Artificial Intelligence in Agriculture 5 (2021) 262–277

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/

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

V.K. Jyothi [a,*], V.N. Manjunath Aradhya [b], Y.H. Sharath Kumar [c], 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), Manday 571438, Karnataka, India

ARTICLE INFO

Article history:
Received 25 August 2020
Received in revised form 6 November 2021
Accepted 6 November 2021
Available online 14 November 2021

Keywords:
Flower region of interest (FRoI)
Linear discriminant analysis (LDA)
Retrieval of flower videos
Multiclass support vector machine

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.

Contents

* Corresponding author.
E-mail addresses: jyothivk.mca@gmail.com (V.K. Jyothi), aradhya@sjce.ac.in (V.N.M. Aradhya), dsg@compsci.uni-mysore.ac.in (D.S. Guru).

https://doi.org/10.1016/j.aiia.2021.11.001

1. Introduction

Due to the ease of availability of recent video capturing devices such as cameras, mobiles, storage media, users can easily capture and store a large number of videos. Video contains a large information than images. A single video can capture the reality better than thousands of images. Recently, video databases have become much larger, hence there is need for automatic analysis and retrieval system with the minimum intervention of human is essentially required. Video has become a significant element of communication environment. 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 (Shen et al., 2016). Videos are categorized into different domains for example sports, news, surveillance, commercials, medical etc., again domain specific videos are categorized into different subcategories/classes (Geetha et al., 2009).

Data acquisition tools with recent technological advancements allowed researchers/scientists to acquire data from different application domains in the form of images and videos, these are a large and complex in nature (Mufti et al., 2021). One of the important aspects of organic life is its outstanding diversity. There exists a very large number of species of flowers in the world and the estimation of flower species ranges between 2,20,000 and 4,20,000 (Chaitra et al., 2021). Specialized knowledge is required to recognize the taxonomic information of flowers. Plant Identification skills and Taxonomic knowledge is restricted to a limited number of individuals (Jyothi et al., 2018, Wäldchen et al., 2018). To address the taxonomist's flower species identification requirement, a significant amount of research work has been carried out in the field of Artificial Intelligence and Video/Image Processing for automatic flower recognition and retrieval.

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. Floriculture is one of the important commercial trades in agriculture (Guru et al., 2010). Day to day there is an increase in the demand for flowers. Floriculture involves nursery, flower trade, seed production from flowers (Guru et al., 2011). Further, it is found useful in horticulture, interest in knowing the flower names for decoration, cosmetics and medicinal use etc., (Das et al., 1999a, 1999b). 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., (Das et al., 1999a, 1999b). 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. 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.

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., (Hu et al., 2011). In the literature we found retrieval of videos based on an object (Morand et al., 2010), frame (Shekar et al., 2016), video (Geetha et al., 2009; Gao et al., 2009; Han et al., 2014; Liang et al., 2012), keywords (Priya and Domnic, 2014). For the retrieval of videos the features and algorithms such as optical flow tensor and Hidden Markov Models (HMMs) (Gao et al., 2009), the multi-modal spectral clustering and ranking algorithm (Han et al., 2014), block wise intensity comparison (Geetha et al., 2009), Scale Invariant Feature Transform (SIFT) (Zhu et al., 2016), Bag-of-Features (Cui et al., 2016), dynamic weighted similarity measure with color and edge descriptors (Liang et al., 2012) 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) (Geetha et al., 2009), Fisher Discriminant Ratio (Shen et al., 2016), Linear Discriminant Analysis (Gao et al., 2009), semi-supervised linear discriminant analysis (Wang et al., 2016), supervised linear dimensionality reduction (Cui et al., 2016), nonparametric discriminant analysis (Khan et al., 2012) 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 (Guru et al., 2018a, 2018b) such as GLCM (Haralick et al., 1973), LBP (Ojala et al., 2002) and SIFT Lowe (2004) 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 (Guru et al., 2018a, 2018b) 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.

a) Creation of a reasonably large dataset of videos of flowers which shall be made available public for research purpose.
b) Proposal of fusion of features strategy to improve the performance of the existing model.
c) Proposal of an algorithmic model for the retrieval of videos of flowers using all flower regions of interest.
d) Proposal of a model for the retrieval of videos of flowers with maximum flower regions of interest
e) Adoption of a dimensionality reduction approach to improve the efficiency of the system.
f) Addressed retrieval of videos of flowers even when a query video contains flowers of more than one class.
g) 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. (See Table 1.)

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

$$X = \{x_{v1}, x_{v2}, x_{v3}, \ldots, x_{vi}, \ldots, x_{vn}\} \tag{1}$$

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

$$x_{vi} = \{F_1, F_2, F_3, \ldots, F_i, \ldots, F_N\} \tag{2}$$

Then the keyframes of the video x_{vi} are selected using GMM cluster based algorithmic model (Guru et al., 2018a, 2018b). 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,

$$K_y = \{k_1, k_2, k_3, \ldots, k_i, \ldots, k_y\} \tag{3}$$

The flowers in keyframes are segmented from their background using statistical region merging algorithm (Nock and Nielsen, 2004). The keyframes after segmentation can be defined as

$$SK_y = \{sk_1, sk_2, sk_3, \ldots, sk_i, \ldots, sk_y\} \tag{4}$$

3.1.1. Gaussian Mixture model (GMM)

Gaussian Mixture model (GMM) is a statistical and unsupervised learning model. GMM (Stauffer and Grimson, 1999), 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 (Chen et al., 2015).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

