Path Based Pairwise Data Clustering with Application to Texture Segmentation

Bernd Fischer, Thomas Zöller and Joachim M. Buhmann

Rheinische Friedrich Wilhelms Universität
Institut für Informatik III, Römerstr. 164
D-53117 Bonn, Germany
{fischerb, zoeller, jb}@cs.uni-bonn.de,
WWW home page: http://www-dbv.informatik.uni-bonn.de

Abstract. Most cost function based clustering or partitioning methods measure the compactness of groups of data. In contrast to this picture of a point source in feature space, some data sources are spread out on a low-dimensional manifold which is embedded in a high dimensional data space. This property is adequately captured by the criterion of connectedness which is approximated by graph theoretic partitioning methods.
We propose in this paper a pairwise clustering cost function with a novel dissimilarity measure emphasizing connectedness in feature space rather than compactness. The connectedness criterion considers two objects as similar if there exists a mediating intra cluster path without an edge with large cost. The cost function is optimized in a multi-scale fashion. This new path based clustering concept is applied to segment textured images with strong texture gradients based on dissimilarities between image patches.

1 Introduction

Partitioning a set of objects into groups and thus extracting the hidden structure of the data set is a very important problem which arises in many application areas e.g. pattern recognition, exploratory data analysis and computer vision. Intuitively, a good grouping solution is characterized by a high degree of homogeneity of the respective clusters. Therefore, the notion of *homogeneity* must be given a mathematically precise meaning which strongly depends on the nature of the underlying data.

In this paper we will deal with an important subclass of partitioning methods namely clustering according to pairwise comparisons between objects. Such data is usually called *proximity* or *(dis)similarity* data respectively. This data modality is of particular interest in applications where object (dis)similarities can be reliably estimated even when the objects are not elements of a metric space. There is a rich variety of clustering approaches developed particularly for this data modality in the literature. Most of them fall in the category of agglomerative methods [13]. These methods share as a common trait that they

start grouping with a configuration composed of exactly one object per cluster and then they successively merge the two most similar clusters. Agglomerative methods are almost always derived from algorithmic considerations rather than on the basis of an optimization principle which often obscures the underlying modeling assumptions.

A systematic approach to pairwise clustering by objective functions as described in [17] is based on an axiomatization of invariance properties and robustness for data grouping. As a consequence of this axiomatic approach we restrict our discussion to intra cluster criteria. Our second important design decision for pairwise clustering replaces the pairwise object comparison by a path-based dissimilarity measure, thereby emphasizing cluster connectedness. The effective dissimilarity between objects is defined as the largest edge cost on the minimal intra cluster path connecting both objects in feature space. Two objects which are assigned to the same cluster are either similar or there exists a set of mediating objects such that two consecutive objects in this chain are similar.

The distinction between compactness and connectedness principles is also addressed by two other recently proposed clustering methods. Tishby and Slonim [20] introduced a Markovian relaxation dynamics where the Markov transition probability is given by object dissimilarities. Iterating such a relaxation dynamics effectively connects objects by sums over minimal paths. The method, however, does not include the constraint that all considered paths have to be restricted to nodes from the same cluster. The other method which was introduced by Blatt, Wiseman and Domany [2] simulates the dynamics of a locally connected, diluted ferromagnet. The partial order at finite temperature is interpreted as a clustering solution in this model.

2 Pairwise Data Clustering

Notational Prerequisites: The goal of data clustering is formally given by the partitioning of n objects $\mathbf{o}_i$, $1 \leq i \leq n$ into k groups, such that some measure of intra cluster homogeneity is maximized. The memberships of objects to groups can be encoded by a $n \times k$ Matrix $\mathbf{M} \in \{0,1\}^{n \times k}$. In this setting the entry $M_{i\nu}$ is set to 1 if and only if the ith object is assigned to cluster ν which implies the condition $\sum_{\nu=1}^{k} M_{i\nu} = 1, \forall i = 1 \ldots n$. The set of all assignment matrices fulfilling this requirement is denoted in the following by $\mathcal{M}_{n,k}^{\text{part}}$.

The dissimilarity between two objects $\mathbf{o}_i$ and $\mathbf{o}_j$ is represented by D_{ij}. These individual dissimilarity values are collected in a matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$. It is worth noting here that many application domains frequently confront the data analyst with data which violates the triangle inequality. Moreover, the self–dissimilarity of objects often is non vanishing, even negative dissimilarities might occur or a certain percentage of dissimilarities is unknown. For our proposed method, we only require symmetry, i.e. $D_{ij} = D_{ji}$. To distinguish between known and unknown dissimilarities neighborhood sets $\mathcal{N}_1, \ldots, \mathcal{N}_n$ are introduced, i.e., $j \in \mathcal{N}_i$ denotes that the dissimilarity D_{ij} is known.

The Objective Function: With these notational preliminaries, we are now able to address the important modeling step of choosing an appropriate cost function. An axiomatization of objective functions for data clustering based on invariance and robustness criteria is given in [17]. This approach makes explicit the intuitively evident properties that a global shift of the data or a rescaling as well as contamination by noise should not sensitively influence the grouping solution. In accordance with this approach we will focus on the following cost function:

$$\mathcal{H}^{\mathrm{pc}}(\mathbf{M},\mathbf{D}) = \sum_{\nu=1}^{k}\sum_{i=1}^{n} M_{i\nu} d_{i\nu}, \quad \text{where} \quad d_{i\nu} = \frac{\sum_{j\in\mathcal{N}_i} M_{j\nu} D_{ij}}{\sum_{j\in\mathcal{N}_i} M_{j\nu}}. \tag{1}$$

$\mathcal{H}^{\mathrm{pc}}$ sums up individual contributions $d_{i\nu}$ for each object $\mathbf{o}_i$ and each group $\mathcal{C}_\nu$ where $d_{i\nu}$ stands for the average dissimilarity between $\mathbf{o}_i$ and objects belonging to cluster $\mathcal{C}_\nu$. $\mathcal{H}^{\mathrm{pc}}$ thus favors intra–cluster compactness. It is the use of this normalization that removes the sensitive dependency of the minimum of $\mathcal{H}^{\mathrm{pc}}$ to constant shifts of the dissimilarity values and makes it insensitive to different cluster sizes.

