International Journal of Research In Science & Engineering e-ISSN: 2394-8299
Volume: 3 Issue:1 January 2017 p-ISSN: 2394-8280

CONTENT BASE INFORMATION RETRIEVAL TECHNIQUE WITH HYBRID NATURE

Divyata Patil *[1], Rasna Sharma[2]

*[1]Research Scholar, CIIT College Indore,M.P,India.
[2]Assistant Professor, CIIT college Indore,M.P,India.
*Department of Computer Science & Engineering.

ABSTRACT

Content-based information retrieval (CBIR) has been an active research area for the last two decades and much progress has been made in that time. However, there are still many challenges to be overcome and in this paper we highlight some of these together with some approaches that we have developed to address these problems. The content-based analysis of such dataset requires effective and efficient techniques for information extraction. In this paper describe the Biometric dataset for establishing the identity of an individual based on the physical, chemical or behavioral attributes of the person. On that biometric dataset we just apply information retrieval technique with hybrid nature. This technique is helpful in classifying the objects based on their extracted pattern. The presented paper introduces discussion about that technique.

Keywords CBIR, Biometric dataset, Feature extraction

1.INTRODUCTION

Biometrics is the science of establishing the identity of an individual based on the physical, chemical or behavioral attributes of the person. The relevance of biometrics in modern society has been reinforced by the need for large-scale identity management systems whose functionality relies on the accurate determination of an individual's identity in the context of several different applications. Examples of these applications include sharing networked computer resources, granting access to nuclear facilities, performing remote financial transactions or boarding a commercial flight. The proliferation of web-based services (e.g., online banking) and the deployment of decentralized customer service centers (e.g., credit cards) have further underscored the need for reliable identity management systems that can accommodate a large number of individuals. Biometrics offers certain advantages such as negative recognition and non-repudiation that cannot be provided by tokens and passwords [32]. Negative recognition is the process by which a system determines that a certain individual is indeed enrolled in the system although the individual might deny it. This is especially critical in applications such as welfare disbursement where an impostor may attempt to claim multiple benefits (i.e., double dipping) under different names.

Non-repudiation is a way to guarantee that an individual who accesses a certain facility cannot later deny using it (e.g., a person accesses a certain computer resource and later claims that an impostor must have used it under falsified credentials). Biometric systems use a variety of physical or behavioral characteristics including fingerprint, face, hand/finger geometry, iris, retina, signature, gait, palm print, voice pattern, ear, hand vein, odor or the DNA information of an individual to establish identity [12, 36]. In the biometric literature, these characteristics are referred to as *traits*, *indicators*, *identifiers* or *modalities*.

e-ISSN: 2394-8299
p-ISSN: 2394-8280

Operation of a biometric system

A biometric system is essentially a pattern recognition system that acquires biometric data from an individual, extracts a salient feature set from the data, compares this feature set against the feature set(s) stored in the database, and executes an action based on the result of the comparison. Therefore, a generic biometric system can be viewed as having four main modules: a sensor module; b quality assessment and feature extraction module; c matching module; and database module. Each of these modules is described below.

- Sensor module: A suitable biometric reader or scanner is required to acquire the raw biometric data of an individual. To obtain fingerprint images, for example, an optical fingerprint sensor may be used to image the friction ridge structure of the fingertip. The sensor module defines the human machine interface and is, therefore, pivotal to the performance of the biometric system.
- Quality assessment and feature extraction module: The quality of the biometric data acquired by the sensor is first assessed in order to determine its suitability for further processing. Typically, the acquired data is subjected to a signal enhancement algorithm in order to improve its quality.
- Matching and decision-making module: The extracted features are compared against the stored templates to generate match scores. In a fingerprint-based biometric system, the number of matching minutiae between the input and the template feature sets is determined and a match score reported. The match score may be moderated by the quality of the presented biometric data.

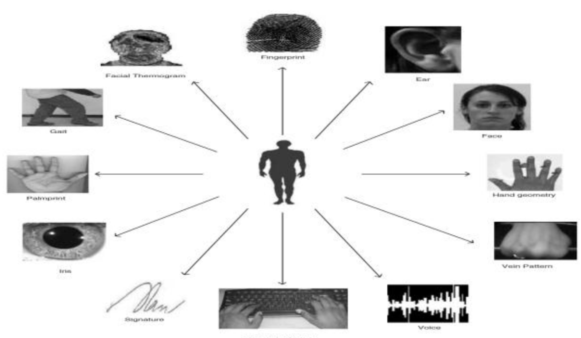


Figure 1 General Categories of Biometric Image of Human Body

- System database module: The database acts as the repository of biometric information. During the enrollment process, the feature set extracted from the raw biometric sample (i.e., the template) is stored in the database (possibly) along with some biographic information (such as name, Personal Identification Number (PIN), address, etc.) characterizing the user. The data capture during the enrollment process may or may not be supervised by a human depending on the application.

