A Survey on Image Data Analysis through Clustering Techniques for Real World Applications

Seema Wazarkar, Bettahally N. Keshavamurthy

PII: S1047-3203(18)30172-X
DOI: https://doi.org/10.1016/j.jvcir.2018.07.009
Reference: YJVCI 2243

To appear in: *J. Vis. Commun. Image R.*

Received Date: 20 October 2017
Revised Date: 4 May 2018
Accepted Date: 14 July 2018

Please cite this article as: S. Wazarkar, B.N. Keshavamurthy, A Survey on Image Data Analysis through Clustering Techniques for Real World Applications, *J. Vis. Commun. Image R.* (2018), doi: https://doi.org/10.1016/j.jvcir.2018.07.009

A Survey on Image Data Analysis through Clustering Techniques for Real World Applications

Seema Wazarkar[a]*, Bettahally N. Keshavamurthy [b]

[a,b]*Department of Computer Science and Engineering, National Institute of Technology Goa, Farmagudi, Ponda-403401, Goa, India*

Abstract

A huge amount of image data is being collected in real world sectors. Image data analytics provides information about important facts and issues of a particular domain. But, it is challenging to handle voluminous, unstructured and unlabeled image collection. Clustering provides groups of homogeneous unlabeled data. Therefore, it is used quite often to access the interesting data easily and quickly. Image clustering is a process of partitioning image data into clusters on the basis of similarities. Whereas, features extracted from images are used for the computation of similarities among them. In this paper, significant feature extraction approaches and clustering methods applied on the image data from nine important applicative areas are reviewed. Medical, 3D imaging, oceanography, industrial automation, remote sensing, mobile phones, security and traffic control are considered applicative areas. Characteristics of images, suitable clustering approaches for each domain, challenges and future research directions for image clustering are discussed.

Keywords: Image Clustering; Feature Extraction; Real World Applications

1. Introduction

These days, a use of Internet is rapidly increasing pave to data explosion, which results into big data. It contains a large amount of image data. For the image data analysis, image processing [AmitKumar et al. 2016; Jain 1989] and data mining [Chen et al. 1996; Han et al. 2012] techniques play an important role. Many software tools [Mikut et al. 2011; Jovic et al. 2014] such as R [Torgo 2010], WEKA [Holmes et al. 1994; Hall et al. 2009], RapidMiner [Burget et al. 2010], SciKit [Berthold et al. 2008], KNIME [Pedregosa et al. 2011], SparkMLlib [Meng et al. 2015] are available to handle different types of data automatically. But, these tools are not sufficient for detailed analysis of a particular data type as it has limited number of data mining techniques [Mikut et al. 2011]. Hence, it may not be able to handle all the types of data with different characteristics. Mikut et al. said that generalized powerful tool for multidimensional data (like image and video) is not available. In 2014, Jovic et al. also studied several tools and concluded that there is no single best tool and each tool has their corresponding strength and limitations [Jovic et al. 2014]. While analysing the image data from large datasets, it is important to partition the given data. It helps to retrieve an important and interesting data easily and efficiently. Clustering is a very useful technique to accomplish this data partitioning task. As image clustering is an unsupervised learning based approach, it is capable to handle unlabelled image data. It also helps in reduction of less significant data from huge dataset. Whereas, it is mainly utilized for image segmentation and Content-Based Image Retrieval (CBIR).