Optimization: The optimization of objective functions like $\mathcal{H}^{\mathrm{pc}}$ is computationally difficult since combinatorial optimization problems of this kind exhibit numerous local minima. Furthermore, most of the data partitioning problems are proven to be $\mathcal{NP}$–hard. For robust optimization, stochastic optimization techniques like *Simulated Annealing* (SA) [11] or *Deterministic Annealing* (DA) ([19,7]) have shown to perform satisfactorily on many pattern recognition and computer vision applications. Effectively these annealing methods fall in the class of homotopy methods with smoothing controlled by a temperature parameter. In the zero temperature limit the comparatively fast local optimization algorithm known as ICM is obtained [1].

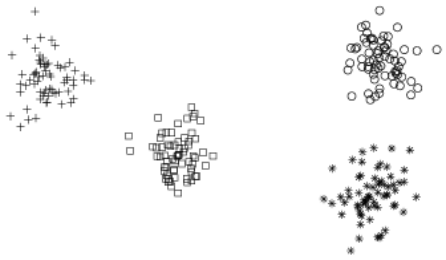


Fig. 1. Prototypical situation for the standard pairwise clustering approach

Drawbacks of the Approach: So far we have discussed a very powerful approach to the pairwise data clustering problem. It is theoretically well founded, and showed to be applicable in a wide range of data analysis problems ranging from texture segmentation [8] and document clustering [17] to structuring of genome data bases [7]. An ideal situation for the cluster compactness criterion is

depicted in figure 1 for a toy data set. All objects have been assigned to clusters in an intuitively correct manner.

However there exist situations, where the exclusive focus on compactness fails to capture essential properties of the data. Two prototypical examples are given in figure 2. It is clear that the human data analyst expects that the ring like structures or the spiral arms are detected as clusters. It is the novel contribution of this paper to propose a way of formalizing this goal by determining effective inter object dissimilarities while keeping the well approved clustering objective function.

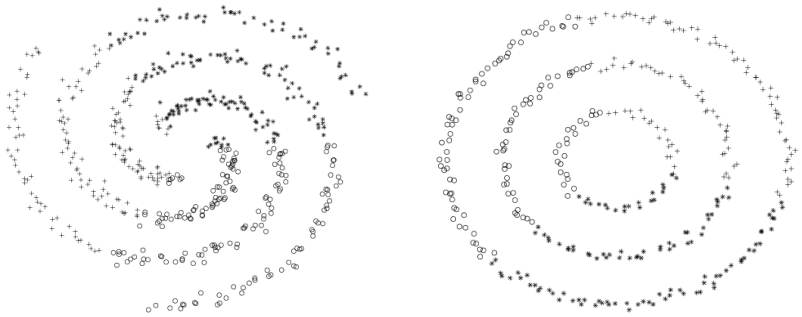


Fig. 2. Data sets on which the naive intra cluster compactness criterion fails

3 Path Based Clustering

In order to motivate our novel modification of the basic pairwise clustering method we will often try to appeal to readers geometric intuition despite the fact that the objects do not necessarily reside in a metric space.

Modeling: As demonstrated before the pairwise clustering approach according to $\mathcal{H}^{pc}$ is not well suited for the detection of elongated structures in the object domain. This objective function sums all intra-cluster dissimilarities whereby similar objects will be grouped together regardless of their topological relations. This is a good solution as long as the data of one group can be interpreted as a scatter around a single centroid (see figure 1). But if the data is a scatter of a curve (c.f. figure 2) or a surface pairwise clustering as defined by $\mathcal{H}^{pc}$ will fail. In this case the effective dissimilarity of two objects should be mediated by peer objects along that curve, no matter how large the extend of that structure may be.

Assume that the objects o_i and o_j belong to the same manifold. Then with high probability there exists a path from o_i to o_j over other objects of this manifold such that the dissimilarities of two consecutive objects are small. The reason for this is that the density of objects coming from one source is expected to be comparatively high. On the other hand if o_i and o_j belong to different,

non-intersecting manifolds then all paths from o_i to o_j have with high probability at least one pair of consecutive objects with high dissimilarity.

As clusters are defined by their coherence one is only interested in paths through the object domain which traverse regions of high density. The part of a path where the density is lowest should thus determine the overall costs of this path. In other words the maximum dissimilarity along a path determines its cost. In order to formalize our ideas about mediating the dissimilarity of any two objects by peers in the same cluster we define the effective dissimilarity D_{ij}^{eff} between objects o_i and o_j to be the length of the minimal connecting path.

$$D_{ij}^{\text{eff}}(\mathbf{M},\mathbf{D}) = \min_{\mathbf{p}\in\mathcal{P}_{ij}(\mathbf{M})} \left\{ \max_{h\in\{1,\ldots,|\mathbf{p}|-1\}} \left\{ D_{\mathbf{p}[h]\mathbf{p}[h+1]} \right\} \right\}, \text{ where} \tag{2}$$

$$\mathcal{P}_{ij}(\mathbf{M}) = \left\{ \mathbf{p}\in\{1,\ldots,n\}^l \,\middle|\, \exists\nu : \prod_{h=1}^{l} M_{p[h]\nu} = 1 \wedge l \leq n \wedge p[1] = i \wedge p[l] = j \right\}$$

is the set of all paths from o_i to o_j through cluster ν if o_i and o_j belong to cluster ν. If both objects belong to different clusters $\mathcal{P}_{ij}(\mathbf{M})$ is the empty set and the effective dissimilarity is not defined.

