

Retrieval of flower videos based on a query with multiple species of flowers

Jyothi V K^{a, *}, V. N. Manjunath Aradhya^b, Sharath Kumar Y H^c, and D S Guru^a

^a Department of Studies in Computer Science, Manasagangothri, University of Mysore, Mysore – 570 006, Karnataka, INDIA

^b Department of Computer Applications, JSS Science and technology University, Mysore, Karnataka, India

^c Department of Information Science and Engineering, Maharaja Institute of Technology Mysore (MITM), Mandya-571438, Karnataka, India

jyothivk.mca@gmail.com, aradhya@sjce.ac.in, sharathyhk@gmail.com, and dsg@compsci.uni-mysore.ac.in

Abstract: Searching, recognizing and retrieving a video of interest from a large collection of a video data is an instantaneous requirement. This requirement has been recognized as an active area of research in computer vision, machine learning and pattern recognition. Flower video recognition and retrieval is vital in the field of floriculture and horticulture. In this paper we propose a model for the retrieval of videos of flowers. Initially, videos are represented with keyframes and flowers in keyframes are segmented from their background. Then, the model is analysed by features extracted from flower regions of the keyframe. A Linear Discriminant Analysis (LDA) is adapted for the extraction of discriminating features. Multiclass Support Vector Machine (MSVM) classifier is applied to identify the class of the query video. Experiments have been conducted on relatively large dataset of our own, consisting of 7788 videos of 30 different species of flowers captured from three different devices. Generally, retrieval of flower videos is addressed by the use of a query video consisting of a flower of a single species. In this work we made an attempt to develop a system consisting of retrieval of similar videos for a query video consisting of flowers of different species.

Keywords: Flower Region of Interest (FRoI), Linear Discriminant Analysis (LDA), retrieval of flower videos, Multiclass Support Vector Machine.

1. Introduction

There is a growth in digital video data in recent years, due to the availability of digital devices such as mobiles and cameras. Users can search and share desired videos due to networking technology, which has made developing an automated system to search and retrieve videos. And it is an interesting and active research [1]. Videos are categorized into different domains for example sports, news, surveillance, commercials, medical etc., again domain specific videos are categorized into different subcategories/classes [2]. To design a video retrieval system, two main prominent methods are used to increase retrieval performance. First is to find more appropriate features to describe videos and second is an appropriate dimensionality reduction method for selecting most discriminative features. Developing a flower video retrieval system is a domain specific with many applications. It is an application in the field of floriculture for commercial trades. Due to the development of technology in business, trader can store a large volume of videos. Instead of visiting the nurseries for their desired flowers, users can analyse the entire flower before purchasing it and its seeds. Also, they can view different species of flowers along with different variants available in each species. Further it finds applications such as medicinal, cosmetics, industrial use for the extraction of oils from flowers and decoration etc., [3]. In such cases, it is essential to develop an automated system to search and retrieve videos of flowers of user's interest. Therefore, the proposed research motivates to design an automated system for the retrieval of users desired videos of flowers. The challenges involved in flower videos to design a retrieval system are illumination: light variations differ from different angles and varied seasonal time; variation in viewpoint: videos with varying viewpoint of flowers changes appearance of the flower in size, shape, pose and rotation; cluttered background, variation among intra class and inter class, multiple instances of flowers in videos etc.

2. Related Works

Generally, the video retrieval system retrieves similar videos based on query by example. An example may be an image, keywords, sketch, object, video, video frame etc., [4]. In the literature we found retrieval of videos based on an object [5], frame [6], video [2, 7-9], keywords [10]. For the retrieval of videos the features and algorithms such as optical flow tensor and Hidden Markov Models (HMMs) [7], the multi-modal spectral clustering and

ranking algorithm [8], block wise intensity comparison [2], Scale Invariant Feature Transform (SIFT) [11], Bag-of-Features [12], dynamic weighted similarity measure with color and edge descriptors [9] are used. When a set of features are used to represent of a video, then the dimension of features may be high. If the dimension of the feature vector is high, the video retrieval system consumes more computational time. It can be reduced with the feature dimensionality reduction techniques. The dimensionality reduction techniques such as Principal Component Analysis (PCA) [2], Fisher Discriminant Ratio [1], Linear Discriminant Analysis [7], semi-supervised linear discriminant analysis [13], supervised linear dimensionality reduction [12], nonparametric discriminant analysis [14] are utilized to reduce the feature dimension in other video retrieval systems.

2.1 Previous work

In proposed work, to design a flower video retrieval system the features of previous work [15] such as GLCM [24], LBP [25] and SIFT [22] are utilized. Instead of extracting features from entire keyframe, features are extracted in two different modes from each keyframe of the video. Initially, from all Flower Region of Interest (FRoI), secondly, from maximum Flower Region of Interest (Max.FRoI). A dimensionality reduction method is introduced for the features extracted from Max.FRoI, to improve the performance of the system with greater extent, which leads the fast accessing of videos. In the previous work [15] the query video consists of a single class of flowers. In the present work along with single class of flower videos query video also consists of multiclass flowers. The dataset considered in the present work is relatively large. The comparative study is made with previous work to show the effectiveness of the proposed work.

2.2 Contributions of the proposed work

The contributions are summarized as follows.

1. Creation of a reasonably large dataset of videos of flowers which shall be made available public for research purpose.
2. Proposal of fusion of features strategy to improve the performance of the existing model.
3. Proposal of an algorithmic model for the retrieval of videos of flowers using all flower regions of interest.

4. Proposal of a model for the retrieval of videos of flowers with maximum flower regions of interest
5. Adoption of a dimensionality reduction approach to improve the efficiency of the system.
6. Addressed retrieval of videos of flowers even when a query video contains flowers of more than one class.
7. Compared the proposed model with earlier proposed model and a deep learning model.

3. Proposed work

The proposed model comprises three stages namely, preprocessing, extraction of features and retrieval. The block diagram of the proposed flower video retrieval system using Flower Region of Interest (FRoI) is as shown in Fig. 1.

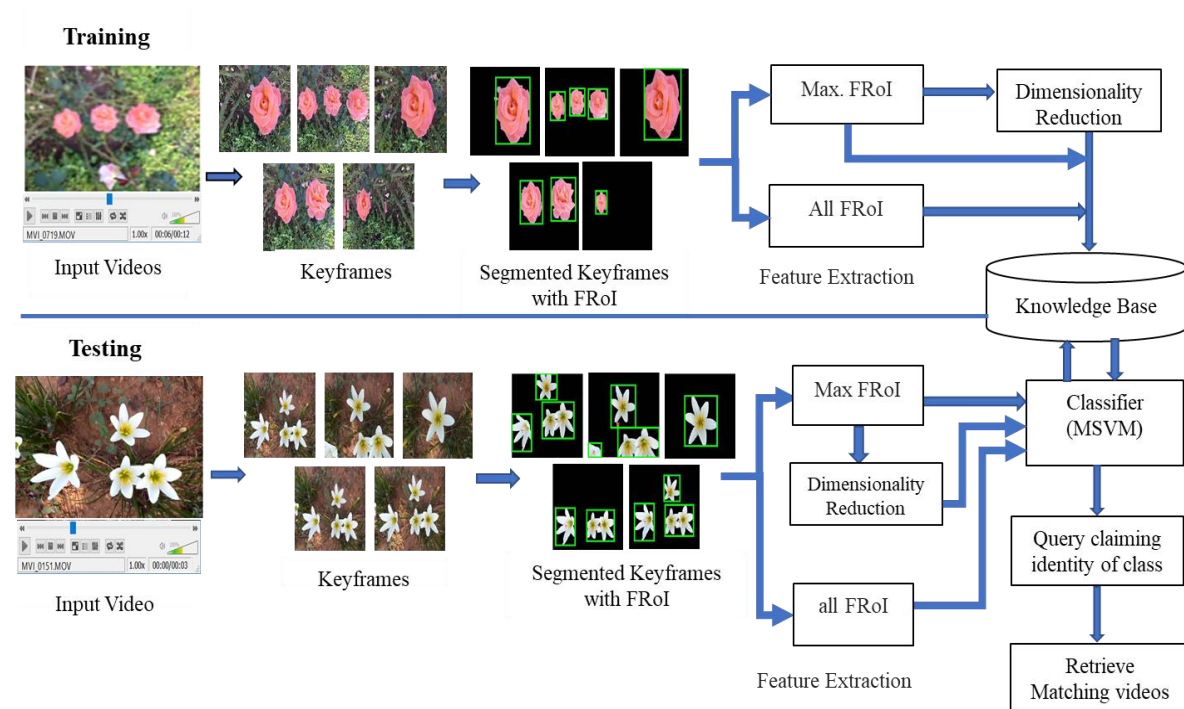


Figure 1: Block diagram of the proposed class based flower video retrieval system

