International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Exploration of Cognition Impact: An Experiment with Cartoon Retrieval through Indexing

Prem Singh M, Sharath Kumar Y H

Submitted: 02/06/2024 **Revised:** 20/07/2024 **Accepted:** 28/07/2024

Abstract. Large-scale Cartoon Image retrieval systems are able to calculate Image-to-Image similarity and accommodate differences in timing, key and tempo. Simple vector distance measure is not adequately powerful to perform Cartoon Image recognition, and expensive solutions such as dynamic time warping do not scale to millions of instances, making Cartoon Image retrieval inappropriate for commercial-scale applications. In this work, the content-based music features of Images are used as input and transformed them into vectors by using the 2D Fourier transform approach. By projecting the Images into a fusion vector of PCA and LDA, the efficient KD Tree and R Tree Indexing algorithm is used to compare the similarity of Images and retrieve the most similar Images from the large-scale database. The proposed system is not only efficient enough to perform scalable content-based music retrieval but can also develop the potential of making similar music recognition applications faster and more accurate.

Keywords: *CNN, MFCC, Deep Learning*

1 Introduction

Cartoon Image Retrieval (CSR) generally refers to the problem of identifying different interpretations of the same musical work. Since 2006, this challenge has been enlisted in the centralized yearly Music Information Retrieval (MIR) evaluation sessions known as the MIR EXchange (MIREX). Ever since, several systems for this nature of task have been submitted and evaluated on a fixed, but undisclosed dataset. As the results obtained by these systems are expressed in the form of general performance numbers, no information is provided that could reveal the influence of specific CSR system component design choices and the composition of the evaluation dataset on the obtained retrieval results. Although CSR appears to be more specific than classic music genre retrieval, Cartoon Images still span a broad range of types, each with their own variants and invariants, posing specific challenges on the design of the CSR system. In order to validate design motivations and identify which system aspects are most critical for performance results, it is necessary to consider CSR systems as combinations of general system components and review performance with respect to these components. Additionally, the design of the evaluation dataset is critical for obtaining true insight into the performance of CSR systems

Many technologies such as data mining, music structure segmentation, machine learning approaches, chroma feature extraction, and content based similarity approaches have been employing for identifying similar twin Images in Indian different languages. Prime aim of the effort is to identify the similar Identical Grid-Images in Indian languages through the technique, further it help toobtain efficient and accurate result. Own dataset of certain number of Images can be exercise to test and train the model. By analyzing techniques that are previously employed to detect Cartoon Images, new efforts are proposed and are applied to identify the twin Images. The proposed work will overcome

Department of Computer Science, Government College (Autonomous) Mandya, Karnataka, India

premmsingh1973@gmail.com

Department of Information Science and Engineering, Maharaja Institute of Technology Mysore, Karnataka, India

sharathyhk@gmail.com

the limitations of the existing system such as Retrieval, Classification and provides the better results. Consequently, obtained results will be helpful to solve the copyright issues while making use of already existing music pattern.

2 Related Works

Interestingly and recently, in the area of Cartoon Image identification, there has been a considerable amount of new approaches [2, 3] try to handle different issues. The typical goal is trying to effect new algorithms or combinations of them in order to improve the results in comparison, but the recent focus by most researchers has been diverted towards scalable strategies [3, 4]. The most common way to calculate the similarity between two different Images is with alignment-based methods and they have shown to be able to produce good results1 (75% MAP in MIREX'2009). However, these methods are computational expensive and, when applied to large databases, they can become impractical: the best performing algorithm [1] in MIREX'2008 implemented a modified version of the Smith-Waterman algorithm and taken approximately 104 hours to compute the results for 1,000 Images. In other words, the algorithm exercised to the Million Image Dataset (MSD), the estimated time would be of 6 years [5]. Martin et al. [3] suggest the use of Basic Local Alignment Search Tool (BLAST), a bioinformatics sequence searching algorithm, as an alternative to dynamic programming solutions. The data is indexed based in similarity between Images, and to compute the similarity value, the best subsequences are chosen, and then compared. Khadkevich et al. [4] extract information about chords and store them using Locality-Sensitive Hashing (LSH). Bertin-Mahieux et al. [3] adopt the 2D Fourier Transform Magnitude for large-scale Cartoon detection, further improved by Humphrey et al. [6], who modified the original work to use a sparse, high-dimensional data-driven component, and a supervised reduction of dimensions. The author Balen et al. [2] extract high-level musical features that describe harmony, melody, and rhythm of a musical piece. In other words those descriptors are stored with LSH, which allows retrieving the most similar Images. Lu and Cabrera [7] employed hierarchical K-means clustering on chroma features to find audio words or centroids. A Image will then be represented by its audio words. Moreover similarity with other Images will be determined by audio words share with the same location.

Outside the field of large-scale Cartoon identification, several solutions regarding distance fusion [4] have been suggested. Salamon et al. [7] extract the melodic line, the bass-line, and HPCP 12-bins for each Image. They explore the fusion of those features in order to disCartoon which results in the best performance. Distance fusion is also the main focus in the work of Degani et al. [8], where they propose a heuristic for distance fusion. The work consists of normalizing all values to [0, 1], computing a refined distance value, and produce a single matrix of results.

The conventional approaches described above calculate the distance between a query and the Images to be compared, and determine that the Image with the nearest distance is highly likely to be a Cartoon. Since process is separate from each query, the result from "another version of the same Cartoon" cannot be taken into account. If it is possible, Images with different lengths can be represented in the same space. Furthermore, if similar/dissimilar Image pairs are known, the metric to measure the Image distance can be optimized, rather than using the Euclidean distance. However, instead of taking the distance matrix directly to rank the similarity, here first perform a transformation using PCA and LDA to rearrange each Image in the high-dimensional space. Subsequently, the distance metric is learned from Image pairs in the new representation and their labels. Similarly select core Images with diverse musical properties and apply them for both embedding and training. In summary, the approach assumes that the distance between the core set and each Image can be a discriminating feature to easily group the same Cartoons. The conceptual illustration of this new representation is shown in Figure 1. Here, effort is to examine whether the K-D tree and R-tree can be effective to retrieve the similarity between Images. Additionally, the best performance is revealed over Cartoon Image retrieval by applying the PCA and LDA to the matrix generated.

3 Proposed Work

Here, an effort to develop a method to build an efficient Image retrieval strategy which can retrieve the similar pattern for a given query. Firstly, the preprocessing pillar attempts to convert image into features for Image, and reduce those through PCA and LDA. Further, reduced features are fusion to a vector. Eventually, tree-indexing structure such as KD-tree and R-tree are exercised to effectively retrieve the more relevant Images.