With the new definition of the effective dissimilarity we are able to define the objective function for path based clustering. It has the same functional form as for pairwise clustering.

$$\mathcal{H}^{\text{pb}}(\mathbf{M},\mathbf{D}) = \sum_{\nu=1}^{k}\sum_{i=1}^{n} M_{i\nu} d_{i\nu}, \quad \text{where} \quad d_{i\nu} = \frac{\sum_{j=1}^{n} M_{j\nu} D_{ij}^{\text{eff}}(\mathbf{M},\mathbf{D})}{\sum_{j=1}^{n} M_{j\nu}}. \tag{3}$$

Thereby the desirable properties of shift and scale invariance of the pairwise clustering cost function are conserved:

$$\forall c, D_0 \in \mathbb{R} : \operatorname*{argmin}_{\mathbf{M}\in\mathcal{M}} \mathcal{H}^{\text{pb}}(\mathbf{M},\mathbf{D}) = \operatorname*{argmin}_{\mathbf{M}\in\mathcal{M}} \mathcal{H}^{\text{pb}}(\mathbf{M}, c\mathbf{D} + D_0), \tag{4}$$

since $\forall\, c, D_0 \in \mathbb{R}$ holds $\mathcal{H}^{\text{pb}}(\mathbf{M}, c\mathbf{D} + D_0) = c\mathcal{H}^{\text{pb}}(\mathbf{M},\mathbf{D}) + N D_0$.

Optimization by Iterated Conditional Mode (ICM): Finding the minimum of $\mathcal{H}^{\text{pb}}(\mathbf{M},\mathbf{D})$ has a high computational complexity. There are many different methods known to avoid a complete search in the assignment configuration space ($|\mathcal{M}| = k^n$). A very effective and simple method is called iterated conditional mode [1]. ICM assigns an object to a cluster under the condition that all other assignments are kept fix. In algorithm 1 the function $s_{\pi[i]}(\mathbf{M}, e_\nu)$ changes the $\pi[i]$-th row of the assignment matrix $\mathbf{M}$ by replacing it with the μth unit vector. This so called single site update is iterated over all objects in a random manner. A complete cycle of visiting all sites is called a sweep. As a common site visitation schedule an arbitrary permutation π of all objects is generated before each sweep in order to avoid local minima due to a fixed visitation order. The algorithm repeats the sweeps until convergence is reached, i.e. no assignment is changed. During each update step the objects are assigned to clusters such that

Algorithm 1 Iterated conditional mode (ICM) for Path Based Clustering

Require: dissimilarity matrix $\mathbf{D}$
number of objects n
number of clusters k
Ensure: $\operatorname{argmin}_{\mathbf{M}\in\mathcal{M}} \mathcal{H}^{\mathbf{pb}}(\mathbf{M},\mathbf{D})$ with high probability
choose $\mathbf{M}$ randomly
repeat
  $\pi = perm(\{1,\ldots,n\})$
  for all $i \in \{1,\ldots,n\}$ **do**
    $\nu^* = \operatorname{argmin}_{\nu\in\{1,\ldots,k\}} \mathcal{H}^{\mathrm{pb}}(s_{\pi[i]}(\mathbf{M},e_\nu),\mathbf{D})$
    where $s_{\pi[i]}(\mathbf{M},e_\nu)$ assigns object $\pi[i]$ to cluster ν
    $\mathbf{M} = s_{\pi[i]}(\mathbf{M},e_{\nu^*})$
  end for
until converged
return $\mathbf{M}$

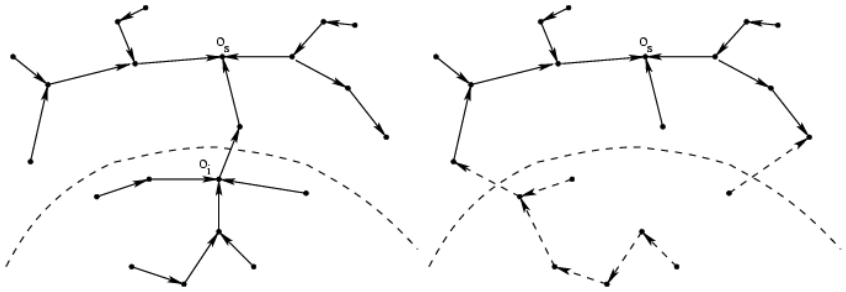
the costs of the resulting configuration are minimal. Therefore it is guaranteed that the costs, or energy respectively, decreases in each sweep and ICM will terminate after a finite number of cycles.

Critical for the running time is the update step. Here the recalculations of the dissimilarity matrix dominate the complexity due to the fact that this computational effort is necessary during each assignment update. Basically what has to be solved is an `ALL-PAIRS-SHORTEST-PATH` problem for each group of objects. For a full graph the algorithm of Floyd has a running time of $\mathcal{O}(n^3)$ [3]. If object o_i is updated, the ICM algorithm tries to find the minimum of the local costs by hypothetically assigning the given object to the various groups. Therefore, the effective dissimilarity to each cluster has to be determined once with object o_i inside and once with o_i outside the cluster. Thus $2k$ different effective dissimilarity matrices are needed. In the next paragraph an efficient implementation of this update step is presented.