3.1. Preprocessing

The preprocessing stage involves the processes such as selection of keyframes, segmentation and extraction of flower region of Interest (FRoI). The proposed system initially converts video to frames. Suppose that the flower video dataset ' X ' consists of ' vn ' number of samples and it is stated as

$$X = \{x_{v1}, x_{v2}, x_{v3}, \dots, x_{vi}, \dots, x_{vn}\} \quad (1)$$

Let the flower video x_{vi} consists of a finite set of ' F_N ' number of frames and it is defined as

$$x_{vi} = \{F_1, F_2, F_3, \dots, F_i, \dots, F_N\} \quad (2)$$

Then the keyframes of the video x_{vi} are selected using GMM cluster based algorithmic model [16]. Here, Block wise entropy feature is extracted from each frame of the video and similar frames are grouped together using Gaussian Mixture Model and the frame near to each cluster centroid are selected as keyframes of the video. GMM is explained in section 3.1.1. When the set of keyframes are selected from x_{vi} , then the video x_{vi} is represented as ' K_y ' number of keyframes and is defined as,

$$K_y = \{k_1, k_2, k_3, \dots, k_i, \dots, k_y\} \quad (3)$$

The flowers in keyframes are segmented from their background using statistical region merging algorithm [17]. The keyframes after segmentation can be defined as

$$SK_y = \{sk_1, sk_2, sk_3, \dots, sk_i, \dots, sk_y\} \quad (4)$$

3.1.1 Gaussian Mixture Model (GMM)

Gaussian Mixture model (GMM) is a statistical and unsupervised learning model. GMM [18], preserves content of the scene, the idea behind GMM is to describe pixels, some of which represent the background while the others represent the foreground in the scene. A finite number of mixtures of Gaussian distributions are used to generate data points. It preserves the sub-sampling property; it leads for clustering data points. The GMM parameters are estimated from data using the maximum expectation algorithm. A GMM is a weighted sum of several Gaussian densities. Therefore, in the present work to create clusters GMM is used for the selection of keyframes. Clusters are created by fitting the Gaussian distribution on data (x) with ' n ' features, the Gaussian function is defined as [19,32, 33].

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where μ is the mean and σ is the standard deviation of data (features) ' x '.

3.1.2 Extraction of Flower Region of Interest (FRoI)

After the process of segmentation of keyframes, all flower regions are selected using connected component analysis and the selected flower regions are named as Flower Regions of Interest (FRoI's) (refer Fig. 1). Then from FRoI's of each keyframe, features such as GLCM, LBP and SIFT are extracted for further processing.

3.2 Extraction of Features

Video visual features such as color, texture, local invariant features, etc., play an important role in the retrieval of videos [20, 21]. Some of the different species of flowers are similar in color. For example, we can find red colored rose, hibiscus, bougainvillea belongs to three different species. Therefore, color feature many not discriminate flowers from one species to another. There exists a large intra class variability and inter class similarity in the dataset. Due to this there are two prime motivations in the selection of features to describe flowers in videos. primarily, the texture of an individual species of flowers are similar, therefore textural features are used to describe the flowers in videos. Secondly, the species consists of variation in view point, illumination of flowers, in such cases for the discrimination of species of flowers the feature called Scale Invariant Feature Transform is well suitable [22].

Texture Features

Texture of an image/frame contain unique visual patterns. Texture features describes the object surface, these features are independent of object color [23]. The videos of flowers consist of large intra class variation such as variation in color of flowers. Therefore, to describe the flower region, texture features play a vital role. In this work, texture features namely, Gray Level Co-occurrence Matrix and Local Binary Pattern are used.

Gray Level Co-Occurrence Matrix (GLCM)

GLCM describes the texture of flower in terms of statistical information. In the current work, the system extracts 14 different gray level co-occurrence of statistical values [24, 34] are extracted from each FRoI. These features are represented as a feature vector.

Local Binary Pattern (LBP)

LBP describes the texture description in terms of local features of the flower region. An approach to recognize local binary patterns of image texture, and their occurrence histogram

proved that LBP is a powerful texture feature [25]. It is robust in terms of variation and transformation of the gray scale. In the proposed work, the system extracts LBP features [25] which are invariant to local grayscale variations in the FRoI. LBP texture features are extracted using 3x3 neighbourhood by the value of centre pixel, the pixels of eight neighbors are thresholded. In 3x3 neighbourhood, the centre pixel LBP value is obtained by thresholded binary values are weighted by powers of two and summed up.

Scale Invariant Feature Transform (SIFT)

SIFT plays a vital role in video retrieval for the analysis of the video content [11]. In SIFT the set of image features are generated in 4 stages [22]. In the first stage, the model searched over all scales and image locations to identify interest points that are invariant to orientation and scale. In the second stage, at each location, model is determined scale and location which is named as keypoint localization. In the third stage, based on local image gradient directions, orientations are assigned to each keypoint location. Finally, at the selected scale in the region around each keypoint, it generates descriptors, with a kernel of 4x4 histogram of 8 bins. These histograms compute the direction and magnitude of the gradient in the region of 16x16 pixels. The histograms results are represented in the form of descriptors. In the current work these feature descriptors are used to describe the FRoI's [22].

To design the proposed model, the features such as Gray Level Co-occurrence Matrix (GLCM) [24], Local Binary Pattern (LBP) [25] and Scale Invariant Feature Transform (SIFT) features proposed by [22] are extracted. Initially we propose to accomplish extracting these features by considering an entire keyframe after segmentation [15]. Subsequently, we employ the extraction of features on all flower regions of each keyframe of the video. And finally, we accomplish extraction of these features by selecting the Maximum Flower Region among all flower regions of the keyframe for the purpose of retrieval.

3.2.1 Entire keyframe

In this method [15], the model extracts the features such as Gray Level Co-occurrence Matrix (GLCM) [24], Local Binary Pattern (LBP) [25] and Scale Invariant Feature Transform (SIFT) [22] from an entire keyframe after segmentation and generates feature vector. Then, in the proposed model these features are fused like GLCM+LBP,

GLCM+SIFT, LBP+SIFT, GLCM+LBP+SIFT to improve the performance of the system. The video x_{vi} is represented as a set of features and is defined as,

$$x_{vi} = \{f_1, f_2, f_3, \dots, f_i, \dots, f_N\} \quad (5)$$

Then, $x_{vi} = F_i M_i$, where $F_i M_i = \{f_1, f_2, f_3, \dots, f_i, \dots, f_N\}$, similarly, features for all videos of a data base 'X' of equation (1) is defined as,

$$\mathfrak{R}^D = \{F_1 M_1(x_{v1}), F_2 M_2(x_{v2}), F_3 M_3(x_{v3}), \dots, F_i M_i(x_{vi}), \dots, F_n M_n(x_{vn})\} \quad (6)$$

Where $F_1 M_1(x_{v1}), F_2 M_2(x_{v2}), F_3 M_3(x_{v3}), \dots, F_i M_i(x_{vi}), \dots, F_n M_n(x_{vn})$ are the feature matrices of the videos $x_{v1}, x_{v2}, x_{v3}, \dots, x_{vi}, \dots, x_{vn}$ respectively in equation (1).

3.2.2 All Flower Regions of Interest

The proposed system extracts features such as GLCM [24], LBP [25] and SIFT [22] from all flower regions of keyframes and is as shown in Fig.2. In the proposed model these features are fused like GLCM+LBP, GLCM+SIFT, LBP+SIFT, GLCM+LBP+SIFT to improve the performance of the system. Let R_r be the number of selected flower regions of a keyframe sk_i in equation (4). Then, sk_i with number of flower regions is defined as

$$sk_i = \{R_1 SK_i, R_2 SK_i, R_3 SK_i, \dots, R_r SK_i\} \quad (7)$$

Then, the feature vector of all regions is represented as

$$R_1 SK_i = [f_{11}, f_{12}, f_{13}, \dots, f_{1M}], R_2 SK_i = [f_{21}, f_{22}, f_{23}, \dots, f_{2M}], \dots, R_r SK_i = [f_{r1}, f_{r2}, f_{r3}, \dots, f_{rM}]$$

Finally, the feature vector of all the regions of a keyframe sk_i as shown in equation (7) is represented as,

$$F_i M_i^d = \{[f_{11}, f_{12}, f_{13}, \dots, f_{1M}], [f_{21}, f_{22}, f_{23}, \dots, f_{2M}], \dots, [f_{r1}, f_{r2}, f_{r3}, \dots, f_{rM}]\}$$

Then, all regions of a keyframe sk_i is defined as,

$$F_i M_i^d = \forall R_i SK_i \in sk_i \quad (8)$$

Where FM_i^d is the feature matrix of the video x_{vi} of the equation (1) consists of $\forall R_i SK_i$ all regions of a keyframe sk_i in equation (7).

Then, the feature vector of all FRoI's of all keyframes of a video x_{vi} can be defined as

$$FM^d(x_{vi}) = \forall F_j M_j^d \in SK_y \quad (9)$$

Where $FM^d(x_{vi})$ is the feature matrix of the video x_{vi} of equation (1) consists of all feature matrices of all 'y' keyframes of a video as shown in equation (4).

