

A GENERALIZED GAMMA DISTRIBUTION FOR RETRIEVING MEDICAL IMAGE

M. Priscilla Dinkar¹, S.Ambika², A.Uday Kumar³

^{1,2,3}Assistant Professor, Department of ECE,

SVCET, Etcherla, Srikakulam, (India)

ABSTRACT

Efficient diagnosis plays a crucial role for treatment. In many cases of criticalness, radiologists, doctors prefer to the usage of internet technologies in order to search for similar cases. Accordingly in this paper an effective mechanism of Content Based Image Retrieval (CBIR) is presented, which helps the radiologists/doctors in retrieving similar images from the medical dataset. The paper is presented by considering brain medical images from a medical dataset. Feature vectors are to be extracted efficiently so as to retrieve the images of interest. In this paper a two-way approach is adopted to retrieve the images of relevance from the dataset. In the first step the Probability Density Functions (PDF) are extracted and in the second step the relevant images are extracted using correlation coefficient. The accuracy of the model is tested on a database consisting of 1000 MR images related to brain. The effectiveness of the model is tested using Precision, Recall, Error rate and Retrieval efficiency. The performance of the proposed model is compared to Gaussian Mixture Model (GMM) using quality metrics such as Maximum distance, Mean Squared Error, Signal to Noise Ratio and Jaccard quotient.

Keywords: *Generalized Gamma Distribution, Content Based Image Retrieval, Relevancy, Correlation, Quality metric.*

I INTRODUCTION

Brain disease is one of the most striking factors for the increase in mortality rates. In India this rates have been increasing continuously [10]. Most of the diseases causing brain tumors may be either Benign or Malignant. The classifications of these are subjected to the size of the tumors. Hence effective diagnosis and identification of these diseases will be of crucial importance. Most of the cases related the treatment of these diseases is subjected to MR imaging. MR imaging is mainly choosing because of its property of non-ionization. Along with the advantage, the grater disadvantage with these techniques is that the reports are generated by the radiologists using the visual perception of the MR images [4]. However as the number of cases of brain related diseases are increasing, the visual perception may lead to incorrect decisions. Another added factor to this is the proportionate increase of the radiologists versus the increase in the number of cases is lagging. Hence the advantage of using automated systems for effective brain image scanning and report generations are of crucial

importance. This paper highlights a methodology which helps to retrieve the relevant images from the databases. This system may help in particular to the paramedics, radiologist residing in remote areas to suggest a basic treatment for patient in residing in rural areas. So that he/she can be shifted to nearby super specialized hospitals.

Many systems have been presented in the literature based on Picture Archive and Communication Systems (PACS) [10]. Also text based retrieval systems are listed in the literature wherein the relevant text/keywords from the lab reports are retrieved against a query. However these systems have their own drawbacks since text annotation is difficult [8]. However, the usage of text will be useful to some of the patients eager to know about some related facts about their disease. Hence, in this paper we present a novel methodology of CBIR using Generalized Gamma Distribution which helps to retrieve similar images based on content as well as text. The updated equations of the model parameters are estimated using Expectation Maximization (EM) algorithm. The main advantage behind the usage of this distribution is of the fact that the human tissues are asymmetric in nature and it can handle the speckles more efficiently.

The CBIR applications are multifold. The help will be in particular to the area of medical imaging wherein based on the features, several images of relevance can be retrieved. These images will be of good information to the specialists for diagnosing/ treating the patients. Many models have been presented in this area based on a global features and local features using the entire MR image and blocks of the image [1]. A retrieval rate of about 94% was reported. In the literature, different models [5][10][3] are presented for CBIR based on segmentation techniques of region, cluster and deformable models.

