



**HAL**  
open science

# Independent Component Clustering for Skin Lesions Characterization

S. K. Tasoulis, C. N. Doukas, I. Maglogiannis, V. P. Plagianakos

► **To cite this version:**

S. K. Tasoulis, C. N. Doukas, I. Maglogiannis, V. P. Plagianakos. Independent Component Clustering for Skin Lesions Characterization. 12th Engineering Applications of Neural Networks (EANN 2011) and 7th Artificial Intelligence Applications and Innovations (AIAI), Sep 2011, Corfu, Greece. pp.472-482, 10.1007/978-3-642-23960-1\_55 . hal-01571494

**HAL Id: hal-01571494**

**<https://inria.hal.science/hal-01571494>**

Submitted on 2 Aug 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Independent Component Clustering for Skin Lesions Characterization

S.K. Tasoulis<sup>1</sup>, C.N. Doukas<sup>2</sup>, I. Maglogiannis<sup>1</sup>, and V.P. Plagianakos<sup>1</sup>

<sup>1</sup> Department of Computer Science and Biomedical Informatics,  
University of Central Greece, Papassiopoulou 2–4, Lamia, 35100, Greece  
{stas, imaglo, vpp}@ucg.gr

<sup>2</sup> Department of Information and Communication Systems Engineering,  
University of the Aegean, Karlovassi, 83200, Samos, Greece  
doukas@aegean.gr

**Abstract.** In this paper, we propose a clustering technique for the recognition of pigmented skin lesions in dermatological images. It is known that computer vision-based diagnosis systems have been used aiming mostly at the early detection of skin cancer and more specifically the recognition of malignant melanoma tumor. The feature extraction is performed utilizing digital image processing methods, i.e. segmentation, border detection, color and texture processing. The proposed method combines an already successful clustering technique from the field of projection based clustering with a projection pursuit method. Experimental results show great performance on detecting the skin cancer.

**Key words:** Pigmented Skin Lesion, Image Analysis, Feature Extraction, Unsupervised clustering, Cluster analysis, Independent Component Analysis, Projection Pursuit, Kernel density estimation.

## 1 Introduction

Several studies found in literature have proven that the analysis of dermatological images and the quantification of tissue lesion features may be of essential importance in dermatology [1, 3]. The main goal is the early detection of malignant melanoma tumor, which is among the most frequent types of skin cancer, versus other types of non-malignant cutaneous diseases. The interest in melanoma is due to the fact that its incidence has increased faster than that of almost all other cancers and the annual incidence rates have increased on the order of 3 – 7% in fair-skinned populations in recent decades [2].

The advanced cutaneous melanoma is still incurable, but when diagnosed at early stages it can be cured without complications. However, the differentiation of early melanoma from other non-malignant pigmented skin lesions is not trivial even for experienced dermatologists. In several cases, primary care physicians underestimate melanoma in its early stage [3]. To deal with this problem in several cases we utilize data mining methods. In particular, using clustering could

be the key step to understand the differences between the types and subtypes of skin lesions.

In this paper, based on an already saucerful clustering technique, we propose a new algorithmic framework for the skin lesion characterization. The paper is organized as follows: in Section 2 we present the image dataset, as well as the preprocessing and segmentation, and feature extraction techniques applied. Next, in Section 4 we present the ICA model. Section 5 is devoted to the proposed method and in Section 6 we investigate the efficiency of the proposed technique. The paper ends with concluding remarks.

## 2 Skin Lesions Image Analysis

The image data set used in this study is an extraction of the skin database that exists at the Vienna and the Athens General Hospital, kindly provided by Dr. Ganster. The whole data set consists of 3631 images, 972 of them are displaying nevus (dysplastic skin lesions), 2590 featuring non-dysplastic lesions and the rest 69 images contain malignant melanoma cases. The number of the melanoma images set is not small considering the fact that malignant melanoma cases in a primordial state are very rare. It is very common that many patients arrive at specialized hospitals with partially removed lesions.

The first step in an image analysis workflow is image segmentation, which in this case concerns the separation of the skin lesion from the healthy skin. For the special problem of skin lesion segmentation, mainly region-based segmentation methods are applied [5, 9]. A simple approach is thresholding, which is based on the fact that the values of pixels that belong to a skin lesion differ from the values of the background.