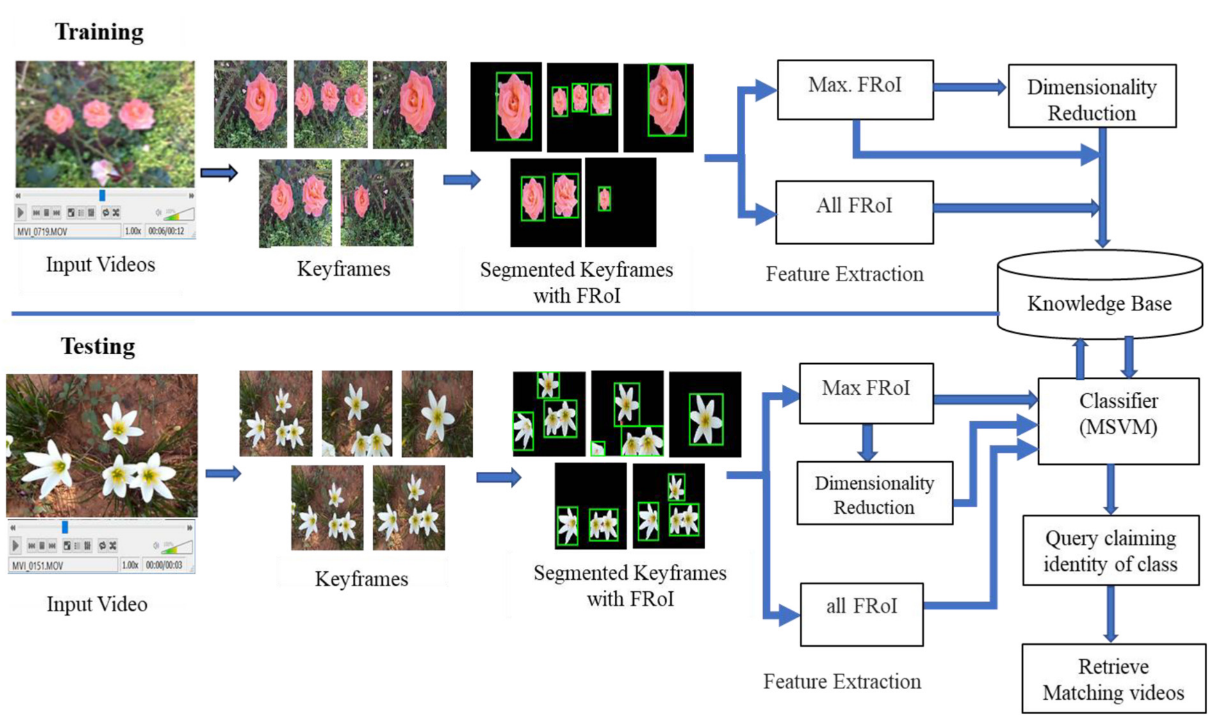


Fig. 1. Block diagram of the proposed class based flower video retrieval system.

Table 1
Summary of mentioned technologies and applications in related works.

Sl. No.	Algorithms	Applications	References
1	optical flow and Hidden Markov Models	retrieval of videos	Gao et al., 2009
2	multi-modal spectral clustering and ranking algorithm	retrieval of videos	Han et al., 2014
3	Principal Component Analysis	Feature dimensionality reduction	Geetha et al., 2009
4.	Fisher Discriminant Ratio, Linear Discriminant Analysis, semi-supervised linear discriminant analysis, supervised linear dimensionality reduction, nonparametric discriminant analysis	Feature dimensionality reduction	Shen et al., 2016 Gao et al., 2009 Wang et al., 2016 Cui et al., 2016 Khan et al., 2012

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 (Hong et al., 2014; Li et al., 2015). 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 flowers in videos consists of variation in view point and illumination, in such cases features extracted from Scale Invariant Feature Transform Lowe (2004) are considered.

3.2.1. Texture features

Texture of an image/frame contain unique visual patterns. Texture features describes the object surface, these features are independent of object color Hu et al. (2011). 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.

3.2.1.1. 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 (Haralick et al., 1973) are extracted from each FRoI. These features are represented as a feature vector.

3.2.1.2. 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 (Ojala et al., 2002). It is robust in terms of variation and transformation of the gray scale. In the proposed work, the system extracts LBP features (Ojala et al., 2002) which are invariant to local grayscale variations in the FRoI. LBP texture features are extracted using 3×3 neighbourhood by the value of centre pixel, the pixels of eight neighbors are thresholded. In 3×3 neighbourhood, the centre pixel LBP value is obtained by thresholded binary values are weighted by powers of two and summed up.

3.2.2. Scale invariant feature transform (SIFT)

SIFT plays a vital role in video retrieval for the analysis of the video content (Zhu et al., 2016). In SIFT the set of image features are generated in 4 stages Lowe (2004). 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 4×4 histogram of 8 bins. These histograms compute the direction and magnitude of the gradient in the region of 16×16 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 Lowe (2004).

To design the proposed model, the features such as Gray Level Co-occurrence Matrix (GLCM) (Haralick et al., 1973), Local Binary Pattern (LBP) (Ojala et al., 2002) and Scale Invariant Feature Transform (SIFT) features proposed by Lowe (2004) are extracted. Initially we propose to accomplish extracting these features by considering an entire keyframe after segmentation (Guru et al., 2018a, 2018b). 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.

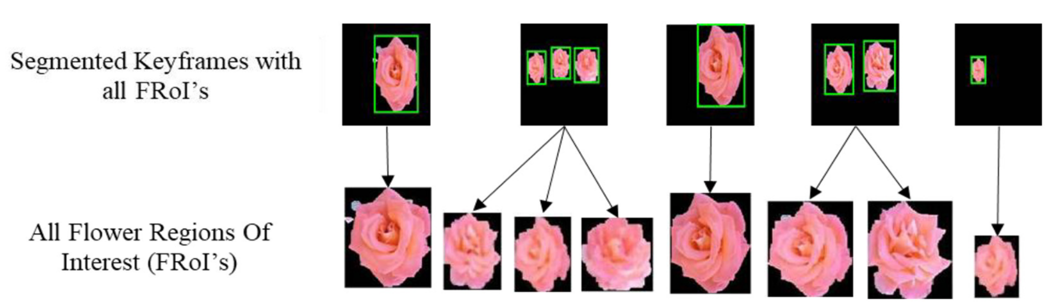


Fig. 2. Extraction of features from all flower regions of interest.

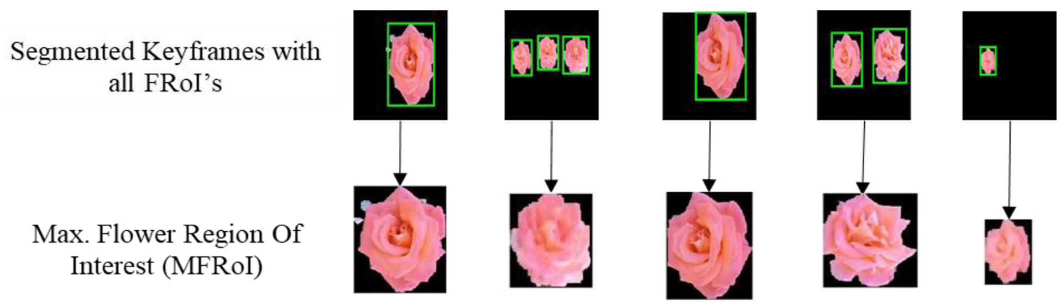


Fig. 3. Extraction of Maximum Flower Region of Interest (Max. FRoI).

Fig. 4. Samples of flower videos with large intraclass variation from 30 classes of videos.

