

Chapter 4.24

Image Mining: A Case for Clustering Shoe Prints

Wei Sun

Monash University, Australia

David Taniar

Monash University, Australia

Torab Torabi

La Trobe University, Australia

ABSTRACT

Advances in image acquisition and storage technology have led to tremendous growth in very large and detailed image databases. These images, once analysed, can reveal useful information to our uses. The focus for image mining in this article is clustering of shoe prints. This study leads to the work in forensic data mining. In this article, we cluster selected shoe prints using k-means and expectation maximisation (EM). We analyse and compare the results of these two algorithms.

INTRODUCTION

With the improvement of computer technology, such as multimedia data acquisition and storage

techniques, the application of multimedia information becomes more and more prevalent. Therefore, the demand for discovering patterns from a great deal of multimedia data is becoming more relevant in many applications (Ordonez & Omiecinski, 1998; Zaine et al, 1998). In this article, we focus on clustering shoe prints. The work presented in this article is part of a larger project on forensic data mining focusing on multimedia data.

The main objectives of this work are (i) to cluster shoe prints, (ii) to analyse the results of each clustering algorithm, (iii) to use a visualisation tool to see how the clusters are affected by changes of input variables, and (iv) to examine the differences in the distributions of variables from cluster to cluster. Our experiments were conducted to cluster a series of shoe prints by using clustering algorithms in Weka.

PRELIMINARY

Figure 1 shows an overview of the shoe prints mining processes described in this article. The shoe prints from image database are first processed to extract the RGB information. With these RGB data, Weka is used to cluster the extract RGB values of the selected shoe prints. The results are then evaluated and interpreted to obtain the final knowledge, which can be applied to forensic applications.

Figure 2 shows some sample clustered shoe prints. Obviously, we can see there are some commonalities in each group of shoe prints. Actually, the contrast values of images in group two are the lowest. On the other hand, the contrast values in group three are the highest, containing of only black and white colour. In addition, the red colour in group four is very remarkable. Finally, the last group has no similarity with the other groups, so it is separated to another cluster. Each shoe print has its own characteristics; these are reflected from their colour, texture, contrast value, homogeneity, and so forth. To find their

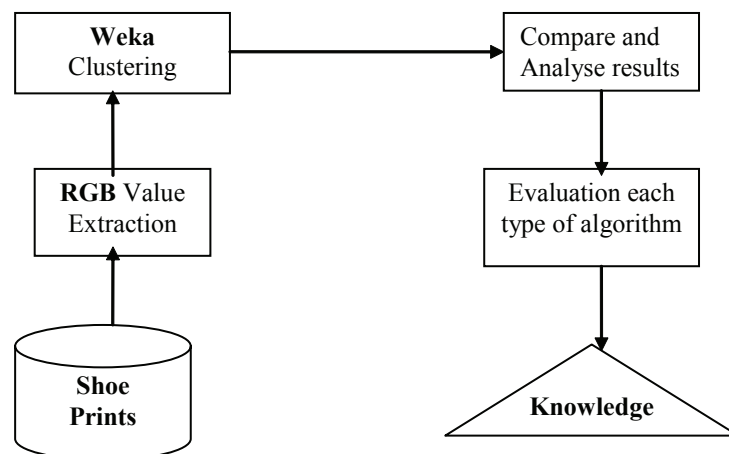
colour characteristics, our focus is to group them and analyse the results.

For the experiment, we will use Weka to cluster the shoe prints, where the sample shoe prints are the selected 29 images. To make the experiment more convincible, the RGB values of images chosen are quite close, so the images are not as identifiable as those in Figure 2.

Clustering Techniques

Image clustering is usually performed in the early stages of the mining process. Feature attributes that have received the most attention for clustering are colour, texture, and shape. Generally, any of the three, individually or in combination, could be used. There is a wealth of clustering algorithms available (Han & Kamber, 2006): density based clustering algorithms (e.g., Dbscan and Optics), partition-based algorithms (e.g., EM), mixture-resolving (e.g., Make Density Based Clusterer) and mode-seeking algorithms (e.g., Cobweb), and nearest neighbour clustering (e.g., k -means and Farthest First). Once the images have been

Figure 1. Shoe prints clustering process



14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/image-mining-case-clustering-shoe/7991

Related Content

Innovative Access and Query Schemes for Mobile Databases and Data Warehouses

Alfredo Cuzzocrea (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 860-871).

www.irma-international.org/chapter/innovative-access-query-schemes-mobile/20772

Improving Sequence Diagram Modeling Performance: A Technique Based on Chunking, Ordering, and Patterning

Thant Synand Dinesh Batra (2013). *Journal of Database Management* (pp. 1-25).

www.irma-international.org/article/improving-sequence-diagram-modeling-performance/100404

Databases Modeling of Engineering Information

Z.M. Ma (2009). *Selected Readings on Database Technologies and Applications* (pp. 65-84).

www.irma-international.org/chapter/databases-modeling-engineering-information/28546

Towards Real-Time Multi-Sensor Golf Swing Classification Using Deep CNNs

Libin Jiao, Hao Wu, Rongfang Bie, Anton Umekand Anton Kos (2018). *Journal of Database Management* (pp. 17-42).

www.irma-international.org/article/towards-real-time-multi-sensor-golf-swing-classification-using-deep-cnns/218925

An Emergent Model of End-users' Acceptance of Enterprise Resource Planning Systems: A Grounded Theory Approach

Fiona Fui-Hoon Nahand Xin Tan (2015). *Journal of Database Management* (pp. 44-66).

www.irma-international.org/article/an-emergent-model-of-end-users-acceptance-of-enterprise-resource-planning-systems/153517