

# AN APPLICATION OF ANT COLONY OPTIMIZATION TO IMAGE CLUSTERING

*Tomas Piatrik and Ebroul Izquierdo*

Multimedia & Vision Research Group, Queen Mary University of London

## ABSTRACT

Content-based image retrieval can be dramatically improved by providing a good initial clustering of visual data. The problem of image clustering is that most current algorithms are not able to identify individual clusters that exist in different feature subspaces. In this paper, we propose a novel approach for subspace clustering based on Ant Colony Optimization and its learning mechanism. The proposed algorithm breaks the assumption that all of the clusters in a dataset are found in the same set of dimensions by assigning weights to features according to the local correlations of data along each dimension. Experiment results on real image datasets show the need for feature selection in clustering and the benefits of selecting features locally.

## 1. INTRODUCTION

In the past decade, there has been a rapid growth in the consumption of digital media, specifically, video and images. As the use of digital media increases, effective retrieval and management techniques become more important. A fundamental step towards content based image annotation and retrieval is classification. The traditional supervised classification techniques assume prior knowledge to be available for all the thematic classes that are present in the considered dataset. On the contrary, unsupervised classification does not require predefined knowledge about the data. The main task of classification with unsupervised learning, commonly known as *clustering*, is to partition a given data set into groups (clusters) such that the data points in a cluster are more similar to each other than points in different clusters [1]. The clustering of images to semantic meaningful classes involves many image attributes including a particular combination of colour, texture or shape features. The problem is that most current image clustering algorithms typically consider all features, or dimensions, of the data in an attempt to learn as much as possible about the objects. However, different low-level image descriptors and similarity measures are not designed to be combined naturally and straightforwardly in a meaningful way. The low-level descriptors used in this work have different scales in different dimensions and they

are based on specific and different visual cues, representing various aspects of the content. The aim for the machine is to learn associations between complex combinations of low-level features and semantic concepts. In the literature, the problem of the selection of important features is solved by *feature selection* techniques [2]. These techniques derive a set of important features for the classification and clustering analysis, however, they can only find one relevant feature subset for all clusters. Therefore, instead of examining the dataset as a whole, subspace clustering algorithms localize their search and are able to uncover clusters that exist in multiple, possibly overlapping subspaces.

In this paper we propose a novel feature selection approach for clustering based on Ant Colony Optimization (ACO). Proposed method clusters the data and assigns weights to features according to the local correlations of data along each dimension. In the next section we present a background on the proposed clustering technique. The evaluation using two image datasets are presented and some conclusions are discussed in Section 3.

## 2. CLUSTERING USING ANTS

Recent developments in the heuristic optimization techniques are concentrated on natural and biological systems. The ACO is one of the most popular optimization techniques in the area of swarm intelligence [3]. The real power of ants resides in their colony brain and pheromone-driven communication within the colony.

In our proposal, The ACO model plays its part in assigning both images and feature weights to a cluster and each ant is giving its own clustering solution [4]. Each ant will assign each image  $x_i$ ,  $1 \leq i \leq n$  to the cluster  $\pi_u$ ,  $1 \leq u \leq k$ , with the probability  $P_{(x_i, \pi_u)}$ . For each centroid  $c_u$ , and for each feature  $F_l$ , new feature weights are computed according to intra-cluster similarities. For each image  $x_i$ ,  $1 \leq i \leq n$ , new generalized centroids are recomputed and the assigned pheromone to each solution is incremented according to quality of the clustering. A widely adopted definition of optimal clustering is a partitioning that the intra cluster similarity is minimized while the inter cluster similarity is maximized. The algorithm stops when all ants achieve the same clustering results.