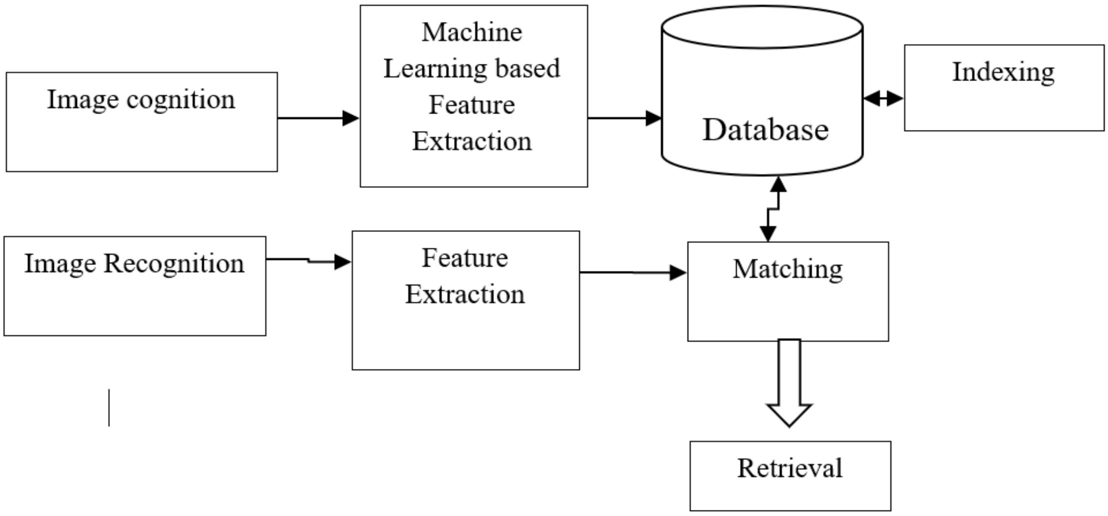


Figure 1: Proposed work

3.2.1 Feature Extraction

In this proposed method the given manuscript is pre-processed. The features like SIFT and SURF are extracted then clustered by using clustering methods, which described following subsection.

3.2.1.1 Scale Invariant Feature Transform

Scale-Invariant Feature Transform (SIFT) [38][39] is normally used to depict neighborhood locales from a picture in a scale and rotational invariant way. More often than not, SIFT alludes to a two-advance procedure including key points identification (utilizing for instance the Difference of Gaussian (DoG) strategy) and calculation of SIFT descriptors around these key points. The initial step can any way be supplanted by some other technique, for example, Harris indicator or a straight forward standard framework of key points. Given a key point (arranges), its scale (characterizing the zone canvassed by the descriptor in the picture) and the principle over whelming direction of the angle inside that zone, neighborhood slope histograms were inspected in eight ways on a 4 × 4 frame work. A 128-dimensional SIFT descriptor was then framed by amassing the 16 histograms of angles in the frame work. At long last, standardization is frequently applied on the subsequent vector.

The subsequent highlights are known to be scale and turn invariant, which implies that a pivoted and scaled picture ought to give fundamentally the same as SIFT highlights than those figured on the first picture. For characterization errands, it enables more robustness to these changes. Strikingly, on account of standard key point location just, tests yield fundamentally less keypoints than symptomatic documents, which bodes well since key points chiefly respond to venation and edges while manifestations likewise trigger numerous different key points at various scales. The scale space for the j^{th} and $(j+1)^{th}$ octave is shown in figure 2. The orientation histograms are relative to the key point orientation as shown in figure 3.

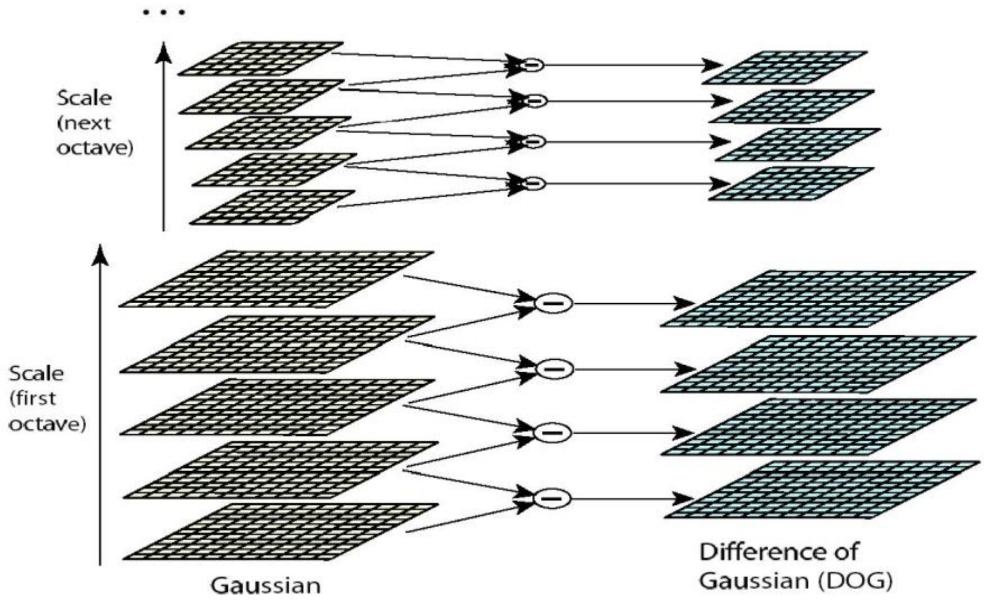


Figure 2. The Gaussian convolved images at different scales, and the computation of the Difference-of-Gaussian images

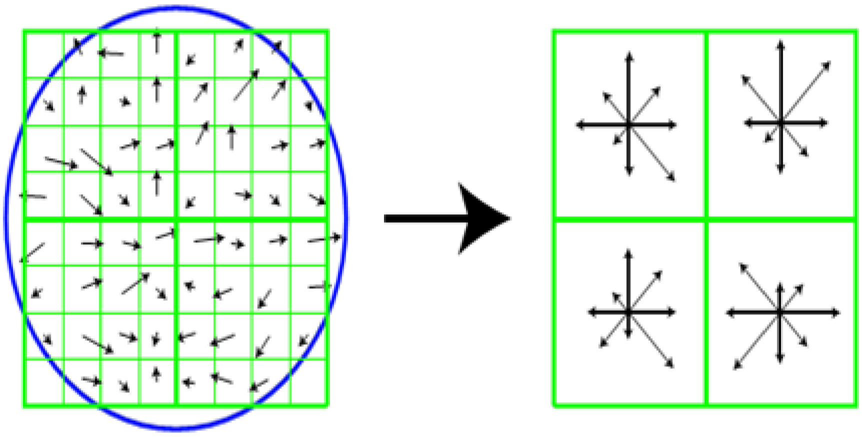


Figure 3. Image gradients and Key point descriptor

3.2.1.2 Speeded Up Robust Features

Speeded Up Robust Features (SURF)[40][41] is a scale invariant and rotation invariant interest point detector and descriptor. This algorithm has been used in face and ear [67] based personal identification systems. This method used for recognizing hand written Nandinagari character [50] and is used in the proposed work because it provides good distinct features and is robust to scale, rotation, illumination and view point changes. This algorithm uses a key point detector and descriptor method which is explained as below.

1. Detecting Keypoints with Fast-Hessian
2. Extracting SURF Descriptor
3. Orientation Assignment
4. Descriptor Components

3.2.2 Retrieval using Bag of Features

The bag of features provides a concise encoding scheme to represent a large collection of images using a sparse set of visual word histograms. The following steps outline the procedure:

- Select the Image Features for Retrieval
- Create a Bag Of Features
- Index the Images
- Search for Similar Images

In recent past, Bag-of-Features (BoF) model is getting popularity in the area of image retrieval, also known as object retrieval, based on local feature, due to the scalability of retrieval system. In this model, local features are trained to build visual vocabulary in advance. This vocabulary is then used to quantify the local features of image, the similar local features are presented in approximation as tier cluster Centre, as visual word. The quantization of visual word affects the retrieval result strongly in the image retrieval based on local features. Moreover, there is a crucial effect of pre-trained bag-of-features on the quantization of visual word.