Efficient Implementation of Update Step: One observes that the instances of the `ALL-PAIRS-SHORTEST-PATH` problems in two consecutive update steps are almost the same. They differ only in one single object: that object, which is to be updated. So it makes sense to check if one effective dissimilarity matrix can be used as a starting point for the calculation of an effective dissimilarity matrices in the next update step. Fortunately, those k dissimilarity matrices which correspond to the current configuration are used again in the next step. What about the other k matrices?

For $k-1$ of them a possible new configuration is given by adding object o_i, whereas for one cluster the new costs without o_i have to be computed. Consider the first case: object o_i is to be inserted in a certain group. For reasons of computational efficiency a complete recalculation of the effective dissimilarity matrix of that cluster is to be avoided. A closer look at Floyds algorithm leads the way: Its first step is given by the initialization of the matrix with the direct distances between the objects which in our case is given by the input dissimilarity

matrix. The goal is to get the shortest path distance between each pair of objects. For any object Floyds algorithm tries to put it on a path between each pair of objects. If now o_i is put into a new cluster only the last iteration of the Floyd algorithm has to be performed in which the current object is put on each of the existing paths between entity pairs. This step has a running time of $\mathcal{O}(n^2)$. In order to do so one has to get effective dissimilarities between o_i and all other objects in the considered cluster. Because of the symmetry of the original input dissimilarity matrix the effective dissimilarities are again symmetric. For that reason it is sufficient to solve one `SINGLE-SOURCE-SHORTEST-PATH` problem in order to arrive at the effective dissimilarities between o_i and all other objects. This can be solved with the Dijkstra's algorithm in a running time of $\mathcal{O}(n^2 \log n)$ [3]. So far we can compute the update step with an overall running time of $\mathcal{O}(n^2 \log n)$.



(a) (b)

Fig. 3. The graphs show the predecessor graph of object o_s before (a) and after (b) object o_i is removed from the cluster. For those objects where the path to o_s does not lead over o_i the effective dissimilarity may not change.

It remains to describe the necessary recalculations in the case where o_i is to be removed from the cluster it has been assigned to in the previous configuration. Again this goal has to be reached with the least possible computational effort. However in this situation the asymptotic running time of $\mathcal{O}(n^3)$ for a complete recalculation of dissimilarities for the whole group of objects can not be decreased. Nevertheless there is a good heuristic for lowering the running time. If there exists a shortest path from o_s to o_t which does not lead over o_i, the effective dissimilarity D_{st} will not change if o_i is removed from the cluster. One can obtain a predecessor matrix for all objects in a given cluster in the same running time as it takes to compute the effective dissimilarities. For each object we can thus determine the shortest path tree to all other objects in a running time of $\mathcal{O}(n)$. If we have the shortest path tree from o_s then only those dissimilarities between o_s and another object o_t have to be updated where t is an index out of the set of all objects in the subtree with root o_i (see figure 3). The total running time for an assignment update step is thus $\mathcal{O}(n^3)$.

Multi-scale optimization: As can be seen from the previous section the proposed optimization scheme is computationally highly demanding. An improvement in performance as well as solution quality is reached by using multi-scale techniques. The idea of multi-scale optimization is to lower the complexity by decreasing the number of considered entities in the object space. For example in image processing one can compute the given optimization tasks on a pyramid of different resolution levels of the image. At the first stage of such a procedure one solves the given computational problem on the subsampled image with lowest resolution, maps the result to the next finer resolution level and starts the optimization again. The expectation is that the result from the coarser level provides a reasonable starting point for the next finer level in the sense that convergence to a good solution can be reached within a few iterations. The probability of obtaining the global minimum is raised even if a local optimization scheme is used and the running time will dramatically decrease. A general overview of multi-scale optimization and its mathematically rigorous description is given in [15]. In order to pursue this approach a proper initialization of the coarse levels is needed. To this end three kinds of mappings have to be defined:

First of all, a function I^ℓ is needed, which maps the objects from the finer level ℓ to the next coarser one (level $\ell+1$). The multi-scale operator I^ℓ is defined as a function

$$I^\ell : \mathbb{O}^\ell \to \mathbb{O}^{\ell+1}, \tag{5}$$

where $\mathbb{O}^\ell$ is the set of objects on the ℓ^{th} resolution level and $\mathbb{O}^0$ is the set of objects on the finest level. So far this is just a formal definition. There is indeed a large design freedom in choosing the concrete form of this mapping. In the general case subsuming highly similar objects is reasonable. For our most prominent application task texture segmentation however we can use the natural topology of square neighborhoods of image sites in order to determine the fine to coarse mapping.

Second, the input dissimilarity matrix for the coarser level has to be defined in such a way that the basic modeling assumptions which underly the objective function are not violated. If the objects which map to one single entity in the coarser level belong to different clusters, the newly formed super–object in the coarser level should belong to one of these. Otherwise two completely different groups will get closer in the coarser level and the structure of the data is lost.

