}, \text{itmax}$ )
 $\mathbf{k} = \text{cov}(\mathbf{t}); [\mathbf{u}, \mathbf{v}] = \text{eigs}(\mathbf{k}); \mathbf{t} = \mathbf{t} * \mathbf{u}; \mathbf{t} = \mathbf{t} - \bar{\mathbf{t}}; \text{it} = 0;$ 
while  $\text{it} < \text{itmax}$ 
   $\mathbf{b} = \mathbf{N} * \mathbf{t}; \mathbf{b} = \mathbf{b} - \bar{\mathbf{b}};$ 
  if  $\text{std}(\mathbf{b}_1) > \text{std}(\mathbf{b}_2)$ 
     $\mathbf{t}_1 = -\text{sign}(\mathbf{b}_1) * \mathbf{b}_2; \mathbf{t}_2 = \mathbf{b}_2;$ 
  else
     $\mathbf{t}_1 = -\text{sign}(\mathbf{b}_2) * \mathbf{b}_1; \mathbf{t}_2 = \mathbf{b}_1;$ 
  end
   $\text{it} = \text{it} + 1;$ 
end
 $\Omega(:, 1) = \mathbf{t}_1; \Omega(:, 2) = \mathbf{t}_2$ 

```

$A : \Omega(:, 1) > 0; B : \Omega(:, 1) \leq 0;$

Therein, we view the matrix \mathbf{N} as the matrix of features and consider that \mathbf{N} is sparse. Further, we view the signs of a pair of columns as two latent (hidden) binary support variables $s_1(+)$ and $s_2(-)$ and thus establish the connection with the Ising model which is a special Markov random field (MRF) [3].

In Table 2 we define $s_1 = \text{sign}(\mathbf{b}_1)$ if $\text{std}(\mathbf{b}_1) > \text{std}(\mathbf{b}_2)$ and $s_2 = \text{sign}(\mathbf{b}_2)$ if $\text{std}(\mathbf{b}_1) \leq \text{std}(\mathbf{b}_2)$, where "std" denotes the standard deviation. Instead of using the Ising model to represent pixels, we work with similarities contained in the normalized sparse matrix \mathbf{N} . We denote the optimized feature vectors by Ω . The initial condition is given by the pair of columns $\mathbf{t}_{1,2}$ which, with the change in notation such that $\mathbf{S} \equiv \mathbf{N}$, can be obtained according to:

$$\mathbf{N} \Rightarrow \mathbf{P} \in \mathbb{R}^{N^2 \times H} \Rightarrow \mathbf{p}_i = \mathbf{p}_i - \bar{\mathbf{p}}_i \Rightarrow \mathbf{R} \Rightarrow \mathbf{c}_{1,2} \Rightarrow \mathbf{t}_{1,2} . \quad (6)$$

Due to memory limitations we did not search for the global minimum on correlation in \mathbf{N} but instead operated on a subset matrix \mathbf{P} of H randomly selected columns. Thus, in the experiment shown in Fig. 4 we selected $H = 100$ in order to process a 321×481 image from the Berkeley database.

