Chapter 8
Image Mining:
Techniques for Feature Extraction

Tuğrul Taşci
Sakarya University, Turkey

ABSTRACT

In today's World, huge multi-media databases have become evident due to the fact that Internet usage has reached at a very-high level via various types of smart devices. Both willingness to come into prominence commercially and to increase the quality of services in leading areas such as education, health, security and transportation imply querying on those huge multi-media databases. It is clear that description-based querying is almost impossible on such a big unstructured data. Image mining has emerged to that end as a multi-disciplinary field of research which provides example-based querying on image databases. Image mining allows a wide variety of image retrieval and image matching applications intensely required for certain sectors including production, marketing, medicine and web publishing by combining the classical data mining techniques with the implementations of underlying fields such as computer vision, image processing, pattern recognition, machine learning and artificial intelligence.

INTRODUCTION

Aggregation of information is undoubtedly the key element inspring to the progress of humankind. Information has been increasing until today with an accelerated manner. The main reason for that is obviously the tendency of utilizing all kinds of information and experience regarding human life in order to deepen knowledge in the same areas or to explore novel ideas in various different areas. For instance, observing the notably silent flight of owl by the virtue of its wings nature can allow engineers to apply such a structure in aircraft design to enable more comfortable and noise-free landing and takeoffs. Also, storage and transfer of information to the new generations gained along the ages through the techonological equipments and methods of each particular age may be evaluated as the complementary factor for the formation of ever-increasing information process.

The rapid advancements of technology today provide an opportunity to acquire massive amount of data in diverse fields including natural and applied sciences, social and human sciences, life and health

DOI: 10.4018/978-1-5225-0983-7.ch008

sciences and various other fields related to daily life experiences. The becoming of internet, mobile connectivity and highly-portable digital image and video storage devices as an integral part of our daily life leads to permanently growing data in almost all fields of daily life. In this context, the convenient analysis of this data in order to obtain useful information for the purpose of facilitating human life may be fairly considered as a usual result of humankind evolution.

Data mining is a method of knowledge discovery with a rather broad definition. To become more specific, data mining can be defined as a process of extraction of understandable and purposive information and associations from massive amount of data in order to utilize in diversified applications. According to Zaki & Meira (2014), the emergence of data mining as a research field has allowed the analysis of all type of patterns and models to be performed with the applications ranging from scientific discovery to business intelligence and analytics. In their study, Linoff & Berry (2011) define data mining as a process of huge data investigation in order to discover significant patterns and rules. Data mining is quotted as one of the most important phases of a knowledge discovery process comprising data cleaning, data integration, data selection and transformation, data mining, pattern evaluation and knowledge representation in the work of Jiawei et.al. (2012).

In today's world, data related to different areas including scientific research, education, economy and demography is stored digitally in databases, especially in the developed countries. Obtaining useful information from those huge databases is a quite complicated problem and implies efficient methods to be applied. Within this context, data mining is a well-established field of research involving various powerful methods and offers availed solutions. However, images and videos accepted as non-standard forms of data have been intensively in use especially in the past decade. Multimedia databases have dramatically grown in association with widely available Internet based services, intensive use of digital recording mobile devices and the large size of images and videos because of high definition file formats. Data mining with traditional methods has become inadequate in handling multi-media data and extracting information efficiently. Thus, a new research field has been formed namely image mining which derives methods and techniques of data mining and other related fields such as image processing, pattern recognition and machine learning in order to discover information robustly from non-standard data sources containing high portion of data produced more particularly in recent past.

Joseph & Wilson (2014) define image mining as the process of information discovery on the image databases. The ultimate purpose in an image mining application is the retrieval of similar images and linking associated data semantically with an example image which can be visually queried from database. For instance, the information of whether the people living in different regions have similar diseases can be obtained by analyzing regional weather satellite images through an image mining system. Ever-increasing information obtained from images and videos implies existence of efficient methods allowing such a visual query to be performed. Hence in their work, Singhai & Shandilya (2010) highlight the claim that tendency to the multi-media retrieval sytems has been increasing collateraly with the rapidly increasing demand of accurate and fast content-based querying.

Image mining is investigated in this chapter by briefly revising the studies referring to the field's place in the literature and revealing the relationship between image and traditional data mining. A generic image mining process algorithm is given with the phases containing the principle operations. The existing applications of image mining are summarized by emphasizing their prominent attributes. A descriptive overview of content-based image retrieval systems is also introduced in terms of functional