Review of Logo Recognition Technology Based on Image Classification and Retrieval

汤鹏飞
21821344

摘要（Abstract）

In recent years, the development of Internet technology provides people with many social platforms for information exchange. At the same time, people can easily use mobile phones, cameras and other portable devices to get pictures, videos and other multimedia information anytime and anywhere, and upload it to various social platforms to share with others. How to effectively identify such vast information is an urgent problem to be solved. Logo recognition is one of the most critical issues in many areas, such as copyright infringement detection, advertising, brand data statistics, augmented reality and so on. This paper analyses the current research status in logo recognition, and expounds some main theories and technologies.

1. Background

Content Based Image Retrieval (CBIR) is a branch of multimedia information processing which many important applications are based on. Logo recognition and retrieval, as a sub-direction of CBIR, has attracted more and more attention in recent years due to its potential commercial value. Logo recognition and retrieval is considered to be a crucial tool in industry, commerce and trademark registration. We can often see various kinds of logos in daily life, especially in TV or social media. Logos impress people in a visualized way, and enterprises naturally gain potential economic benefits

As a special visual media, graphical logo is of great significance in identifying the objects it represents. In the industrial and commercial fields, logo can make people connect it with a particular product or service. This kind of correlation with economic interests prompts enterprises to design a symbolic image that can represent the image of the enterprise. When designing logo trademarks, if we can find similar logos from a logo query identification system, we can avoid possible commercial disputes in the future. In addition, with the wide application of Intelligent Transport System (ITS), Logo which symbolizes the brand of automobile, has many specific applications based on Logo recognition, such as guiding consumers' propensity of purchase, vehicle tracking and so on. In the field of smartphones, Logo recognition has shown greater commercial value. Specifically, people can use mobile phones to take photos of a brand of Logo and upload, and then get the relevant information of the enterprise.

2. State of art

CBIR refers to finding images with certain features or specific contents directly from a large number of databases. The core of CBIR is to use visual features of images to retrieve images. The traditional text-based retrieval needs to label the image manually first, and then use the method of text retrieval according to the label information of the image. But because the manual labeling needs a lot of manpower and material resources and everyone has different understanding of the image, it cannot achieve satisfactory results. Therefore, in recent years, CBIR technology has attracted extensive attention in computer vision, and has rapidly become one of the hotspots of research domestically and abroad.

Compared with traditional text-based retrieval, CBIR system integrates various technology such as digital image processing and pattern recognition. Some key technologies are applied such as image feature extraction and processing, effective representation of features, similarity measurement between images based on different methods, and efficient indexing. The process of feature extraction and indexing can be automatically completed by computer, which avoids the subjectivity of manual description. In addition, it has the following characteristics: directly analyzing image content and extracting features; using information retrieval technology to index features on the basis of massive databases; using approximate matching between features to refine gradually; users can modify and submit retrieval requests according to the results returned by the system, which has strong interactivity.

The basic process of CBRI system is: firstly, proper preprocessing is done before image retrieval. Second, combined with practical application, the required features are extracted based on specified algorithms, such as color, texture, sketch and shape, which are stored in the feature database. At the same time, the feature database is indexed to improve the query. Then, according to the query conditions set by users, based on one or several feature combinations, the similarity matching algorithm is used to calculate the similarity between the input image and the image features in the database, and the similar images matched with the retrieval image in the database are returned to users according to the calculation results. Users can choose whether to modify the query conditions according to the feedback results to achieve more satisfactory results.

We can naturally conclude that the system is composed of two important parts: the off-line feature library generation system and the on-line query system. Feature base generation stage, as the premise and basis of CBIR, mainly includes pre-processing of image, extracting image features, building index and storing features into database. Among them, image preprocessing include image enhancement, histogram equalization, sharpening, binarization and a series of related operations. Feature extraction, as the core module of the library generation system, includes texture features, color features, shape features and so on. The query subsystem mainly consists of three parts: query interface, identification search engine and browser that displays result. The query interface enable users to input relevant data (such as some image features desired by users and some restrictive parameters), identifies the similarity between features and ranks them according to the similarity size by the search engine, and presents the sorted results one by one through the display result browser.