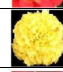
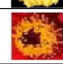
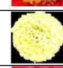
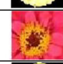
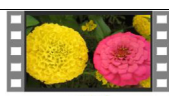
Sl. No.	No. of Classes in a video	Multi class Flower Video	Flower Region of Interest	Correctly Identified?
1	2	 12.MOV		Yes
				Yes
2	2	 9.MOV		Yes
				No
3	2	 10.MOV		Yes
				Yes
4	2	 16.MOV		Yes
				No
5	2	 18.MOV		Yes
				Yes
6	2	 22.MOV		Yes
				Yes
7	2	 33.MOV		Yes
				Yes
8	3	 4.MOV		Yes
				Yes
				No

Fig. 5. Query acquiring the identity of the class for multiclass flower video.

3.2.3. Entire keyframe

In this method (Guru et al., 2018a, 2018b), the model extracts the features such as Gray Level Co-occurrence Matrix (GLCM) (Haralick et al., 1973), Local Binary Pattern (LBP) (Ojala et al., 2002) and Scale Invariant Feature Transform (SIFT) Lowe (2004) 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,

$$x_{vi} = \{f_1, f_2, f_3, \ldots, f_i, \ldots, f_N\} \tag{6}$$

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

$$\mathfrak{R}^D = \{F_1M_1(x_{v1}), F_2M_2(x_{v2}), F_3M_3(x_{v3}), \ldots, F_iM_i(x_{vi}), \ldots, F_nM_n(x_{vn})\} \tag{7}$$

Where $F_1M_1(x_{v1}), F_2M_2(x_{v2}), F_3M_3(x_{v3}), \ldots, F_iM_i(x_{vi}), \ldots, F_nM_n(x_{vn})$ are the feature matrices of the videos $x_{v1}, x_{v2}, x_{v3}, \ldots, x_{vi}, \ldots, x_{vn}$ respectively in Eq. (1).

3.2.4. All flower regions of interest

The proposed system extracts features such as GLCM (Haralick et al., 1973), LBP (Ojala et al., 2002) and SIFT Lowe (2004) 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 Eq. (4). Then, sk_i with number of flower regions is defined as.

$$sk_i = \{R_1SK_i, R_2SK_i, R_3SK_i, \ldots, R_rSK_i\} \tag{8}$$

Then, the feature vector of all regions is represented as

$$R_1SK_i = [f_{11}, f_{12}, f_{13}, \ldots, f_{1M}], R_2SK_i = [f_{21}, f_{22}, f_{23}, \ldots, f_{2M}], \ldots, R_1SK_i = [f_{r1}, f_{r2}, f_{r3}, \ldots, f_{rM}]$$

Finally, the feature vector of all the regions of a keyframe sk_i as shown in Eq. (8) is represented as,

$$F_1M_1{}^d = \{[f_{11}, f_{12}, f_{13}, \ldots, f_{1M}], [f_{21}, f_{22}, f_{23}, \ldots, f_{2M}], \ldots, [f_{r1}, f_{r2}, f_{r3}, \ldots, f_{rM}]\}$$

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

$$F_1M_1^d = \forall R_iSK_i \in sk_i \tag{9}$$

Where $F_1M_1^d$ is the feature matrix of the video x_{vi} of the Eq. (1) consists of $\forall R_iSK_i$ all regions of a keyframe sk_i in Eq. (8).

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

$$FM^d(x_{vi}) = \forall F_jM_j^d \in SK_y \tag{10}$$

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

$$\mathfrak{R}^D = FM^d(x_{v1}), FM^d(x_{v2}), FM^d(x_{v3}), \ldots, FM^d(x_{vi}), \ldots, FM^d(x_{vn}) \tag{11}$$

3.2.5. Maximum flower region of interest

In this method, features such as GLCM (Haralick et al., 1973), LBP (Ojala et al., 2002) and SIFT Lowe (2004) 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 Eq. (4). Therefore, the feature vector defined in Eq. (8) can be redefined in this case as,

$$MF_iM_i^d = Max(R_iSK_i) \in sk_i \tag{12}$$

Where $MF_iM_i^d$ is the feature matrix of the video x_{vi} of the Eq. (1) consists of $Max(R_iSK_i)$ maximum flower region of a keyframe sk_i in Eq. (8).