Assuming that the shape and texture features play a vital role in identifying the contents more effectively, in this model identification of Malignant and non-Malignant tissues is experimented. Other models based on auto refreshing [6], wavelets [10], semantic gap [9], relevance feedback [10], Gabor filter [4], K-means algorithm [3], Genetic algorithm [8], KD-Tree [7], binary splitting [2], quad trees [10] have been proposed in the literature. However, these models lack in efficiency while retrieving the images of relevance because of several disadvantages like K-means algorithm is sensitive to initial clusters which may not be efficient for handling medical images where the data is both continuous and discrete. Models like binary splitting, KD-Tree, Genetic algorithms are subjected to parametric dependence and complexity. Hence efficient methods to retrieve the images of interest will be of a great advantage.

1.1 The related work

Models based on Gamma, Log-Normal and Nakagami distributions are also highlighted in the literature for handling asymmetric distributions. However, Nakagami distribution overrules the other distributions while characterizing the speckles in tissues [5]. Nevertheless this Nakagami distribution fails in handling larger impulse response of the speckles generated by human speckles compared to Generalized Gamma Distribution.

This paper presents a methodology using Generalized Gama Distribution. The organizing structure of the paper is as follows. The proposed methodology is presented in Section 2. The experimentation, results together with performance analysis and evaluation are highlighted in Section 3, the final Section 4 concludes the paper.

II PROPOSED METHODOLOGY USING GENERALIZED GAMMA DISTRIBUTION

The proposed methodology is depicted in the block diagram shown in Figure 1, and it will be very much useful for medical image analysis for certain remote areas, where only the minimum first aid is available. Using the developed methodology, procedures can be evaluated and necessary treatments can be derived.

These methods are deployed mainly on the brain image data collected from University of Rennes1 database. Every brain image consists of vital information such as white matter (WM), Gray Matter (GM) and Cerebro Spinal Fluid (CSF). In order to assess the damage of the brain tissues, effective methodologies are to be developed to discriminate WM, GM, and CSF.

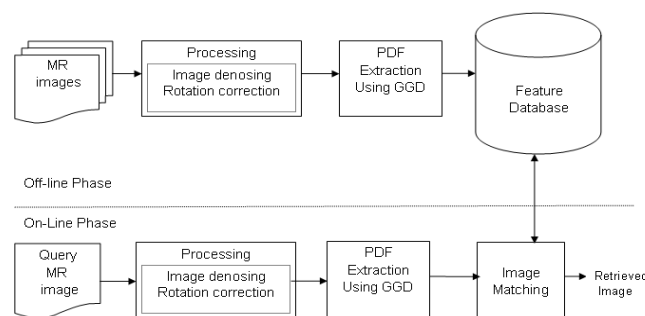


Fig 1. Block diagram of the proposed model

2.1 Generalized Gamma Distribution

Generalized Gamma Distribution is proposed for the classification of the images in this paper. In general, in Content Based Medical Image Retrieval, the damaged tissues are to be recognized. These damaged pixels behave in a different manner compared to the non-damaged pixels and the intensity levels of these pixels will be more i.e. the number of non-damaged pixels are more than the damaged pixels. And these damaged pixels will have their peaks at the origin and hence the damaged medical images will have their distributions with long tails. In order to interpret these distributions, tail distributions will be more appropriate and to cater these distributions the best suited distribution is Generalized Gamma Distribution. Generalized Gamma Distribution includes several other distributions such as Weibull distribution, Gamma distribution, Log-normal distribution, Chi-square distribution and Exponential distribution as particular cases. The main advantage of using Generalized Gamma Distribution is that, it can sensitize the medical image features in presence of noise and with minimum variance more effectively.

The Probability Density Function of the Generalized Gamma Distribution is of the form

$$f(x, k, c, a, b) = \frac{c(x-a)^{ck-1} e^{-\left(\frac{x-a}{b}\right)^c}}{b^{ck} \Gamma(k)} \quad \dots (1)$$

Where 'a', 'b' and 'x' are called gamma variants and 'c' and 'k' are called shape parameters. By varying the value of the shape parameters, the particular cases of gamma distribution can be modeled.