Although the automatic selection of the stopping criterion is still an ongoing work, we note that one possibility to obtain the optimal partition is to analyze the dynamic oscillatory behavior of the correlation coefficient $\rho(\mathbf{t}_1, \mathbf{t}_2)$ (first column in Fig. 4) and we observed that the optimal figure ground cut occurs at the change in phase of ρ . The binary A (figure) and B (ground) partitions have to be computed twice for the two successive iterations corresponding to the ρ transition. The final segmentation result F combines the intermediate results at

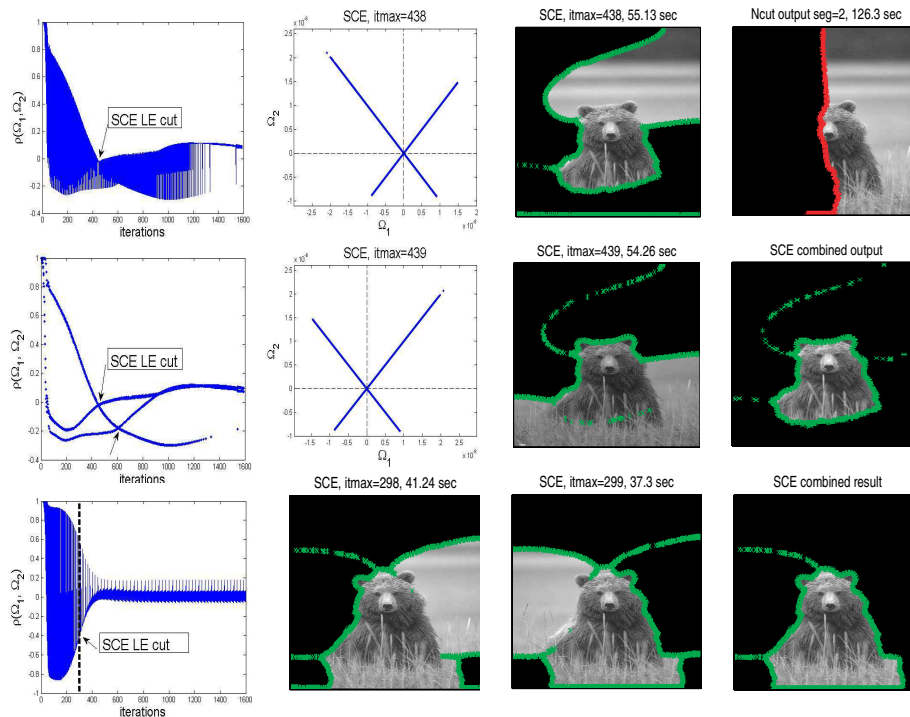


Fig. 4. Concept of the Ising-based SCE. We analyze the dynamic oscillatory behavior of the correlation coefficient to find the optimal transition. For the two iterations near the transition point we compute the binary classification and combine the results to yield the optimal figure ground segmentation. The transition is given by the point where the hyperplanes are flipped around the Ω_1 (strongest) axis as can be seen in the second column. On our computer SCE runs faster (see the computational time above the diagrams) than *Ncut* and returns more perceptually meaningful binary segmentation.

different iterations such that $F = F_1 \cap F_2$, where $F_1 = A_1 \cup B_1$ and $F_2 = A_2 \cup B_2$ (see Fig. 4 last row). Random subsampling of \mathbf{N} (6) explains somewhat different, but consistent with perceptual meaning, segmentation results in the second and the third row of Fig. 4, where we used different random subsets of \mathbf{N} . The results in Fig. 4 show that not only the Ising-based SCE detects the foreground innovations in the analyzed image but also has a factor 2 speed-up compared to *Ncut*.

6 Conclusions and Future Work

In this paper we developed an efficient alternative to eigenvector-based feature classification. We started by examining the conditions of the feature space equivalence between the proposed SCE and the kernel-PCA outputs. We further have shown that the proposed algorithm reduces the dimension of the feature space

while improving classification performance and thus results in a better projection and higher inter-cluster separability than kernel-PCA. In regard to image segmentation, we demonstrated that the proposed method has potential to replace eigenvector-based computation at least for applications considering the detection of foreground innovations. Our future work will concentrate on generalizing SCE to multiclass problems as well as investigating the regularization and stopping criteria of the proposed Ising-based SCE extension.

Although the Ising model takes SCE beyond the equivalence pursuit, it shows that segmentation without eigenvectors is a more flexible framework than that offered by the standard spectral clustering.

Acknowledgments

This research was supported by the National Biophotonics and Imaging Platform Ireland funded under the HEA PRTL Cycle 4, co-funded by the Irish Government and the European Union - *Investing in your future*. Our special thanks go to the anonymous referees whose invaluable suggestions and comments have helped us to improve this paper.