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

$$FM^d(x_{vi}) = \forall \left(MF_jM_j^d\right) \in SK_y \tag{13}$$

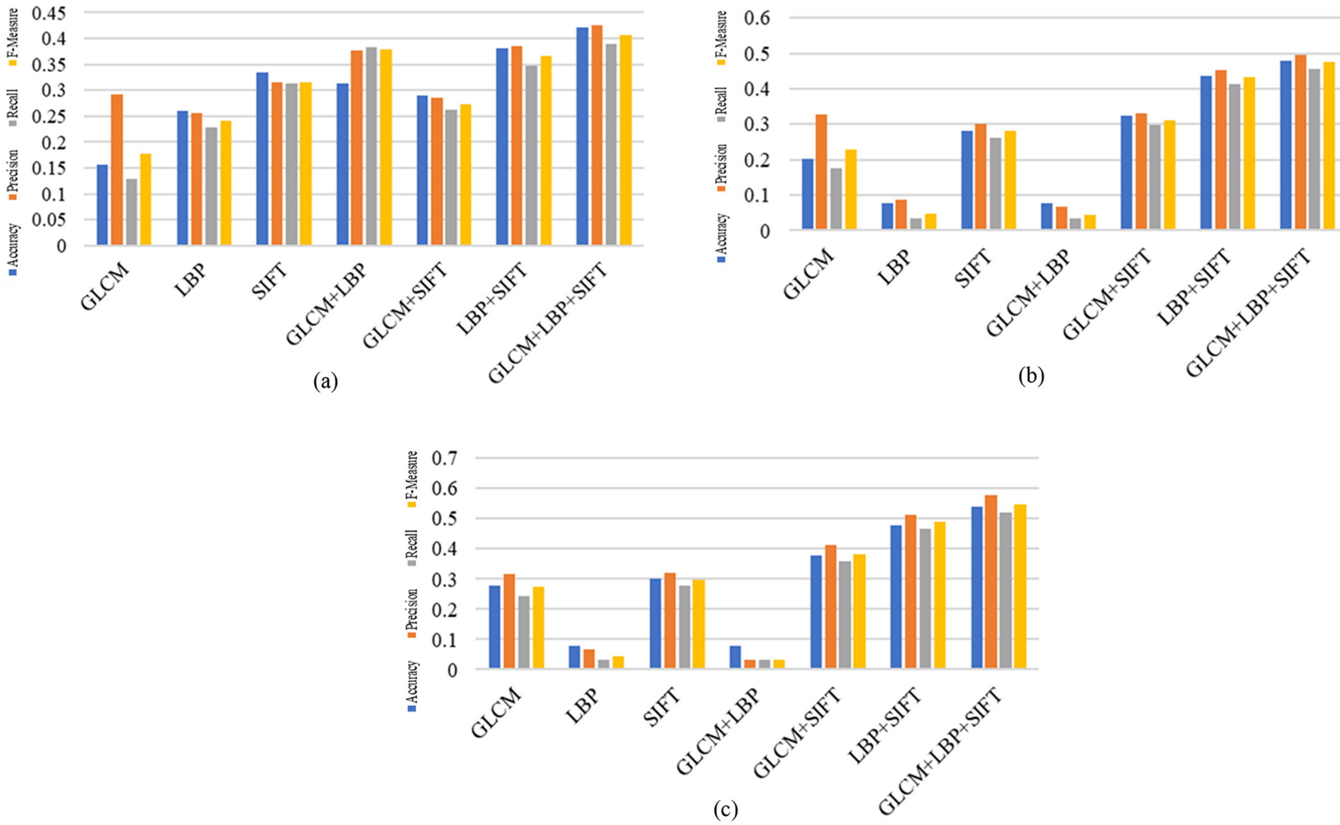


Fig. 6. Features extracted from all FRoI's for SGGP dataset: (a) 30% Train – 70% Test, (b) 50% Train – 50% Test, (c) 70% Train – 30% Test.

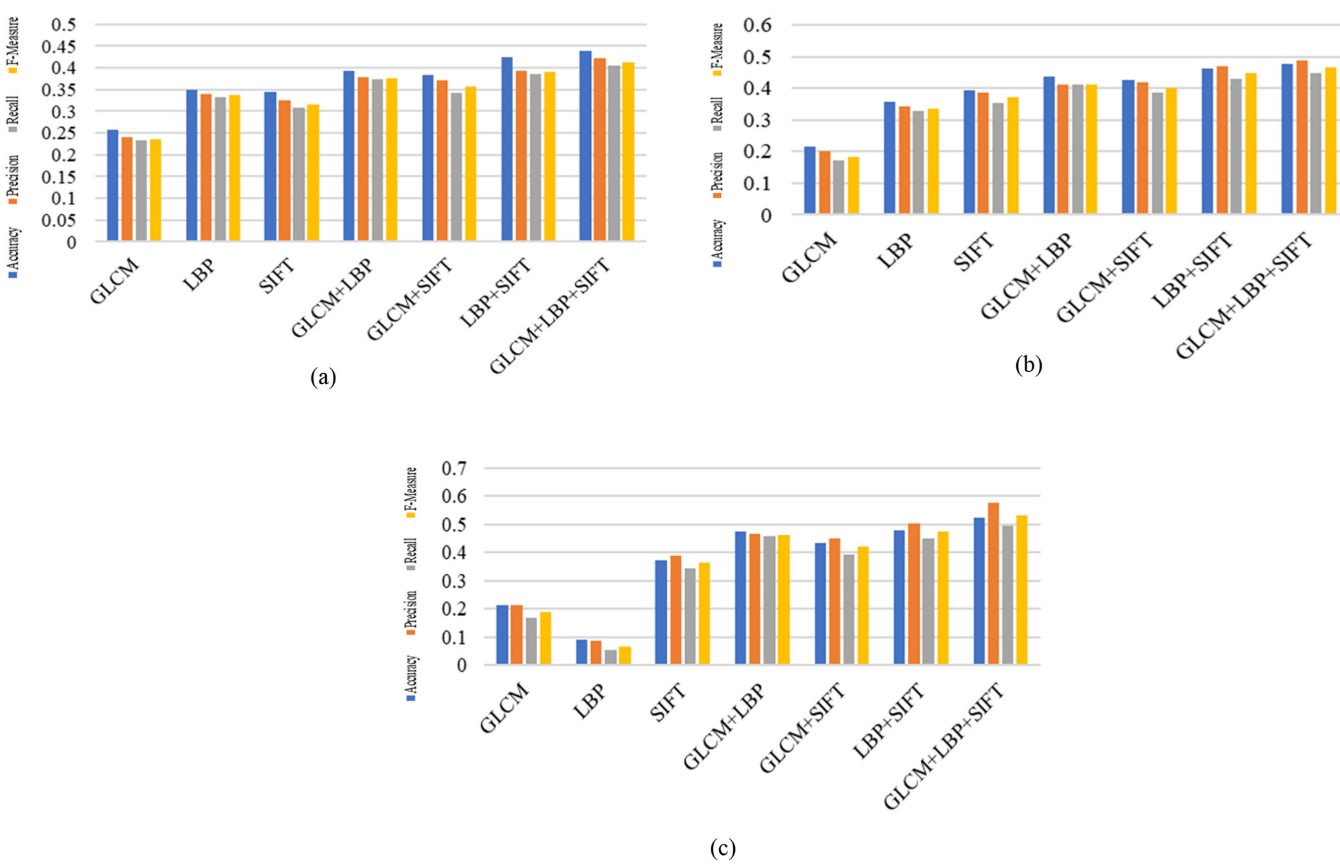


Fig. 7. Features extracted from all FRoI's for Sonycyber Shot dataset: (a) 30% Train - 70% Test, (b) 50% Train - 50% Test, (c) 70% Train - 30% Test.