2.2 Expectation Maximization (EM) Algorithm for Generalized Gamma Distribution

In order to retrieve the relevant images more efficiently and effectively, the parameters of the medical images are to be estimated effectively. The parameters, Expectation Maximization algorithm is used for obtaining the final estimates. The final updates derived in [2] and are presented here.

$$\alpha_k^{(l+1)} = \frac{1}{N} \sum_{s=1}^N E^{(l)}[t_k(Z_s, \theta^{(l)})] \quad \dots (2)$$

$$c = \frac{1}{\frac{1}{\Gamma} \frac{\partial \Gamma}{\partial c} - k \log\left(\frac{x-a}{b}\right) + \frac{(x-a)^c}{b^c \log\left(\frac{x-a}{b}\right)}} \quad \dots (3)$$

$$a = x - \frac{1-ck}{\frac{1}{\Gamma} \frac{\partial \Gamma}{\partial a} - \frac{c}{b} \left(\frac{x-a}{b}\right)^{c-1}} \quad \dots (4)$$

$$b = \frac{ck}{\frac{c}{b^{c+1}} (x-a)^c - \frac{1}{\Gamma} \frac{\partial \Gamma}{\partial b}} \quad \dots (5)$$

$$k = 1 + \frac{\left[\int_0^\infty e^{-t} (\log_e t) t^{k-1} dt \right]}{\Gamma(k-1) \left[c \log\left(\frac{x-a}{b}\right) - \frac{1}{\Gamma} \frac{\partial \Gamma}{\partial k} \right]} \quad \dots (6)$$

2.3 Image matching using correlation

In this paper we present a novel methodology for identifying the relevancy between the images in the database and feature or a query image is considered for relevance. The main advantage of this methodology is that it can be used to test whether a subpart of the image belongs to the image in the database or not and it is given by

$$C(L, K) = \frac{\sum_{i=1}^n L * K}{\sqrt{\text{sum}(L_1 \cdot L_2) * \text{sum}(K_1 \cdot K_2)}} \quad \dots (7)$$

Where L is the image from the database and K is the query image (Note: K and L should be of same dimension) and we say that the query image K is the part of the image L, if the correlation approaches to near 1.

2.4 CBIR Algorithm for the proposed methodology

- Step 1: Obtain histogram of all the images in the database, in order to model the various particular cases of Generalized Gamma distribution
- Step 2: Calculate PDFs of Generalized Gamma Distribution
- Step 3: Find the relevant image based on correlation using the formula

$$C(L, K) = \frac{\sum_{i=1}^n L * K}{\sqrt{\text{sum}(L_1 \cdot L_2) * \text{sum}(K_1 \cdot K_2)}}$$

Using the correlation value, most relevant images can be extracted.

- Step 4: Retrieve similar images based on correlation value of step 3

III EXPERIMENTATION AND PERFORMANCE EVALUATION

The brain images from the image dataset are given as inputs to the Generalized Gamma Distribution proposed in section 2.1. The PDF of the images is retrieved and stored. The query image is processed using section 2.2. The relevant images that are matched are compared against the PDF.

The input query image is processed using the methodology under-laid in Section 2.3 and the relevant images are retrieved. The performance of the developed methodology is tested using metrics like Precision, Recall, Error rate and Retrieval efficiency. In order to present this paper we have considered an image dataset obtained from University of Rennes1. The relevant images from the data set against the query are matched based on Region of Interest (ROI) using the equation 7.

The methodology is tested by varying the images ranging count from 200, 300, 500 and 1000 images in the database. The next level of analysis is to identify the match between the ROI and the images considered with above sizes. Since it is mandatory to test the appropriateness of the developed methodology, we have tested the performance on different operating systems by varying image sizes as mentioned above. The time elapsed against each of the images under different operating system environments like Windows, Linux, UNIX is considered.

