International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 01 | Jan 2021 www.irjet.net p-ISSN: 2395-0072

# Survey on Feature Extraction Methods

**Poornima Shetagar[1], Sehraantaj Sadarbhai[2], Vaibhavi Kulkarni[3], Vijayalaxmi Patil[4]**

[1]*Poornima Shetagar, Student, Department of Information Science and Engineering, SDMCET, Dharwad, Karnataka, India.*
[2]*Sehraantaj Sadarbhai, Student, Department of Information Science and Engineering, SDMCET, Dharwad, Karnataka, India.*
[3]*Vaibhavi Kulkarni, Student, Department of Information Science and Engineering, SDMCET, Dharwad, Karnataka, India.*
[4]*Vijayalaxmi Patil, Student, Department of Information Science and Engineering, SDMCET, Dharwad, Karnataka, India.*

---

**Abstract –** *The immense development in image processing has given remarkable results in medical image analysis.* Image processing *is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. In any image processing image feature plays very significant role. Image feature extraction methods are very popular, since they give more accurate results and also it has increased work efficiency of doctors. In this survey paper different types of feature extraction methods such as shape-based feature extraction and texture-based feature extraction, its various forms and applications are discussed.*

***Key Words*: Feature extraction, Texture extraction, Shape extraction, Image processing, Medical image analysis.**

## 1. INTRODUCTION

Feature extraction is a type of dimensionality reduction i.e., transformation of high dimensionality space to low dimensionality space where a large number of pixels of the image are efficiently represented in such a way that interesting parts of the image are captured effectively. It has broad range of applications such as Machine Learning, Image Processing etc. [5]. It has huge scope in Pattern Recognition in image processing. Hence it has given immense results in medical image analysis including counting of red blood cells, white blood cells and biopsy recognition of cancer cell; visceral size, shape, and anomaly detection, etc. However, because of the complexity of medical images, methods of feature extraction in medical image analysis applications have some limitations [1]. In the feature extraction methods, there are four types of image features often used: color (Gray) image features, texture features, shape features and spatial relations. Through the image feature extraction, features should be able to describe objects abstractly and concretely.[1] For any image processing images should have characteristics such as Uniqueness ,Integrity ,Invariance under the geometric Structure, Agility, Abstractness[1] Various Feature Extraction methods are developed in the last decade but each method has its own advantage and disadvantages some are easy to implement and on other hand some have low computational complexity and some method have fast computation speed and so on.

This paper is organized as follows: the literature survey of various methods is presented in section 2. The Analysis and discussion are presented in section 3. Finally, this paper is concluded in section 4. The papers referred are in section 5.

## 2. Literature Survey

In this section we have discussed various methods to develop a content-based image retrieval system for medical image analysis. Here we have discussed about the feature extraction methods namely shape based and texture-based methods with their types and varieties along with some of their applications.

### 2.1 Shape Based Feature Extraction Methods

Shape features can be divided into two categories, one is based on the characteristics of the border, and the other is based on regional characteristics. Accordingly, the methods of image shape feature extraction are also divided into boundary-based feature extraction methods and region-based feature extraction methods [1].

#### 2.1.1 The Shape Feature Extraction Methods Based on the Boundary

**[A] Some Simple Descriptor:**

***Perimeter of the Boundary***

The perimeter of the boundary is the length of the smallest outer boundary contour of the connected region. Calculated as follows: First, the thinning method. The outline of connected regions is calculated by morphological methods. The length of the outline is calculated directly. In the X, Y direction, the length of path point distance is set to 1. And the length of diagonal distance is set to 2. The computing speed of this method is fast. Second, the chain code method. The contour length of connected region is calculated by chain code. When the chain code of outer contour is even, set its

number to M. When the chain code of outer contour is odd number, set its number to N. So that, the perimeter of the connected region is

P=M+ 2 N. (1)

The operation of this way is complex and the speed is slow. This has slow speed but complexed operation. Third, the boundary tracing method. When the outer contour is unknown, its length can be calculated by the boundary tracing method. The program design of this method is complex and the operation speed is medium.[1]

***Diameter of Boundary***

The Diameter of bounder is longest distance between two points. The diameter of a boundary B is defined as,

Diam (B) = max [ D ( $p_i$ , $p_j$ ) ] (2)

Where D is a distance measure which is any of Euclidean distance, block distance and chessboard distance, and pi and pj are points on the boundary. The value of the diameter and the orientation of a line segment major axis of the boundary are useful descriptors of a boundary.[1]

***Eccentricity***

The line perpendicular to major axis is minor axis of a boundary. The ratio of the major to the minor axis is called the eccentricity of the boundary.[1]

***Curvature***

In the continuous case, the curvature is the rate of change of slope. In the discrete space, the curvature means the rate of the total number of the border (the border perimeter) pixels to the number of boundary pixels whose boundary direction change significantly.[1]

***BIC (Border/Interior classification)***

The method classifies the pixels of the image as interior or border. Then two histograms for the pixels classifies as edge and interior are generated [2].