For that reason a mapping between each object in the coarse level and a representative object in the next finer level has to be defined as the third ingredient of the multi-scale approach. The function R^ℓ will denote the representative object in level ℓ for each object in level $\ell+1$.

$$R^\ell : \mathbb{O}^{\ell+1} \to \mathbb{O}^\ell \quad \text{with} \quad R^\ell(o_I) \in \left\{ o_i | I^\ell(o_i) = o_I \right\} \tag{6}$$

One possibility of defining such a representative is to choose the object nearest to the center of mass of the set. We are now able to define the input dissimilarity matrix $\mathbf{D}^{\ell+1}$ on the $(\ell+1)^{\text{th}}$ level.

$$\mathbf{D}^{\ell+1}(I, J) = \mathbf{D}^\ell(R^\ell(o_I), R^\ell(o_J)) \tag{7}$$

Having assembled all the necessary parts which constitute the resolution pyramid, the optimization can proceed as described in the beginning of this section. The intermediate coarse to fine mappings of computational results is achieved in a straightforward manner. A pictorial description of the multi-scale pyramid construction is given in figure 4.

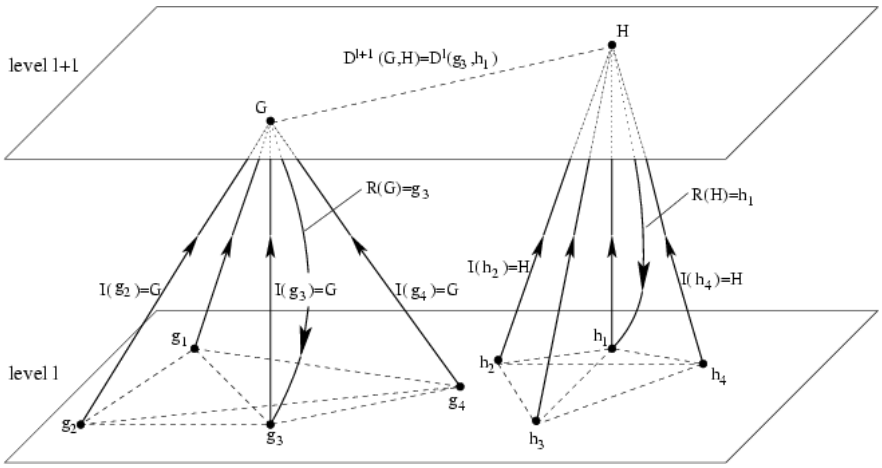


Fig. 4. Illustration of the multi-scale coarsening operators: I is mapping the objects on the finer level to the corresponding super–object in the coarser level, R is back-projecting to the representative in the finer level in order to enable the computation of the coarse level dissimilarity matrix.

4 Texture Segmentation by Pairwise Data Clustering

In order to pose the segmentation of images according to texture content as a pairwise data clustering problem, we follow the approach of Hofmann et al. [8]. A suitable image representation for texture segmentation is given by a multi-scale family of Gabor filters

$$G(\mathbf{x},\sigma,\mathbf{k}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mathbf{x}^t\mathbf{x}}{2\sigma^2} + i\mathbf{k}^t\mathbf{x},\right) \tag{8}$$

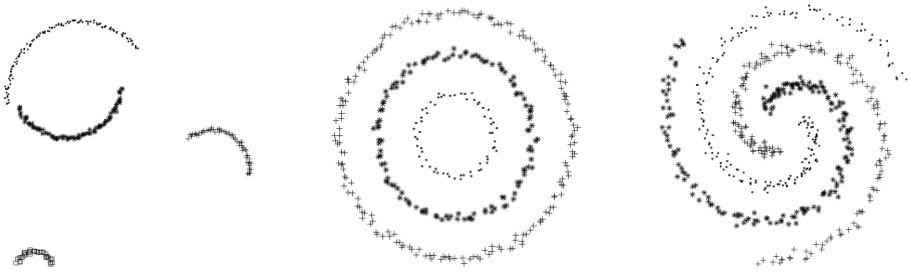
which are known to have good discriminatory power for a wide range of textures [8] [10]. In good agreement with psychophysical experiments [4] these Gabor filters extract feature information in the spatial correlation of the texture. The moduli of a bank of such filters at three different scales with octave spacing and four orientations are used as the representation of the input image. Thus the resulting twelve dimensional vector of modulus values $I(\mathbf{x})$ for each image location $\mathbf{x}$ comprises our basic texture features.

By its very nature, texture is a non local property. Although $I(\mathbf{x})$ contains information about spatial relations of pixels in the neighborhood of $\mathbf{x}$, it may not suffice to grasp the complete characterization of the prevalent texture. Therefore the image is covered with a regular grid of image sites. Suppose a suitable binning $t^* = 0 < t_1 < \ldots < t_L$ is given. For each site $i, i = 1 \ldots N$ the empirical feature distribution $f_i^{(r)}$ is computed for all Gabor channels r in order to arrive at a texture description for these spatially extend image patches. The dissimilarity between textures at location i and j is then computed independently for each of the channels by a χ^2 statistic.

$$D_{ij}^{(r)} = \chi^2 = \sum_{k=1}^{L} \frac{(f_i^{(r)}(t_k) - \hat{f}^{(r)}(t_k))^2}{\hat{f}^{(r)}(t_k)},$$
$$\text{where } \hat{f}^{(r)}(t_k) = [f_i^{(r)}(t_k) + f_j^{(r)}(t_k)]/2. \tag{9}$$

In order to combine the different values $D_{ij}^{(r)}$ into one dissimilarity value for each pair of objects the L_1–norm is used: $D_{ij} = \sum_r D_{ij}^{(r)}$. As a comparative study shows, this norm outperforms all norms in the L_p family [18].