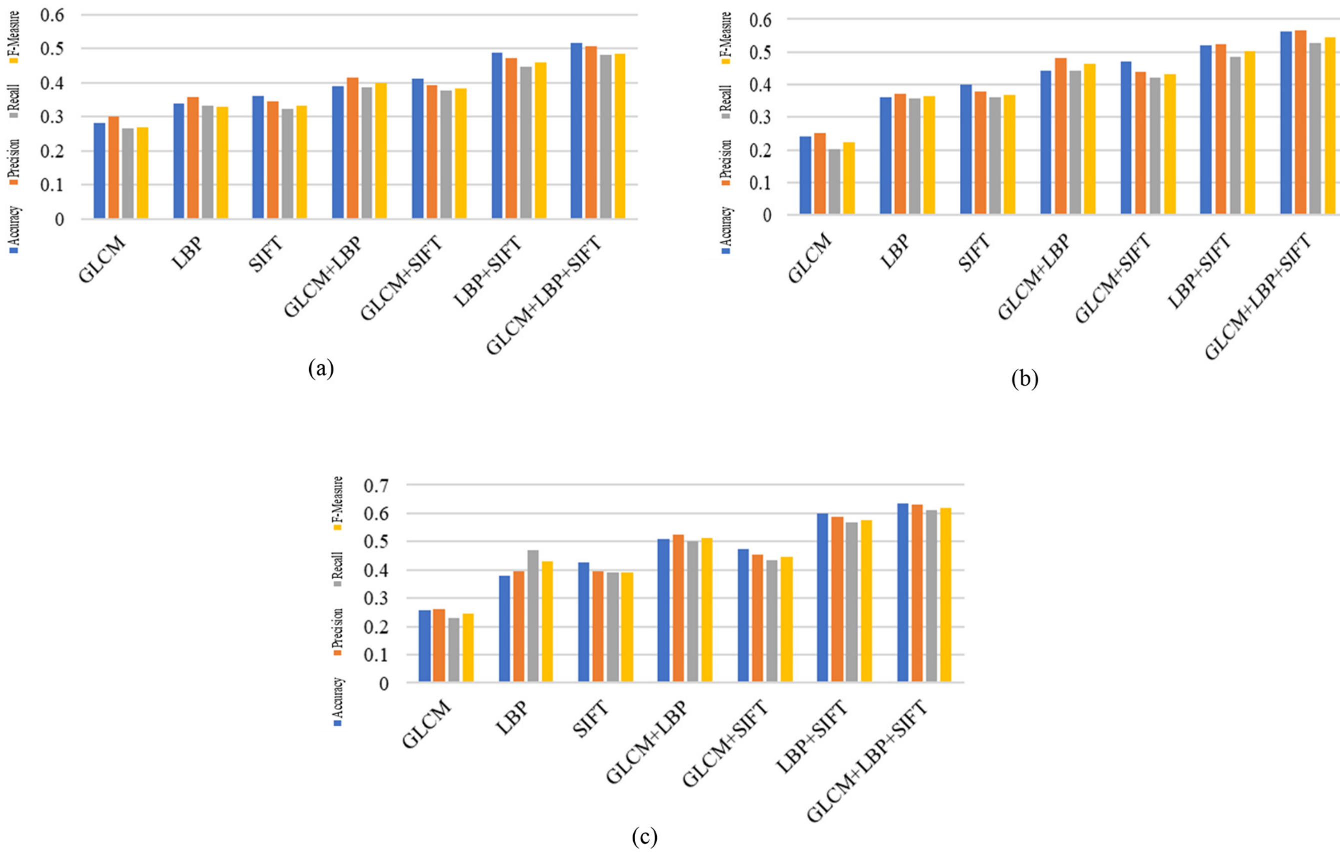


Fig. 8. Features extracted from all FRoI's for Canon dataset: (a) 30% Train – 70% Test, (b) 50% Train – 50% Test, (c) 70% Train – 30% Test.

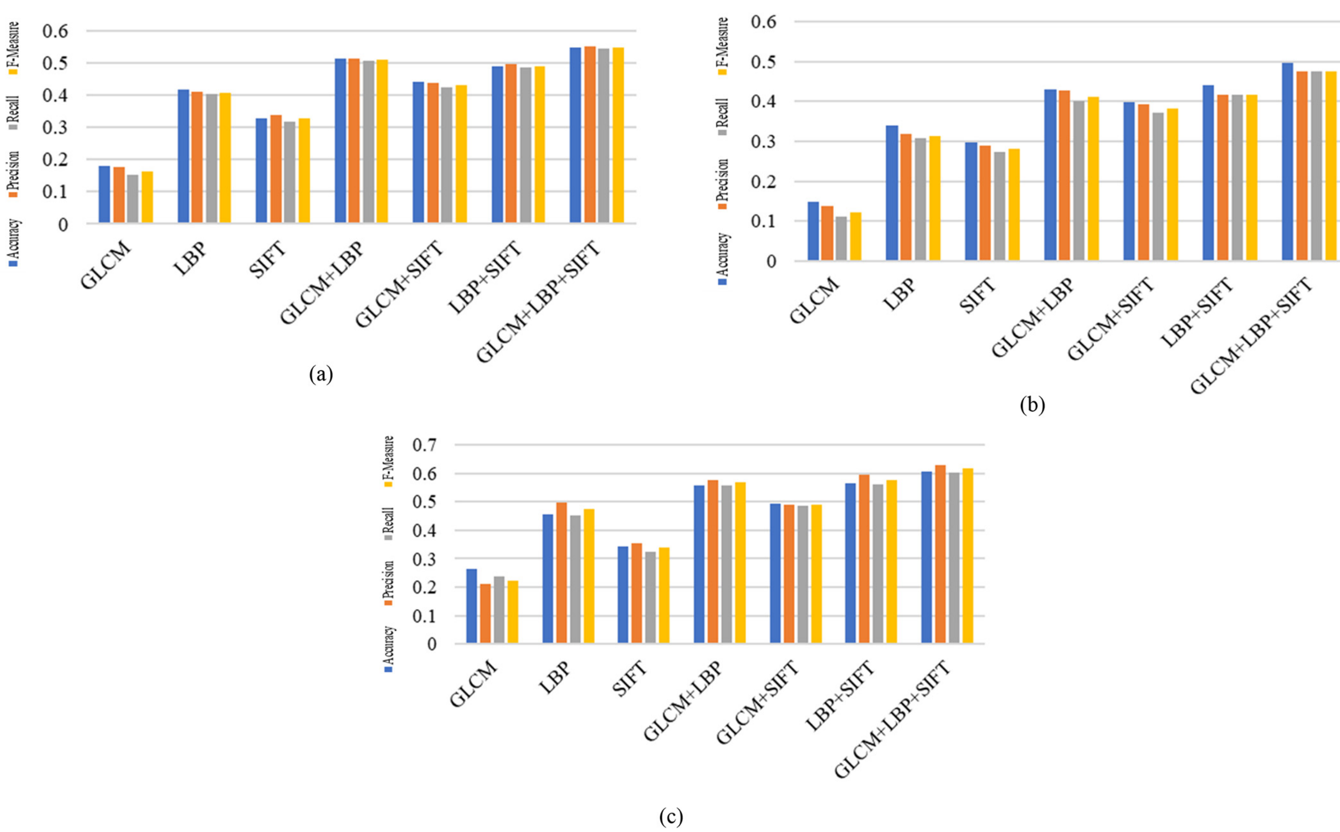


Fig. 9. Features extracted from Max FRoI for SGGP dataset: (a) 30% Train – 70% Test, (b) 50% Train – 50% Test, (c) 70% Train – 30% Test.