In this study, a more sophisticated approach of a local/adaptive thresholding technique was adopted, where the window size, the threshold value and degree of overlap between successive moving windows were the procedure parameters. The details of this method may be found in [8]. Image analysis and feature extraction is performed by measurements on the pixels that represent a segmented object allowing non-visible features to be computed. Several studies have also proven the efficiency of border shape descriptors for the detection of malignant melanoma on both clinical and computer based evaluation methods [7, 11]. Three types of features are utilized in this study: Border Features which cover the A and B parts of the ABCD-rule of dermatology, Color Features which correspond to the C rules and Textural Features, which are based on D rules. More specifically the extracted features are as follows:

### *Border features*

- Thinness Ratio measures the circularity of the skin lesion defined as:  
 $TR = 4\pi \text{Area}/(\text{perimeter})^2$ .
- Border Asymmetry is computed as the percent of non-overlapping area after a hypothetical folding of the border around the greatest diameter or the maximum symmetry diameters.

- The variance of the distance of the border lesion points from the centroid location.
- Minimum, maximum, average and variance responses of the gradient operator, applied on the intensity image along the lesion border.

#### Color Features

- Plain RGB color plane average and variance responses for pixels within the lesion.
- Intensity, Hue, Saturation Color Space average and variance responses for pixels within the lesion:  $I = \frac{R+G+B}{3}$ ,  $S = 1 - \frac{3}{R+G+B}[\min(R, G, B)]$ , and

$$H = \begin{cases} W & , G > B, \\ 2\pi - W & , G < B, \\ 0 & , G = B, \end{cases} \text{ and } W = \arccos\left[\frac{R(1 - \frac{1}{2}(G + B))}{(R - G)^2 + (R - B)(G - B)^{\frac{1}{2}}}\right].$$

- Spherical coordinates LAB average and variance responses for pixels within the lesion:  
 $L = \sqrt{R^2 + G^2 + B^2}$ ,  $\text{AngleA} = \cos^{-1}[\frac{B}{L}]$ ,  
and  $\text{AngleB} = \cos^{-1}[\frac{R}{L \sin(\text{AngleA})}]$ .

#### Texture features

- Dissimilarity,  $d$ , which is a measure related to contrast using linear increase of weights as one moves away from the GLCM (gray level co-occurrence matrix) diagonal:  $d = \sum_{i,j=0}^{N-1} P_{i,j} \|i - j\|$ , where  $i$  and  $j$  denote the rows and columns, respectively,  $N$  is the total number of rows and columns, and  $P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}}$  is the normalization equation in which  $V_{i,j}$  is the DN value of the cell  $i, j$  in the image window.
- Angular Second Moment, ASM, which is a measure related to orderliness, where  $P_{i,j}$  is used as a weight to itself:  $ASM = \sum_{i,j=0}^{N-1} i P_{i,j}^2$ .
- GLCM Mean,  $\mu_i$ , which differs from the familiar mean equation in the sense that it denotes the frequency of the occurrence of one pixel value in combination with a certain neighbor pixel value and is given by  $\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j})$ . For the symmetrical GLCM, holds that  $\mu_i = \mu_j$ .
- GLCM Standard Deviation,  $\sigma_i$ , which gives a measure of the dispersion of the values around the mean:  $\sigma_i = \sqrt{\sum_{i,j=0}^{N-1} P_{i,j} (i - \mu_i)^2}$ .

### 3 Clustering Background

The “divisive” hierarchical clustering techniques produce a nested sequence of partitions, with a single, all-inclusive cluster at the top. Starting from this all-inclusive cluster the nested sequence of partitions is constructed by iteratively splitting clusters, until a termination criterion is satisfied. Any divisive clustering algorithm can be characterized by the way it chooses to provide answers to the following three questions:

- $Q_1$ : Which cluster to split further?  
 $Q_2$ : How to split the selected cluster?  
 $Q_3$ : When should the iteration terminate?

The projection based divisive clustering algorithms in particular, projects the high dimensional data onto a lower dimensional subspace to provide answers to the questions above, in a computationally efficient manner. Note also that certain answers to one of these questions may render obsolete one of the others. However, this is not always the case.

To formally describe the manner in which projection based divisive clustering algorithms operate, let us assume the data is represented by an  $n \times a$  matrix  $D$ , whose each row represents a data sample  $d_i$ , for  $i = 1, \dots, n$ . Finally, if  $A$  is the matrix with columns the vectors that denote the targeted subspace, then

$$D_{n \times k}^P = D_{n \times a} A_{a \times k},$$

is the projection of the data onto the lower  $k$ -dimensional subspace defined by the matrix  $A$ . The most studied technique for such data analysis is the Principal Component Analysis (PCA) [21]. PCA can be viewed as one of many possible procedures for projection pursuit [17] and is able to compute meaningful projections of high dimensional data [4, 13, 15, 26].