In BoF based image retrieval system, to improve the visual word training, a k-means clustering algorithm is used. The distribution of features data on each dimension is analyzed and the distance method, which is used to partition the data space in high- dimensional indexing according to the data distribution adaptively, is combined to obtain the initial clustering centers. Finally the visual vocabulary is obtained by using k-means algorithm to cluster the feature data and train the visual words.

4 Dimensionality Reductions

In many applications such as machine learning and data mining, one is often challenge with very high dimensional data. High dimensionality increases both time and space requirements for processing the data [18]. In the presence of many irrelevant and redundant features, learning methods tend to over-fit and become less predictable. A common way to resolve this problem is by dimensionality reduction criteria. Generally, dimensionality reduction can be achieved by either feature selection [8] or by subspace learning. Moreover, modern applications have widely expanded the use of complex and high dimensional data. The enormous, high dimensional image datasets generated by domains of science and industry to demand new techniques for dimensionality reduction, feature extraction, analysis, and visualization.

Further, dimensionality reduction aims to translate high dimension to a low dimension representation such that similar input objects are mapped to all nearby points. However, most of the existing dimensionality reduction techniques have few limitations. Firstly, they do not produce a function from the input to manifold that can be applied to new points, whose relationship to training points are undetermined. Secondly, the existence of distance metric in the input space.

4.1 Principal Component Analysis (PCA)

Generally, local features are able to contribute and represent to certain extent possible. In other words, robustly representative features through broader features such as global features are essential. In this context, to erect those features through PCA [17] is the suitable way. On the other hand, it is a powerful technique for extracting the global structure from the high dimensional dataset to reduce the dimensionality and to extract the abstract features of the cartoon images. Thus, to identify the patterns and highlights the similarities. PCA is an effective effort for the dimensionality reduction and aims in reducing the dimensions of an n × p matrix X. It refers to the principal components, and the subsequent use of these components in understanding the data. On the other hand, PCA suffers from a number of shortcomings [17], such as its implicit assumption of Gaussian distributions and its restriction to linear combinations.

Algorithm

Step 1: Subtract the mean from each of the data dimensions. This produces a data set whose mean is zero.

Step 2: Calculate the covariance matrix

$$C^{m\times n} = \left(C_{i,j}, C_{i,j} = Cov\left(Dim_i, Dim_j\right)\right)$$

where $C^{m\times n}$is a matrix, each entry is the result of covariance between two separate dimensions.

Step 3: Calculate the eigenvectors and values of the covariance matrix.

Step 4: Choose components and form a feature vector: once eigenvectors are found from the covariance matrix, he next step is to order them by eigen value, highest to lowest. The number of eigenvectors will be the number of dimensions of the new data set. The objective is to construct a feature vector

FeatureVector = (eig_1, eig_2,....,eig_n)

Step 5: Derive the new data set via transpose of the FeatureVector and multiply it on the left of the original data

4.2 Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) is a classical technique in pattern recognition [17], it is used to find a linear combination of features, which distinguishes or separates two or more classes of the objects, and the resulting combination can be used as a linear classifier.

LDA is supervised subspace learning method based on the Fisher Criterion [19]. Fisher criterion plays a vitalizing role in dimensionality reduction, which aims in finding a feature representation by which the within-class distance is made minimum and the between-class distance is maximum.

Further, it aims to find a linear transformation W∈$R^{d\times m}$that maps $X_i$in the d-dimensional space to a m-dimensional space, in which the between class scatter is maximized while the within-class scatter is minimized, i.e.

$$argmax_w tr\left((W^T S_w W)^{-1}(W^T S_b W)\right) \tag{1}$$

Where $S_b$and $S_w$are the between-class scatter matrix and within-class scatter matrix respectively, which are defined as:

$$S_b = \sum_{k=1}^{c} n_k(\mu_k - \mu)(\mu_k - \mu)^T \tag{2}$$

$$S_w = \sum_{k=1}^{c} \sum_{i\in c_k}^{1} (x_i - \mu_k)\,(x_i - \mu_k)^T \tag{3}$$

Where$C_k$is the index set of the k-th class, $\mu_k$and$n_k$are mean vector and size of k-th class respectively in the input data space, It is easy to show that Eq. (1) is equivalent to:

$$argmax_w tr\left((W^T S_t W)^{-1}(W^T S_b W)\right) \tag{4}$$

Where$S_t$is the total scatter matrix, defined as follows,

$$S_t = \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T \tag{5}$$

when the total scatter matrix $S_t$is non-singular, then the solution of Eq. (4) consists of the top eigenvectors of the matrix $S_t^{-1}S_b$corresponding to nonzero eigenvalues.

Algorithm

Step 1: Compute the d-dimensional mean vectors for the different classand datasets.

Step 2: Compute the scatter matrices between-class and within-class scatter matrix.

Step 3: Compute the eigenvectors(e_1, e_2, ...,e_d) and corresponding eigenvalues (λ_1, λ_2, ..., λ_d) for the scatter matrices.

Step 4: Sort all the eigenvectors by decreasing eigenvalues and then choose k-eigenvectors with the largest eigenvalues to form a $d\times k$-dimensional matrix W in which every column represents an eigenvector.

Step 5: Use this $d\times k$ eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the equation $Y = X \times W$

5. KD-TREE

There is a dearth of general approach to effectively retrieve the Images from the large databases. Most of the current approaches are designed to handle a large dataset. Here the size of the database is taken in to consideration. The KD-tree (K- dimensional) indexing schema supports the range search with good pruning and lays its efficiency from the point of search time.

Construction

A special case of the binary space partitioning trees is KD-tree, which is constructed in a recursive fashion. At the root, all the data points are split into two equal halves by a partition hyper plane. Then each half is assigned to one of the child node and is recursively splitted to create a balanced binary tree. The leaf node may contain a single point or more than one point in different implementations. Each node in the constructed KD-tree corresponds to a cell in R_k , bounded by a set of partition hyper planes. A partition hyper plane is perpendicular to a partition axis and it is decided by a partition value. The partition axis in conventional KD-tree is the coordinate axis which has the greatest variance and partition value is the median of the projections of all the data points through the partition axis.

Search

A top-down searching is done from the root to leaf nodes to find the nearest neighbour of a query point. At each internal node, it is required to check which side of the partition hyper plane the query point lies, and then the associated child node is accessed accordingly. The descent down process requires the comparisons of $\log_2 n$ height of the KD-tree to reach a leaf node. The data points associated with the first leaf node becomes the first element for the nearest neighbour, which may not be the true nearest neighbour. It must be follow a process of backtracking procedure or iterative search to other leaf nodes for searching the better elements. The widely used scheme with high chance to find the true nearest neighbour is a priority search based on priority queue, in which all the cells are searched in order of distance from the query point. The search process will terminate when there are no more cells within the distance which is defined as the best point.

The nearest neighbour search in the high-dimensional may require traversing a very large number of nodes and processing the cost linear time. Alternatively, an approximate nearest neighbour search is usually performed through an advanced search termination scheme. Even after searching for required number of nodes or if the distance from the closest cell to the query exceeds $\delta = d(\mathbf{x}_p, \mathbf{x}_q) / (1+\varepsilon)$, where$\mathbf{x}_q$is the query point, $\mathbf{x}_p$is the nearest neighbour found so far and it is a positive termination parameter. It guarantees that there is no subsequent point to be found and can be closer to q than δ. In this manner, the search is guaranteed to have some probability to obtain the true nearest neighbour.

