Techniques of Feature Extraction Based on Contents of Image

Mrs. Aparna Joshi
M.Sc. (CS)
Pratibha College of Commerce & Computer Studies, Chinchwad, Pune India
aajoshi2@gmail.com

Mrs. Hemalata Chavan
MCA (Eng.)
Pratibha College of Commerce & Computer Studies, Chinchwad, Pune India
m_hemalata2000@yahoo.com

Abstract - In various image processing applications, feature extraction techniques are helpful. When input data to an algorithm is too large and redundant (much data, but not much information) to be processed then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. Image features are useful in classifying and recognition of images. Feature extraction is finding the set of parameter that define the image precisely and uniquely. It also used in pattern recognition, feature matching, image segmentation, image fusion, video processing, medical diagnosis, remote sensing etc.

Keywords - Feature Extraction, Local Feature, Global Feature, Geometrical Features

I. INTRODUCTION

As the images grow complex and diverse, retrieval of correct image information becomes difficult, represented by numeric values called features or descriptors that are meant to represent properties of the image. Image information may contain text or content feature of the image and it extracts in the form of pixel information, intensity, contrast, features like - color, texture, shape etc. Image retrieval systems have been developed based on text and content of the images. Text-based image retrieval system based on keyword and text descriptors, which are manually annotated. This textual image information cannot capture perceptual saliency and conceptual features with certain entities or events and hence fail at essence of perfection and visual interpretation of picture content. This retrieval systems concern about only one language and text described by humans, instead of looking into the content of images. The content-based image retrieval (CBIR) systems is a set of properties (similarity features) able to effectively characterize the content of images which minimize errors in information retrieval. CBIR system help users to retrieve relevant images based on similarities of visual contents like colors, shapes, textures, objects, brightness etc. which are considered the lower level features of an image. The similarity between the users query and images are computed. In this each image stored in the database, has its features extracted and compared to the features of the query image. It provides a method to find images in large databases by using unique descriptors from a trained image. This image descriptors are known as features of the image. The image descriptors or features of image include texture, color, intensity and shape of the object inside an image.

II. FEATURE DETECTION

Feature is a distinguishing characteristic of an image. Feature extraction techniques are applied on images to get features that will be useful in classifying and recognition of images. Features define the behavior of an image and it shows its place in terms of storage taken, efficiency in classification and obviously in time consumption also. Feature extraction describes relevant shape information contained in a pattern so that task of classifying the pattern is made easy by a formal procedure. Feature extraction is a special form of dimensionality reduction. The main goal of feature extraction is to obtain the most relevant information from the original data and represent that information in a lower dimensionality space. In this process relevant features are extracted from objects/ alphabets to form feature vectors. These feature vectors are then used by classifiers to recognize the input unit with target output unit. It becomes easier for the classifier to classify between different classes by looking at these features as it allows fairly easy to distinguish.[1]

Feature extraction is done after the preprocessing phase or techniques which may involve binarization, thresholding, resizing, normalization, restoration, enhancement or just representation of the data etc. are applied on the sampled image. [1,3]

Criteria to choose features given by Lippman are:

"Features should contain information required to distinguish between classes, be insensitive to irrelevant variability in the input, and also be limited in number, to permit, efficient

computation of discriminant functions and to limit the amount of training data required" [1]

III. TYPES OF FEATURES

General features in feature extraction are application independent such as color, speed, texture, and shape. According to the abstraction level, they can be further divided into:

1. **Pixel-level Features** :-
   These are also known as, low-level features. These features are calculated at each pixel, e.g. color, location. Low-level features are basic features that can be extracted automatically from an image without any shape information.
2. **Local Features** :-
   Features calculated over the results of subdivision of the image band on image. These features are usually geometric (e.g. concave/convex parts, number of endpoints, branches, Joints etc).
3. **Global Features** :-
   Features calculated over the entire image or just regular sub-area of an image. These are usually topological (connectivity, projection profiles, number of holes, etc) or statistical (invariant moments etc.)
4. **Domain-specific Features** :-
   Application dependent features such as human faces, fingerprints, and conceptual features.[2,3,4,5]

Several feature-extraction techniques like Average RGB, Color Moments, Co-occurrence, Local Color Histogram, Global Color Histogram and Geometric Momentsare.

IV. CLASSIFICATION OF FEATURE EXTRACTION

Feature extraction methods are classified into three major groups as:

1. Statistical Features
2. Global Transformation and Series Expansion Features
3. Geometrical and Topological Features

1. **Statistical Features** :
   These features are derived from the statistical distribution of points. They provide high speed and low complexity and take care of style variations to some extent. They may also be used for reducing the dimension of the feature set.[1,4,6,7]
   Major statistical features are – zoning, Characteristic Loci, Crossing and Distances.
2. **Global Transformation and Series Expansion Features** :-
   These features are invariant to global deformations like translation and rotations. A continuous signal generally contains more information that needs to be represented for the purposes of classification. One way to represent a signal is by a linear combination of a series of simpler well defined functions. The coefficients of the linear combination provide a compact encoding known as series expansion. Common transform and series expansion features are: Fourier Transforms, Walsh Hadamard Transform, Rapid transform, Hough Transform, Gabor Transform, Wavelets, Karhunen Loeve Expansion, Moments. [4,7,8]
3. **Geometrical and Topological Features**
   These features may represent global and local properties of characters and have high tolerances to distortions and style variations. These topological features may encode some knowledge about the contour of the object or may require some knowledge as to what sort of components make up that object. It contains features like - Strokes, Stroke Directions and Bay, Chain Codes, End points intersections of line segments and loops, Strokes relations and angular properties.[6,8,9]

The most widely features are color, edge, shape and texture.

