



# INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 2)

Available online at: [www.ijariit.com](http://www.ijariit.com)

## A hybrid model for CBIR classification using texture feature selection

Manpreet Singh

[aseem.1726@gmail.com](mailto:aseem.1726@gmail.com)

Shaheed Bhagat Singh State Technical Campus,  
Firozpur, Punjab

Sonika Jindal

[aseem.1726@gmail.com](mailto:aseem.1726@gmail.com)

Shaheed Bhagat Singh State Technical Campus,  
Firozpur, Punjab

### ABSTRACT

*Content-based image retrieval has been an active analysis space in past years. Many alternative solutions are planned to boost the performance of retrieval, however, the massive a part of these works have targeted on sub-parts of the retrieval drawback, providing targeted solutions just for individual aspects (i.e., feature extraction, similarity measures, indexing, etc.). The implementation of the CBIR model using the Tamura texture features will be implemented along with classification method features in this project. This model will produce the efficient content-based image retrieval (CBIR) based on robust Tamura texture feature descriptors for the high performance. This model will enable the CBIR query search based upon encrypted feature descriptors using the early termination based method. The CBIR model in the project would be improved by using the multivariate feature descriptors in the perfect amalgamation to enhance the performance of the implemented model.*

**Keywords:** CBIR, Feature selection, Classification, SVM, Tamura features.

### 1. INTRODUCTION

Data mining is a method of discovering fascinating data cherish patterns; associations, changes, anomalies and vital structures, from a great amount of knowledge, keep in info, knowledge warehouse or different info repositories. Knowledge to the wide accessibility of big quantity of knowledge in electronic type and at hand want for turning such data into helpful information and knowledge for broad application as well as market research, business management, and call support, data processing has attracted a good deal of attention in info trade in a recent year.

- Data mining has popularly treated as the equivalent word of information discovery in info, though some researchers read data processing as an important step of information discovery. A data discovery method contains an unvaried sequence of the following step:
- Data improvement that handles uproarious, erroneous, missing or moot knowledge.
- Data integration, wherever multiple, heterogeneous knowledge supply is also integrated into one.
- Data choice, wherever knowledge relevant to analysis task is retrieved from info.
- Data transformation, wherever knowledge is remodeled or consolidated into from acceptable for mining by activity combination operations.
- Data mining, that is crucial method wherever intelligent ways are applied so as to extract knowledge patterns.
- Pattern analysis, that is to spot the really fascinating pattern a represent data supported some power live.
- Knowledge presentation, wherever visualization and data illustration techniques are wont to gift the deep-mined data to the user.

#### 1.1 Major tasks of information discovery

In general, data processing tasks will be classified into 2 categories: descriptive data processing and prophetic data processing. The previous describes the information set in apothegmatic outline manner and presents fascinating general properties of knowledge.

A knowledge mining system might accomplish one or additional of the subsequent data processing tasks.

- **Class/Category description:** Class description provides an apothegmatic and account of an assortment of knowledge and distinguishes it from others. The account of an assortment of knowledge is named category characterization, the comparison between 2 or additional collections of knowledge are named comparison or discrimination.
- **Association:** Association is the discovery of association relationships or correlations among a group of things. There are numerous association analysis algorithms like Apriority search, mining multiple levels, multi dimensional association, mining association for numerical.
- **Classification:** Classification analyzes a group of coaching knowledge (a set of the object whose category label is known) and constructs a model for every category supported the options within the knowledge. A decision tree or set of classification rule is generated by such a classification method. There are several classification methodologies developed within the field of machine learning, static, database, neural network.
- **Prediction:** This mining operate predicts the doable price of some missing knowledge and also the price distribution of certain attributes in an exceedingly set of objects. It involves the finding of a set of attributes relevant to the attribute of interest and predicting the worth distribution supported set of knowledge just like choose an object.
- **Clustering:** Clustering analysis is tool-established clusters embedded in knowledge wherever a cluster could be an assortment of knowledge object that's just like each other. Similarity will be specific by a user of consultants.
- **Time Series Analysis:** statistic analysis is to investigate a massive set of your time series knowledge to seek out sure regularities and fascinating characteristics, as well as explore for similar sequences and sub sequences, mining ordered patterns, cyclist, trends, and deviation.

## 1.2 Content-based Image Retrieval

In this computing age, just about all spheres of human life as well as commerce, government, academics, hospitals, crime interference, police work, engineering, design, journalism, fashion and graphic style, and historical analysis use pictures for economical services. An outsized assortment of pictures is remarked as image information. a picture information may be a system wherever image knowledge square measure integrated and keep [1]. Image knowledge embraces the raw pictures and data extracted from pictures by machine-controlled or laptop motor-assisted image analysis.