**[B] Fourier Descriptors**

Fourier transform can access the characteristics of the object through changing the sensitivity and direct representation into the frequency domain. The Fourier descriptors are insensitive to translation, rotation, scale changes and the starting point [1].

**[C] Statistical Moments**

Statistical moment is a statistical form of the image pixel. Shape of boundary segments can be described quantitatively by using simple statistical moments, such as the mean, variance, and higher-order moments. Moment is calculated by pixel, and it is affected little by noise. To describe the amplitude of gas discrete variable v, and form histogram on p(vi), i=0,1, 2, A-1. Here, A is the number of discrete amplitude values. p (vi) is the probability valuation of generate value vi. So that the first n-mean estimate is [1]

$m_{n\,(V)} = \sum_{i=0}^{A-1} ( V_i - m )^n p ( v_i )$ (3)

In the formula: m is the mean of v, μ2 is its variance.

$m = \sum_{i=0}^{A-1} V_i p( V_i )$ (4)

**2.1.2 Shape Feature Extraction Methods Based on Region**

**[A] Some Simple Descriptors**

***Regional Area***

In the digital image, the area of a region is calculated as the number of pixels in the region. Mark the pixels within the region as f(x,y)=1,outside the region as f(x,y)=0. The area can be calculated as:

$A = \sum_{(x,y) \in R} 1$ (5)

***Roundness***

Roundness is known as compactness. Using roundness measure the shape degree the graphics tend to circle. Formula for roundness calculation is,

$C = (4\pi A) / P2$ (6)

In the formula, P is the perimeter of regional bounder, A is the regional area. When the region is circular, C takes the maximum value 1. When the region is long and thin strip or more complex, C value is smaller.[1]

***Regional Focus***

Regional focus is also a kind of shape characteristics. The coordinates calculated according to all the points belonging to the region are the coordinates of the regional focus. [1]

$\bar{x} = 1/A \sum_{(x,y) \in R} x \; \bar{y} = 1/A \sum_{(x,y) \in R} y$ (7)

In formula, x,y is the points within the region, A is the regional area.

**[B] Topological Descriptors**

Topological features are useful for the overall description of the image plane region. The commonly used topological descriptors are the number of the holes H, connected components C and the Euler number E. [1]

$E = C - H$ (8)

### 2.1.3 Applications of Shape Feature Extraction Methods in Medical Image Analysis

Accuracy and efficiency of doctor's diagnosis are improved. Image shape feature extraction methods were widely used in leucocyte image feature extraction, CT brain tumor image extraction, edge extraction head CT, lung cancer liver cancer medical image feature extraction, the extraction of human parasite eggs in the image recognition and so on.[1]

### 2.2 Texture Feature Extraction Methods

### 2.2.1 Statistical Methods

***Contrast***

Contrast feature is a measure of the image contrast or the number of local variations present in an image.

$$C = \sum_{i=0}^{L-1} \sum_{j=1}^{L-1} (i-j)^2 P_{i,j} \quad (9)$$

If there is a large amount of variation in an image the P[i,j]'s will be concentrated away from the main diagonal and contrast will be high.

***Entropy***

Entropy measures the disorder of an image and randomness of intensity distribution. It achieves its largest value when all elements in P matrix are equal [4]

$$H(X) = -\sum_{i=1}^{L} P(i) \log P(i) \quad (10)$$

Mean (X)

It represents amount of brightness present in an image [4]. The mean calculates the average value of gray level intensities. If mean value is high then image is bright and if mean value is low then image is dark. Mean of an image may be defined [4] as:

$$\bar{X} = \sum_{i=1}^{L} i\, P(i) \quad (11)$$

***Energy***

It is also known as consistency measures the image uniformity [4] i.e. intensity level distribution present in the image. If energy value is high then intensity level distribution is small in the image. Energy can be defined as:

$$E = \sum_{i=1}^{L} [P(i)]^2 \quad (12)$$

***Haralick***

Haralick Texture is used to quantify an image based on texture. The fundamental concept involved in computing Haralick Texture features is the Gray Level Co-occurrence Matrix, the basic idea is that it looks for pairs of adjacent pixel values that occur in an image and keeps recording it over the entire image.

### 2.2.2 Transform Based Methods

***Haar Wavelet***

Haar Wavelets [3] are fastest to compute and simplest to implement. The **Haar wavelet** is a sequence of rescaled "square-shaped" functions which together form a wavelet family or basis. Wavelet analysis is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal basis. The Haar sequence is now recognized as the first known wavelet basis and extensively used as a teaching example.

***Gabor Transform***

It is used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time. The function to be transformed is first multiplied by a Gaussian function, which can be regarded as a window function, and the resulting function is then transformed with a Fourier transform to derive the time-frequency analysis.