The feature dimension of a video x_{vi} i.e., $FM^d(x_{vi})$ consists of the features extracted from all regions of each keyframe of the video x_{vi} . Similarly, the feature vectors obtained for all videos of a database 'X' can be defined as,

$$\mathcal{R}^D = \{FM^d(x_{v1}), FM^d(x_{v2}), FM^d(x_{v3}), \dots, FM^d(x_{vi}), \dots, FM^d(x_{vn})\} \quad (10)$$

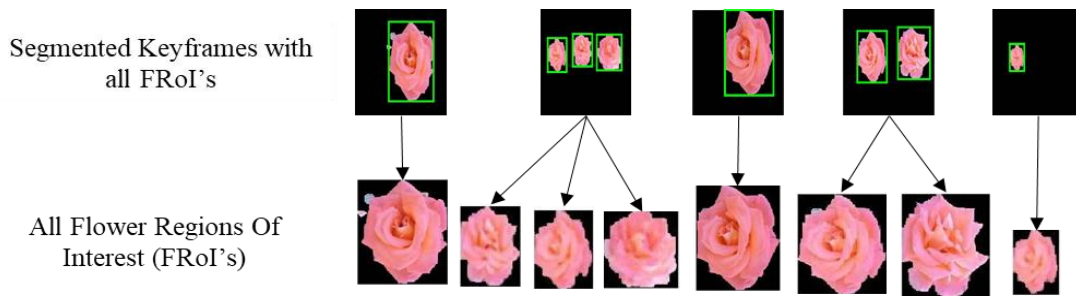


Figure 2: Extraction of features from all flower regions of interest

3.2.3 Maximum Flower Region of Interest

In this method, features such as GLCM [24], LBP [25] and SIFT [22] are extracted from the Maximum Flower Region of Interest (Max. FRoI) among all flower regions of each keyframe of a video, then features are fused like GLCM+LBP, GLCM+SIFT, LBP+SIFT, GLCM+LBP+SIFT to improve the performance of the system.. Fig. 3 shows the selected flower region. Max. FRoI is obtained by selecting the maximum flower region i.e., the flower region having high density of pixels and is the largest area among all the regions in each keyframe. When there is only one flower region in the keyframe then that will be considered as Max. FRoI as shown in Fig. 3. It reduces the dimension of the features of the proposed retrieval system as compared with all FRoI's. Through Max. FRoI model, the efficiency can be improved. The features are extracted, after selecting Max. FRoI from each keyframe of

equation (4). Therefore, the feature vector defined in equation (8) can be redefined in this case as,

$$MF_i M_i^d = \text{Max}(R_i SK_i) \in sk_i \quad (11)$$

Where $MF_i M_i^d$ is the feature matrix of the video x_{vi} of the equation (1) consists of $\text{Max}(R_i SK_i)$ maximum flower region of a keyframe sk_i in equation (7).

Then the feature matrix of Max. FRoI of all keyframes of a video x_{vi} can be defined as

$$FM^d(x_{vi}) = \forall (MF_j M_j^d) \in SK_y \quad (12)$$

Where $FM^d(x_{vi})$ is the feature matrix of the video x_{vi} of equation (1) consists of maximum flower region feature matrices of all 'y' keyframes of a video as shown in equation (4).

The feature dimension of a video x_{vi} i.e., $FM^d(x_{vi})$ in equation (12) consists of the features extracted from maximum flower region of each keyframe of the video x_{vi} . Similarly feature vectors for all videos of database 'X' are obtained and are defined as,

$$\mathfrak{R}^D = \{FM^d(x_{v1}), FM^d(x_{v2}), FM^d(x_{v3}), ..., FM^d(x_{vi}), ..., FM^d(x_{vn})\} \quad (13)$$

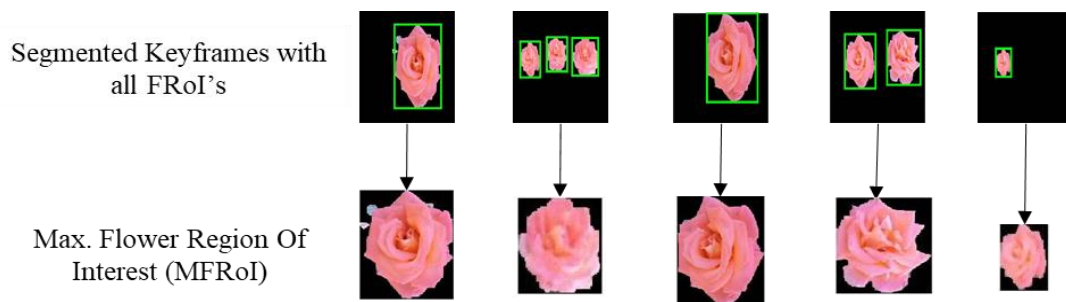


Figure. 3. Extraction of Maximum Flower Region of Interest (Max. FRoI)

Further, even though the Max.FRoI reduces the dimension of the features of the proposed retrieval system as compared with all FRoI's, to improve the efficiency of the retrieval system, the most discriminant features from Max. FRoI are obtained using LDA and is discussed in section 3.2.4. The feature dimension of a video x_{vi} as shown in equation (12) is

represented as the reduced discriminant features obtained from Max. FRoI using LDA and it can be defined as

$$FM^d(x_{vi}) = DR(\forall(MF_j M_j^d) \in SK_y) \quad (14)$$

Where $j=1$ to 'y' keyframes of a video x_{vi} as shown in equation (4).

Finally, the reduced feature vectors for all videos of the database 'X', are defined as,

$$\mathfrak{R}^D = \{DR(FM^d(x_{v1})), DR(FM^d(x_{v2})), DR(FM^d(x_{v3})), \dots, DR(FM^d(x_{vi})), \dots, DR(FM^d(x_{vn}))\} \quad (15)$$

3.2.4 Linear Discriminant Analysis (LDA)

LDA is a supervised dimensionality reduction method [26]. Ronald Fisher in 1936 proposed discriminant analysis, to find a new feature space from original feature space. LDA plays a vital role in order to maximize class separability and preserves the within class similarity. It maximizes the distance between the projected data of inter classes and minimizes the distance between the predictable data of the intra class [27,13] and hence in the current work we have applied LDA for the reduction of feature dimension.

The reduced dimension of the feature vector is defined as follows,

$$DR(FMV_i) = \{f_1, f_2, f_3, \dots, f_{dr}\} \quad (16)$$

For the retrieval of videos, the proposed model utilizes reduced features obtained after dimensionality reduction. The reduced feature vector of FMV_i consists of 30 features.

3.3 Retrieval: Query claiming identity of class

Initially, for a given query video 'Q_v', the system acquires the identity of the class using Multiclass Support Vector Machine (MSVM). Then the similar videos are retrieved from the predicted class. For the retrieval of a query video, the model is trained with two different set of features explained in section 3.2.2 and section 3.2.3 and the experimental results are shown in section 4.

Support vector machine (SVM) is a computationally powerful tool for supervised learning [28, 14, 35]. Support vector machine is a vector-space-based classification method for both

linear and non-linear data. The fundamental idea of SVM classifier is to find the optimal separating hyperplane between two classes. For more information please refer [29, 30].

4. Experiments and Results

4.1 Datasets

Dataset is a fundamental requirement to test the efficiency of any automatic system designed. To conduct experiments, relatively large dataset is required. Since the standard flower video dataset is not publicly available, we created flower video datasets. To create flower video datasets, we used three devices namely, Samsung Galaxy Grand Prime (SGGP) mobile, Sony Cyber Shot camera and Canon camera. SGGP dataset consists of 2611 videos of 8 mega pixels. Sony Cyber Shot camera dataset consists of 2521 videos of 14.1 mega pixels. And Canon camera consists of 2656 videos of 16 mega pixels. Videos captured with the duration ranges from 4 to 60 seconds. We have captured 30 different species of flowers from all the three devices. There exists a small inter class and large intra class variations. Videos captured in the real environment during summer, rainy and winter seasons. Videos involved the challenges such as viewpoint variations, illumination, cluttered background, and multiple instances of the flowers. Flower video samples with large intra-class variations from the dataset we created are shown in Fig. 13.

Along with the above mentioned three datasets, we created a dataset with multiple classes of flowers in a video for querying. The dataset contains two and three different classes of flowers. The samples of these flower videos are shown in Fig. 14.

