Applied Soft Computing 62 (2018) 667–701

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Novel hybrid object-based non-parametric clustering approach for grouping similar objects in specific visual domains

Kaya Kuru[a,*], Wasiq Khan[b]

[a] School of Engineering, University of Central Lancashire, Preston PR1 2HE, UK
[b] School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester M15 6BH, UK

ARTICLE INFO

Article history:
Received 30 October 2015
Received in revised form 25 October 2017
Accepted 5 November 2017
Available online 10 November 2017

Keywords:
Cluster analysis
Visual object detection
Feature extraction
Image processing
Pairwise similarity comparison
Non-parametric clustering

ABSTRACT

Current widely employed clustering approaches may not yield satisfactory results with regard to the characteristics and distribution of datasets and number of clusters to be sought, especially for visual domains in multidimensional space. This study establishes a novel clustering methodology using a pairwise similarity matrix, *Clustering Visual Objects in Pairwise Similarity Matrix (CVOIPSM)*, for grouping similar objects in specific visual domains. A dimensionality reduction and feature extraction technique, along with a distance measuring method and a newly established algorithm, *Clustering in Pairwise Similarity Matrix (CIPSM)*, are combined to develop the CVOIPSM methodology. CIPSM utilizes both *Rk-means* and an *agglomerative, contractible, expandable* (ACE) technique to calculate a membership degree based on maximizing inter-class similarity and minimizing intra-class similarity. CVOIPSM has been tested on several datasets, with average success rates on downsized subsamples between 87.5% and 97.75% and between 81% and 87% on the larger datasets. The difference in the success rates for small and large datasets is not statistically significant ($p > 0.01$). Moreover, this method automatically determines the likely number of clusters without any user dictation. The empirical results and the statistical significance test on these results ensure that CVOIPSM performs effectively and efficiently on specific visual domains, disclosing the interrelated patterns of similarities among objects.

1. Introduction

Cluster analysis divides a set of nearby objects into categories in which objects are more homogeneous than in other categories, based on their features and intrinsic similarities (i.e., without referring to pre-trained datasets). Clustering is of increasing importance owing to the quickly growing number of large databases in every discipline, most of which are in multidimensional space and beyond human comprehension. Clustering algorithms are applied in a wide variety of domains and in a remarkable number of different disciplines – such as astronomy, medicine and genetics, biology and zoology, marketing, and geography – that require similarities to be distinguished from dissimilarities. Furthermore, we live in a digital world in which the number of visual datasets is increasing logarithmically and thus, the demand for better clustering techniques specific to visual datasets has risen considerably. Researchers are often more interested in grouping visual datasets based on the specific objects they contain, instead of on whole images. Therefore, new methodologies should be established to handle the particularities of working with rapidly growing datasets, i.e., first, clustering a set of objects into the desired number of groups without supervised dictation and second, correctly accommodating objects in these groups. Our methodology is presented in phases and its merits are demonstrated for sample visual objects belonging to different databases whose characteristics differ in several respects in order to quantify the results.

Broadly, our method combines conceptual understanding of a novel hybrid methodology using a dimensionality reduction technique for feature extraction and selection and a distance measuring technique for calculating the similarities among objects are merged with a new method, *Clustering in Pairwise Similarity Matrix (CIPSM)*, for unsupervised partitioning of objects. The notation used throughout the paper for the proposed hybrid methodology is *Clustering Visual Objects in Pairwise Similarity Matrix (CVOIPSM)*. The methodology is presented, along with a general framework and a roadmap for clustering objects in visual domains, and its implementation is separated into two components: the first acquires features and the distances/similarities between these features and the second clusters objects using these similarities. A variety of

* Corresponding author at: Fylde Rd, *CT* Building, CM 225, School of Engineering, Preston, Lancashire PR1 2HE, UK.
E-mail address: kkuru@uclan.ac.uk (K. Kuru).

https://doi.org/10.1016/j.asoc.2017.11.007
1568-4946/

techniques is required for good object detection, feature selection, and distance measurement among objects in different visual datasets. Thus, these steps are separated from the clustering component to allow users to cluster datasets whose features and distances are acquired using different methods. Therefore, the user can choose other techniques for the first part based on the characteristics of the visual domain being studied. Moreover, the clustering component, CIPSM, is general and can be performed alone for unsupervised clustering of not only visual datasets, but also any dataset. This methodology is tested on four datasets from three databases. Clustering success is evaluated using an external quality measure technique that compares the clusters produced by our methodology to known classes already in the related dataset domains.

The structure of this paper is as follows: Section 2 discusses related studies, Section 3 unveils the conceptual framework of the proposed methodology, Section 4 presents the datasets and the study design for evaluating the proposed methodology, the results are explained in Section 5, and finally, Sections 6 and 7 discuss and conclude the findings in this study.