7.2 R-Tree

Based on the previous research, there are a number of spatial data index methods proposed. One such example of these methods is the R-tree,it is used in spatial database as a spatial access method. The aim of spatial access method is improving query performance by incorporating an additive data structure since the high query performance is one of the key features of successful retrieval systems [21].

R-tree is composed of the root, intermediate node, and leaf node. Each leaf node is created in the form of Minimum Bounding Rectangle (MBR) [22]. Leaf node does not store the actual spatial objects, but it stores the minimum bounding rectangle of the actual spatial objects. Figure 2 shows some rectangles, organized to form a corresponding R-tree. Therefore, the Region R1 contains three labeled objects and R2 with three and finally R3 with two.

R-Tree Steps: Query algorithm of R-tree is similar to B- tree. It uses top-down query from the root node, but MBR of R-tree will overlap to result in the non-uniqueness of query path. Since the B-tree uses only the address, the B-Tree results will depend on the existence of the required query. R-tree searches the database using the miles space from the required query. The user searches for a specific Image in the system database. If the system finds this data, it directly compares it with the current database, then if this data is found; finally it will retrieve the data that is in the database on the map. However, if the search data was not found in the current database or web database, there will not be any data to be given to the user.

Usage of R-Tree: The R-tree algorithm is one of the indexing algorithms. It is flexible since it is organized according to the data index structure [22]. The index can be established without the prediction of the entire spatial extent because it is a natural extension of B-Tree, perhaps has similar structure and characteristics. One of the main usages of R-Tree is to integrate with traditional relational database. Hence, many R-Tree spatial databases are chosen for spatial index. However, when the R-Tree node entries exceed M, the node must be split.

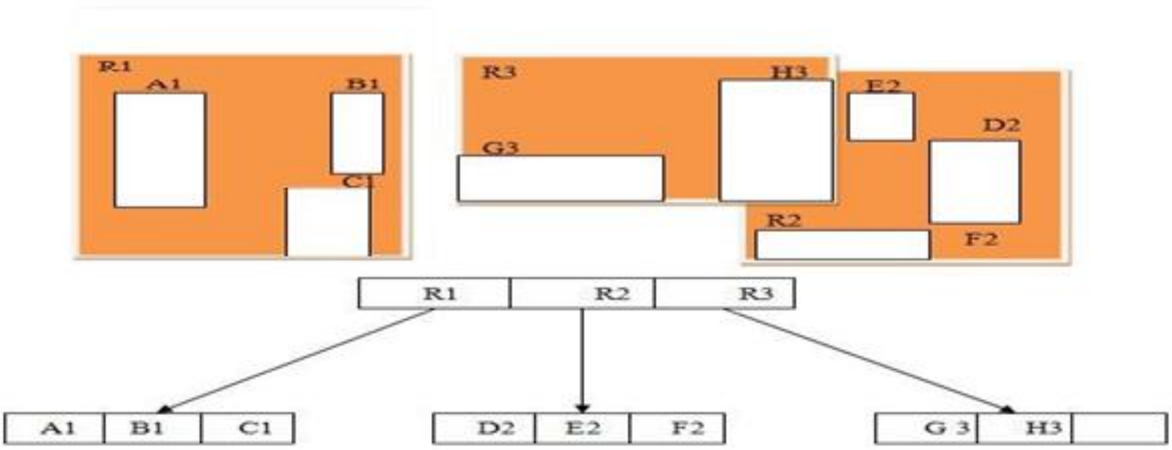


Figure 4: R-tree Rectangles

7.3 Matching

The Image have some set of features, once the feature set is obtained, it is stored in the database. However, storing in the database in an efficient manner is required such that the possible candidate list is selected the matching process should be effective. Thus, a backend tool of indexing mechanism is used, which stores the data in some pre-defined manner so that during the matching phase only a few likely candidates are selected. Hence, the effort enforce KD-tree and R-tree based approach for indexing the obtained features. The following subsection provides the overview of KD-tree indexing principle.

8 Datasets

The created dataset Cartoons30 contains 30 sets of original and Cartoon Images - spanning across genres, styles, live and recorded music, the dataset is biased towards regional languages. Most Images contain a Cartoon version however some Images contain up to three. Similarly, the extended Cartoons80 [12] proposed at MIREX 2007 to benchmark Cartoon Image recognition systems. The dataset contains 80 sets of original and Cartoon Images, 166 in total –which comprises genres, styles, live and recorded music. In other words Cartoons80 predominantly, oriented towards western music.

9 Experimentation

Here, extended the experiment over created and real datasets in order to reveal the capability of proposed criteria. The work has been implemented in a Matlab R2013a using an Intel Pentium 4 processor, 2.99 GHz Windows PC with 2 GB of RAM.

Subsequently, the reduced features vectors through PCA and LDA are combined into a single vector. However, the fusion vector is made to compute the combination of PCA and LDA as a dominant feature via dataset of Cartoon 80 and Cartoon 30. In addition, to reveal performances of KD tree and R tree indexing by varying retrieval ranking from 10 to 50. Precisely, pick Images randomly from the databases and experimentation is conducted more than five iterations. The effort outcomes are emphasizing to witness through maximum accuracy obtained in all cases. In view of further appreciation and correlated the supremacy of indexing method, the attempt has been performed on database under varying size from 30 to 70 percent of database in various instances.

Moreover, the results obtained for reduced PCA and LDA set 10, 20, 30, 40 and 50 over Cartoons 80 dataset with the appropriate training such as 30, 50 and 70 percentage plots are shown in figures 5 and 6. Further, those are tabulated for both ranking and reduction method with varying training samples. However, analysis of graphical representation can be noticed that the fusion-reduced features with top ranking retrieval achieve relatively higher accuracy in all cases. In addition, outcomes of Fusion approach has portrayed maximum accuracy when compare to other classifiers due to its computational excellence.

Further, to enhance the proposed with the reduced the features of both reduction methods are fused into one vector logical OR operation. Thus, fused vector is called as decision level fusion which contains the actual class from both the vector. The result of fusion vectors is shown in figure 7 and figure 8. Thus, to evaluate the erected effort with the difference between the maximum and minimum accuracy has calculated and tabulated in table 1 and 2 for all the reduction methods.