LITERATURE REVIEW

B. S. Manjunathet al. [1] proposed the technical details of color and texture descriptors currently in the MPEG-7 standard. The color descriptors include two histogram-based descriptors, the SCD and the CSD, the dominant color descriptor, and the CLD. The histogram descriptors capture the global distribution of color where as the dominant color descriptor represents the dominant colors present. The CLD captures the spatial distribution or layout of the colors in a compact representation. While MPEG-7 standards accommodate different color spaces, most of the color descriptors are constrained to one or a limited number of color spaces for ensuring inter-operability.

P. Gunhanet al. [2] proposed a ranking algorithm using dynamic clustering for CBIR system which retrieves and ranks images according to a similarity function based on a feature vector distance model. A new property in deciding ranking of results has been defined which uses dynamic clustering methods about retrieved images, they make relevant groups that contain similar images. Using the groups, the author has analyze similarity relationship of retrieved results and the query image. The ranking and the similarity value of retrieved images are adjusted according to the cluster analysis.

Sung-Bae Cho et al. [3] proposed an approach that searches an image with human preference and emotion using GA. They have use wavelet transform to extract image features and IGA to search the image that the user has in mind. When the user gives appropriate fitness to what he or she wants, the system provides the images selected based on the user's evaluation. They have conducted several experiments to evaluate the performance of this system. These results show that their approach allows one to search not only an explicitly expressed image, but also an abstract image such as "cheerful impression image," "gloomy impression image," and so on. However, a couple of problems remain which need to be devised better encoding methods to express the emotion of images better and apply several genetic operators to improve the performance.

S.Mummanaet al. [4] observed that finally to precise, the main feature of NPRF is to efficiently optimize the retrieval quality of interactive CBIR. On one hand, the navigation patterns derived from the users' long term browsing behaviors are used as a good support for. First, in view of very large data sets, they have scaled their proposed method by utilizing parallel and distributed computing techniques. The experimental results reveal that the proposed approach NPRF is very effective in terms of precision and coverage. Within a very short term of relevance feedback, the navigation patterns can assist the users in obtaining the global optimal results. Moreover, the new search algorithm NPRFSearch can bring out more accurate results than other well-known approaches.

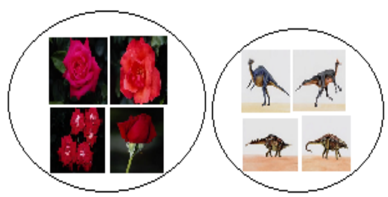
E. Tiakaset al. [5] useda promising indexing scheme MSIDX, which analyzes the image content according to the value cardinalities that appear on the dimensions of the respective descriptor vectors. The proposed scheme supports the desired functionalities of modern applications, since it is capable of performing accurate content- based retrieval in low search time and handles the dynamic operations of insertions and deletions in real-time. Through extensive experimental evaluation of MSIDX in five different collections of image descriptor vectors, they have showed the superiority of the proposed indexing scheme against other state-of-the-art hashing methods, also suitable for approximate similarity search. In particular, they illustrated how MSIDX preserves low search time, by also shattering the glass ceiling of the hashing method's limited accuracy. The proposed method is affected by the parameter, since a trade-off between accuracy and search time does exist, where the increase of results in "paying" more search time in order to achieve higher accuracy. Since the parameter is crucial for the performance of MSIDX, a detailed deterministic and probabilistic analysis was provided to set the accepted bounds of. Furthermore, based on the experimental evaluation, we verified that high accuracy is achieved, even for small values of and a further increase of is not required.

E. R. Hruschka et al. [6] present a survey of evolutionary algorithms designed for clustering tasks. Much of stress has been given on to partitioned algorithms that look for hard clustering of data. It provides an up-to-date overview of evolutionary algorithms for clustering, covering things like multi-objective and ensemble-based evolutionary clustering. It also deals the taxonomy which highlights some very important aspects in the context of evolutionary data clustering i.e. fixed or variable number of clusters, cluster-oriented or non-oriented operators, context-sensitive or context-insensitive operators, guided or unguided operators, binary, integer or real encodings, centroid-based, medoid-based, label-based, tree-based or graph-based representations, among others.