The performance of the proposed model is analysed in different modes of extraction of features. Results of the features extracted from all FRoI's as shown in the section 4.2, the features extracted from Maximum FRoI (MFRoI) is as shown section 4.3 and the features extracted from Maximum FRoI (MFRoI) with LDA is as shown in section 4.4. And also, the results obtained in previous work of extracting features from entire keyframe [15] are shown in section 4.5. The dataset we created is used to conduct experiments. In order to evaluate the system, metrics such as accuracy, precision, recall and F-measure are used and are given below. The results are tabulated with varying training and testing videos.

$$Accuracy = \frac{\text{Sum of videos retrieved correctly}}{\text{Total number of query videos}} \quad (17)$$

$$Precision = \frac{\text{Total number of videos retrieved are relevant}}{\text{Total number of videos retrieved}} \quad (18)$$

$$Recall = \frac{\text{Total number of videos retrieved are relevant}}{\text{Total number of similar videos in the database}} \quad (19)$$

$$F - Measure = \frac{2 * Precision * Recall}{(Precision + Recall)} \quad (20)$$

4.2 All FRoI's

The result analysis of proposed retrieval system trained with the features extracted from all FRoI's are shown in the following figures Fig. 4, Fig. 5 and Fig. 6 for SGGP, Sonycyber Shot and Canon datasets respectively. From the results we can observe that the accuracy of the system in this approach achieved 53.83% for SGGP dataset, 52.36% for Sonycyber Shot and 63.56% for Canon dataset for 70% training and 30% testing.

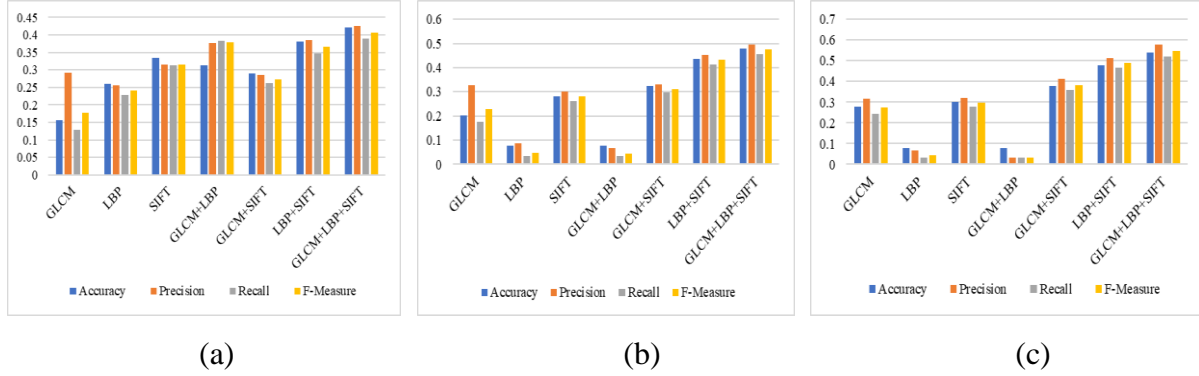


Fig. 4. Features extracted from all FRoI's for SGGP dataset: (a) 30 % Train – 70% Test, (b) 50 % Train – 50% Test, (c) 70 % Train – 30% Test

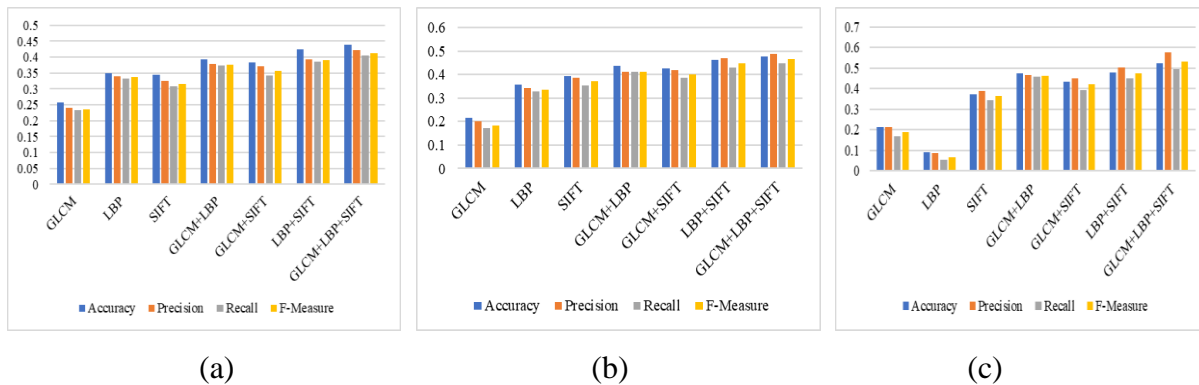


Fig. 5. Features extracted from all FRoI's for Sonycyber Shot dataset: (a) 30 % Train - 70% Test, (b) 50 % Train - 50% Test, (c) 70 % Train - 30% Test

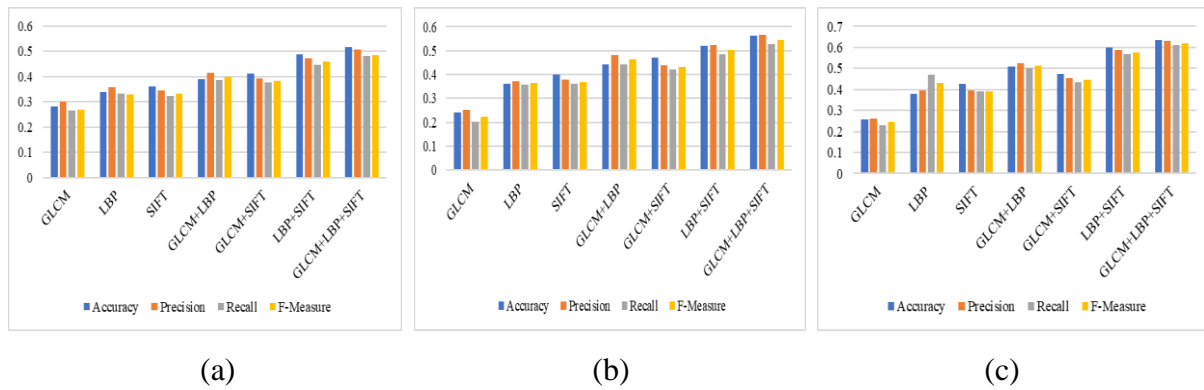


Fig. 6. Features extracted from all FRoI's for Canon dataset: (a) 30 % Train – 70% Test, (b) 50 % Train – 50% Test, (c) 70 % Train – 30% Test

4.3 Max. FRoI

The result analysis of proposed retrieval system trained with the features extracted from maximum flower region of interest are shown in the following figures Fig. 7, Fig. 8 and Fig. 9 for SGGP, Sonycyber Shot and Canon datasets respectively. From the results we can observe that the accuracy of the system in this approach is achieved 60.59% for SGGP dataset, 67.07% for Sonycyber Shot dataset and 75.79% for Canon dataset for 70% training and 30% testing. Further, from the results we can observe that the Max. FRoI give improved results than all FRoI's for all the three datasets.

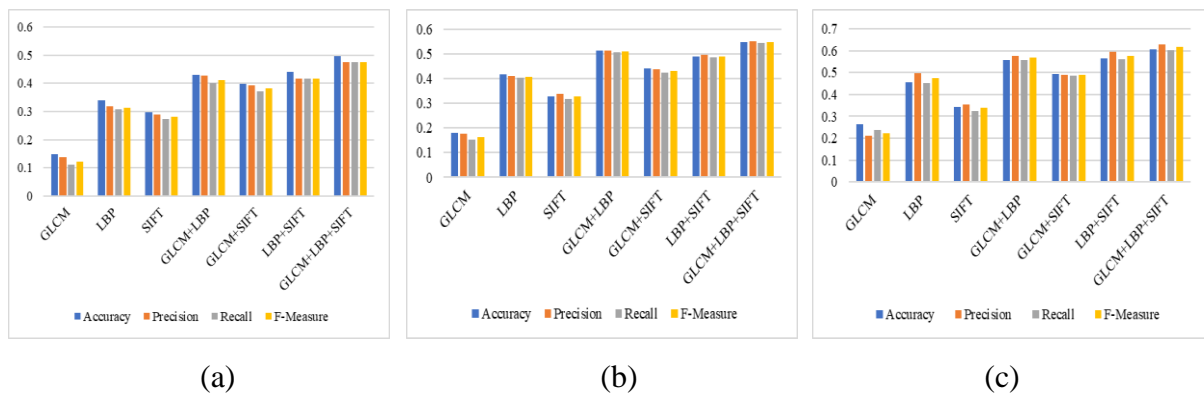


Fig. 7. Features extracted from Max FRoI for SGGP dataset: (a) 30 % Train – 70% Test, (b) 50 % Train – 50% Test, (c) 70 % Train – 30% Test