V. COLOR FEATURES

As the color image carries more information than grey image, color has been successfully applied to retrieve images, and also it has very strong correlations with the underlying objects in an image. The color feature is robust to background complications, scaling, orientation, perspective, and size of an image.[5,8,10]

Images characterized by color features have many advantages:

i. Robustness –
   The color histogram is invariant to rotation of the image on the view axis, and changes in small steps when rotated otherwise or scaled. It is also insensitive to changes in image and histogram resolution and occlusion.
ii. Effectiveness-
   There is high percentage of relevance between the query image and the extracted matching images.
iii. Implementation simplicity –
   The construction of the color histogram is a straightforward process, including scanning the image, assigning color values to the resolution of the histogram, and building the histogram using color components as indices.[3,8,9,10]

For color feature extraction techniques applied – color histogram.

VI. SHAPE FEATURES

Shape is an important visual feature and it is one of the primitive features for image content description. Shape content description is difficult to define because measuring the similarity between shapes is difficult. Shape descriptors can be divided into two main categories: region based and contour-based methods. Region-based methods use the whole area of an object for shape description and calculate the feature from the entire region. While contour-based methods use only the information present in the contour of an object and calculate the feature from the boundary and ignore its interior [2,3,6]

Edge detection

Edge can be defined as a set of connected pixels that lies in the boundary between two regions. Separation of high frequency information means edge detection. An edge or boundary is the

external information of image. The internal features in an image can be found using segmentation and texture.[3,,9,10]

Texture detection

Texture is a powerful regional descriptor that helps in the retrieval process. Texture feature extraction is very robust technique for a large image which contains a repetitive region. The texture is a group of pixel that has certain characterize. The texture feature methods are classified into two categories: spatial texture feature extraction and spectral texture feature extraction. Texture, on its own does not have the capability of finding similar images, but it can be used to classify textured images from non-textured ones and then be combined with another visual attribute like color to make the retrieval more effective. Texture of an image is quantitatively described by its coarseness. The coarseness index is related to the spatial repetition period of the local structure.[3,6,7,8,9]

Basically, texture representation methods can be classified into two categories: structural; and
statistical. Statistical methods, including Fourier power spectra, co-occurrence matrices, shift invariant principal component analysis (SPCA), Tamura features, World decomposition, Markov
random field, fractal model, and multi-resolution filtering techniques such as Gabor and wavelet
transform, characterize texture by the statistical distribution of the image intensity.[3,7,8,11]

VII. CONCLUSION

Feature plays an important role in image processing. According to the application of image processing, different features of an image and feature extraction techniques are applied. These extracted features are further used in selection and classification process in image analysis.

The widely used feature extraction methods are Template matching, Deformable templates, Unitary Image transforms, Graph description, Projection Histograms, Contour profiles, Zoning, Geometric moment invariants, Zernike Moments, Spline curve approximation, Fourier descriptors, Gradient feature and Gabor features, Average RGB, Color Moments, Co-occurrence, Local Color Histogram.

REFERENCES

1. Gaurav Kumar, Pradeep Kumar Bhatia, “A Detailed Review of Feature Extraction in Image Processing Systems”, Fourth International Conference on Advanced Computing & Communication Technologies, 2014.
2. Ryszard S. Choras, “Image Feature Extraction Techniques and Their Applications for CBIR and Biometircs Systems”, International Journal of Biology and Biomedical Engineering. Issue 1, Vol. 1, 2007.
3. R. Revathi, Dr. M. Hamalatha, “An Emerging Trend of Feature Extraction Method In Video Processing” Academy & Industry Research Collaboration Center (AIRCC), CSCP , Vol. 2, http://airccj.org/CSCP/vol2/csit2209.pdf
4. Aman Chadha, Shshmit Mallik, Ravdeep Johar, “Comparative Study and Optimization of Feature-Extraction Techniques for Content Based Image Retrievel”, International Journal of Computer Applications (0975-8887), Volume 52- No. 20, August 2012.
5. Jayamala K. Patil, Raj Kumar, “Feature Extraction & Retrieval of Plant Leaf Diseases Using Image Histogram”, International Science Press, India, Vol. 5, Number 2, July-December 2013, pp. 143-151.
6. Seyyid Ahmed Medjabed, “A Comparative Study of Feature Extraction Methods in Images Classification”, I.J. Image Graphics and Signal Processing, 2015, 3, 16-23, DOI:10.5815
7. D. P. Tian and B. Shaanxi, “A Review on Image Feature Extraction and Representation”, Techniques International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 4, pp. 385-396, 2013
8. L. Li and P. W. Fieguth, "Texture Classification from Random Features," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 3, pp. 574-586, 2012.
9. T. Yu, V. Muthukkumarasamy, B. Verma and M. Blumenstein, "A Texture Feature Extraction Technique Using 2D-DFT and Hamming Distance,", Fifth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'03), 2003.
10. Ribo Ge, Hailong Shen, “ Researching on Feature Extraction of Brain CT Image”, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol 6, No. 5, (2013), pp 39-48.
11. Miss Meenu Sahu, Prof. Amit Saxena, Dr. Manish Manoria, “Application of Feature Extraction Technique : A Review”, International Journal of Computer Science and Information Technologies, Vol. 6 (3), 2015, 3014-3016.