## 4 Independent Component Analysis

Independent component analysis (ICA) [12, 14] is a technique that finds underlying factors or independent components from multivariate (multidimensional) statistical data by maximizing the statistical independence of the estimated components. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear or nonlinear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed non-gaussian and mutually independent, and they are called the independent components of the observed data. These independent components can be found by ICA. We can define ICA as follows. Let  $x = (x_1, \dots, x_n)$  be the random vector that represents the data and  $s = (s_1, \dots, s_n)$  be random vector that represents the components. The task is to transform the data  $x$ , using a linear static transformation  $W$ , as  $s = Wx$ , into maximally independent components  $s$  measured by some function of independence. The definition of independence for ICA that we utilize in this work is the maximization non-gaussianity. The most used measure of non-gaussianity is kurtosis and the second measure is given by negentropy. Kurtosis is zero for a gaussian random variable and nonzero for most non-gaussian random variables. Negentropy is based on the information-theoretic quantity of (differential) entropy. The entropy of a random variable can be interpreted as the degree of information that the observation of the variable gives. The more random and unstructured the variable is, the larger its entropy.

#### 4.1 Relation to Projection Pursuit

The critical attribute of the ICA model is that we can use it to find directions for which the 1-dimensional projected data onto these directions show the least Gaussian distribution. It has been argued by Huber [18] and by Jones and Sibson [23] that the Gaussian distribution is the least interesting one, and that the most interesting directions are those that show the least Gaussian distribution. Interesting distribution can be a consider a distribution that captures the structure of the data. As such ICA can be considered as an 1-dimensional projection pursuit technique for finding directions of maximum non-gaussianity.

#### 4.2 The FastICA Algorithm

To find the direction for maximum non-gaussianity, we utilize a well known fixed point algorithm. The FastICA algorithm [20] is a very efficient method for maximizing the objective function with respect to the selected measure of non-gaussianity. For this task it is assumed that the data is preprocessed by centering and whitening. Whitening can be achieved with principal component analysis or singular value decomposition. Whitening ensures that all dimensions are treated equally, before the algorithm is run. In this work, we only make use of the FastICA algorithm for one unit. The FastICA for one unit finds a direction  $w$  such that the projection  $D_{n \times k}^P = D_{n \times a} w_{a \times 1}$ , maximizes nongaussianity. Nongaussianity is here measured by the approximation of negentropy [19].

### 5 The Proposed Framework

A new clustering algorithm that was proposed recently [27], incorporates information about the true clusters in the data from the density of their projections on the principal components. Based on that principle, we introduce an new algorithmic scheme that utilize the ICA model to find optimal directions to project the data and then splits the data based on the density of their projections on these directions. In contrast to the previous approach, in this work, we choose as a splitting point the maximum of all the local minima of the projections density denoted by  $x^*$ . Finally, to estimate the ICA model, the FastICA algorithm is utilized. The proposed algorithmic scheme mICDC (refer to Algorithm 1) utilizes the following criteria:

- (Stopping Criterion) *ST*: Let  $\Pi = \{\{C_i, P_i\}, i = 1, \dots, k\}$  a partition of the data set  $\mathcal{D}$  into  $k$  sets  $C_i$ , and the assorted projections  $P_i$  of them onto the direction of maximum nongaussianity. Let  $\mathcal{X}$ , be the set  $x_i^*$  of the density estimates  $\hat{f}(x_i^*; h)$  of the projection  $P_i$  of the data of each  $C_i \in \Pi$ ,  $i = 1, \dots, k$ . Stop the procedure when the set  $\mathcal{X}$  is empty.
- (Cluster Selection Criterion) *CS*: Let  $\Pi = \{\{C_i, P_i\}, i = 1, \dots, k\}$  a partition of the data set  $\mathcal{D}$  into  $k$  sets  $C_i$ , and the assorted projections  $P_i$  of them onto the direction of maximum nongaussianity. Let  $\mathcal{F}$  be the set of the density estimates  $f_i = \hat{f}(x_i^*; h)$  of  $x_i^*$  for the projection  $P_i$  of the data of each

$C_i \in \Pi$ ,  $i = 1, \dots, k$ . The next set to split is  $C_j$ , with  $j = \arg \max_i \{f_i : f_i \in \mathcal{F}\}$ .