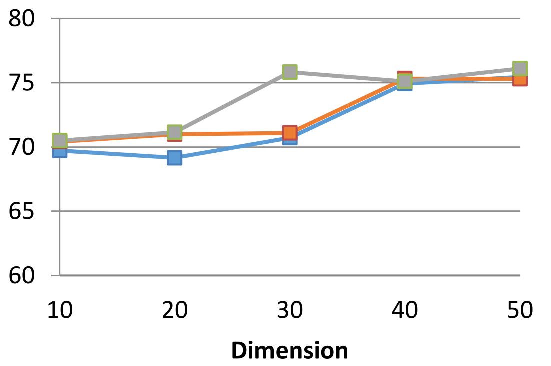
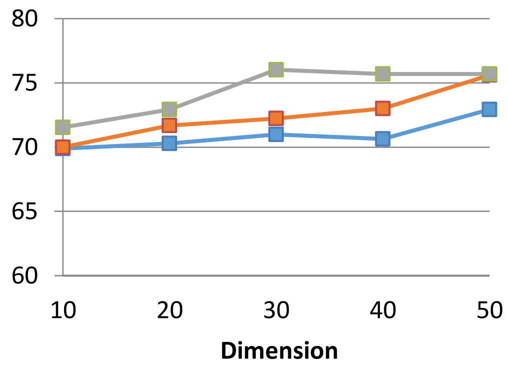
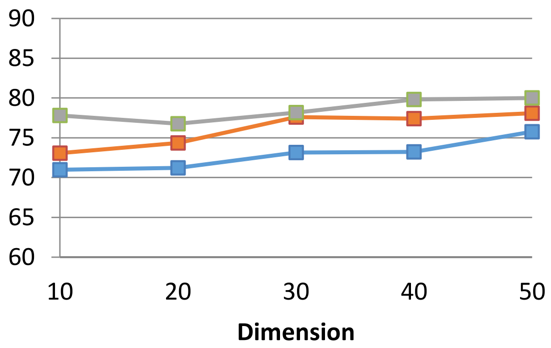
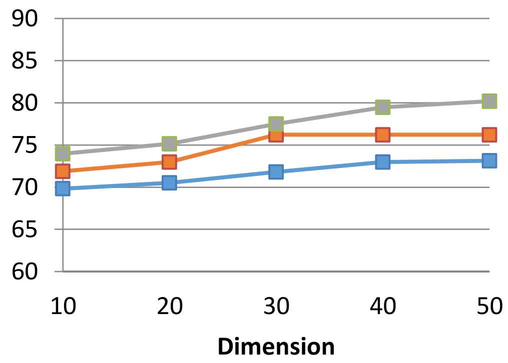
The experimental outcomes have relieving via KD tree supported for Cartoon 30 and Cartoon 80 are shown convincing performance through empirical display of accuracy. Additionally the experimentation with R-tree effort perhaps enhance the accuracy when compare to KD-tree. Furthermore, the critical observation of potentiality of fused the vector of reduced features PCA and LDA are more encourage the accuracy in all the cases of training percentage. Figure 9, 10, 11 and 12 shows the comparison of all the three methods

However, accuracy improves as features increases. Also progress saturates for even after the 50 and 70 dimension of training. Further, accuracy spectrum potential difference is 6 in both cases for PCA. In LDA effort, feature dimension emphasize the snail phase improvement in accuracy for various dimensions and saturated after 40 and hardly a matter training impact. On the other hand, accuracy patch spread value approximated empirically 11.32 and smaller value is 4. However, the smaller value indicate the feature fusion and decision fusion are predominantly support the improvement of accuracy is as shown in table 1 and 2.

10 Conclusion

The new framework of Cartoon Image retrieval of viewing the scores as features, performing feature normalization and training has led to a sizable increase in performance compared to existing Cartoon Image retrieval through exhibiting the high score. Moreover, feature standardization has made Cartoon Image retrieval move from just determining high scores to general retrieval system, while using supervised training offers further increases in performance. Despite these gains, there still exists a space for improvement. In this work, also compare the reduction strategy with retrieval indexing as KD and R tree respectively.

Summarily, here observe that when individual reduced vector is provided to indexing is not enough when compare to fusion of reduced features. Moreover, the emphasizing feature fusion and decision fusion enhance the accuracy empirically for varying the training samples. The fusion vector achieves 95% as maximum accuracy.

	KD-tree	**R-tree**
PCA		
LDA		

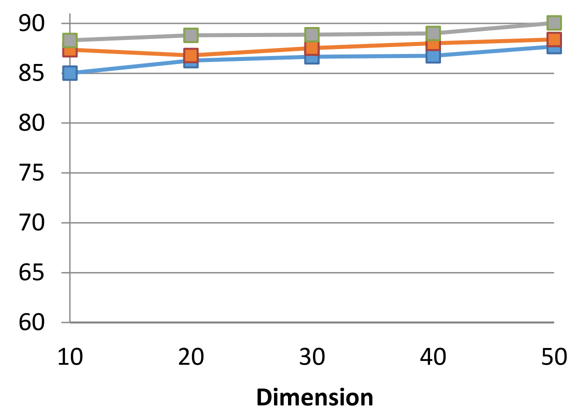
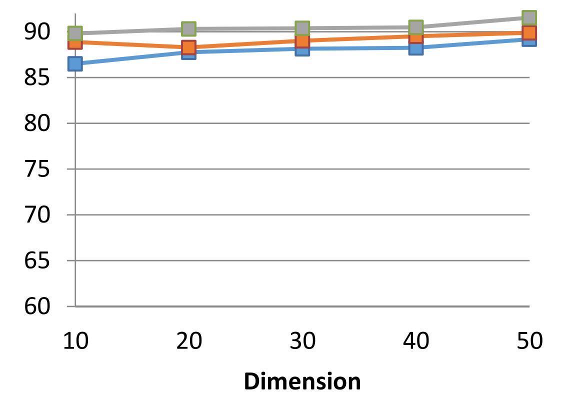
FUSION		

Figure 5: Accuracy for Cartoon 30 of Fusion Feature

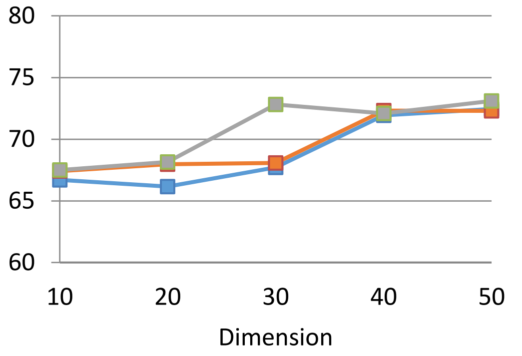
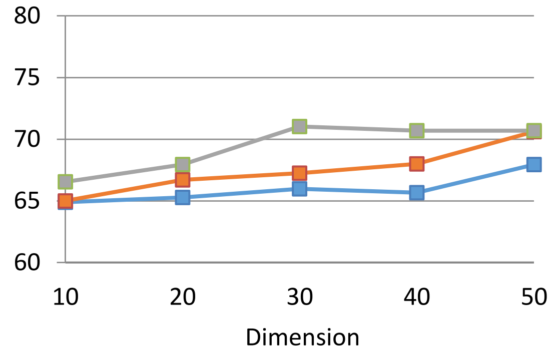
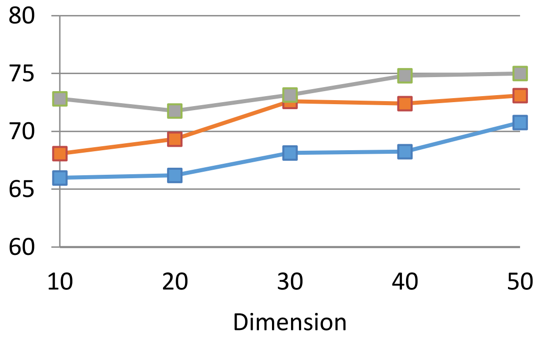
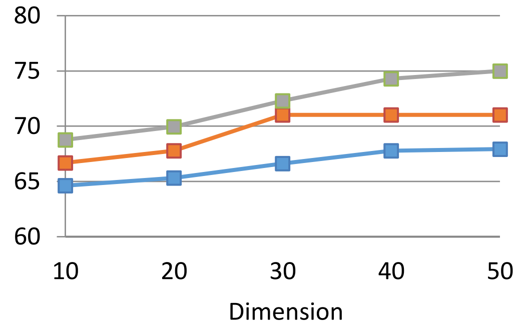
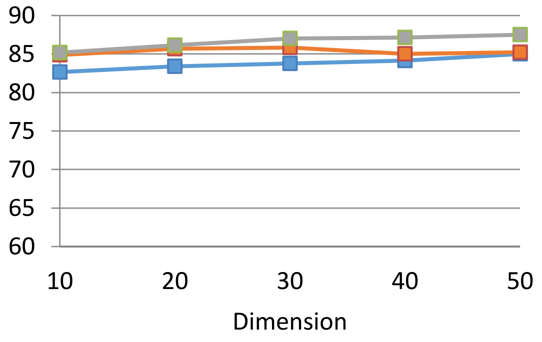
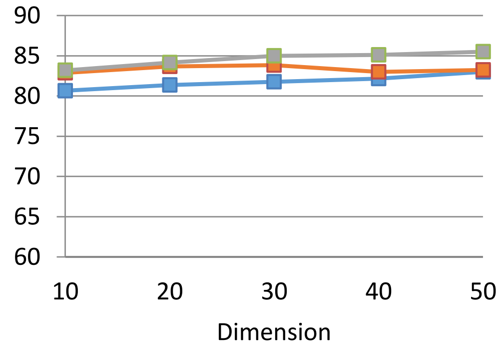
	KD-tree	R-tree
PCA		
LDA		
FUSION		