Personal Computer (PC) with Windows operating system with 200 images in the database is considered and the result was successful and the time elapsed is 1.8 seconds. The tests were conducted in the same environment on

databases of sizes 300, 500 and 1000 and the time elapsed for retrieval is 2.5 seconds, 2.75 seconds and 2.9 seconds respectively. The results were tested under the other operating systems with the above dataset and time taken is as follows.

The Table 1 presents the effectiveness of the model where it is tested in different Operating System environments by varying the image sizes. In both the cases of Operating Systems the model performs better retrieval accuracy in minimum time.

Table 1. Performance of the system by varying the image sizes

Operating System	Images Size	Time elapsed in seconds
Linux	200	1.85
	300	2.6
	500	2.8
	1000	2.85
UNIX	200	1.8
	300	2.65
	500	2.75
	1000	2.8

The query medical image is considered and is to be matched with the similar images in the database, to identify whether the patient is normal or suffering from disease. A local database is maintained and the query image considered is processed with these images in the local database for relevance. The Query image is processed based on Query based image technique presented in Section 2.2 and the relevant retrieved images are obtained based on relevance method and the retrieved images are obtained and are presented in Figure 2 and Figure 3.

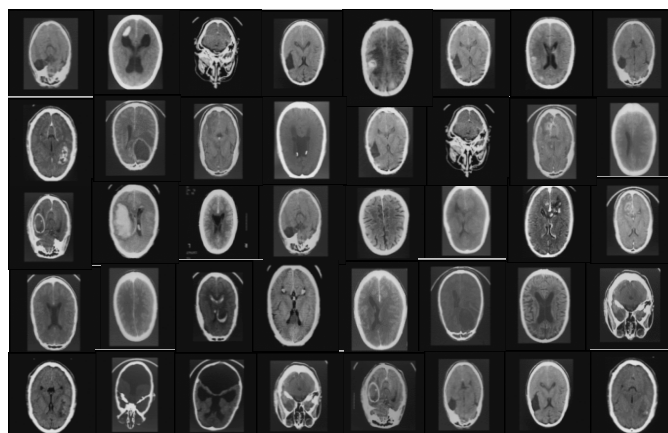


Fig 2. Image Dataset

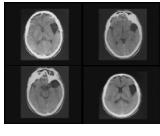
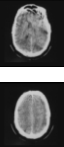
Disease	Query image	Relevant images	Non-Relevant images
Brain tumor			

Fig 3. Input Query Image versus Retrieved Relevant and Non- Relevant images

3.1 Evaluation

The outputs derived against the query processed are evaluated using metrics like Precision, Recall, Error rate and Retrieval efficiency.

3.1.1. Precision

It is the ratio of the number of relevant images retrieved to the total number of irrelevant and relevant images retrieved. It is usually expressed as a percentage.

$$\text{Precision} = (A / (A + C)) * 100;$$

A: Number of relevant images retrieved.

C: Number of irrelevant images retrieved.

A + C: Total number of irrelevant + relevant images retrieved.

3.1.2. Recall

It is the ratio of the number of relevant images retrieved to the total number of relevant images in the database. It is usually expressed as a percentage.

$$\text{Recall} = (A / (A + B)) * 100$$

A: Number of relevant images retrieved

B: Number of relevant images not retrieved

A + B: The total number of relevant images

3.1.3. Error rate

$$\text{Error rate} = \text{Number of non-relevant images retrieved} / \text{Total number of images retrieved.}$$

3.1.4. Retrieval efficiency

Retrieval efficiency = Precision = Number of relevant images retrieved/ Total number of images retrieved (If number of retrieved images > number of relevant images), otherwise

$$\text{Retrieval efficiency} = \text{Number of relevant images retrieved} / \text{Total number of relevant images.}$$

The outputs derived using the proposed metrics are presented in Table 2 and in Figure 4.