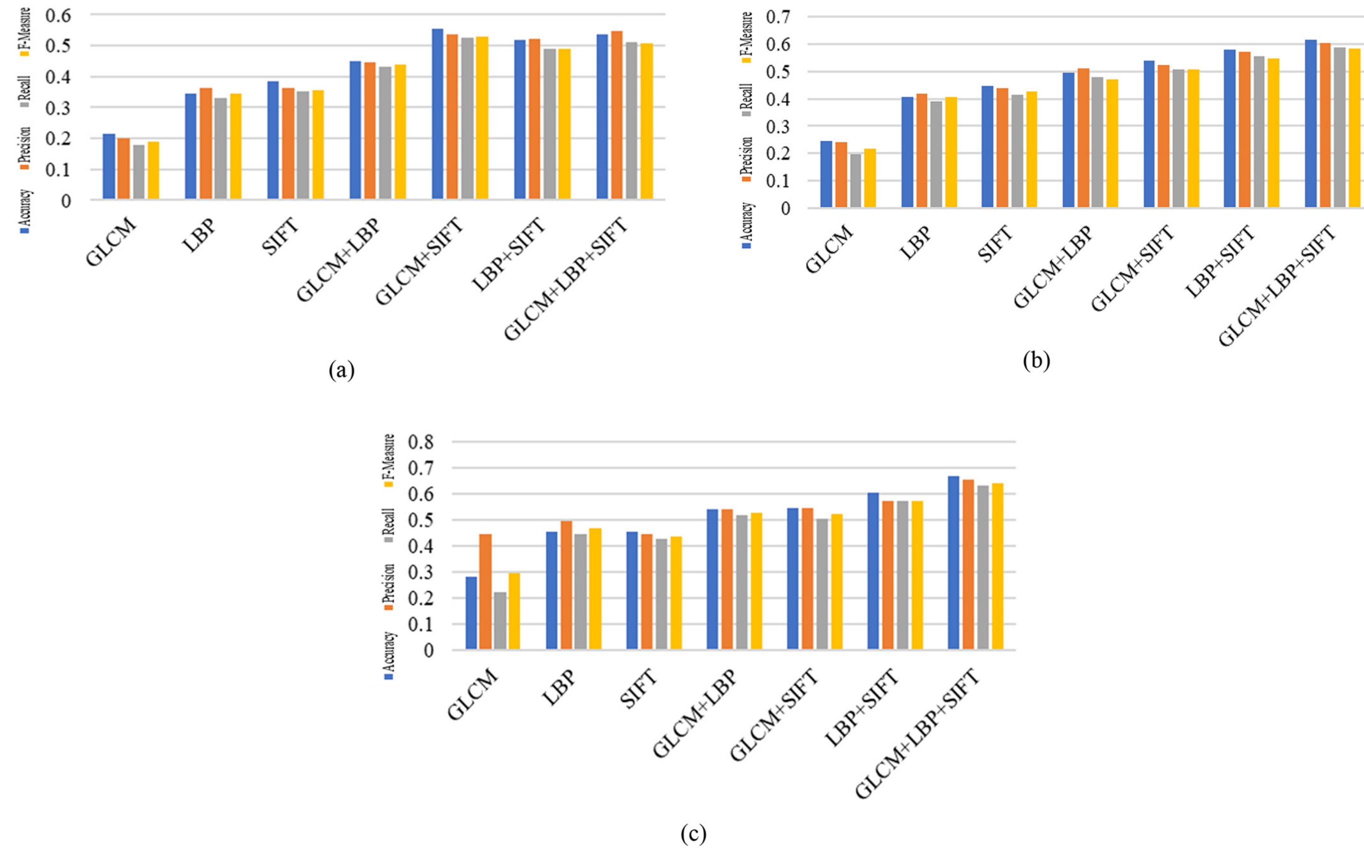


Fig. 10. Features extracted from Max FRoI for Sonycyber Shot dataset: (a) 30% Train–70% Test, (b) 50% Train – 50% Test, (c) 70% Train – 30% Test.