- (Splitting Criterion) *SPC*: Let  $\hat{f}'(x; h')$  be the kernel density estimation of the density of the projections  $p_i \in \mathcal{P}$ , and  $x^*$  the maximum of the minima. Then construct  $P_1 = \{d_i \in \mathcal{D} : p_i \leq x^*\}$  and  $P_2 = \{d_i \in \mathcal{D} : p_i > x^*\}$ .

```

1 Function mICDC ( $\mathcal{D}$ )
2 Get  $u^p$  the direction of maximum nongaussianity of  $\mathcal{D}$ 
3 Calculate  $P = Du^p$  the projection of  $\mathcal{D}$  to  $u^p$ 
4 Set  $\Pi = \{\{\mathcal{D}, P\}\}$ 
5 repeat
6   Select an element  $\{\mathcal{C}, P^C\} \in \Pi$  using Cluster Selection Criterion CS
7   Split  $\mathcal{C}$  into two sub-sets  $\mathcal{C}_1, \mathcal{C}_2$ , using Splitting Criterion SPC
8   Remove  $\{\mathcal{C}, P^C\}$  from  $\Pi$  and set  $\Pi \rightarrow \Pi \cup \{\{\mathcal{C}_1, P^{C_1}\}, \{\mathcal{C}_2, P^{C_2}\}\}$ , where
    $P^{C_1}, P^{C_2}$  are the projections of  $\mathcal{C}_1, \mathcal{C}_2$  on the direction of maximum
   nongaussianity  $u^{p_1}, u^{p_2}$  of  $\mathcal{C}_1$  and  $\mathcal{C}_2$ , respectively
9 until Stopping Criterion ST is not satisfied;
10 return  $\Pi$  the partition of  $\mathcal{D}$  into  $|\Pi|$  clusters

```

**Algorithm 1:** The mICDC algorithm

The computational complexity of this approach, using a brute force technique, would be quadratic in the number of samples. However, it has been shown [16, 28] that using techniques like the Fast Gauss Transform linear running time can be achieved for the Kernel Density Estimation, especially for the one dimensional case. To find the maximum of the minima, we only need to evaluate the density at  $n$  positions, in between the projected data points, since those are the only valid splitting points.

## 6 Experimental Analysis

As defined in Section 2, the image dataset used in our experiments consists of 3631 images; 972 of them are displaying nevus (dysplastic skin lesions), 2590 featuring non-dysplastic lesions, and the rest 69 images contain malignant melanoma cases. The number of the melanoma images is not so small considering the fact that malignant melanoma cases in a primordial state are very rare.

In our first experiment, in an attempt to effectively retrieve the malignant melanoma class, we consider the problem as a two class situation. Since we know beforehand the actual cluster number, we can control the number of clusters that the algorithm retrieves by setting a proper value of the bandwidth parameter for the density function. In [27] the bandwidth parameter was set by choosing a multiple of the  $h_{opt}$  bandwidth (“normal reference rule”), which is the bandwidth

that minimizes the Mean Integrated Squared Error (MISE). This is given by:

$$h_{opt} = \sigma \left( \frac{4}{3n} \right)^{1/5}, \quad (1)$$

where  $\sigma$  is the standard deviation of the data. The starting multiplier value is set to 2 and then we set greater values in an attempt to get good clustering results with fewer clusters. Table 1 reports the clustering results with respect to the multiplier values. 100 experiments have been made for each case and the mean values and the respective standard deviation are presented.

**Table 1.** Results with respect to the mean clustering purity and V-measure (with the observed standard deviation in parenthesis) for several multiplier values for the full dataset.

		mICDC		
Multi.	Purity	V-measure	Clusters	
2	0.9959(0.00)	0.2788(0.07)	9.40(3.06)	
4	0.9974(0.00)	0.6670(0.19)	5.28(2.81)	
8	0.9986(0.00)	0.9005(0.06)	3(0.92)	
		K-means		
	0.9979(0.00)	0.0673(0.00)	13(0)	
	0.9978(0.00)	0.1013(0.00)	7(0)	
	0.9878(0.00)	0.0913(0.06)	4(0)	

To assess the quality of a data partition, additional external information not available to the algorithm, such as class labels, are used [22, 24]. Consequently, the degree of correspondence between the resulting clusters and the classes assigned a priori to each object can be measured. For a dataset  $\mathcal{D}$ , let  $\mathcal{L}$  be a set of labels  $l_i \in \mathcal{L}$ , for each point  $d_i \in \mathcal{D}$ ,  $i = 1, \dots, n$ , with  $l_i$  taking values in  $\{1, \dots, L\}$ . Let a  $k$ -cluster partitioning  $\Pi = \{\mathcal{C}_1, \dots, \dots, \mathcal{C}_k\}$ . The purity of  $\Pi$  is defined as:

$$p(\Pi) = \frac{\sum_{j=1}^k \max \{ |\{p_i \in \mathcal{C}_j : l_i = 1, \dots, L\}| \}}{n}, \quad (2)$$

so that  $0 \leq p(\Pi) \leq 1$ . High values indicate that the majority of vectors in each cluster come from the same class, so in essence the partitioning is “pure” with respect to class labels.