CBIR can be classified into corresponding types according to the different features it extracts, such as color-based image retrieval, texture-based image retrieval, shape-based image retrieval and so on. Among them, color-based retrieval and texture-based retrieval are the most widely used and the most basic means of image recognition and retrieval. In the past research and study, some researcher also put forward many algorithms about Logo recognition and retrieval. But there exists some limitations and cannot get a practical product. Joly and Buisson proposed a method based on SIFT feature matching to search for images with specific logo, but because of the diversity of logo types, they only use locally invariant features, which cannot achieve satisfactory results, and they do not locate logo in images while positioning is very necessary for many applications. David S. Doermann and his colleague try to use some constraints. For example, they add some constraints on the relationship between circles and lines in logos to achieve target recognition. Philbin J uses geometric verification such as RANSAC in the pre-processing stage, and then defines the boundary of the matching region as the location of the target object. Invariants are also used in the pre-processing stage, but single invariant is difficult to cope with different types of Logo in the database. Hassanzadeh and Purghassem mainly focus on file images. They simply use binarization to separate foreground objects as objects of interest when extracting foreground objects. Obviously, there exists potential errors because the logo to be recognized may exist as a background in a image. Behmo and Marcombes use machine learning to reduce the errors caused by Naïve Bayes Nearest Neighbor (NBNN) classifier In the target matching stage while it is only available when each feature point in the image is independent of other feature points. However, because the structure of each logo image is fixed, that is, there is a corresponding spatial relationship between feature points, the Naive Bayesian hypothesis cannot be satisfied, thus it is not suitable for Logo image recognition. Shiliang Zhang and Qingming Huang regard local feature points as a group when modelling spatial relationship. They first use bag of-visual words to get some visual words, and then regard the visual words pairs that frequently appear in a certain distance as visual words. However, if the visual words in one group do not match those in the other, the matching accuracy between the two groups will be reduced because of the quantization errors introduced by the above mismatches.

3. Theory and technology

CBIR usually involves the following key technologies: Logo recognition, clustering classification, image feature extraction, image normalization, word bag model, similarity calculation, etc.

3.1. Logo recognition

Logo recognition, as a branch of image recognition, is mostly based on pattern recognition. It involves two modules: Logo detection and Logo recognition. In Logo detection, the related technologies of digital image processing are widely used, including image transformation, image enhancement, image segmentation and so on. In Logo recognition, it is based on different types of features, which can be divided into the following two types: (1) Logo recognition based on global features: global features include color features, GIST features, texture features and so on. Logo recognition based on global features can overcome the problem of low image quality and noise degradation caused by limited acquisition conditions; (2) Logo recognition based on local features: local features include edge detection, corner features, SIFT descriptors and other local information. Local features make less use of global statistical information, which can overcome the problem of missing or redundant logo caused by inaccurate logo detection results.

3.2. Data mining

Data Mining is a step of Knowledge Discovery in Data Bases (KDD). It refers to the process of automatically searching out information hidden in a large amount of data.

3.2.1 Unsupervised clustering

Clustering is a technology that aims at discovering some intrinsic structures by browsing data. It organizes all data instances into some groups, each of which is called a cluster. Because there is no class or label of data in clustering, clustering technology is often referred to as Unsupervised Learning. There are many clustering algorithms, including K-means clustering, DBSCAN clustering and so on.

K-means clustering is the most widely used clustering algorithm because of its simplicity and efficiency. Given a set of data points and the number of clusters K (K is specified by the user), K-means clustering divides the data into K clusters repeatedly according to a distance function.

Let the set of data instances (points) D be (x_l, x_2,... and x_n}, where x_i= (x_{i1}, x_{i2},..., x_{ir}) is a vector in real space $X \subseteq R$, and r represents the number of attributes of data (dimension of data space). K-means algorithm divides the given data into K clusters. The mean of all data points in each cluster is called Cluster Centroid, which usually represents this class.

The process of the clustering algorithm is as follows: first, K data points are randomly selected from set D as the initial clustering center, then the distance between each data point and each seed clustering center is calculated, and each data point is allocated to the nearest clustering center. The cluster center and the data points assigned to it represent a cluster. Once all data points are allocated, the clustering centers are recalculated according to the existing data points in each cluster. The process is repeated until the clustering centers are no longer changed, or no data points are redistributed to different clusters, or local error of sum of squares is minimized.

3.2.2 Supervised classification

Classification refers to generating a classification /prediction function of a set of associative attribute values A and a set of class labels C from a data set D with class labels. The function to predict the class labels of a new set of attributes (data instances) can be abbreviated as a classifier. There are various kinds of classification algorithm including decision tree, association rules, Bayesian model, support vector machine and so on.

3.3. Image feature extraction

Feature extraction is a concept in computer vision and image processing. It performs primary operations on images in order to check whether each pixel represents a feature. The result of feature extraction is to divide the points on the image into different subsets, which usually consist of isolated points, continuous curves or continuous regions.

3.3.1 Initial feature extraction

·Feature extraction based on color.

Color is a kind of global feature, which describes the surface properties of the scene of the corresponding image or image area. Because color is insensitive to the direction and the size of the image or image region, it cannot capture the local features of the object in the image perfectly. In addition, if the database is large, many unexpected images will be retrieved when only using color feature query. Color histogram is commonly used to express color features as it is not affected by the changes of image rotation and translation, even image scale changes by means of normalization. The main disadvantage is that it does not express the information of color spatial distribution.

·Feature extraction based on texture.

Texture is also a kind of global feature. It also describes the surface properties of the scene of the corresponding image or image area. As a statistical feature, texture features are often insensitive to rotation and noise. However, texture is only a feature of the surface of an object, which cannot fully reflect the intrinsic properties of the object. Therefore, only using texture features cannot obtain high-level image content, and texture features are usually unavailable when the resolution of the image changes. Thus the calculated texture may deviate.

·Feature extraction based on shape.