References

1. Rodgers, J. L. and Nicewander, W. A.: Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, vol. **42**, pp. 59–66, (1988)
2. Tran, H. T., Romanov, D. A. and Levis, R. J.: Control Goal Selection Through Anticorrelation Analysis in the Detection Space. *Journal of Physical Chemistry A*, vol. **110**, pp. 10558–10563, (2006)
3. Cevher, V., Duarte, M. F., Hegde, C. and Baraniuk, R. G.: Sparse Signal Recovery Using Markov Random Fields. *Neural Information Processing Systems (NIPS)*, pp. 257–264, (2008)
4. Shi, J. and Malik, J.: Normalized Cuts and Image Segmentation. *Computer Vision and Pattern Recognition (CVPR)*, pp. 731–737, (1997)
5. Perona, P. and Freeman, W. T.: Factorization Approach to Grouping. *European Conference on Computer Vision (ECCV)*, pp. 655–670, (1998)
6. Weiss, Y.: Segmentation Using Eigenvectors: A Unifying View. *International Conference on Computer Vision (ICCV)*, (1999)
7. Shi, J. and Malik, J.: Normalized Cuts and Image Segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. **22**, pp. 888–905, (2000)
8. Ng, A. Y., Jordan, M. I. and Weiss, Y.: On Spectral Clustering: Analysis and an Algorithm. *Neural Information Processing Systems (NIPS)* 14, pp. 849–856, (2001)
9. Roth, V., Laub, J., Kawanabe, M. and Buhmann, J. M.: Optimal Cluster Preserving Embedding of Nonmetric Proximity Data. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. **25**, pp. 1540–1551, (2003)
10. Constantine, A. G. and Gower, J. C.: Graphical Representation of Asymmetric Matrices. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. **27**, pp. 297–304, (1978)
11. Jolliffe, I. T.: *Principal Component Analysis*, Springer-Verlag, (2002)
12. Anderberg, M. R.: *Cluster Analysis for Applications*, Academic Press Inc., (1973)

13. Zheng, N. and Xue, J.: Statistical Learning and Pattern Analysis for Image and Video Processing, Springer-Verlag London Limited, (2009)
14. Laub, J., Roth, V., Buhmann, J. M. and Müller, K.-R.: On the Information and Representation of Non-Euclidean Pairwise Data. *Pattern Recognition*, vol. **39**, pp. 1815–1826, (2006)
15. Alzate, C. and Suykens, J. A. K.: Multiway Spectral Clustering with Out-of-Sample Extensions Through Weighted Kernel PCA. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. **32**, pp. 335–347, (2010)
16. Monteiro, F. C. and Campilho, A. C.: Spectral Methods in Image Segmentation: A Combined Approach. *LNCS*, vol. **3523**, pp. 191–198, (2005)
17. Srebro, N. and Jaakkola, T.: Linear Dependent Dimensionality Reduction. *Advances in Neural Information Processing Systems (NIPS) 16*, pp. 145–152, (2003)
18. Sanguinetti, G.: Dimensionality Reduction in Clustered Data Sets. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. **30**, pp. 535–540, (2008)
19. Talwalkar, A., Kumar, S. and Rowley, H.: Large-Scale Manifold Learning. *Computer Vision and Pattern Recognition (CVPR)*, (2008)
20. Pełkalska, E., Harol A., Duin, R. P. W., Spillmann, B. and Bunke, H.: Non-Euclidean or Non-Metric Measures Can Be Informative. *LNCS*, vol. **4109**, pp. 871–880, (2006)
21. Belabbas, M.-A. and Wolfe, P.: Spectral Methods in Machine Learning and New Strategies for Very Large Datasets. *Proceedings of National Academy of Sciences (PNAS) of the USA*, vol. **106**, pp. 369–374, (2009)
22. Chang, H. and Yeung, D.-Y.: Robust Path-based Spectral Clustering. *Pattern Recognition*, vol. **41**, pp. 191–203, (2008)
23. Fowlkes, C., Belongie, S., Chung, F. and Malik, J.: Spectral Grouping Using the Nystrom Method. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. **26**, pp. 214–225, (2004)
24. Tung, F., Wong, A. and Clausi, D. A.: Enabling Scalable Spectral Clustering for Image Segmentation. *Pattern Recognition*, vol. **43**, pp. 4069–4076, (2010)
25. Chen, W.-Y., Song, Y., Bai, H., Lin, C.-J. and Chang, E. Y.: Parallel Spectral Clustering in Distributed Systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. **33**, pp. 568–586, (2010)
26. Cour, T., Yu, S. and Shi, J.: *Ncut* demo software, <http://www.cis.upenn.edu/~jshi/software>
27. Schölkopf, B. and Smola, A. J.: *Learning with Kernels*. MIT Press, (2002)
28. Stewart, G. W.: *Matrix Algorithms, Volume II: Eigensystems*. SIAM, (2001)