The police maintain image information of criminals, crime scenes, and take things. Within the health profession, X-rays and scanned image information are unbroken for identification, monitoring, and analysis functions. In discipline and engineering style, image information exists for style comes, finished comes, and machine elements. In business and advertising, journalists produce image databases for varied events and activities resembling sports, buildings, personalities, national and international events, and merchandise advertisements. In historical analysis, the image databases square measure created for archives in areas that embrace arts, sociology, and drugs. During a little assortment of pictures, straightforward browsing will determine a picture. This can be not the case for giant and varied assortment of pictures, wherever the user encounters the image retrieval drawback. a picture retrieval drawback is that the drawback encountered once looking out and retrieving pictures that square measure relevant to a user's request for a information. to unravel this drawback, text-based and content-based are the 2 techniques adopted for search and retrieval of a picture information.

## 2. PREVIOUS STUDY

[2010] Wang et al. [8] delineated that an answer for stratified single-keyword search relating to sure relevancy score. For the primary time this paper outline and solve the matter of effective nevertheless secure stratified keyword search over encrypted cloud information. Stratified search greatly enhances system usability by returning the matching files in a very stratified order relating to sure relevancy criteria

[2012] Penatti, Otávio AB et. [7] al. has performed a comparative study of world color and texture descriptors for net image retrieval. This paper presents a comparative study of color and texture descriptors considering the net because of the setting of use. The authors have taken into consideration the range and large-scale aspects of the net considering an oversized variety of descriptors (24 color and twenty-eight texture descriptors, together with each ancient and recently projected ones). The analysis is formed on 2 levels: a theoretical analysis in terms of algorithms complexities and an experimental comparison considering potency and effectiveness aspects. The experimental comparison contrasts the performances of the descriptors in small-scale datasets and in a very massive heterogeneous info containing quite 230 thousand pictures. Though there's a big correlation between descriptors performances within the 2 settings, there are notable deviations that should be taken into consideration once choosing the descriptors for large-scale tasks.

[2012] Abraham et al. [6] complete that to shield the privacy, users got to cipher their sensitive information before outsourcing it to the cloud. However, the normal coding schemes are inadequate since they create the applying of assortment and looking operations tougher tasks. Consequently, searchable coding systems are developed to conduct search operations over a group of encrypted information. Sadly, these systems solely enable their shoppers to perform a precise search however not search approximate, a very important would like for all these info retrieval systems. Recently, the increased attention has been paid to the approximate searchable coding systems to search out keywords that match the submitted queries or so. Our work focuses on constructing a versatile secure index that permits the cloud server to perform the approximate search operations while not revealing the content of the question trapdoor or the index content. Specifically, the foremost recently science primitive, order protective stellate coding (OPSE), has been utilized to shield our keywords. Our projected theme divides the search operation into 2 steps.

The primary step finds the candidate list in terms of secure pruning code. Specifically, we've developed 2 ways to construct these pruning codes. The second step uses a semi-honest third party to see the simplest matching keyword counting on secure similarity performs. We have a tendency to shall reveal as very little info as attainable thereto third party. We have a tendency to hope that developing such a system can enhance the use of retrieval info systems and build these systems additional easy.

[2014] Xia et al. [5] delineated that the results may come back not solely the precisely matched files, however additionally the files together with the terms semantically relate to the question keyword. Within the projected theme, a corresponding file data is made for every file. Then each the encrypted data set and file assortment is uploaded to the cloud server. With the data set, the cloud server builds the inverted index and constructs linguistics relationship library (SRL) for the keywords set. once receiving a question request, the cloud server 1st finds out the keywords that are semantically relating to the question keyword consistent with SRL. Then each the question keyword and also the donative words are accustomed retrieve the files. The result files are coming back so as consistent with the full relevancy score. Eventually, careful security analysis shows that our resolution is privacy preserving and secure below the previous searchable stellate coding (SSE) security definition. The experimental analysis demonstrates the potency and effectiveness of the theme.

[2014] Cao et al. and rule et al. [4] projected that theme for multi-keyword stratified search, wherever "Inner product similarity" is employed for result ranking. This paper for the primary time defines and solves the difficult downside of privacy-protective multi-keyword stratified search over encrypted cloud information.

### **3. PROBLEM FORMULATION**

The existing model utilizes the early rejection model, which utilizes the term frequency based methodology along with other features descriptors. The term frequency is the method to compute the repetitiveness of the objects within the given data and uses the high priority object selection based procedure, which is prone to the elimination of the useful features due to the lower frequency. Also, the feature in the existing model is based upon the Tamura texture features, which extracts the feature vector based upon the texture features only. The image data can be effectively matched using the color and low-level features, which may be obtained using the specific color based feature descriptor such as a histogram of gradients (HoG), principle component analysis (PCA), etc. For the low-level features, the matrix decomposition based many feature descriptors are utilized such as the scale invariant feature descriptor (SIFT), speeded up robust features (SURF), etc. The use of the components may improve the overall performance of the existing model.