Shape-based retrieval can effectively retrieve objects of interest in images. But there are also some problems. For example, when the object is deformed, the retrieval results are not reliable, and many shape features only describe the local features of the object, and require too much for time and space when describing the object in an all-round way.

·Feature extraction based on space.

Spatial relationship refers to the spatial position or relative direction relationship between multiple objects segmented in an image. These relationships can be divided into connection/adjacency relationship, overlap/overlap relationship and inclusion/inclusion. There are two methods to extract the spatial relationship features of images: one is to automatically segment the image, divide the object or color region contained in the image, and then extract the image features according to these regions, and establish index; the other is to simply divide the image into several regular sub-blocks evenly, and then extract the features from each sub-block of the image, and establish the index.

Spatial relationship features can enhance the ability of describing and distinguishing image content, but spatial relationship features are often sensitive to image or target rotation, inversion, scale change and so on. In addition, in practical applications, it is often not enough to use only spatial information, which cannot effectively and accurately express scene information.

3.3.2 Feature transformation

The dimension may be still large and there may exist some irrelevant or redundant features for the preliminary image features. Future transformation can effectively reduce the dimension of feature space and eliminate possible correlation between features and useless information in features.

·Principal Component Analysis (PCA)

PCA aims at obtaining a set of new features with the greatest variance from a set of features by solving the optimal orthogonal transformation which are linear combinations of the original features and not related to each other. Then, the new features is ranked according to importance and the first principal components are selected. Using fewer principal components to represent the data can reduce the dimension of the feature and eliminate the noise in the data. This algorithm does not consider the class information of samples.

·Linear Discriminant Analysis (LDA)

The basic idea of LDA is to project the high-dimensional data samples into the optimal discriminant vector space to extract the classification information and compress the dimension of feature space. After projection, the data samples are guaranteed to have the maximum distance between classes and the minimum distance between classes in the new subspace, that is to say, the sample data has the best separability in this space. Fisher linear discriminant analysis is the most representative LDA.

·Kernel Principal Component Analysis (KPCA)

Through KPCA, the samples are transformed nonlinearly, and the principal component analysis in the transformed space is used to realize the Non-linear Principal Component Analysis in the original space. According to the properties of the renewable Hilbert space, the covariance matrix in the transformation space can be operated by the kernel function in the original space, thus avoiding complex non-linear transformation. Appropriate type of kernel function is chosen according to practical problems. Different types of kernels reflect different assumptions about data distribution. They can also be regarded as introducing a non-linear distance measure to data.

4. Summary

Logo recognition, as an important branch of multimedia information processing, has attracted more and more attention in recent years due to its potential commercial value. This article introduced a research review of Content Based Image Retrieval (CBIR), a major method for logo recognition. We introduced the state-of-art of CBIR and some related theory and technology.

References

[1] Phan R, Androutsos D. Content-Based Retrieval of Logo and Trademarks in Unconstrained Color Image Databases using Color Edge Gradient Cooccurrence Histograms[J], Computer Vision and Image Understanding(CVIU), 2010, 114(1): 66-84.

[2] Psyllos A P, Anagnostopoulos C N E, Kayafas E. Vehicle Logo Recognition Using a SIFT-Based Enhanced Matching Scheme[J]. Intelligent Transportation Systems, IEEE Transactions on(TITS), 2010, 11(2):322-328.

[3] Li W, Duan L, Xu D, et al. Text-Based Image Retrieval Using Progressive Multi-Instance Learning[A]. ICCV[V], 2011: 2049-2055.

[4] Li Z, Schulte-Austum M, Neschen M. Fast Logo Detection and Recognition in Document Images[A]. Pattern Recognition(ICPR), 2010 20th International Conference on. IEEE[C], 2010: 2716-2719.

[5] Joly A, Buisson O. Logo Retrieval with a Contrario Visual Query Expansion[A] Proceedings of the 17th ACM international conference on Multimedia, ACM[C], 2009: 581-584.

[6] Philbin J, Chum O, Isard M, et al. Object Retrieval with Large Vocabularies and Fast Spatial Matching[A]. Computer Vision and Pattern Recognition(CVPR)[C], 2007: 1-8.

[7] Hassanzadeh S, Pourghassem H. A Fast Logo Recognition Algorithm in Noisy Document Images[A]. Intelligent Computation and Bio-Medical Instrumentation(ICBMI)[C], 2011:64-67.

[8] Behmo R, Marcombes P, Dalalyan A, et al. Towards Optimal Naive Bayes Nearest Neighbor[M], Computer Vision-ECCV 2010. Springer Berlin Heidelberg, 2010:171-184.

[9] Boiman O, Shechtman E, Irani M. In Defense of Nearest -Neighbor Based Image Classification[A]. Computer Vision and Pattern Recognition(CVPR)[C], 2008:l-8.

[10] Zhang S, Huang Q, Hua G et al. Building Contextual Visual Vocabulary for Large-Scale Image Applications[A]. Proceedings of the international conference on Multimedia, ACM[C], 2010: 501-510.