2. Related literature

In this section, we briefly analyze related subjects to provide an overview of the general concepts and techniques that have so far been deployed in image and object analysis. Our evaluation of related studies formed the basis for our study.

2.1. Object detection in images

The specific objects or regions of interest (ROI) in an image have to be roughly distinguished from other objects and acquired in order to perform object-based clustering. Several techniques have been used to segment specific objects in images. JSEG uses color quantization and spatial segmentation to divide images into regions; a combination of color and texture features is widely used for image segmentation and discerning different regions in an image [1]. This technique is successfully employed for general texture-based image clustering. However, it is less successful in detecting specific visual objects in images, and consequently, for cluster analysis based on specific objects in images.

To this end, object-based image retrieval is employed for acquiring specific regions from images; examples of this technique can be found in many studies, such as [2–4]. Some of the widely used automatic object detection techniques are template matching, scale-invariant feature transform (SIFT), speeded-up robust features (SURF), features from accelerated segment test (FAST), binary robust independent elementary features (BRIEF), oriented FAST and rotated BRIEF (ORB), maximally stable extremal regions (MSER), binary robust invariant scalable keypoints (BRISK). Deep convolutional networks are being used for object detection as well [5]. Most of these techniques give a similarity value regarding the specified number of most important keypoints and this value is utilized to decide if there is a similarity between two objects given a threshold value. In addition, a pairwise similarity table can be established using the similarity values acquired from these techniques for visual domains as well. Haar Cascade files are most commonly employed for detecting an object in an image. A Haar Cascade file can be trained on a few hundred samples of a particular object (e.g., a leaf or a face) – which is a time consuming process – or pre-trained Haar Cascade files in the public domain can be deployed. In particular, using Haar Cascade files makes it easier to detect more than one similar object in an image easily and quickly. In most of these object detection techniques, a reference object is compared to the objects in images to find similar objects using a resemblance threshold value. Likewise, we implement region-based object detection in our study, as explained in Section 3, to discern the objects in images. More specifically, we employ Haar Cascade files where possible, a Harris algorithm and local similar neighborhood points to detect the objects in our visual domains, as explained in Section 3.1.1. Cropping objects in images for further analysis after designating their ROIs using automatic object detection can be implemented, with or without background removal, as explained in Section 3.1.1.

2.2. Feature extraction and selection, and distance measurement

Some of the feature extraction and selection techniques for visual objects are kernel PCA, independent component analysis, probability density estimation, local feature analysis, elastic graph matching (EGM), multi-linear analysis, kernel discriminant analysis, Gabor wavelet (GW), Fisher's linear discriminant analysis (FLDA), and support vector machines. In particular, EGM, FLDA, GW, and PCA have been widely employed to extract features from an object in an image region [6]. The high accuracy of these methods for extracting features and subsequently revealing patterns has been demonstrated in many studies. For instance, these methods can perform feature extraction and selection with accuracy rates of up to 96% [6]. With PCA, good success rates can be obtained by detecting patterns from images captured in ideal environments particularly with good illumination – or by employing several image processing techniques to enhance images before feature extraction. In addition, it is computationally efficient compared with other similar methods [7] because dimensionality reduction can be performed easily to accelerate the calculations. Thus, we employ PCA to extract the most important features from visual objects, because of its extensive and successful application to many datasets. In other words, we use PCA for dimensionality reduction by removing less important information (e.g., noise and redundant datasets) before applying our clustering approach, CIPSM. Interested readers are referred to our previous study for more information about how to implement PCA [6].

Some of the well-known approaches for measuring the distance between two points in a features dataset are Bayesian decision theory, multiple similarity, city block, subspace, angular separation, Pearson correlation, and Mahalanobis, Minkowski, Canberra, Chebychev, and Euclidean distances [8]. The Mahalanobis- and Euclidean-based distance measurement techniques are the most widely used of these approaches [8]. We tested these two methods on the features in our visual datasets to determine the better one to use, and found that the Euclidean-based technique outperformed the Mahalanobis one. Thus, this matching technique was selected for our study. Interested readers can find more information about measuring the distance between the features of two visual objects in Calva's study [9].

2.3. Cluster analysis

Cluster analysis is a very broad and wide-ranging field that cannot be covered completely in this section. Therefore, we provide a brief summary of the general concepts of current clustering schemes (especially their shortcomings with regard to cluster analysis of visual objects). Several well-known clustering algorithms are examined in-depth by Bishop [10], Witten [11] and Everitt [12] in their books, as well as in other journal publications. In general terms, centroid-based (e.g., k-means, XMeans, FarthestFirst, FilteredClusterer and c-means), hierarchical (e.g., cobweb, CLINK, SLINK, CURE, and BIRCH), incremental, density-based (e.g., DBSCAN and OPTICS), and probability-based (e.g., EM and Bayesian clustering) schemes predominate clustering methods.

The performance of centroid-based clustering schemes depends on the number of clusters specified [13], initial centroid selection,