5 Results

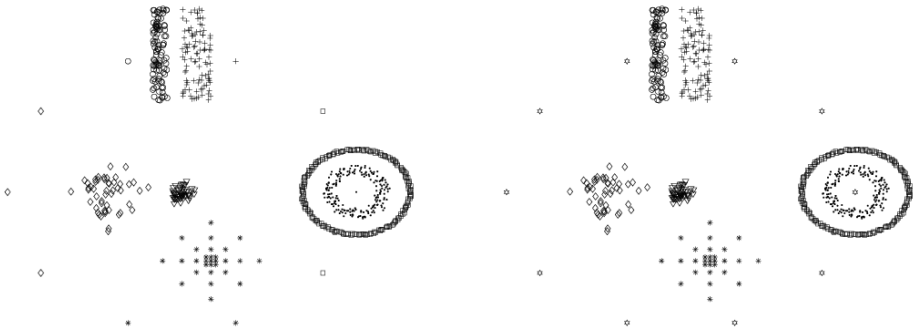


(a) (b) (c)

Fig. 5. Results of the novel path based clustering criterion on three toy data sets: a) arcs b) circular structure c) spiral arms

Artificial Data Sets: In section 2 some drawbacks of the formerly developed clustering method were addressed. As the novel path based approach was especially designed to cure such deficits, we first demonstrate its performance on some artificially generated data sets which would pose challenging problems for the conventional method. In figure 5 the results for three different toy data sets are depicted. Evidently the elongated structures apparent in these data sets are grouped in an intuitively appealing manner by our new algorithm.

Recently another interesting paradigm for pairwise data clustering has been proposed by Fred [5]. In this agglomerative procedure clusters are combined if the dissimilarity increment between neighboring objects is sufficiently small



(a) (b)

Fig. 6. Comparison between path based pairwise clustering (a) and the most competitive agglomerative procedure (b)

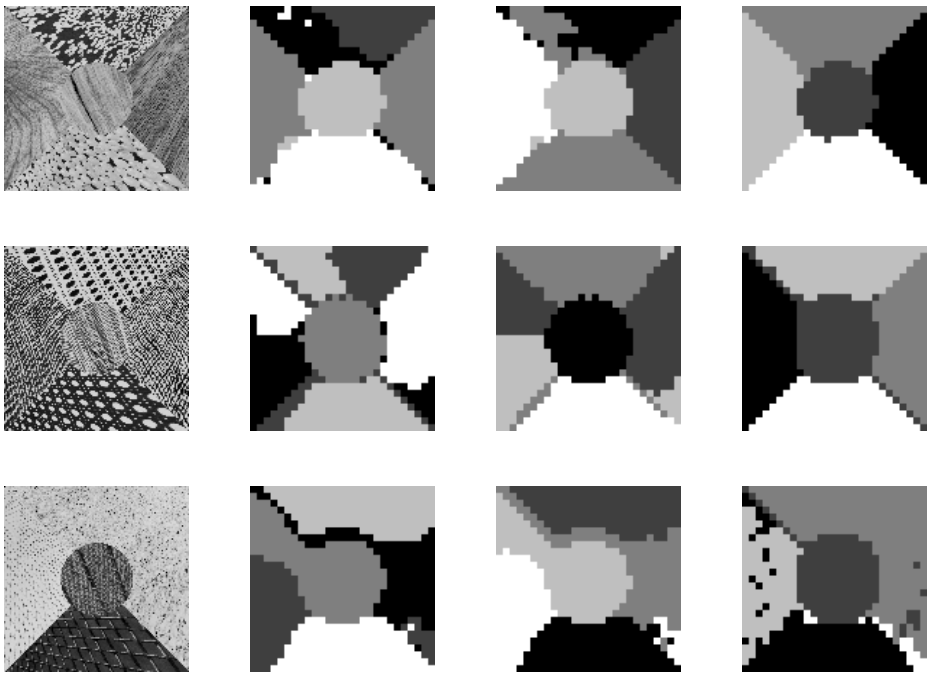
according to some smoothness requirement. Our and Fred's results are depicted in figure 6. In contrast to the approach in [5] the outer ring like structure of this data set does not constitute a cluster in the sense of our objective function and we have, therefore, inferred a solution with seven groups. Apart from this deviation, our method is able to match the competitors performance.

Texture Segmentation: A core application domain for pairwise data clustering techniques is the unsupervised segmentation of textured images. In order to obtain test images with known ground truth a set of mixtures of five textures each, so called Mondrians, has been constructed on the basis of micro textures from the Brodatz album.

(a) (b)

Fig. 7. Multidimensional scaling visualization of texture histograms illustrating the texture drift phenomenon: a) frontal b) tilted view on five textures

Before we come to discuss our results a word about the motivation for path based clustering for texture segmentation is in order. Textures in real world images often exhibit a gradient or drift of the describing features due to perspective distortions. This lack of translational and rotational invariance has been recognized early by Gibson [6]. To our knowledge, Lee et al. [14] were the first to address this problem as an important factor in designing models for texture segmentation. The issue is illustrated by figure 7. Here the texture histograms have been treated as vectors. In order to visualize their structural properties the dimensionality reduction technique *multi–dimensional scaling* (MDS) has been applied to construct a low dimensional embedding of the given vectors while faithfully preserving the inter vector distances. Figure 7 has been generated by using a deterministic annealing implementation of MDS as described in [12]. The left figure shows the case of a frontal view on five exemplary textures whereas the right one depicts the histogram embedding of the same textures when tilting the viewing angle. Clearly, the distorted textures form elongated structures in the feature space, whereas the non-inclined ones are characterized by compact groups.