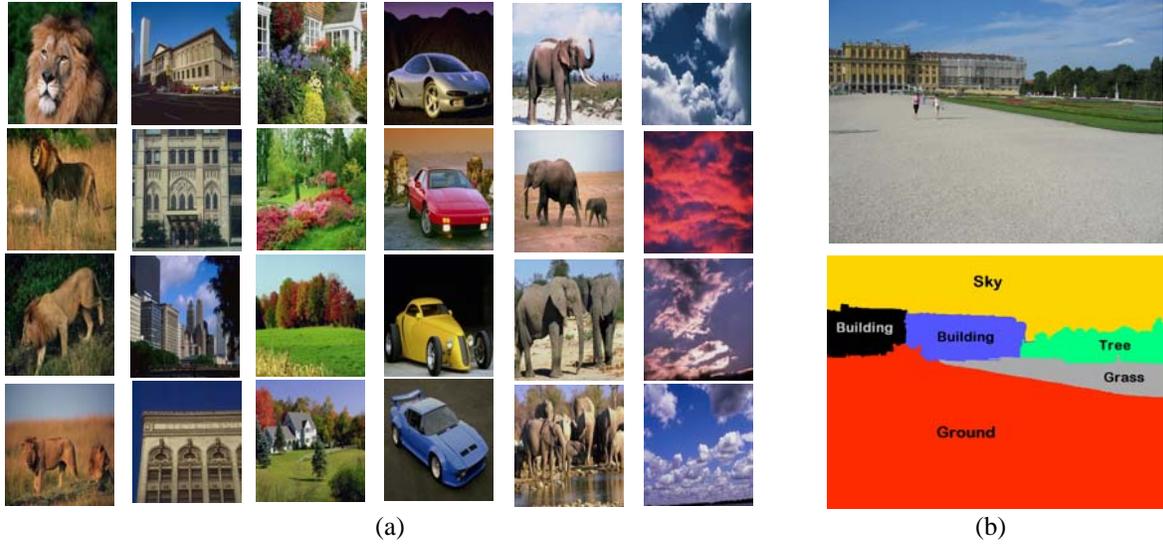


Fig. 1. Several representative images for different categories taken from the a) Corel Image dataset, b) Flickr dataset

### 3. EXPERIMENTAL EVALUATION AND CONCLUSION

In this section we evaluate the proposed algorithm using real-world image datasets. We compare our proposed subspace clustering algorithm based on ACO (denoted by SC-ACO) with PROCLUS, and global feature selection based on K-Means (denoted by GFS-K-Means), and K-Means without feature selection. For visual representation of images, following low-level features (descriptors) are used in this paper: Colour Layout (CLD), Colour Structure (CSD), Dominant Colour (DCD), Edge Histogram (EHD) and Grey Level Co-occurrence Matrix (GLC). First dataset was obtained from The Corel Image database and includes 600 images divided to 6 categories, each consist of 100 images. Second dataset consist of 500 images taken from Flickr which are segmented into regions and manually annotated [5].

Method	Average Error Rates	
	Corel set	Flickr set
SC-ACO	0.35±0.02	0.25±0.03
PROCLUS	0.37±0.05	0.3±0.06
GFS-K-Means	0.47±0.07	0.45±0.07
K-Means	0.52±0.09	0.41±0.08

Table 1. Average error rates for clustering results

Representative samples of images from both datasets are depicted in Figure 1. From experimental results in Table 1 can be seen that our proposed algorithm performs the best for both datasets. All tested methods depend on initialization of centroids what causes unstable clustering. From experimental results on both synthetic and real datasets can be seen that the ACO makes the clustering

algorithm less dependent on the initial parameters; hence it makes it more stable. Furthermore, our experiment results on real-world image datasets show the need for local feature selection. In our future work we will investigate the issue of automatically finding the optimal number of clusters.

### 4. ACKNOWLEDGMENT

The research leading to this paper was supported by the European Commission under contract FP6-027026, Knowledge Space semantic inference for automatic annotation and retrieval of multimedia content – K-Space.

### 5. REFERENCES

- [1] Xu, R., and Wunch, D. 2005. Survey of Clustering Algorithms. IEEE Trans. Neural Network, Vol.6, No.3. 645-678
- [2] Molina, L. C., Belanche, L., and Nebot, A. 2002. Feature Selection Algorithms: A Survey and Experimental Evaluation. Technical Report LSI-02-62-R (Universitat Politècnica de Catalunya, Barcelona, Spain, 2002)
- [3] Colomi, A., Dorigo, M., Maniezzo, V. 1991. Distributed optimization by ant colonies. In Proceedings of ECAL'91 European Conference on Artificial Life (Amsterdam, Netherlands 1991). Elsevier Publishing, 134-142
- [4] Piatrik, T. and Izquierdo, E. 2006. Image Classification using an Ant Colony Optimization approach. In Proceedings of 1st International Conference on Semantic and Digital Media Technologies (Athens, Greece, December 6-8, 2006)
- [5] Papadopoulos, T. G., Chandramouli, K., Mezaris, V., Kompatsiaris, I., Izquierdo, E., and Strintzis, M. G. 2008 A Comparative Study of Classification Techniques for Knowledge-Assisted Image Analysis. In Proc. 9th International Workshop on Image Analysis for Multimedia Interactive services.