METHODOLOGY

The process flow of CBIR application

1. Start the CBIR.
2. We create the Indexing or clustering in this process we set the indexing path of images directory or database and then set the number of clusters and maximum iterations.



Cluster of flowers Cluster of Dragons

3. If we want to view the data then we done this after the indexing process in this select the number of cluster and view.
4. Select the image that wants to search in our database, using *k*-means clustering algorithm and rough filtering constant (α).

   K-means clustering is a method that is commonly used to automatically partition a data set into *k* groups. It proceeds by selecting k initial cluster centers and then iteratively refining them as follows:

   a. Each instance d_i is assigned to its closest cluster center.
   b. Each cluster center C_j is updated to be the mean of its constituent instances.

International Journal of Research In Science & Engineering e-ISSN: 2394-8299
Volume: 3 Issue:1 January 2017 p-ISSN: 2394-8280

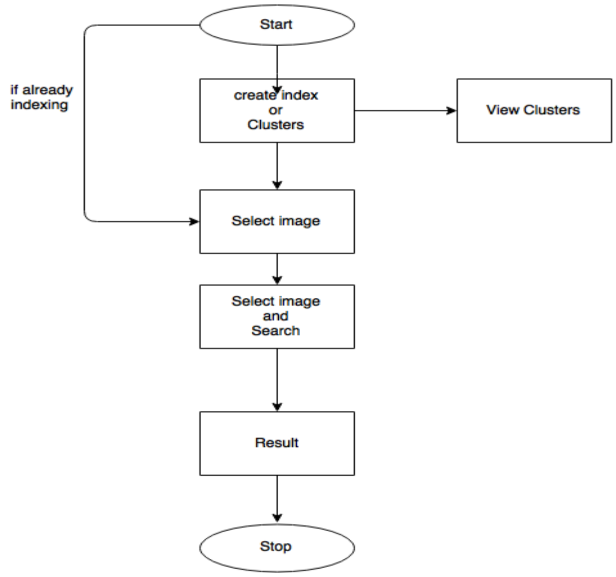


Figure 2: Process of application

Rough filtering constant (α) values indicate the similarity between the query image and the images in the database. The retrieved images are sorted according to their alpha values (degree of similarity).

5. Resultant images are similar to select image or query image.
6. Stop.

RESULT:

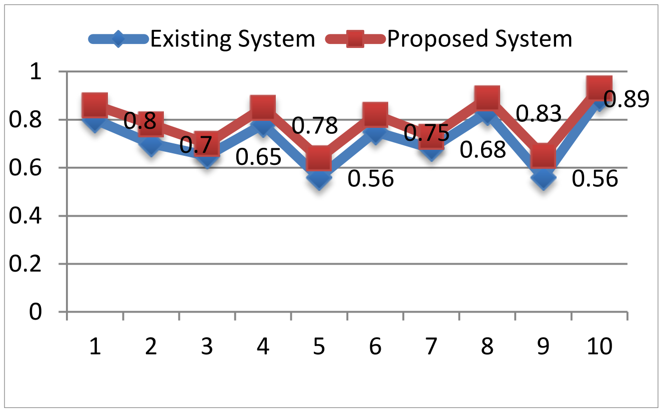
We implemented proposed algorithm based on indexing and clustering. Proposed system implemented using swing (java Desktop) application, JFreeChart library for plot graph. We calculate recall, precision and accuracy of Existing and Proposed system for different queries.

		Reality	
		Actually Good	**Actually Bad**
Prediction	**Rated**	True Positive (tp)	False Positive (fp)
	Rated	False Negative (fn)	True Negative

Recall:

Recall is the probability that a (randomly selected) relevant document is retrieved in a search. Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved. Formula for finding the value is given below

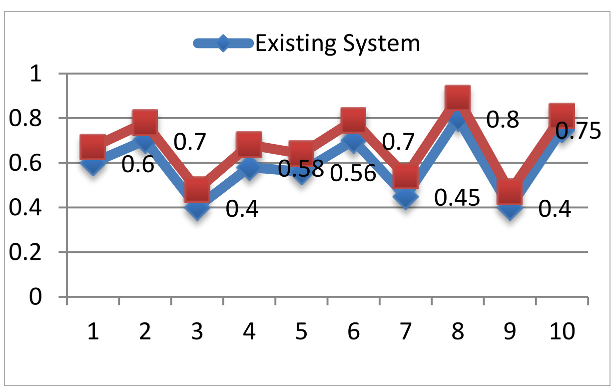
$$\text{recall} = \frac{\text{tp}}{\text{tp+fn}} \ldots\ldots\ldots\ldots\ldots\ldots (1)$$



