

Chapter VII

Making Image Retrieval and Classification More Accurate Using Time Series and Learned Constraints

Chotirat “Ann” Ratanamahatana
Chulalongkorn University, Thailand

Eamonn Keogh
University of California, Riverside, USA

Vit Niennattrakul
Chulalongkorn University, Thailand

ABSTRACT

After the generation of multimedia data turning digital, an explosion of interest in their data storage, retrieval, and processing, has drastically increased in the database and data mining community. This includes videos, images, and handwriting, where we now have higher expectations in exploiting these data at hand. We argue however, that much of this work’s narrow focus on efficiency and scalability has come at the cost of usability and effectiveness. Typical manipulations are in some forms of video/image processing, which require fairly large amounts for storage and are computationally intensive. In this work, we will demonstrate how these multimedia data can be reduced to a more compact form, that is, time series representation, while preserving the features of interest, and can then be efficiently exploited in Content-Based Image Retrieval. We also introduce a general framework that learns a distance measure with arbitrary constraints on the warping path of the Dynamic Time Warping calculation. We demonstrate utilities of our approach on both classification and query retrieval tasks for time series and other types of multimedia data including images, video frames, and handwriting archives. In addition, we show that incorporating this framework into the relevance feedback system, a query refinement can be used to further improve the precision/recall by a wide margin.

1 INTRODUCTION

Much of the world's data is in the form of time series, and many other types of data, such as video, image, and handwriting, can also be trivially transformed into time series. Generally, we can use various image processing techniques (Deselaers, Keysers, & Ney, 2003; Käster, Wendt, & Sagerer, 2003; Krishnamachari & Abdel-Mottaleb, 1999; Wang, Yang, & Acharya, 1997; Yeung & Liu, 1995) to complete multimedia data mining tasks, by measuring similarities among the raw images, using certain features such as color, texture, or shape. However, time series representation of these data can significantly help speed up the process since they can be compared much easier and faster using distance measurements. This fact has fueled enormous interest in time series retrieval in the database and data mining community. We argue, however, that much of this work's narrow focus on efficiency and scalability has come at the cost of usability and effectiveness. For example, the lion's share of previous work has utilized the Euclidean distance metric, presumably because it is very amenable to indexing (Agrawal, Lin, Sawhney, & Shim, 1995; Chan, Fu, & Yu, 2003; Faloutsos, Ranganathan, & Manolopoulos, 1994). However, there is increasing evidence that the Euclidean metric's sensitivity to small differences in the time axis makes it unsuitable for most real world problems (Aach & Church, 2001; Bar-Joseph, Gerber, Gifford, Jaakkola, & Simon, 2002; Diez & González, 2000; Kadous, 1999; Schmill, Oates, & Cohen, 1999).

It has long been known that Dynamic Time Warping (DTW) is superior to Euclidean distance for classification and clustering of time series. However, until lately, most research has utilized Euclidean distance because it is more efficiently calculated. A recently introduced technique that greatly mitigates DTW demanding CPU time has sparked a flurry of research activities. However, the technique and its many extensions still only allow DTW to be applied to moderately large datasets. In addition, almost all of the research on DTW has focused exclusively on speeding up its calculation; there has been relatively little work done on approving its accuracy.

In this work, we target the accuracy aspect of correct classification and introduce a new framework that learns arbitrary constraints on the warping path of the DTW calculation. Apart from improving the accuracy of content-based image retrieval, our technique as a side effect speeds up DTW by many orders of magnitude as well. We show the utility of our approach on datasets from diverse domains and demonstrate significant gains in accuracy and efficiency. Moreover, some additional training or human intervention can also be further incorporated into the classic information retrieval technique of relevance feedback to achieve even much superior results.

1.1 The Ubiquity of Time Series Data

In this section, we wish to expand the readers' appreciation for the ubiquity of time series data. Rather than simply list the traditional application domains, i.e., stock market, weather data, electrocardiograms, etc., we will consider some less obvious applications of multimedia/image data that can benefit from efficient and effective retrieval.

Video retrieval: Video retrieval is one of the most important issues in multimedia database management systems. Generally, research on content-based video retrieval represents the content of the video as a set of image frames, leaving out the temporal features of frames in the shot. However, for some domains, including motion capture editing, gait analysis, and video surveillance, it may be fruitful to extract time series from the video, and index *just* the time series (with pointers back to the original video). Figure 1 shows an example of a video sequence that is transformed into a time series. There are

24 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/making-image-retrieval-classification-more/4155

Related Content

Moving Vehicles Detection in Traffic Video Using Modified SXCS-LBP Texture Descriptor

Arun Kumar H. D. and Prabhakar C. J. (2015). *International Journal of Computer Vision and Image Processing* (pp. 14-34).

www.irma-international.org/article/moving-vehicles-detection-in-traffic-video-using-modified-sxcs-lbp-texture-descriptor/159805

A Novel Algorithm for Segmentation of Parasites in Thin Blood Smears From Microscopy Using Type II Fuzzy Sets and Inverse Gaussian Gradient

Madhu Golla (2019). *International Journal of Computer Vision and Image Processing* (pp. 1-22).

www.irma-international.org/article/a-novel-algorithm-for-segmentation-of-parasites-in-thin-blood-smears-from-microscopy-using-type-ii-fuzzy-sets-and-inverse-gaussian-gradient/233491

Emotion Recognition Using Facial Expression

Santosh Kumar, Shubam Jaiswal, Rahul Kumar and Sanjay Kumar Singh (2018). *Computer Vision: Concepts, Methodologies, Tools, and Applications* (pp. 1768-1787).

www.irma-international.org/chapter/emotion-recognition-using-facial-expression/197024

A Deep Learning-Based Approach to Classification of Baby Sign Language Images

Sulochana Nadgeri and Arun Kumar (2022). *International Journal of Computer Vision and Image Processing* (pp. 1-18).

www.irma-international.org/article/a-deep-learning-based-approach-to-classification-of-baby-sign-language-images/283964

Intelligent Traffic Monitoring System through Auto and Manual Controlling using PC and Android Application

Paromita Roy, Nivedita Patra, Amartya Mukherjee, Amira S. Ashour, Nilanjan Dey and Satya Priya Biswas (2017). *Applied Video Processing in Surveillance and Monitoring Systems* (pp. 244-262).

www.irma-international.org/chapter/intelligent-traffic-monitoring-system-through-auto-and-manual-controlling-using-pc-and-android-application/168910