Table 2. Precision, Recall, Error rate and Retrieval efficiency based on relevant and non-relevant images

No. of relevant images in the database	No. of non-relevant images in the database	Ratio of relevance to non-relevance	Precision	Recall	Error rate	Retrieval efficiency
120	750	0.16	80	6	94	40
98	700	0.14	74	26	74	45
78	650	0.12	62	35	65	46
60	600	0.1	53	42	58	50
44	550	0.08	42	53	47	53
30	500	0.06	35	62	38	62
18	450	0.04	26	74	26	74
8	400	0.02	6	80	20	80

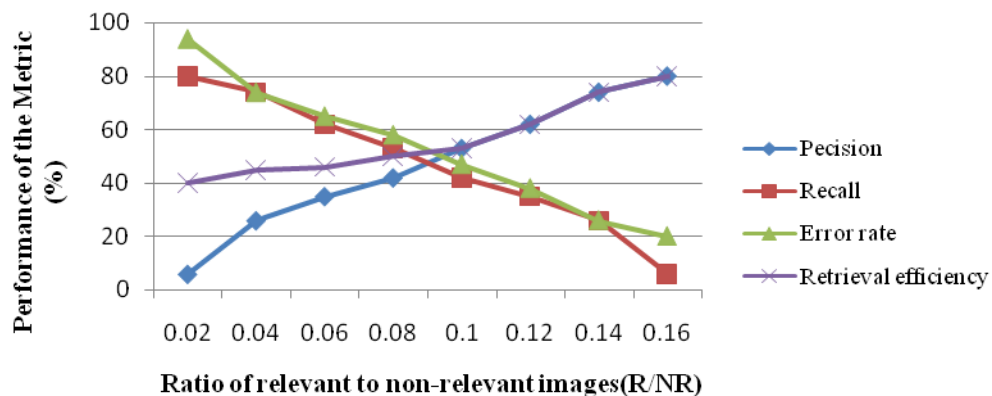


Fig 4. Variation of Precision, Recall, Error rate and Retrieval efficiency values with relevant to non-relevant images ratio

3.2 Analysis of the methodology against the time taken

Table 3 describes the recognition accuracy against the time taken. From this table it can be clearly seen that the relevant images are retrieved effectively.

Table 3 showing the accuracy of Present model

	Brain tumor	Brain Hemorrhage	Fractures
Accuracy (%)	98	96	97
Time taken in seconds	1.8	2.0	2.2

Quality metric	Formula to Evaluate
Maximum Distance	$Max\{ F(j,k) - \hat{F}(j,k) \}$
Mean Squared error	$\frac{1}{MN} \sum_{j=1}^M \sum_{k=1}^N [O\{F(j,k)\} - O\{\hat{F}(j,k)\}]^2 / \sum_{j=1}^M \sum_{k=1}^N [O\{F(j,k)\}]^2$ Where M,N are image matrix rows and columns
Signal to noise ratio	$20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right)$ Where, MAX_I is maximum possible pixel value of image, MSE is the Mean squared error
Jaccard quotient	$\frac{ X \cap Y }{ X \cup Y } = \frac{a}{a+b+c}$ Where, $a = X \cap Y $, $b = \left \frac{X}{Y} \right $, $c = \left \frac{Y}{X} \right $, $d = \overline{X \cup Y} $ and X, Y are input and output image intensities

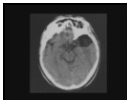
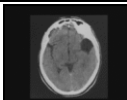
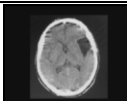
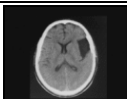
3.3 Performance Evaluation

The retrieved efficiency, in case of medical images against the query image is evaluated using Image quality metrics. The various metrics considered for the work include, Maximum distance, Mean Squared Error, Signal to Noise Ratio and Jaccard quotient. The formulas for computing the above quality metrics are presented in Table 4.

Using these formulae the performance evaluation is carried out and the proposed model is compared to GMM and the results obtained are tabulated and presented in Table 5.