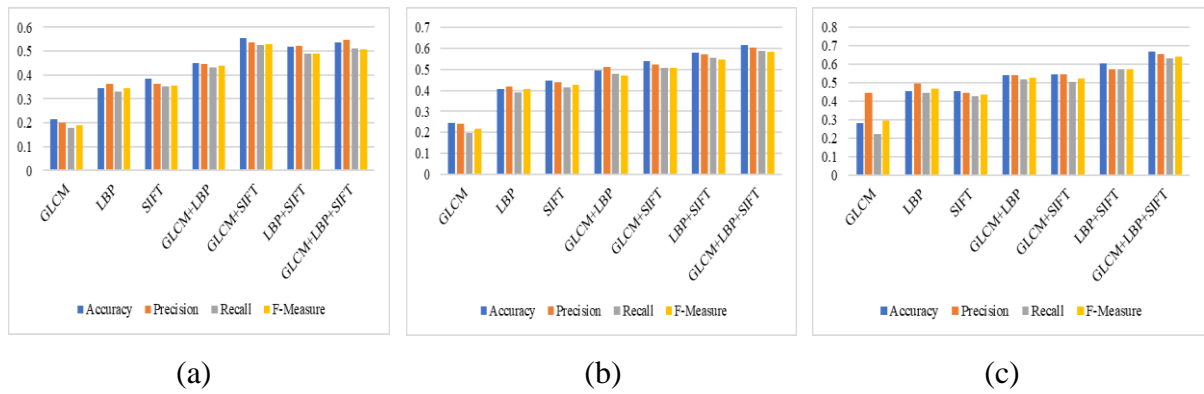


Fig. 8. Features extracted from Max FRoI for Sonycyber Shot dataset: (a) 30 % Train–70% Test, (b) 50 % Train – 50% Test, (c) 70 % Train – 30% Test

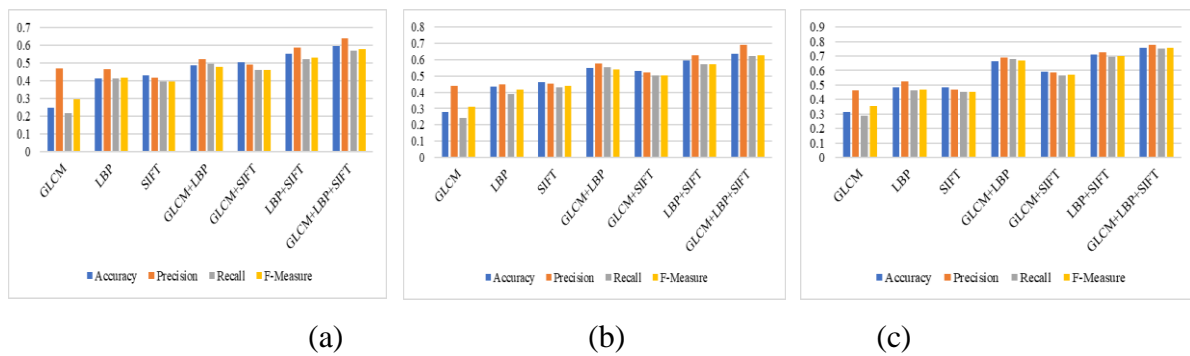


Fig. 9. Features extracted from Max FRoI for Canon dataset: (a) 30 % Train – 70% Test, (b) 50 % Train – 50% Test, (c) 70 % Train – 30% Test

4.4 Max. FRoI with LDA

In this section we obtain discriminant features from Max. FRoI using LDA are passing to the model. It improves the retrieval performance by identifying the class of the query video. Table 1(a) to Table 1(c), Table 2(a) to Table 2(c) and Table 3(a) to Table 3(c) show the result analysis of proposed retrieval system trained with the reduced features extracted from Max. FRoI is as shown in equation (13) for SGGP, Sonycyber Shot and Canon datasets respectively. Further, the tables show that the results obtained from Max. FRoI with LDA gives good results than the results obtained from other proposed modes.

Table 1 (a). SGGP Dataset: Train 30% - Test 70%

Features	Accuracy	Precision	Recall	F-Measure
GLCM [24]	0.2	0.18	0.15	0.17
LBP [25]	0.13	0.11	0.07	0.09
SIFT [22]	0.99	0.99	1	0.99
GLCM+LBP	0.16	0.15	0.1	0.12

GLCM+SIFT	0.99	0.99	0.99	0.99
LBP+SIFT	1	1	1	1
GLCM+LBP+SIFT	0.99	0.99	0.99	0.99

Table 1 (b). SGGP Dataset: Train 50% - Test 50%

Features	Accuracy	Precision	Recall	F-Measure
GLCM [24]	0.22	0.21	0.18	0.19
LBP [25]	0.13	0.12	0.08	0.09
SIFT [22]	0.99	0.99	0.99	0.99
GLCM+LBP	0.19	0.18	0.12	0.14
GLCM+SIFT	0.99	0.99	0.99	0.99
LBP+SIFT	1	1	1	1
GLCM+LBP+SIFT	0.98	0.98	0.98	0.98

Table 1 (c). SGGP Dataset: Train 70% - Test 30%

Features	Accuracy	Precision	Recall	F-Measure
GLCM [24]	0.33	0.31	0.3	0.29
LBP [25]	0.15	0.16	0.1	0.11
SIFT [22]	0.99	0.99	1	0.99
GLCM+LBP	0.2	0.2	0.1	0.16
GLCM+SIFT	0.99	0.99	1	0.99
LBP+SIFT	1	1	1	1
GLCM+LBP+SIFT	0.99	0.99	1	0.99

Table 2 (a). Sonycyber Shot Dataset: Train 30% -Test 70%

Features	Accuracy	Precision	Recall	F-Measure
GLCM [24]	0.21	0.2	0.17	0.18
LBP [25]	0.38	0.41	0.37	0.37
SIFT [22]	0.99	0.99	0.99	0.99
GLCM+LBP	0.45	0.44	0.43	0.42
GLCM+SIFT	0.99	0.99	0.99	0.99
LBP+SIFT	1	1	1	1
GLCM+LBP+SIFT	0.97	0.99	0.96	0.97

Table 2 (b). Sonycyber Shot Dataset: Train 50% -Test 50%

Features	Accuracy	Precision	Recall	F-Measure
GLCM [24]	0.24	0.24	0.2	0.22
LBP [25]	0.37	0.38	0.4	0.35
SIFT [22]	0.99	0.99	1	1

GLCM+LBP	0.49	0.51	0.5	0.47
GLCM+SIFT	0.99	0.99	1	1
LBP+SIFT	1	1	1	1
GLCM+LBP+SIFT	0.99	0.99	1	0.99

Table 2 (c). Sonycyber Shot Dataset: Train 70% -Test 30%

Features	Accuracy	Precision	Recall	F-Measure
GLCM [24]	0.28	0.44	0.2	0.3
LBP [25]	0.39	0.37	0.5	0.43
SIFT [22]	0.99	0.99	1	1
GLCM+LBP	0.54	0.54	0.5	0.53
GLCM+SIFT	0.99	0.99	1	1
LBP+SIFT	1	1	1	1
GLCM+LBP+SIFT	0.99	0.99	1	1

Table 3 (a). Canon Shot Dataset: Train 30% -Test 70%

Features	Accuracy	Precision	Recall	F-Measure
GLCM [24]	0.54	0.53	0.51	0.5
LBP [25]	0.63	0.68	0.64	0.64
SIFT [22]	0.99	0.99	0.99	0.99
GLCM+LBP	0.81	0.83	0.78	0.8
GLCM+SIFT	0.99	0.99	0.99	0.99
LBP+SIFT	1	1	1	1
GLCM+LBP+SIFT	1	1	1	1

Table 3 (b). Canon Shot Dataset: Train 50% -Test 50%

Features	Accuracy	Precision	Recall	F-Measure
GLCM [24]	0.56	0.55	0.52	0.53
LBP [25]	0.66	0.7	0.67	0.67
SIFT [22]	0.99	0.99	0.99	0.99
GLCM+LBP	0.82	0.86	0.8	0.82
GLCM+SIFT	0.99	1	0.99	0.99
LBP+SIFT	1	1	1	1
GLCM+LBP+SIFT	1	1	1	1

Table 3 (c). Canon Shot Dataset: Train 70% -Test 30%

Features	Accuracy	Precision	Recall	F-Measure
GLCM [24]	0.63	0.64	0.61	0.6
LBP [25]	0.69	0.73	0.71	0.7

SIFT [22]	0.99	0.99	0.99	0.99
GLCM+LBP	0.85	0.87	0.86	0.86
GLCM+SIFT	0.99	1	0.99	0.99
LBP+SIFT	1	1	1	1
GLCM+LBP+SIFT	1	1	1	1

4.5 Comparative study between proposed work and previous work

In the previous work [15] the features such as Gray Level Co-occurrence Matrix (GLCM) [24], Local Binary Pattern (LBP) [25] and Scale Invariant Feature Transform (SIFT) [22] are extracted from entire keyframe. With the fusion of these features the model achieved good performance. The retrieval accuracy of previous work [15] achieved 53.83%, 60.18% and 65.73% are shown in Fig. 10, Fig. 11 and Fig. 12 for SGGP, Sonycyber Shot and Canon datasets respectively. In the proposed work, to further improve the retrieval performance, GLCM [24], LBP [25] and SIFT [22] features are extracted in two different modalities as mentioned in section 3.2.2 and 3.2.3. The features extracted from proposed retrieval system using, Max. FRoI and Max. FRoI with LDA these two methods give good results when compared to the previous work [15]. The comparison between the results obtained from previous and proposed approaches namely, features extracted from an entire keyframe, all FRoI's, Max. FRoI and Max.FRoI with LDA are summarized in Table 4 for all datasets.