Precision

Precision is the probability that a (randomly selected) retrieved document is relevant. Precision is the fraction of retrieved documents that are relevant to the query. We apply precision formula for finding the value is given below

$$\text{precision} = \frac{\text{tp}}{\text{tp+fp}} \ldots\ldots\ldots\ldots\ldots\ldots (2)$$



Accuracy

Accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined. We apply accuracy formula for finding the value is given below

$$\text{accuracy} = \frac{tp+tn}{tp+tn+fp+fn} \quad \ldots\ldots\ldots\ldots\ldots(3)$$

Where:

- tp=true positive value
- tn=true negative value
- fp=false positive value
- fn=false positive value

Conclusion

The proposed work broadly focuses on the efficient storage and retrieval of some image-based biometric system based on physiological characteristics. These captured biometric images often contain noise and systematic variations due to different environments and user habit. Therefore, these datasets needs proper preprocessing before the application of data mining techniques. After the preprocessing step, the content descriptors of these biometric images need to be decided, and image features needs to be extracted and represented for effective retrieval of such large biometric databases. In order to achieve these requirements, intelligent methods may be incorporated in data mining approach to narrow down the search space.

REFERENCES

[1]Manjunath, B. S., Ohm, J.R.., Vasudevan,V. V.,and Yamada, A., 2001, Color andTexture Descriptors, Circuits and Systems for Video Technology, IEEE, p. 701-715, doi .1109/76.927424.

[2]Park, G., Baek, Y., and Lee, H-K., 2002, A Ranking Algorithm using Dynamic Clustering forContent-Based Image Retrieval, Springer-Verlag, LNCS 2383, pp. 328-337.

[3]Cho, S.B. and Lee, J.Y., 2002, A Human-Oriented Image Retrieval System using Interactive Genetic Algorithm, Systems, Man, And Cybernetics, IEEE, p.452-458, doi10.1109/TSMCA.2002.802812

[4]Mummana,S., Koundinya,S., 2012, Navigation-Pattern-Based Relevance Feedback For Content Based Image Retrieval,IJESAT, Volume-2, Issue-5, p. 1241 – 1246, ISSN: 2250–3676

[5] Tiakas,E.,Rafailidis,D.,Dimou,A., and Daras,P., 2013, MSIDX: Multi-Sort Indexing for Efficient Content-Based Image Search and Retrieval,Multimedia, IEEE, 15(6), p. 1415-1430,
doi. 10.1109/TMM.2013.2247989

[6] Hruschka, E.R., and Campello, R.J.G.B., 2009, A Survey of Evolutionary Algorithm for clustering, IEEE Transections on Systems, Man, and Cybernetics, 39(2), pp. 133-155.

[7]Yuan, H., and Zhang,X.P., 2010, Statistical Modeling in the WaveletDomain for Compact Feature Extraction and Similarity Measure of Images, Circuits and Systems for Video Technology, IEEE, p. 439-445doi. 10.1109/TCSVT.2009.2031396

[8]Quellec, G., Lamard, M., Cazuguel, G., Cochener, B., and Roux, C., 2010, AdaptiveNonseparable Wavelet Transform via Lifting and itsApplication to Content-Based Image Retrieval, Image Processing, IEEE, p. 25-35, doi. 10.1109/TIP.2009.2030479

[9]Nguyen, N. V., Boucher, A., Ogier, j., and Tabbone, S., 2010, Clusters- Based Relevance Feedback for CBIR : A Combination of Query Movement and Query Expansion,Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), IEEE,
P. 1-6, doi. 10.1109/RIVF.2010.5633118

[10] Schaefer, G.,2010, Content-based retrieval from image databases: colour, compression, and browsing,Information Retrieval & Knowledge Management, IEEE, p. 5-10, doi. 10.1109/INFRKM.2010.5466891

[11]Zhao,L., and Tang, J.,2010,Content-Based Image Retrieval Using Optimal FeatureCombination and Relevance Feedback, Computer Application and System Modeling, IEEE, p.436-442, doi 10.1109/ICCASM.2010.5620791

[12]Zhang, X., Jin-ye, p., Re-ranking algorithm using clustering and relevance feedback for image retrieval, 2010, Educational and Network Technology, IEEE, p. 237-239, doi. 10.1109/ICENT.2010.5532183

[14]Lai, C., and Chen, Y. C., 2011, A User-Oriented Image Retrieval System Based on Interactive Genetic Algorithm, Instrumentation And Measurement, IEEE, P. 3318-3325, doi.10.1109/TIM.2011.2135010