### 2.2.3 Learning Based Approach

Texture based extraction can be done by Learning based approach. It can be divided in 3 subsections i.e., Deep learning methods, Vocabulary based method and Extreme Learning method.[5]

***Vocabulary Methods***

vocabulary learning methods – also called visual dictionary methods – imply the learning of a visual dictionary. Dictionary learning achieves therefore higher flexibility than the wavelet transforms. However, dictionary learning requires more complex computation.[5]

***Deep Learning Methods***

Deep learning – and in particular convolutional neural networks (CNNs) – have recently been used in the computer vision field. The CNN model is a learning method that has recently been successfully implemented in a large number of applications because of its excellent performances of feature representation. Andrearczyk [230] used the CNN approach for the recognition of malignant lymphomas and the

classification of mouse liver tissue based on age, gender, and diet.[5]

***Extreme Learning Method***

Texture signature has also been drawn with extreme learning machine (ELM) [243]. An ELM is nothing but single-hidden layer of neural network with a high-speed learning algorithm. ELM gives fast computation speed and good generalization performance. The method has shown good results in texture classification. [5]

**2.2.4 Application of combining the shape feature extraction methods and the texture feature extraction methods**

The result accuracy was only 80% by using feature extraction methods in the classification of blood leukocytes. The significance in error identification was due to the classification of granular cells. By combining both methods result in shortage, so it yielded more accuracy in recognition rate as high as 90%. In lung cancer and liver cancer CT images of the computer-aided diagnosis researches, both methods were implemented. The analysis of data was highly improved. [1]

**3. Analysis and discussion**

We have reviewed different feature extraction methods and classified them into 2 different classes as shape-based feature extraction and Texture based feature extraction. The shape based is further divided into two categories as method based on boundary and method based on region.

Texture based feature extraction is classified into three different categories as Statistical Methods, Transform Based Methods, Learning Based Approach.

For each method, the concept, and examples of applications have been given. Although there are many feature extraction methods, they still have limitations in solving the problems. And we have selected some of those techniques, at present the issue is we need to combine different methods and need to propose new methods and this is an important field of research.

**4. Conclusion**

Feature extraction is an important link of the image analysis, which received extensive attention in medical image analysis as well. In this paper, we had a literature survey on various methods to develop a content-based image retrieval system for medical image analysis. In this paper we have discussed about the feature extraction methods namely shape based and texture-based methods with their types and varieties along with some of their applications. These methods have shown immense results in classifying leucocyte image features extraction, CT brain tumor image extraction, edge extraction, head CT, lung cancer, liver cancer medical image feature extraction, the extraction of human parasite eggs etc. We have also discussed some texture-based extraction can be done by learning based approach which has high computation speed few have high complex computation, but Combining Shape based and texture-based method gives more accuracy than individual methods [1]

**ACKNOWLEDGEMENT**

We would like to take this opportunity to express our profound gratitude and deep regard to everyone who helped us in completing the paper. We sincerely thank **Dr J. D. Pujari**, Head of the Department, Department of Information Science and Engineering, SDMCET, for his exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project.

We also thank Dr **Vandana S. Bhat**, Department of Information Science and Engineering, SDMCET, our project coordinator for her valuable suggestions were of immense help throughout my project work. Her perceptive criticism kept us working to make this paper in a much better way. Working under her was an extremely knowledgeable experience for us.

I would also like to give my sincere gratitude to all the friends and colleagues who filled in the survey, without which this research would be incomplete.

**REFERENCES**

[1] Jianhua Liu, Yanling Shi. Image Feature Extraction Method Based on Shape Characteristics and Its Application in Medical Image Analysis. Department of Sofeware Institute, Department of Information Engineering. pp. 172–178 (2011).

[2] Herbert Chuctaya, Christian Portugal, Cesar Beltr ´ an, Juan Gutierrez, Cristian L ´ opez, Yv ´ an T ´ upac ."M-CBIR: A medical content-based image retrieval system using metric data-structures", Cathedra Concytec in TIC National University of San Augustin, UNSA Arequipa, Peru(2011).

[3] Ashish Oberoi and Manpreet Singh. "Content Based Image Retrieval System for Medical Databases (CBIR-MD) - Lucratively tested on Endoscopy, Dental and Skull Images" Department of Computer Science & Engineering, M.M. Engineering College, M.M. University Mullana, Ambala, Haryana, PIN-133 207, India(2012).

[4] Karam Singh1, Kiran Jot Singh2, Divneet Singh Kapoor3 . "Image Retrieval for Medical Imaging Using Combined Feature Fuzzy Approach", Dept. of Electronics & Comm. Engg Chandigarh University, Punjab, India(2014).

[5] ANNE HUMEAU-HEURTIER Laboratoire Angevin de Recherche en Ingénierie des Systèmes, University of Angers: Texture Feature Extraction Methods: A Survey January 23 (2019).