4.6 Result analysis and Discussion

We have the following observations from the proposed system of approaches namely, features extracted from an entire keyframe, all FRoI's, Max. FRoI and Max.FRoI with LDA.

1. Features extracted from entire keyframes of a video provide good results with the fusion of the features GLCM+LBP+SIFT as shown in Fig. 10 to Fig. 11.
2. All FRoI's approach generates almost similar results for the combination of features GLCM+LBP+SIFT as compared to the features extracted from an entire keyframe as shown in Fig. 4 to Fig. 6 for SGGP, Sonycyber Shot and Canon datasets respectively.
3. Max. FRoI's approach generates good results for the combination of features GLCM+LBP+SIFT as shown in Fig. 7 to Fig. 9 for SGGP, Sonycyber Shot and Canon datasets respectively. From the results we can observe that, this approach generates improved results than the features extracted from entire keyframe.

4. The proposed approach Max. FRoI with LDA results show the effectiveness of the selection of more discriminating feature subset from original set using LDA. The efficiency of the proposed system using Max. FRoI with LDA is improved and achieved 100% performance for SGGP, Sonycyber Shot and Canon datasets. Table 1 and Table 2 show the combination of features LBP+SIFT achieves good performance for SGGP and Sonycyber Shot datasets. Table 3 show that the combinations of features LBP+SIFT and GLCM+LBP+SIFT achieves good performance for Canon dataset.

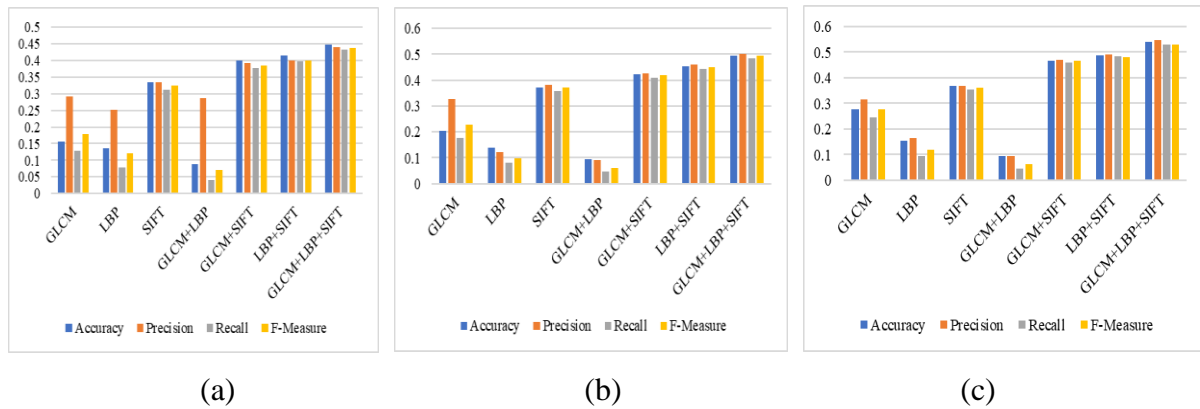


Fig. 10. Features extracted from entire keyframe for SGGP dataset: (a) 30 % Train – 70% Test, (b) 50 % Train – 50% Test, (c) 70 % Train – 30% Test

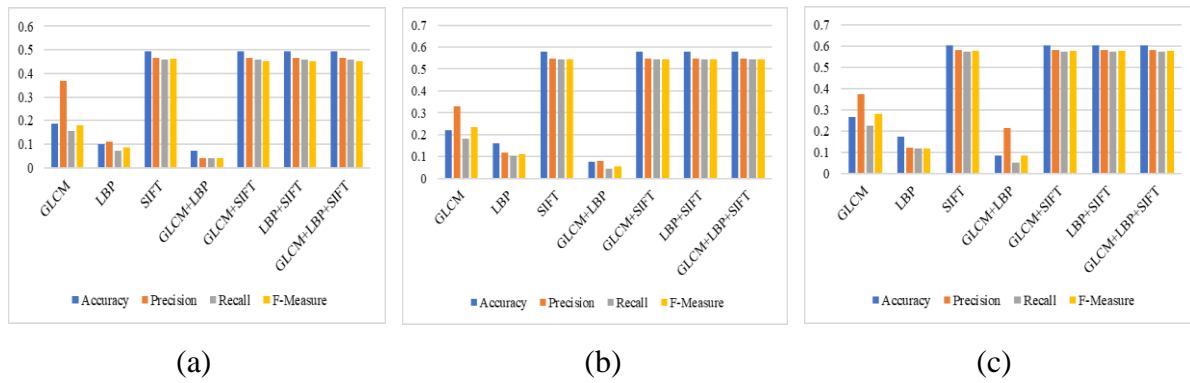


Fig. 11. Features extracted from entire keyframe for Sonycyber Shot dataset (a)30%Train-70%Test (b)50%Train-50% Test (c)70%Train-30% Test

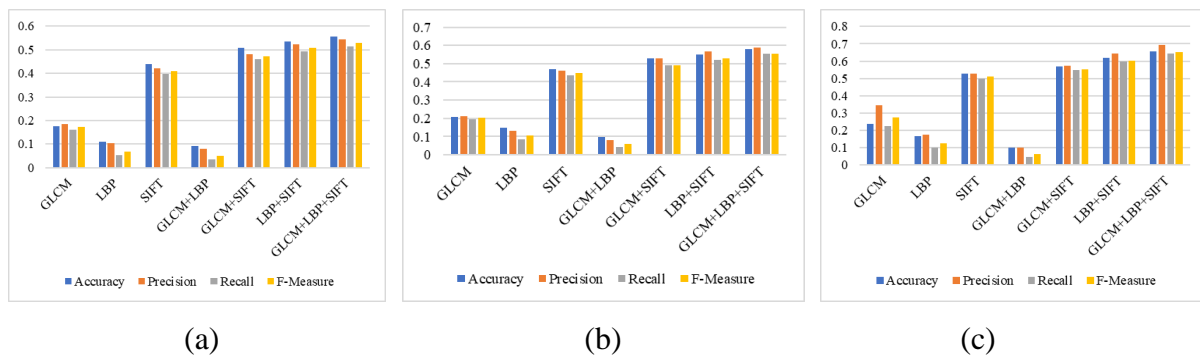


Fig. 12. Features extracted from entire keyframe for Canon dataset: (a) 30 % Train – 70% Test, (b) 50 % Train – 50% Test, (c) 70 % Train – 30% Test

Table 4: Accuracy obtained for feature combinations with different modes of extraction of features with 70% training and 30% testing.

Sl. No.	Modes of extraction of features	Feature Combination	Datasets (results in %)		
			SGGP	Sonycyber Shot	Canon
1	Entire keyframe	GLCM+LBP+SIFT	53.83	60.18	65.73
2	All FRoI's	GLCM+LBP+SIFT	53.83	63.56	52.36
3	Max. FRoI	GLCM+LBP+SIFT	60.59	67.07	75.79
4	Max. FRoI with LDA	LBP+SIFT	100	100	100

4.7 Query with multiple class flowers in videos

Query video may contain multiple classes of flowers in video. There are two cases at this point. First, a query video consists of multiclass flowers in all frames, then the system retrieves similar videos from database by considering Flower Regions of Interest. Second, a query video consists of multiclass flowers not in the same frame, video consists of one class in some duration and then other classes in next duration. In such case, we manually split (cut) the video into shots based on class boundary, then for each shot the system retrieves similar videos based on FRoI using MSVM. Fig. 14 shows the query acquiring the identity of the class for multiclass flowers in a video.

5. Comparative study between proposed work and deep learning model

In [31], authors have proposed a flower video retrieval system using deep learning approach, here the similar videos for a given query video are retrieved using Multiclass Support Vector Machine. For the extraction of features in [31], authors have proposed three different modalities; entire keyframe, segmented flower region of a keyframe, and the gradient of flower region are considered for feature extraction using Deep Convolutional Neural Network of AlexNet architecture. Among these three modalities, the segmented flower region

of a keyframe is achieved better results for smaller dataset. In [31], the query video consists of a single class of flowers. In the present work along with single class of flower videos query video also consists of multiclass flowers. The dataset considered in the present work is relatively large. The presented model is compared against deep learning model [31] which reveals that the proposed one is superior to the existing in terms of retrieval results. The proposed system Max. FRoI with LDA is improved and achieved 100% performance for larger sized datasets namely SGGP, Sonycyber Shot and Canon. The retrieval results in terms of Accuracy, Precision, Recall and F-measure of existing work [31] are compared with present work and the results are shown in Fig. 13, Fig. 14 and Fig. 15 for SGGP, Sonycyber Shot and Canon datasets respectively.