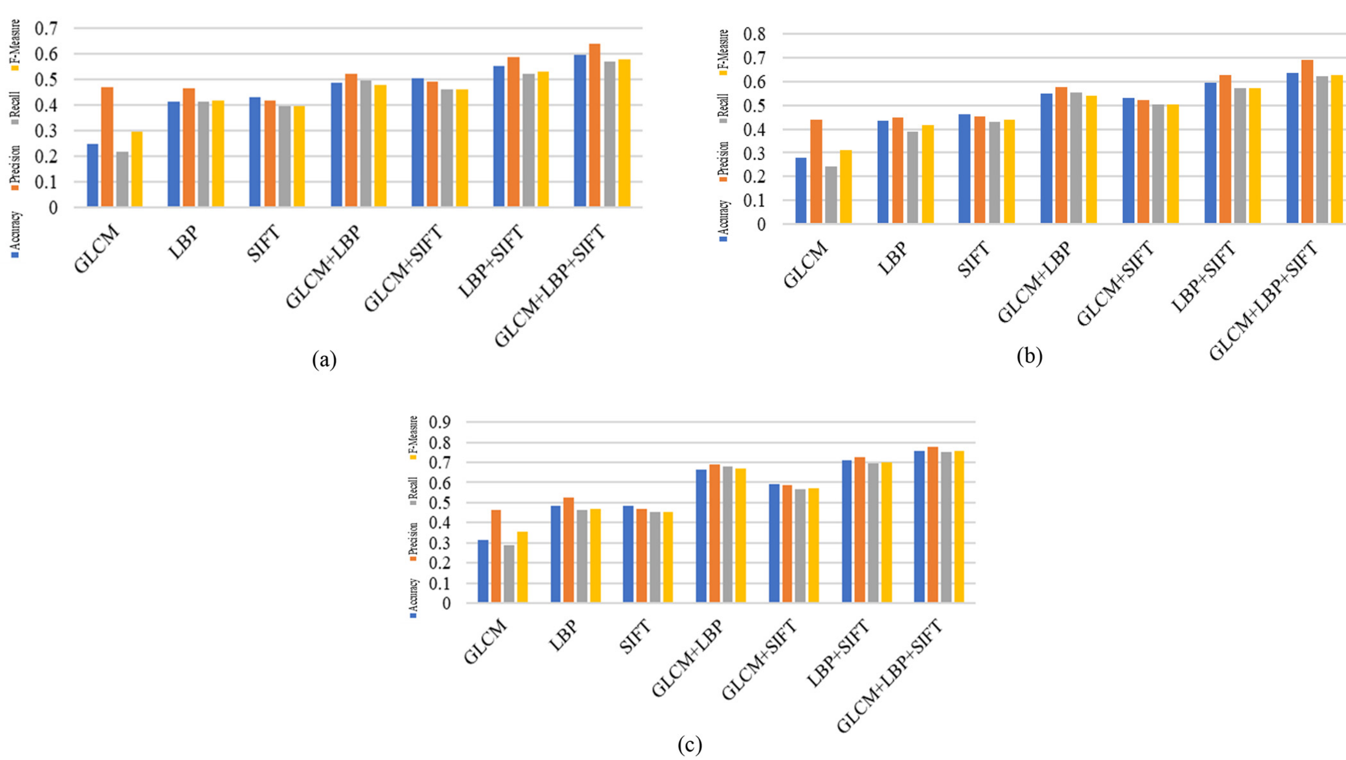


Fig. 11. Features extracted from Max FRoI for Canon dataset: (a) 30% Train – 70% Test, (b) 50% Train – 50% Test, (c) 70% Train – 30% Test.

Table 2
(a). SGGP Dataset: Train 30% - Test 70%. (b). SGGP Dataset: Train 50% - Test 50%. (c). SGGP Dataset: Train 70% - Test 30%.

(a)

Features	Accuracy	Precision	Recall	F-Measure
GLCM (Haralick et al., 1973)	0.2	0.18	0.15	0.17
LBP (Ojala et al., 2002)	0.13	0.11	0.07	0.09
SIFT Lowe (2004)	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

(b)

Features	Accuracy	Precision	Recall	F-Measure
GLCM (Haralick et al., 1973)	0.22	0.21	0.18	0.19
LBP (Ojala et al., 2002)	0.13	0.12	0.08	0.09
SIFT Lowe (2004)	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

(c)

Features	Accuracy	Precision	Recall	F-Measure
GLCM (Haralick et al., 1973)	0.33	0.31	0.3	0.29
LBP (Ojala et al., 2002)	0.15	0.16	0.1	0.11
SIFT Lowe (2004)	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 3
(a). Sonycyber Shot Dataset: Train 30% -Test 70%. (b). Sonycyber Shot Dataset: Train 50% -Test 50%. (c). Sonycyber Shot Dataset: Train 70% -Test 30%.

(a)

Features	Accuracy	Precision	Recall	F-Measure
GLCM (Haralick et al., 1973)	0.21	0.2	0.17	0.18
LBP (Ojala et al., 2002)	0.38	0.41	0.37	0.37
SIFT Lowe (2004)	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

(b)

Features	Accuracy	Precision	Recall	F-Measure
GLCM (Haralick et al., 1973)	0.24	0.24	0.2	0.22
LBP (Ojala et al., 2002)	0.37	0.38	0.4	0.35
SIFT Lowe (2004)	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

(c)

Features	Accuracy	Precision	Recall	F-Measure
GLCM (Haralick et al., 1973)	0.28	0.44	0.2	0.3
LBP (Ojala et al., 2002)	0.39	0.37	0.5	0.43
SIFT Lowe (2004)	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 4
(a). Canon Shot Dataset: Train 30% -Test 70%. (b). Canon Shot Dataset: Train 50% -Test 50%. (c). Canon Shot Dataset: Train 70% -Test 30%.

(a)

Features	Accuracy	Precision	Recall	F-Measure
GLCM (Haralick et al., 1973)	0.54	0.53	0.51	0.5
LBP (Ojala et al., 2002)	0.63	0.68	0.64	0.64
SIFT Lowe (2004)	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

(b)

Features	Accuracy	Precision	Recall	F-Measure
GLCM (Haralick et al., 1973)	0.56	0.55	0.52	0.53
LBP (Ojala et al., 2002)	0.66	0.7	0.67	0.67
SIFT Lowe (2004)	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

(c)

Features	Accuracy	Precision	Recall	F-Measure
GLCM (Haralick et al., 1973)	0.63	0.64	0.61	0.6
LBP (Ojala et al., 2002)	0.69	0.73	0.71	0.7
SIFT Lowe (2004)	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

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

The feature dimension of a video x_{vi} i.e., $FM^d(x_{vi})$ in Eq. (13) 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,

$$\Re^D = FM^d(x_{v1}), FM^d(x_{v2}), FM^d(x_{v3}), \ldots, FM^d(x_{vi}), \ldots, FM^d(x_{vn}) \quad (14)$$

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 Eq. (13) is represented as the reduced discriminant features obtained from Max. FRoI using LDA and it can be defined as

$$FM^d(x_{v1}) = \forall\left(MF_j M_j^d\right) \in SK_y \quad (15)$$

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

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

$$\Re^D = \Big\{DR\left(FM^d(x_{v1})\right)DR\left(FM^d(x_{v2})\right)DR\left(FM^d(x_{v3})\right)\ldots DR\left(FM^d(x_{vi})\right) \ldots DR\left(FM^d(x_{vn})\right)\Big\} \quad (16)$$

3.2.6. Linear discriminant analysis (LDA)

LDA is a supervised dimensionality reduction method (Belhumeur et al., 1999). 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