However, cluster purity does not address the question of whether all members of a given class are included in a single cluster and therefore is expected to increase monotonically with the number of clusters in the result. For this reason, criteria like the V-measure [25] have been proposed. The V-measure tries to capture cluster homogeneity and completeness, which summarizes a clustering solution’s success in including every point of a single class and no others. Again, high values corresponds to better performance. For details on how these are calculated, the interested reader should refer to [25].

As shown even for the case of a high multiplier value where the algorithm finds very few clusters, the performance remains at high levels. For comparison



purposes, we also report the performance of well known k-means algorithm. For the results to be comparable, the number of clusters for the k-means algorithm is set to the values found by the mICDC algorithm. For the computation of k-means, we employ the Matlab function “kmeans”.

To better understand the clustering results, firstly we employ the confusion matrices for the 4 cluster case (Table 2). Class 1 refers to the malignant melanoma class and class 2 to the rest of the dataset.

**Table 2.** Confusion Matrices of the mICDC and k-means algorithms for 4 clusters

	mICDC			k-means	
	Class 1	Class 2		Class 1	Class 2
cluster1	0	3558	cluster1	6	2436
cluster2	62	0	cluster2	62	0
cluster3	4	0	cluster3	1	884
cluster4	3	0	cluster4	0	238

As shown in this case, the mICDC algorithm does not split the much bigger class 2; so the rest of the clusters are considered to belong to class 1 (melanoma samples). This is a very important result, since one can easily conclude that the clusters with the much fewer samples constitute the malignant melanoma class. As shown in Table 3, the malignant melanoma class can also be found even when the algorithms retrieves more clusters.

**Table 3.** Confusion Matrices of the mICDC algorithm for 7 clusters

mICDC					
	Class 1			Class 2	
	Class 1	Class 2		Class 1	Class 2
cluster1	5	0	cluster5	5	0
cluster2	46	0	cluster6	0	3558
cluster3	6	0	cluster7	3	0
cluster4	4	0			

In our next experiment, we will perform clustering on the dataset containing only the displaying nevus and the malignant melanoma classes. The lack of samples of the non-dysplastic class makes the projection pursuit problem a bit more difficult, due loss of information. The results are presented in Table 4. In Table 5, we exhibit a high performance case of the mICDC algorithm. In this case the two classes have been perfectly separated and the majority of the displaying nevus samples are at the same cluster. However, there also exists a small cluster that does not belong to the malignant melanoma class.

## 7 Conclusions

In this paper, a clustering technique for the recognition of pigmented skin lesions in dermatological images is proposed. The images are preprocessed and feature extraction is performed utilizing digital image processing methods, i.e. segmentation, border detection, color and texture processing. The proposed clustering methodology combines an already successful clustering technique from the

**Table 4.** Results with respect to the mean clustering purity and V-measure (with the observed standard deviation in parenthesis) for different multiplier values for the dataset containing only nevus and the malignant melanoma classes.

mICDC			
Multi.	Purity	V-measure	Clusters
2	0.9809 (0.02)	0.3710 (0.23)	7.30 (2.98)
4	0.9520 (0.02)	0.2490 (0.40)	2.20 (0.63)
K-means			
	0.9908 (0.00)	0.1718 (0.01)	10 (0)
	0.9452 (0.02)	0.1420 (0.02)	3 (0)

**Table 5.** Confusion Matrices of the mICDC algorithm for 7 clusters

mICDC					
	Class 1	Class 2		Class 1	Class 2
cluster1	55	0	cluster4	6	0
cluster2	4	0	cluster5	0	52
cluster3	0	920	cluster6	4	0

field of projection based clustering with a projection pursuit method. The new framework utilizes the ICA model to find optimal projections, and then it incorporates information from the density of the projected data to effectively retrieve the malignant melanoma class. Experimental results show great performance on detecting the skin cancer.

**Acknowledgments:** The authors thank the European Social Fund (ESF), Operational Program for EPEDVM and particularly the Program Herakleitos II, for financially supporting this work.