Figure 6: Accuracy for Cartoon 80 of Fusion Feature

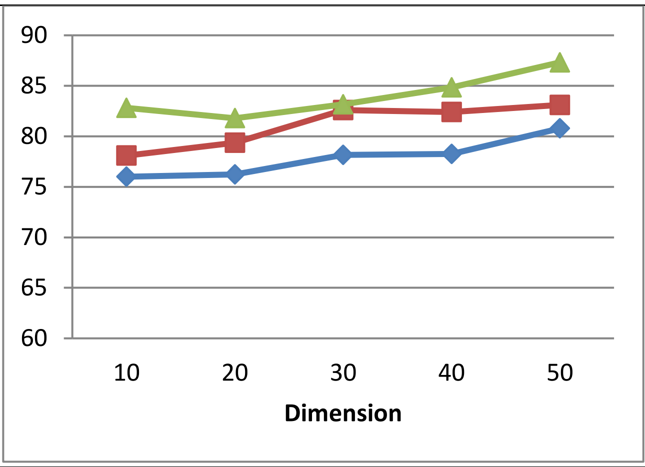
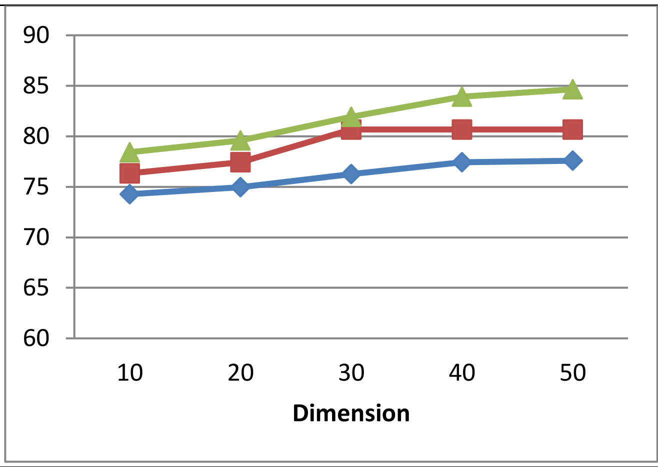
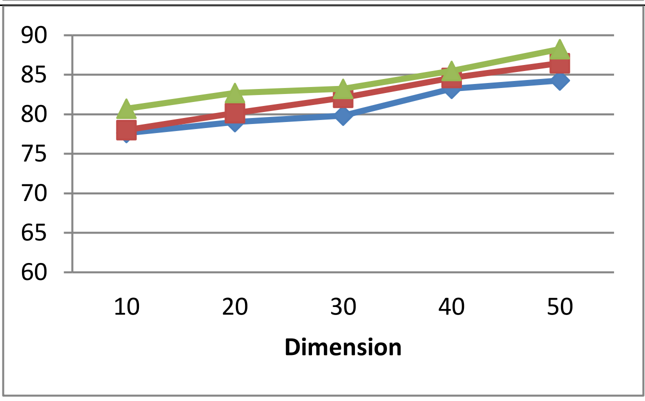
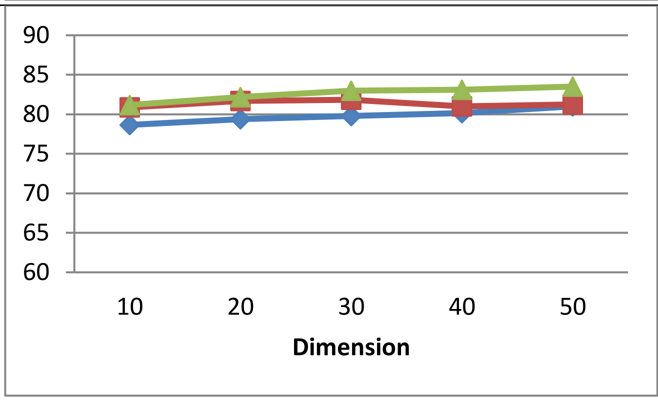
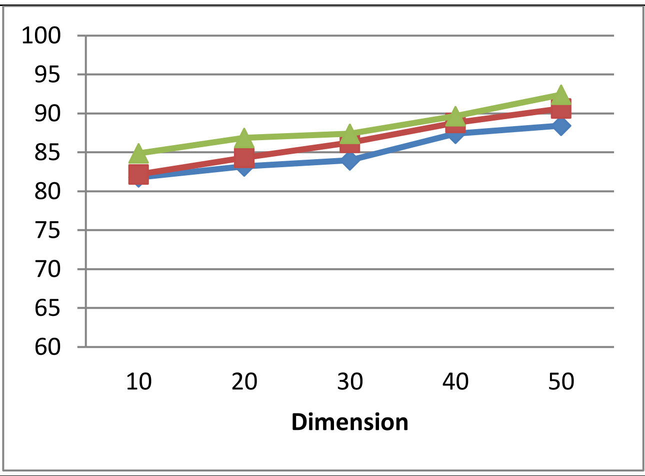
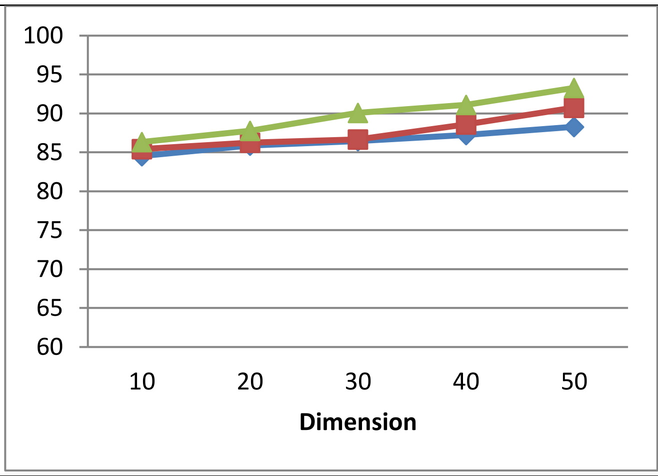
	KD-tree	**R-tree**
PCA		
LDA		
FUSION		

