Robust Face Recognition for Data Mining

Brian C. Lovell
The University of Queensland, Australia

Shaokang Chen
NICTA, Australia

Ting Shan
NICTA, Australia

INTRODUCTION

While the technology for mining text documents in large databases could be said to be relatively mature, the same cannot be said for mining other important data types such as speech, music, images and video. Multimedia data mining attracts considerable attention from researchers, but multimedia data mining is still at the experimental stage (Hsu, Lee & Zhang, 2002). Nowadays, the most effective way to search multimedia archives is to search the metadata of the archive, which are normally labeled manually by humans. This is already uneconomic or, in an increasing number of application areas, quite impossible because these data are being collected much faster than any group of humans could meaningfully label them — and the pace is accelerating, forming a veritable explosion of non-text data. Some driver applications are emerging from heightened security demands in the 21st century, postproduction of digital interactive television, and the recent deployment of a planetary sensor network overlaid on the internet backbone.

BACKGROUND

Although they say a picture is worth a thousand words, computer scientists know that the ratio of information contained in images compared to text documents is often much greater than this. Providing text labels for image data is problematic because appropriate labeling is very dependent on the typical queries users will wish to perform, and the queries are difficult to anticipate at the time of labeling. For example, a simple image of a red ball would be best labeled as sports equipment, a toy, a red object, a round object, or even a sphere, depending on the nature of the query. Difficulties with text metadata have led researchers to concentrate on techniques from the fields of Pattern Recognition and Computer Vision that work on the image content itself. Although pattern recognition, computer vision, and image data mining are quite different fields, they share a large number of common functions (Hsu, Lee & Zhang, 2002).

An interesting commercial application of pattern recognition is a system to semi-automatically annotate video streams to provide content for digital interactive television. A similar idea was behind the MIT MediaLab Hypersoap project (The Hypersoap Project, 2007; Agamanolis & Bove, 1997). In this system, users touch images of objects and people on a television screen to bring up information and advertising material related to the object. For example, a user might select a famous actor and then a page would appear describing the actor, films in which they have appeared, and the viewer might be offered the opportunity to purchase copies of their other films. In the case of Hypersoap, the metadata for the video was created manually. Automatic face recognition and tracking would greatly simplify the task of labeling video in post-production — the major cost component of producing such interactive video.

With the rapid development of computer networks, some web-based image mining applications have emerged. SIMBA (Siggelkow, Schael, & Burkhardt, 2001) is a content-based image retrieval system performing queries based on image appearance from an image database with about 2500 images. RIYA (RIYA Visual Search) is a visual search engine that tries to search content relevant images from the context input. In 2007, Google added face detection to its image search engine (Google Face Search). For example, the URL *http://images.google.com/images?q=bush&imgtype=face* will return faces associated with the name "Bush" including many images of recent US presidents. While the application appears to work well, it does not actually identify the face images. Instead it relies on the associated text metadata to determine identity.

None of the above systems support the input of face images as a query to retrieve similar images of the same person. A robust face recognition method is needed for such kind of systems. Now we will focus on the crucial technology underpinning such a data mining service—automatically recognizing faces in image and video databases.

MAIN FOCUS

Challenges for Face Recognition

Robust face recognition is a challenging goal because differences between images of the same face (intraclass variation) due to nuisance variations in lighting conditions, view point, pose, age, health, and facial expression are often much greater than those between different faces (interclass variation) (Adinj, Moses & Ulman*,* 1997, Zhao, Chellappa, Philips, & Rosenfeld, 2003) An ideal face recognition system should recognize new images of a known face and be insensitive to nuisance variations in image acquisition. Most systems work well only with images taken under constrained or laboratory conditions where lighting, pose, and camera parameters are strictly controlled. This requirement is much too strict to be useful in many data mining situations when only few sample images are available such as in recognizing people from surveillance videos or searching historic film archives. The following is a list of three key problems existing for current face recognition technology:

- Overall accuracy, particularly on large databases
- Sensitivity to changes in lighting, expression, camera angle, pose
- Computational load of searches

Illumination Invariant Face Recognition

Recent research has been focused on diminishing the impact of nuisance factors on face recognition. Two main approaches have been proposed for illumination invariant recognition. The first is to represent images with features that are less sensitive to illumination change (Gao & Leung, 2002) such as using the edge maps of an image. However, the locality problem of edge representation due to small rotation or location errors will degenerate the performance greatly. Yilmaz and Gokmen (2000) overcome the locality problem by using hills to spread the edge profile. These methods assume that features do not change dramatically with variable lighting conditions. Yet this is patently false as edge features generated from shadows may have a significant impact on recognition. Experiments done by Adinj et al. (Adinj, Moses, & Ulman*,* 1997) show that even when using the best illumination insensitive features for image presentation, the classification error is more than 20%. The second main approach is to construct a low dimensional linear subspace for the images of faces taken under different lighting conditions. This method is based on the assumption that images of a convex Lambertian object under variable illumination form a convex cone in the space of all possible images (Belhumeur & Kriegman, 1998). Ignoring the effect of shadows, this subspace has three dimensions (Zhao & Yang 1999). To account for attached shadows 5 to 9 dimensions are needed (Basri & Jacobs, 2003; Georghiades, Belhumeur, & Kriegman, 2001). All these methods assume that the surface of human face is Lambertian reflective and convex, and thus cannot describe cast shadows. Furthermore, such systems need several images of the same face taken under different lighting source directions to construct a model of a given face — in data mining applications it is often impossible to obtain the required number of images.

Expression Invariant Face Recognition

As for expression invariant face recognition, this is still an open problem for machine recognition and it is also quite a difficult task for humans. The approach adopted in the work of Black, Fleet, and Yacoob (2000) is to morph images to the same expression as the one used for training. A problem is that not all images can be morphed correctly. For example an image with closed eyes cannot be morphed to a standard image because of the lack of texture inside the eyes. Liu, Chen, and Kumar (2003) proposed using optical flow for face recognition with facial expression variations. However, it is hard to learn the motion within the feature space to determine the expression changes, since the way one person expresses a certain emotion is normally

Related Content