## References

1. Maglogiannis, I., Doukas, C.: Overview of advanced computer vision systems for skin lesions characterization. *IEEE Transactions on Information Technology in Biomedicine* 13(5), 721–33 (2009)
2. Marks, R.: Epidemiology of melanoma. *Clin Exp Dermatol* 25, 459–463 (2000)
3. Pariser, R., Pariser, D.: Primary care physicians errors in handling cutaneous disorders. *J Am Acad Dermatol* 17(3), 239–245 (1987)
4. Nilsson, M.: Hierarchical Clustering using non-greedy principal direction divisive partitioning. *Information Retrieval* 5(4), 311–321 (2002)
5. Zhang, Z., Stoecker, W., Moss, R.: Border detection on digitized skin tumor image. *IEEE Transactions on Medical Imaging* 19(11), 1128–1143 (2000)
6. Maglogiannis, I.: Automated segmentation and registration of dermatological images. *Springer* 2, 277–294 (2003)
7. Maglogiannis, I., Pavlopoulos, S., Koutsouris, D.: An integrated computer supported acquisition, handling and characterization system for pigmented skin lesions in dermatological images. *IEEE Transactions on Information Technology in Biomedicine* 9(1), 86–98 (2005)

8. Maglogiannis, I., Zafiroopoulos, E., Kyranoudis, C. : Intelligent Segmentation and Classification of Pigmented Skin Lesions in Dermatological Images. *Advances in Artificial Intelligence: 4th Hellenic Conference on AI, Lecture Notes in Computer Science 3955*, 214–223 (2006)
9. Chung, D.H., Sapiro, G.: Segmenting skin lesions with partial-differential-equations-based image processing algorithms. *IEEE Transactions on Medical Imaging* 19(7), 763–767 (2000)
10. Nachbar, F., Stolz, W., Merkle, T., Cognetta, A.B., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O., Plewig, G.: The abcd rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. *J. Amer. Acad. Dermatol.* 30(4), 551–559 (1994)
11. Stoecker, W.V., Li, W.W., Moss, R.H.: Automatic detection of asymmetry in skin tumors. *Computerized Med. Imag. Graph* 16(3), 191–197 (1992)
12. A. Hyvärinen, J.K.E.O.: *Independent Component Analysis*. John Wiley & Sons. (2001)
13. Boley, D.: Principal direction divisive partitioning. *Data Mining and Knowledge Discovery* 2(4), 325–344 (1998)
14. Comon, P.: Independent component analysis, a new concept? *Signal Process.* 36, 287–314 (April 1994)
15. Dhillon, I.S.: Co-clustering documents and words using bipartite spectral graph partitioning. In: *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 269–274. ACM New York, NY, USA (2001)
16. Greengard, L., Strain, J.: The fast gauss transform. *SIAM J. Sci. Stat. Comput.* 12(1), 79–94 (1991)
17. Huber, P.: Projection Pursuit. *Annals of Statistics* 13, 435–475 (1985)
18. Huber, P.J.: Projection pursuit. *Annals of Statistics* 13(2), 435–475 (1985)
19. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. *Neural Networks* 13(4-5), 411–430 (2000)
20. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks* 10, 626–634 (1999)
21. Jain, A.K., Dubes, R.C.: *Algorithms for clustering data* (1988)
22. Jain, A., Dubes, R.: *Algorithms for Clustering Data*. Prentice Hall (1988)
23. Jones, M.C., Sibson, R.: What is projection pursuit? *Journal of the Royal Statistical Society. Series A (General)* 150(1), 1–37 (1987)
24. Kogan, J.: *Introduction to Clustering Large and High-Dimensional Data* (2007)
25. Rosenberg, A., Hirschberg, J.: V-measure: A conditional entropy-based external cluster evaluation measure. In: *2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. pp. 410–420 (2007)
26. Tasoulis, S., Tasoulis, D.: Improving principal direction divisive clustering. In: *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), Workshop on Data Mining using Matrices and Tensors*. Las Vegas, USA (2008)
27. Tasoulis, S., Tasoulis, D., Plagianakos, V.: Enhancing Principal Direction Divisive Clustering. *Pattern Recognition* 43, 3391–3411 (2010)
28. Yang, C., Duraiswami, R., Gumerov, N.A., Davis, L.: Improved fast gauss transform and efficient kernel density estimation. In: *Ninth IEEE International Conference on Computer Vision, 2003. Proceedings*. pp. 664–671 (2003)