Figure 7: Accuracy for Cartoon 30 of Decision Fusion

	KD-tree	**R-tree**

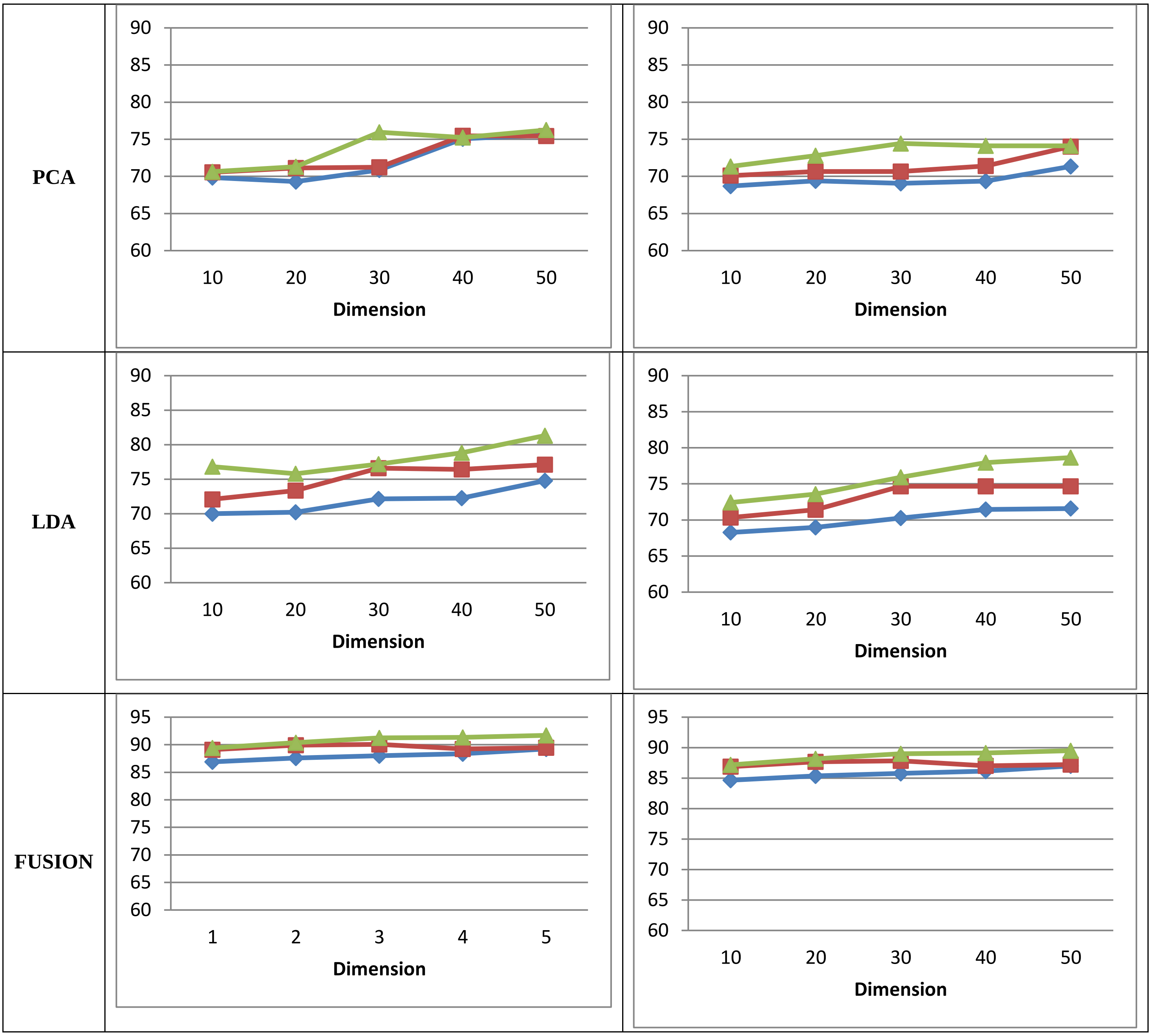


Figure 8: Accuracy for Cartoon 80 Decision Fusion

Table 1: Shows approximated accuracy difference between the lowest and highest Feature Fusion

	Cartoon 30		Cartoon 80	
	KD-Tree	R-Tree	KD-Tree	R-Tree
PCA	9.01	10.39	6.94	6.07
LDA	11.32	10.05	9.1	10.39
Fusion	4.86	5.20	4.85	4.85

Table 2: Shows approximated accuracy difference between the lowest and highest Decision Feature