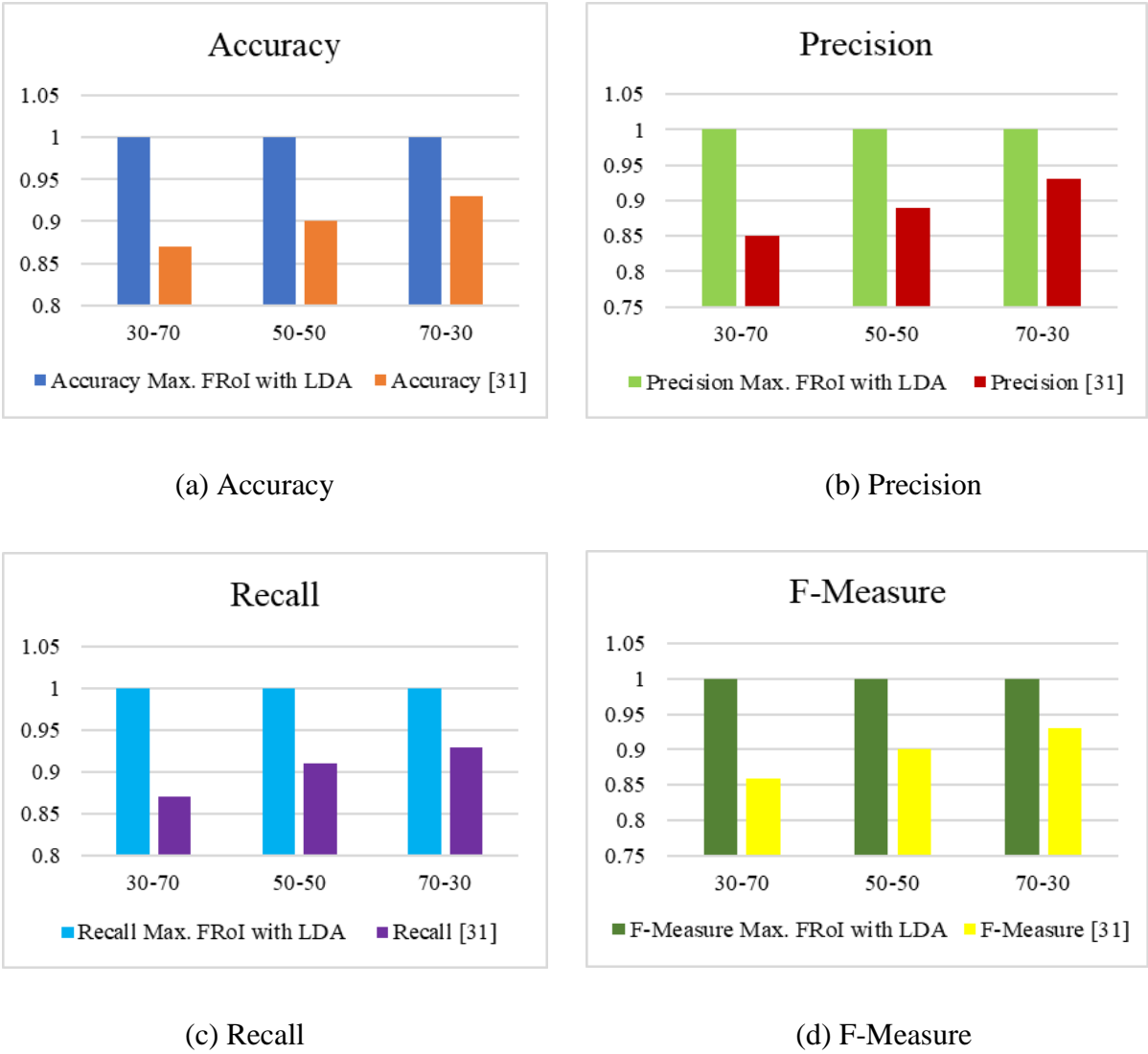
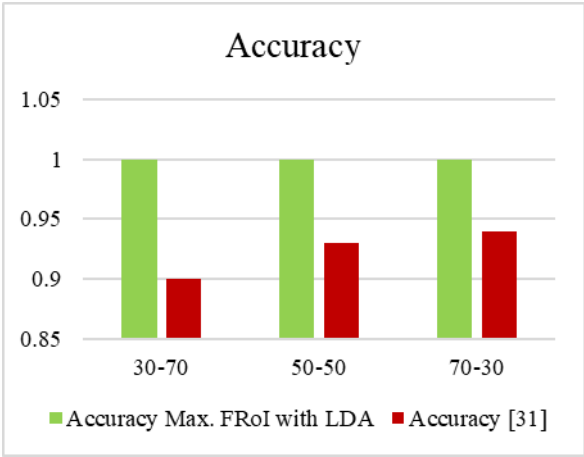
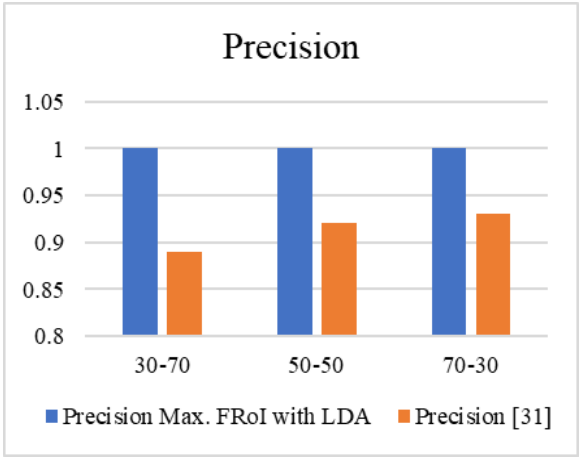


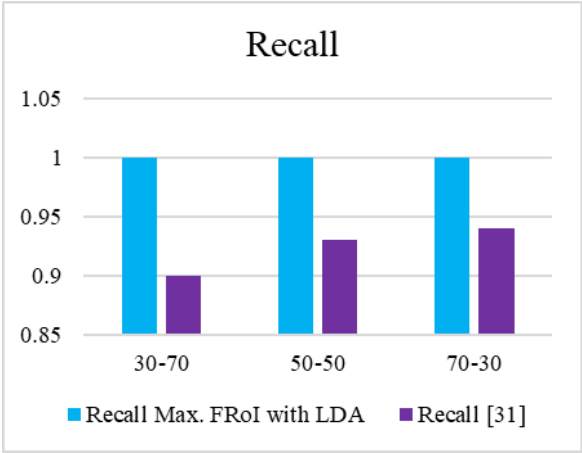
Fig 13: Comparative study between proposed work and deep learning model [31] for SGGP dataset



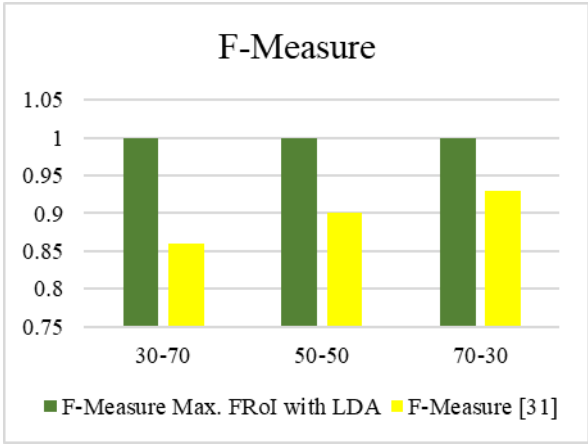
(a) Accuracy



(b) Precision

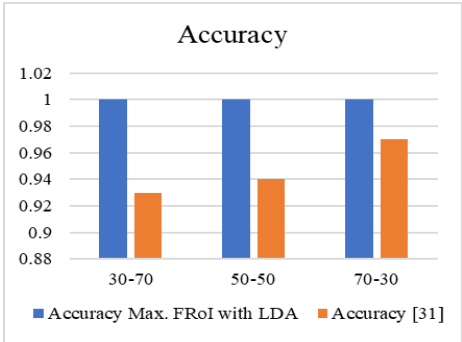


(c) Recall

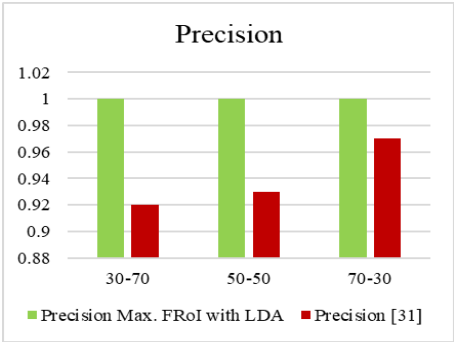


(d) F-Measure

Fig 14: Comparative study between proposed work and deep learning model [31] for Sonycyber Shot dataset