### **4. RESEARCH GAPS**

The existing model is based upon the efficient content-based image retrieval (CBIR) based on early rejection model and robust feature descriptors for the high performance, which offers the insecure early feature matching for early rejection model, which can be improved by using the safe query system.

The existing CBIR model utilizes only one feature descriptor, which may affect the accuracy as well as the overall performance of the CBIR model.

Also, the early termination process carries a room for improvement because it calculates the term-frequency based compressed feature calculation, which reduces the overall retrieval accuracy. The term frequency does not define the overall feature set of an image but the summary of the feature only. The term-frequency must be avoided and other effective short-feature descriptor must be incorporated in order to improve the retrieval accuracy.

The existing model uses the one term-at-a-time model for the CBIR ranking. The one-term-at-a-time method scan one property of the image at once, which can be improved by one-image all-terms at once method. A quick response hybrid feature descriptor will be realized in order to improve the speed of the CBIR system.

### **5. PROPOSED MODEL**

The implementation of the CBIR model using the Tamura texture features will be implemented along with classification methods in this project. This model will produce the efficient content-based image retrieval (CBIR) based on robust Tamura texture feature descriptors for the high performance. This model will enable the CBIR query search based upon encrypted feature descriptors using the early termination based method. The CBIR model in the project would be improved by using the multivariate feature descriptors in the perfect amalgamation to enhance the performance of the implemented model.

### **6. CONCLUSION**

In this project, as the initial try, a secure linguistics enlargement primarily based similar search theme over encrypted cloud knowledge is planned. The question is submitted victimization single keyword search. The planned theme may come not solely the specifically matched files, however additionally the files as well as the terms semantically regarding the question keyword. The owner outsources the encrypted files and data set to the server. With the file data, the cloud builds the inverted index and constructs linguistics relationship library (SRL) for the keywords. The co-occurrence of terms is employed to capture the linguistics relationship of keywords within the lexicon that offers acceptable linguistics distance between terms to accomplish the question keyword extension. So as to possess higher results, the user will discard the unwanted results for future, which can ne'er be shown to him. For this purpose fuzzy linguistics connection matrix is employed. The rule is analogous and comparable to the expert mechanism of the human brain based neural network computing and has an initial learning mechanism.

## **7. REFERENCES**

- [1] Bansal, Ajay Kumar, and Swati Mathur. "CBIR Feature Extraction Using Neuro-Fuzzy Approach." In *Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing*, pp. 535-541. Springer India, 2016.
- [2] Meena, Mamta, and Vinayak A. Bharadi. "Hybrid Wavelet-Based CBIR System Using Software as a Service (SaaS) Model on Public Cloud." *Procedia Computer Science* 79 (2016): 278-286.
- [3] Ahmad, Jamil, Muhammad Sajjad, Irfan Mehmood, Seungmin Rho, and Sung Wook Baik. "Describing Colors, Textures, and Shapes for Content-Based Image Retrieval-A Survey." *arXiv preprint arXiv: 1502.07041* (2015).
- [4] Cao, N., Wang, C., Li, M., Ren, K., & Lou, W. (2014). Privacy-preserving multi-keyword ranked search over encrypted cloud data. *Parallel and Distributed Systems, IEEE Transactions on*, 25(1), 222-233.
- [5] Xia, Yu, Shouhong Wan, and Lihua Yue. "A New Texture Direction Feature Descriptor and Its Application in Content-Based Image Retrieval." In *Proceedings of the 3rd International Conference on Multimedia Technology (ICMT 2013)*, pp. 143-151. Springer Berlin Heidelberg, 2014.
- [6] Abraham, Ajith, Dharma Agrawal, Siby Abraham, Emilio Corchado, Preeti Bajaj, and Geetam Tomar. "Information and Communication Technologies.", vol. 1, issue 1, pp. 512-518, IEEE 2012.
- [7] Penatti, Otávio AB, Eduardo Valle, and Ricardo da S. Torres. "Comparative study of global color and texture descriptors for web image retrieval." *Journal of Visual Communication and Image Representation* 23, no. 2 (2012): 359-380.
- [8] Wang, Cong, Ning Cao, Jin Li, Kui Ren, and Wenjing Lou. "Secure ranked keyword search over encrypted cloud data." In *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on*, pp. 253-262. IEEE, 2010.