	Cartoon 30		Cartoon 80	
	KD-Tree	R-Tree	KD-Tree	R-Tree
PCA	5.25	6.01	6.11	4.48

LDA	11.12	7.16	8.45	9.1
Fusion	10.44	8.14	4	4.87

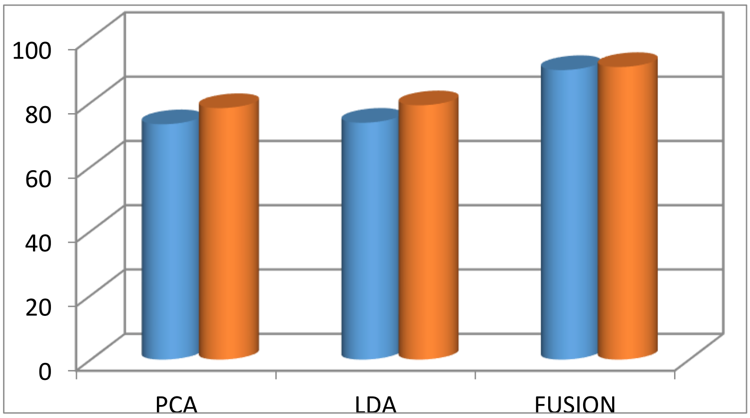


Figure 9: The approximated accuracy comparison for Cartoon 30 Feature Fusion

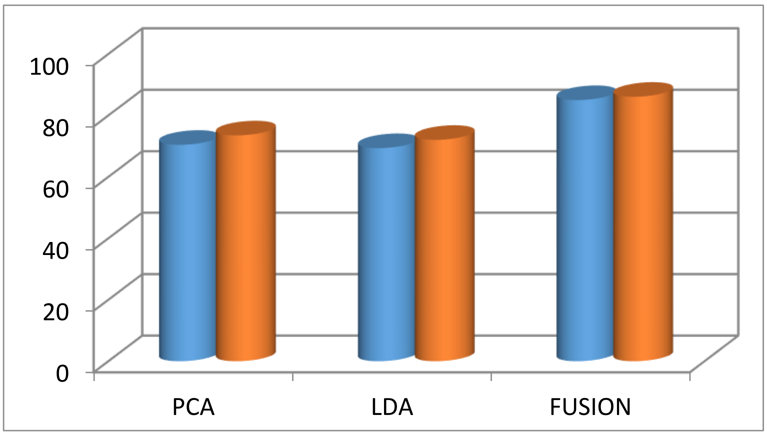


Figure 10: The approximated accuracy comparison for Cartoon 80 Feature Fusion

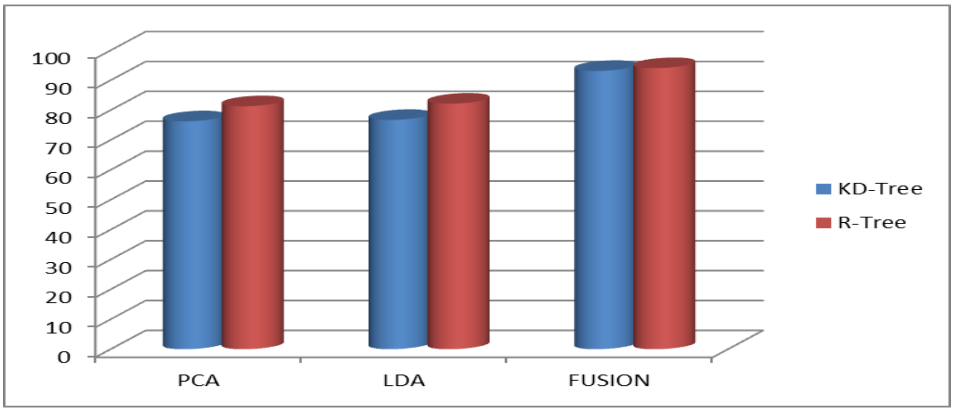


Figure 11: The approximated accuracy comparison for Cartoon 30 Decision Fusion

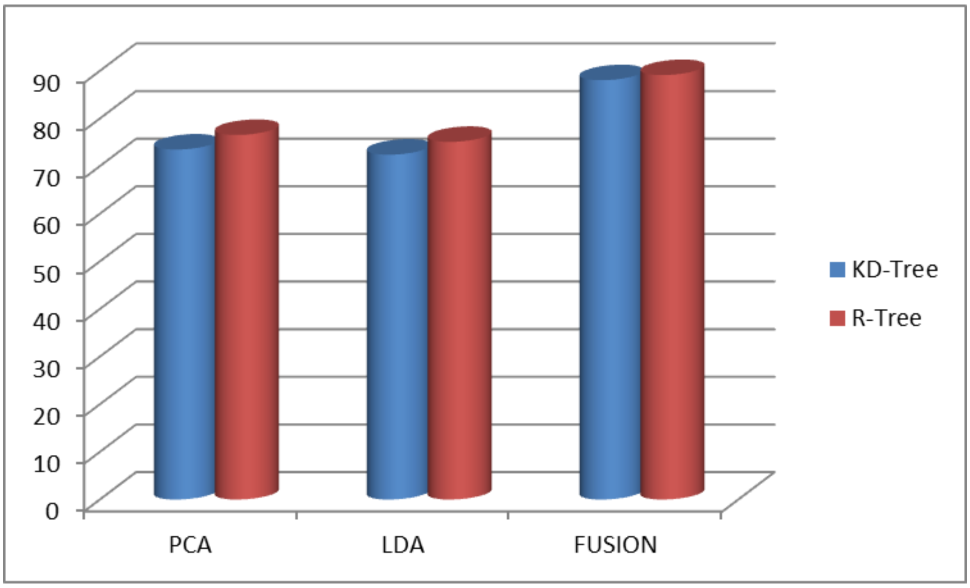


Figure 12: The approximated accuracy comparison for Cartoon 80 Decision Fusion

References

[1] M. Haseyama and A. Matsumura, "A trainable retrieval system for cartoon character images," in *Proc. ICME*, Jul. 2003, pp. 393–396.

[2] Y. Yang, Y. Zhuang, D. Xu, Y. Pan, D. Tao, and S. Maybank, "Retrieval based interactive cartoon synthesis via unsupervised bi-distance metric learning," in *Proc. ACM Multimedia*, 2009, pp. 311–320.

[3] D. Sykora, J. Burianek, J. Zara, Sketching cartoons by examples, In Proceedings of The 2nd Eurographics Workshop on Sketch-Based Interfaces and Modeling, 27–34, 2005

[4] Tiejun Zhang, D. Tao, D. Xu, J. Yu, and J. Luo," Recognizing cartoon image gestures for retrieval and interactive cartoon clip synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1745–1756, Dec. 2010.

[5] Jun Yu, Dongquan Liu, Dacheng Tao and Hock Soon Seah "On Combining Multiple Features for Cartoon Character Retrieval and Clip Synthesis", *IEEE Trans.* Cybernetics, Vol. 42, no. 5, pp.1413-1427, Oct. 2012

[6] H. Kang, S. Lee, C. K. Chui, Coherent line drawing, In Proceedings of the 5th International Symposium on Non-photorealistic Animation and Rendering, 43–50, 2007

[7] S. Belongie, J. Puzicha and J. Malik Shape Matching and Object Recognition Using Shape Contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24 (24): 509-521

[8] K bozas, Qi Han, Handan Hou, Xiamu Niu " Local Invariant Shape Feature for Cartoon Image Retrieval", IEEE Trans. 2013 Second International Conference on Robot, Vision and Signal Processing

[9] R Zhou ,Y. Yang, S. Xiang, and C. Zhang, "Neighborhood MinMax projections," in *Proc. IJCAI*, 2007, pp. 993–998

[10] D. A. Forsyth, J. Ponce. Computer Vision: A Modern Approach, Prentice Hall, 2002

[11] C. D. Juan and B. Bodenheimer, "Cartoon textures," in *Proc. Symp.Comput. Animation*, 2004, pp. 267–276.

[12] R Hu and S. Lee, "Human action recognition using shape and CLG-motion flow from multi-view image sequences," *Patt. Recog.*, vol. 41, no. 7, pp. 2237–2252, 2008.

[13] Y. Houssem, D. Zhang, G. Lu, and W. Ma, "A survey of content-based image retrieval with high-level semantics," *Patt. Recog.*, vol. 40, no. 1, pp. 262–282, 2007.

[14] T.Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Known. Data*

Eng., vol. 21, no. 9, pp. 1299–1313, Sep. 2009.

[15] Dalal, N., Triggs B., 2005. “Histogram of Oriented Gradients for human detection” in : CVPR. pp. 886-893

[16] Lowe, David G. (1999). “ Object recognition from local Scale invariant features”. Proceedings of the international conference on computer vision 2. pp .1150-1157

[17] E. Oja. *Subspace methods of pattern recognition*, volume 6 of Pattern recognition and image processing series. John Wiley & Sons, 1993.

[18] D.Gleich[Online].Available:http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=12422&objectType=FILE

[19] S. Balakrishnama, A. Ganapathiraju, *Linear Discriminant Analysis.*

[20] Deng Liqiong, Chen Danwen, Yuan Zhimin, Wu Lingda. Attributebased Cartoon Scene Image Search System, Advanced Materials Research, v268-270, p1030~1035, 201

[21] SongHai Zhang, Tao Chen, YiFei Zhang, ShiMin Hu and Ralph R. Martin. Vectorizing Cartoon Animations. IEEE Transactions on Visualization and Computer Graphics, 2009,15 (4).

[22] Yuxiang Xie, Xidao Luan, Xin Zhang, Chen Li, Liang Bai. A Cartoon Image Annotation and Retrieval System Supporting Fast Cartoon Making. 2014 IEEE 17th International Conference on Computational Science and Engineering.

[23] Mehendale, N. Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. 2, 446 (2020).

[24] Mishra, A., Rai, S.N., Mishra, A., Jawahar, C.V. (2016). IIIT-CFW: A Benchmark Database of Cartoon Faces in the Wild. In: Hua, G., Jégou, H. (eds) Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science(), vol 9913. Springer, Cham.

[25] Shukla, P., Gupta, T., Singh, P., Raman, B. (2020). CARTOONNET: Caricature Recognition of Public Figures. In: Chaudhuri, B., Nakagawa, M., Khanna, P., Kumar, S. (eds) Proceedings of 3rd International Conference on Computer Vision and Image Processing. Advances in Intelligent Systems and Computing, vol 1022. Springer, Singapore.

[26] Ming, Z., Burie, JC. & Luqman, M.M. Cross-modal photo-caricature face recognition based on dynamic multi-task learning. IJDAR 24, 33–48 (2021).

[27] W. Zheng, L. Yan, C. Gou, W. Zhang and F. -Y. Wang, "A Relation Network Embedded with Prior Features for Few-Shot Caricature Recognition," (2019) IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 2019, pp. 1510-1515, doi: 10.1109/ICME.2019.00261.