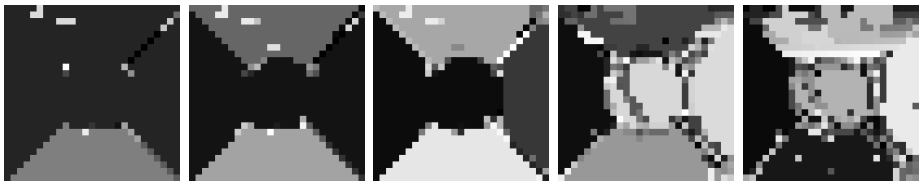


(a) (b) (c) (d)

Fig. 8. Segmentation results for texture Mondrians a) input image b) ACM c) PWC d) Path Based PWC

In order to give an impression of the performance of the novel path based clustering approach three typical grouping solutions on mixtures of tilted Brodatz textures are shown in figure 8. For comparison the results of the conventional pairwise clustering approach (PWC) and another recently proposed and highly competitive histogram clustering method known as *Asymmetric Clustering Model* (ACM) [16] are also shown. All of these results were reached by multi-scale techniques. In this context it is interesting to shed some light on the different topologies in the object and spatial domain. Whereas the novel clustering algorithm presumes a certain topological structure in the object realm, namely that of elongated data sources, the spatial relations of two–dimensional images yield another interesting starting point in terms of the multi-scale coarsening strategy. Evidently spatially neighboring sites on the image grid are likely to belong to the same texture category. Therefore combining four by four regions of image sites can be used for defining the multi-scale pyramid in the case of texture segmentation.

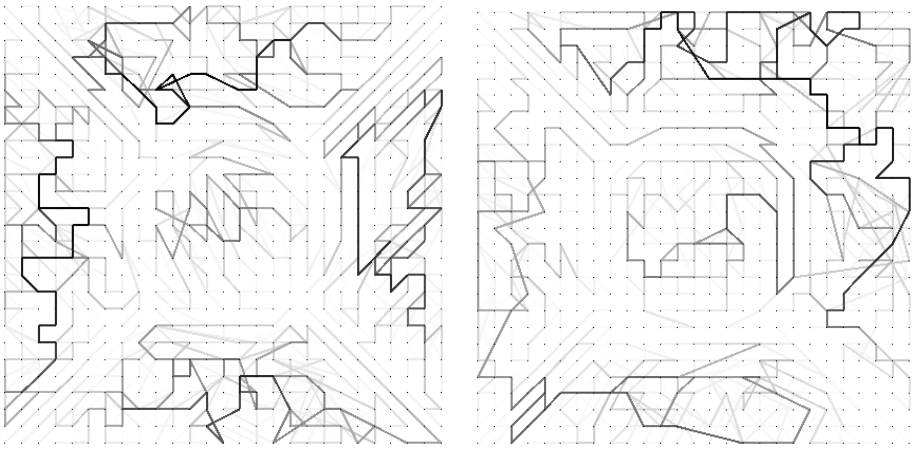
As can be seen in figure 8, path based pairwise data clustering clearly outperforms its competitors on this testbed. The results show that the other algorithms tend to split perceptually homogenous regions due to the fact that they cannot handle texture gradients properly. Moreover the image regions in which an edge between adjacent textures occurs are notoriously difficult to handle. The mixed statistics of such image parts do not pose so much of a difficulty for our novel algorithm because often there are links in terms of mediating paths to textures on either side of the edge. Thus the region of concern will then be adjoined to one of these instead of being considered as a group in its own right. However in some cases (c.f. the second row in figure 8) even the novel method fails to achieve the expected results. Thus the problem of mixed statistics can not be considered completely resolved. Another interesting example is given by the last row of figure 8. Here the same texture has been used twice, once in the upper part and once on the right side of the Mondrian. Our method groups these two regions together thereby recognizing the textural similarity whereas the competing approaches separate this texture in different segments.



(a) (b) (c) (d) (e)

Fig. 9. Segmentation results for the texture Mondrian from example 8(a) with Minimum Spanning Tree Clustering using a) 15 cluster b) 26 c) 42 d) 145 and e) 210 cluster

Path Based Clustering is related, but not identical to the agglomerative minimum spanning tree clustering algorithm (MST) (c.f. [9]). If outliers are present far away from all clusters, MST will put them in single clusters, whereas PBC assigns them to the nearest cluster. Figure 9 shows some results of MST applied to texture segmentation. The agglomerative algorithm has been stopped at different levels. The result with 26 clusters, for instance, contains only 3 groups with more than 3 elements. The result with 145 clusters is the first to distinguish the 5 different texture segments. However this solution suffers from a large amount of noise near the texture boundaries.

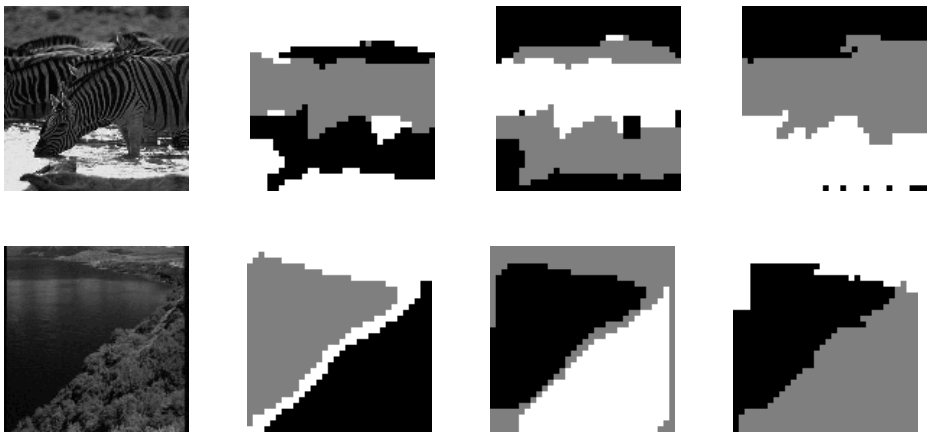


(a) (b)