Table 5. Evaluated Image Quality Metrics

Image	Quality Metric	GMM	Generalized Gamma Distribution	Standard Limits	Standard Criteria

	Maximum Distance	0.422	0.9325	-1 to 1	Closer to 1
	Mean Squared error	0.04	0.094	0 to 1	Closer to 0
	Signal to noise ratio	17.41	33.89	$-\infty$ to ∞	As big as Possible
	Jaccard quotient	0.089	0.703	0 to 1	Closer to 1
	Maximum Distance	0.221	0.912	-1 to 1	Closer to 1
	Mean Squared error	0.2404	0.2019	0 to 1	Closer to 0
	Signal to noise ratio	14.45	39.85	$-\infty$ to ∞	As big as Possible
	Jaccard quotient	0.0677	0.7921	0 to 1	Closer to 1
	Maximum Distance	0.345	0.807	-1 to 1	Closer to 1
	Mean Squared error	0.22	0.2123	0 to 1	Closer to 0
	Signal to noise ratio	19.88	39.71	$-\infty$ to ∞	As big as Possible
	Jaccard quotient	0.043	0.7143	0 to 1	Closer to 1
	Maximum Distance	0.224	0.971	-1 to 1	Closer to 1
	Mean Squared error	0.24	0.1192	0 to 1	Closer to 0
	Signal to noise ratio	21.42	37.41	$-\infty$ to ∞	As big as Possible
	Jaccard quotient	0.045	0.874	0 to 1	Closer to 1

IV CONCLUSION

In this paper an efficient methodology for brain MRI image retrieval against the query image is proposed. This proposed system helps to retrieve the relevant images from the image dataset effectively. The proposed model exhibits good recognition rates of above 90%. The evaluation of the developed model is carried out by comparing with the existing models based on GMM, by using quality metrics. The results show that, this developed algorithm outperforms the existing algorithm. This methodology suits very well in applications wherein assistance is needed for the doctors at remote areas.

REFERENCES

- [1] Ahmad Fauzi M. F., and Ahmad W., "Efficient block based matching for content based retrieval of CT head images," *In Proceedings of the IEEE 10th workshop on multimedia signal processing*, Cairns, Qld, pp. 142-147, 2008.
- [2] Babu G. P., and Murty M. N., "A near-optimal initial seed value selection in k-means means algorithm using a genetic algorithm," *Pattern Recognition Letters*, vol. 14, no. 10, pp. 763-769, 1993.
- [3] Linde Y., Buzo A., and Gray R., "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84-95, 1980.
- [4] MadhusudhanaRao T. V., PallamSetty S., and Srinivas Y., "Content based Image Retrievals for Brain Related Diseases," *International Journal of Computer Applications*, vol. 85, no. 11, pp. 0975-8887, 2014.

- [5] Ng H. P., Ong S. H., Foong K. W. C., Goh P. S., and Nowinski W. L., "Medical image segmentation using k-means clustering and improved watershed algorithm," *In Proceedings of the IEEE Southwest symposium on image analysis and interpretation*, Denver, CO, pp. 61-65, 2006.
- [6] Prasad B. G., and Krishna A. N., "Statistical texture feature-based retrieval and performance evaluation of CT brain images," *In Proceedings of the 3rd international conference on electronics computer technology*, Kanyakumari, India, pp. 289-293, 2011.
- [7] Sanjay S., and Trimbak S., "An Effective Mechanism to Neutralize the Semantic Gap in Content Based Image Retrieval (CBIR)," *The International Arab Journal of Information Technology*, vol. 11, no. 2, pp. 124-133, 2014
- [8] Shahabi C., and Safar M., "An experimental study of alternative shape-based image retrieval techniques," *Multimedia Tools and Applications*, vol. 32, no. 1, pp. 29-48, 2007.
- [9] Trpovski Z., "Content based image retrieval: from pixels to semantics," *In Proceedings of the 9th symposium on neural network applications in electrical engineering, NEUREL*, Belgrade, pp. 7-12, 2008.
- [10] World Health Organization Cancer Fact Sheets, Available at <http://www.who.int/mediacentre/factsheets/fs297/en/index.html>, Accessed