(a) Accuracy



(b) Precision

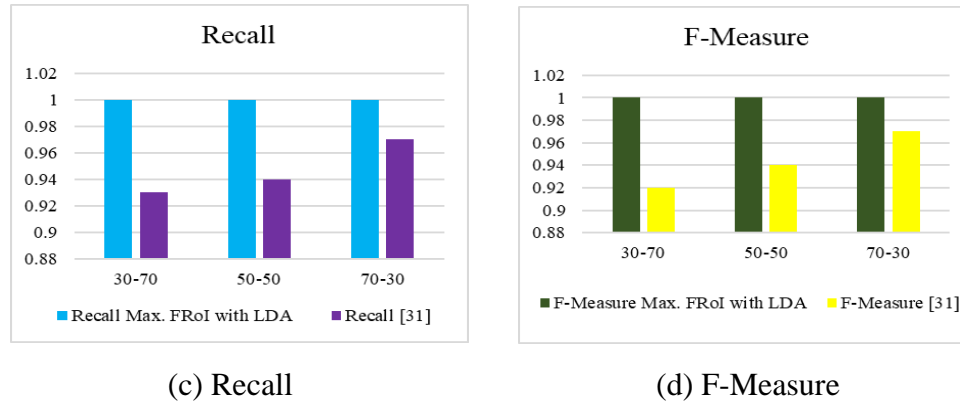


Fig 15: Comparative study between proposed work and deep learning model [31] for Canon dataset

6. Conclusion

The main aim of this work is to discover the solution to a problem of retrieval of videos of flowers through query by video mechanism. The presented system works based on keyframes represented for each video. Features extracted in three different modalities namely, all regions of flowers in the keyframe, maximum flower region in the keyframe and finally, maximum flower region with a set of discriminating features generated by LDA. The presented system is compared against our previous work and the deep learning retrieval system, which reveals that the proposed system with Max. FRoI and LDA is superior to the existing models in terms of retrieval results. Further, the proposed system retrieves similar videos when the query video consists of multi class flowers.

Future work

The research work presented in this paper can be further extended in following ways:

1. An attempt on shot boundary or class boundary detection when a video consists of multiple species of flowers can be further explored.
2. The current research work limits the species of flowers to 30. There is scope for extending the class size and to explore different methodologies to retrieve flower videos in real time.





















Fig. 13 Samples of flower videos with large intraclass variation from 30 classes of videos

References

1. Shen X-J., Mu. L., Li Z., Wu H-X., Gou J-P., and Chen X., 2016. Large-scale support vector machine classification with redundant data reduction. *Neurocomputing* 172, pp.189-197.
2. Geetha M K., Palanivel S., and Ramalingam V., 2009. A novel block intensity comparison code for video classification and retrieval. *Expert Systems with Applications* 36, pp. 6415-6420.
3. Das M, Manmatha R and Riseman E., 1999. Indexing flower patent images using domain knowledge. *IEEE intelligent systems* 14 (5), pp. 24-33.
4. Hu W., Xie N., Li L., Xianglin Z and Maybank S., 2011. A Survey on Visual Content-Based Video Indexing and Retrieval. *IEEE transactions on System, Man and Cybernetics-Part C: Applications and Reviews*, Vol. 41(6), pp.797-819.
5. Morand C., Benois-Pineau J., Domenger J. P., Zepeda J., Kijak E., Guillemot C., 2010. Scalable object-based video retrieval in HD video databases. *Signal Processing: Image Communication*. Vol. 25, pp.450-465.

6. Shekar B H, Uma K P., and Holla K R., 2016. Video clip retrieval based on LBP variance. *Procedia Computer Science* 89, pp. 828-235.
7. Gao X., Xuelong L., Jun F., and Dacheng T., 2009. Shot-based video retrieval with optical flow tensor and HMMs. *Pattern Recognition Letters* 30, pp.140-147.
8. Han J., Ji X., Hu. X., Han. J., Liu T., 2014. Clustering and retrieval of video shots based on natural stimulus fMRI. *Neurocomputing* 144, pp.128-137.
9. Liang B., Xiao W., and Liu X., 2012. Design of video retrieval system using MPEG-7 descriptors, *Procedia Engineering* 29, pp.2578-2582.
10. Priya G G L., and Domnic S., 2014. Shot based keyframe extraction for ecological video indexing and retrieval. *Ecological Informatics* 23, pp.107-117.
11. Zhu Y., Huang X., Huang Q., Tian Q., 2016. Large-scale video copy retrieval with temporal concentration SIFT. *Neurocomputing* 187, pp.83-91.
12. Cui M., Cui J., Li H., 2016. Dimensionality reduction for histogram features: A distance-adaptive approach. *Neurocomputing* 173, pp.181-195.
13. Wang S., Lu J., Gu X., Du H., Yang J., 2016. Semi-supervised linear discriminant analysis for dimension reduction and classification. *Pattern Recognition* 57, pp.179-189.
14. Khan N. M., Ksantini R, Ahmad I. S., and Boufama B., 2012. A novel SVM+NDA model for classification with an application to face recognition. *Pattern Recognition*, Vol. 45, pp.66-79.
15. Guru D S, Jyothi V K, and Kumar Y H S, 2018. Features Fusion for Retrieval of Flower Videos. *Proceedings of DAL 2018, LNNS* 43, pp.221-234.
16. Guru D. S., Jyothi V. K., Kumar Y. H. S., 2018. Cluster based approaches for keyframe selection in natural flower videos. *springer AISC 736, ISDA 2018*, pp.474-484.
17. Nock R., and Nielsen F., 2004. Statistical region merging. *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, No.11, pp. 1-7.
18. Stauffer C., and Grimson W.E.L., 1999. Adaptive background mixture models for real-time tracking. in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.
19. Chen W., Tian Y., Wang Y., and Huang T., 2015. Fixed-point Gaussian mixture model for analysis friendly surveillance video coding. *Computer vision and image understanding*, Vol.142, pp. 65-79
20. Hong R., Pan. J, Hao. S., Wang. M., Xue F., Wu X., 2014. Image quality assessment based on matching pursuit. *Inf. Sci.* 273, pp.196-211.
21. Li J., Li X., Yang B., Sun X., 2015. Segmentation-based image copy move forgery detection scheme. *IEEE Transaction, Inf. Forensics Secur.* 10 (3), pp. 507-518.
22. Lowe D G, 2004. Distinctive Image Features from Scale-Invariant Keypoint. *International Journal of Computer Vision*, Vol. 60, pp.91-110.
23. Hu W., 2011. A survey on visual content-based video indexing and retrieval, *IEEE transactions on systems, MAN, and Cybernetics-Part C: Applications and reviews*, Vol-41, No.6, pp. 797-819.
24. Haralick R M., Shanmugam. K and Dinstein I., 1973. Textural features for image classification. *IEEE Transaction on Systems, man and Cybernetics*, Vol.SMC-3(6), pp.610-621.
25. Ojala T, Pietikainen M and Maenpaa T, 2002. Multiresolution Gray Scale and Rotation Invariant Texture classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume-24, Issue-7, pp.971-987.
26. Belhumeur P. N., Hespanha J. P., and Kriegman D. J., 1999. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 711-720.
27. Gyamfi K. S., Brusey J., Hunt A., Gaura E., 2018. Linear dimensionality reduction for classification via a sequential Bayes error minimisation with an application to flow meter diagnostics. *Expert Systems with Applications*, vol. 91, pp.252-262.

28. Kumar M. A., and Gopal M., 2011. A hybrid SVM based decision tree. Pattern Recognition, Vol. 43, pp. 3977-3987.
29. Vapnik V. N., 1998. Statistical learning theory. John wiley and sons, New York, USA.
30. Duda R.O., Hart P.E., and Stork D.G., 1997. Pattern classification, Second edition.
31. Jyothi V K., D S Guru., and Sharath Kumar Y H., 2018. Deep Learning for Retrieval of Natural Flower Videos, Elsevier Procedia computer science, Vol. 132, pp. 1533-1542.
32. V.N. Manjunath Aradhya, Ashok Rao, G Hemantha Kumar, Language independent skew estimation technique based on Gaussian mixture models: a case study on South Indian scripts, International Conference on Pattern Recognition and Machine Intelligence, 2007, pp. 487-494.
33. TM Rajesh, VN Manjunath Aradhya, An application of GMM in signature skew detection, i-manager's Journal on Pattern Recognition, Vol. 2, Issue No. 3, 2015, pp.8.
34. Lohithashva, B.H., Manjunath Aradhya, V.N., Guru D S, Violent video event detection based on integrated LBP and GLCM texture features, Revue d'Intelligence Artificielle, Vol. 34, No. 2, 2020, pp. 179-187.
35. VN Manjunath Aradhya, G Hemantha Kumar, S Nousath, Multilingual OCR system for South Indian scripts and English documents: An approach based on Fourier transform and principal component analysis, Engineering Applications of Artificial Intelligence, Vol. 21, No.4, 2008, pp. 658-668.

Sl. No.	No. of Classes in a video	Multi class Flower Video	Flower Region of Interest	Correctly Identified?
1	2			Yes
				Yes
2	2			Yes
				No
3	2			Yes
				Yes
4	2			Yes
				No
5	2			Yes
				Yes
6	2			Yes
				Yes








7	2	 33.MOV		Yes
				Yes
8	3	 4.MOV		Yes
				Yes
				No

Fig. 14 Query acquiring the identity of the class for multiclass flower video