Fig. 10. Visualization of frequencies with which a given edge is lying on an optimal path between some objects. a) example one, b) example 3 of figure 8

Another interesting insight in our novel approach is given by looking at the frequencies with which an edge between objects is lying on an optimal path between any two other objects. Such a visualization for the first and last example of figure 8 is given in figure 10. Here the objects are given by the image sites lying on a homogeneous grid. The darker the depicted edge, the more often this particular link has appeared on an optimal path. In the case of well separated groups (example 1 of figure 8) the visible edges are all in the interior of the five segments. On the other hand there is a number of frequently used edges forming a chain which traverses the border between Mondrian segments in the case of the merged textures (example three in figure 8).

Apart from artificially created scenarios the ultimate test of a texture segmentation method are real-world images. In this case texture drift occurs as a natural consequence of perspective distortion. Here some results on photographic images taken from the COREL photo gallery are shown in figure 11.



(a) (b) (c) (d)

Fig. 11. Segmentation results for real world images: a) input image b) ACM c) PWC d) Path Based PWC

Again the path based approach to pairwise data grouping performs best. Furthermore the problem of the competing algorithms with image regions on texture edges becomes apparent again. Our novel method yields satisfactory results not introducing mixed statistics groups.

6 Conclusion

In this contribution a novel algorithm for pairwise data clustering has been proposed. It enhances the conventional approach by redefining inter object dissimilarities on the basis of mediating paths between those entities which belong to the same group or cluster. Moreover an efficient multi-scale optimization scheme for the new clustering approach has been developed. The ability of path based pairwise data clustering to generate high quality grouping solutions has been demonstrated on artificially created data sets as well as for real world applications.

However we consider the new grouping criterion to be work in progress. First of all, the current technique for reducing the number of actually considered dissimilarity values is given by simple regular subsampling. Better alternatives in the sense of Gibbs sampling or methods based on the histogram of dissimilarities should be developed. Furthermore, better stochastic optimization methods for our novel clustering approach like simulated annealing have to be formulated and embedded in the multi-scale framework.

Acknowledgement: The authors are grateful to A. Fred for providing the data set in fig. 6.

References

1. J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48:25–37, 1986. 3, 5
2. M. Blatt, M. Wiesman, and E. Domany. Data clustering using a model granular magnet. *Neural Computation*, 9:1805–1842, 1997. 2
3. T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. MIT Press, 1989. 6, 7
4. I. Fogel and D. Sagi. Gabor filters as texture discriminators. *Biological Cybernetics*, 61:103–113, 1989. 9
5. A. L. N. Fred and J. M. N. Leitão. Clustering under a hypothesis of smooth dissimilarity increment. In *Proc. of the 15th International Conference on Pattern Recognition*, volume 2, pages 190–194. IEEE Computer Society, 2000. 10, 11
6. J. J. Gibson. *The Ethological Approach to Visual Perception*. Houghton Mifflin, 1979. 12
7. T. Hofmann and J. M. Buhmann. Pairwise data clustering by deterministic annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(1):1–14, 1997. 3
8. T. Hofmann, J. Puzicha, and J. M. Buhmann. Unsupervised texture segmentation in a deterministic annealing framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):803–818, 1998. 3, 9
9. A. Jain and R. Dubes. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ 07632, 1988. 14
10. A. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991. 9
11. S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. 3
12. H. Klock and J. M. Buhmann. Data visualization by multidimensional scaling: A deterministic annealing approach. *Pattern Recognition*, 33(4):651–669, 1999. partially published in EMMCVPR '97. 12
13. G. Lance and W. Williams. A general theory of classification sorting strategies: Ii. clustering systems. *Computer Journal*, 10:271 – 277, 1969. 1
14. T. L. Lee, D. Mumford, and A. Yuille. Texture segmentation by minimizing vector–valued energy functionals : The coupled–membrane model. In *Proceedings of the European Conference on Computer Vision (ECCV'92)*, volume 588 of *LNCS*, pages 165–173. Springer Verlag, 1992. 12
15. J. Puzicha and J. M. Buhmann. Multiscale annealing for unsupervised image segmentation. *Computer Vision and Image Underst.*, 76(3):213–230, 1999. 8
16. J. Puzicha, T. Hofmann, and J. M. Buhmann. Histogram clustering for unsupervised segmentation and image retrieval. *Pattern Recognition Letters*, 20:899–909, 1999. 13
17. J. Puzicha, T. Hofmann, and J. M. Buhmann. A theory of proximity based clustering: Structure detection by optimization. *Pattern Recognition*, 2000. 2, 3
18. J. Puzicha, Y. Rubner, C. Tomasi, and J. M. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. In *Proceedings of the International Conference on Computer Vision (ICCV) '99*, pages 1165–1173, 1999. 10
19. K. Rose, E. Gurewitz, and G. Fox. A deterministic annealing approach to clustering. *Pattern Recognition Letters*, 11:589–594, 1990. 3
20. N. Tishby and N. Slonim. Data clustering by markovian relaxation and the information bottleneck method. In *Advances in Neural Information Processing Sytems*, volume 13. NIPS, 2001. to appear. 2