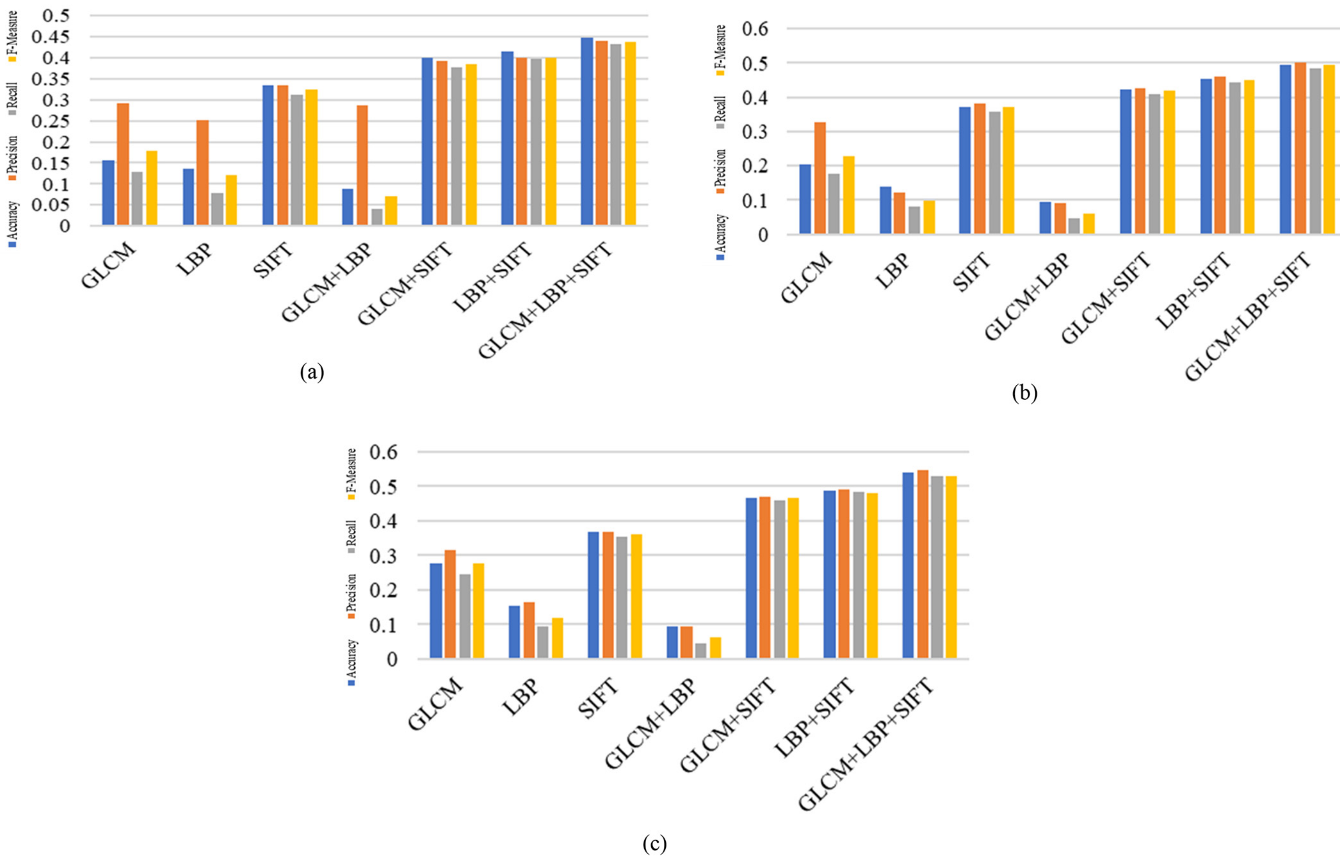


Fig. 12. Features extracted from entire keyframe for SGGP dataset: (a) 30% Train – 70% Test, (b) 50% Train – 50% Test, (c) 70% Train – 30% Test.

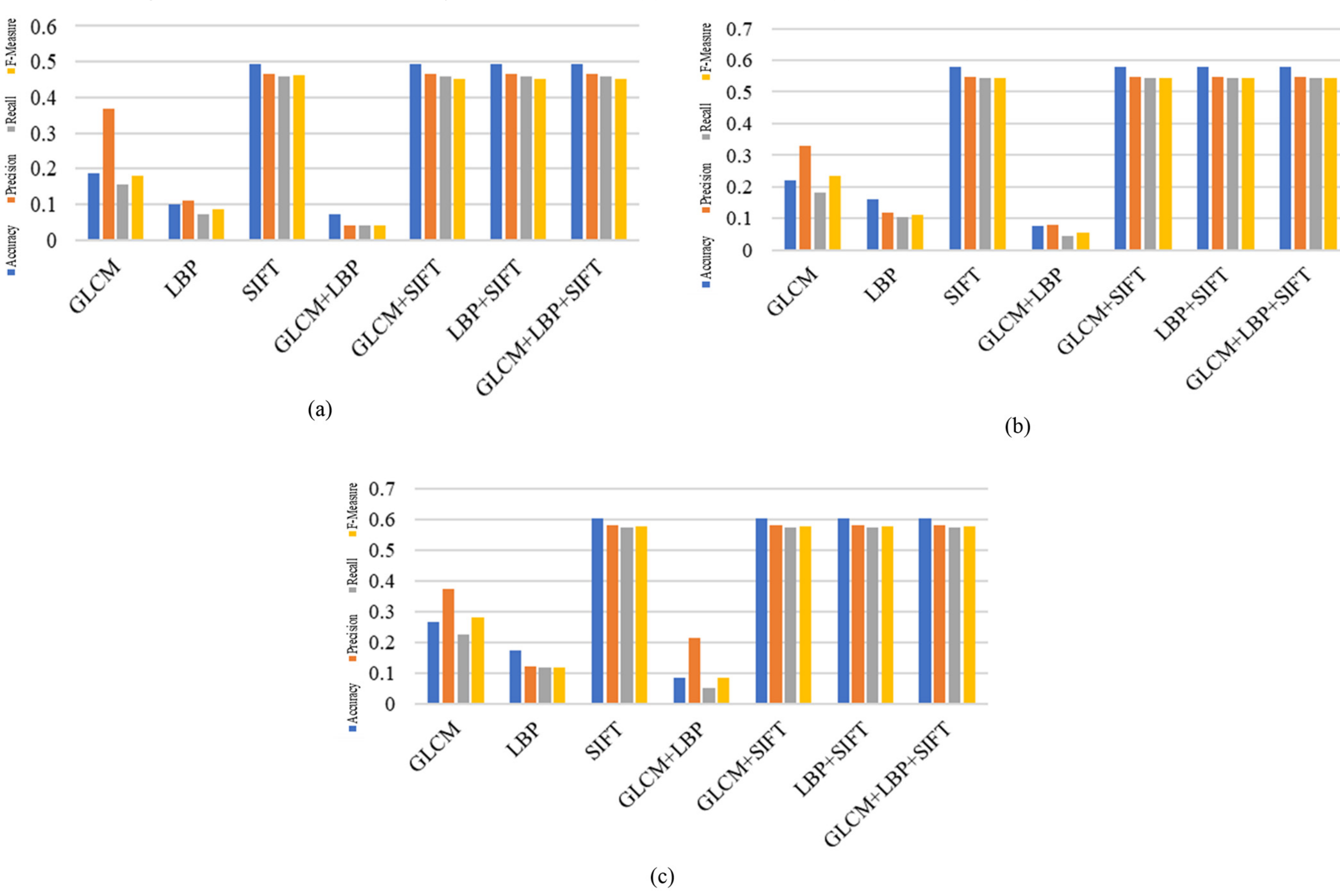


Fig. 13. Features extracted from entire keyframe for Sonycyber Shot dataset (a)30%Train-70%Test (b)50%Train-50% Test (c)70%Train-30% Test.

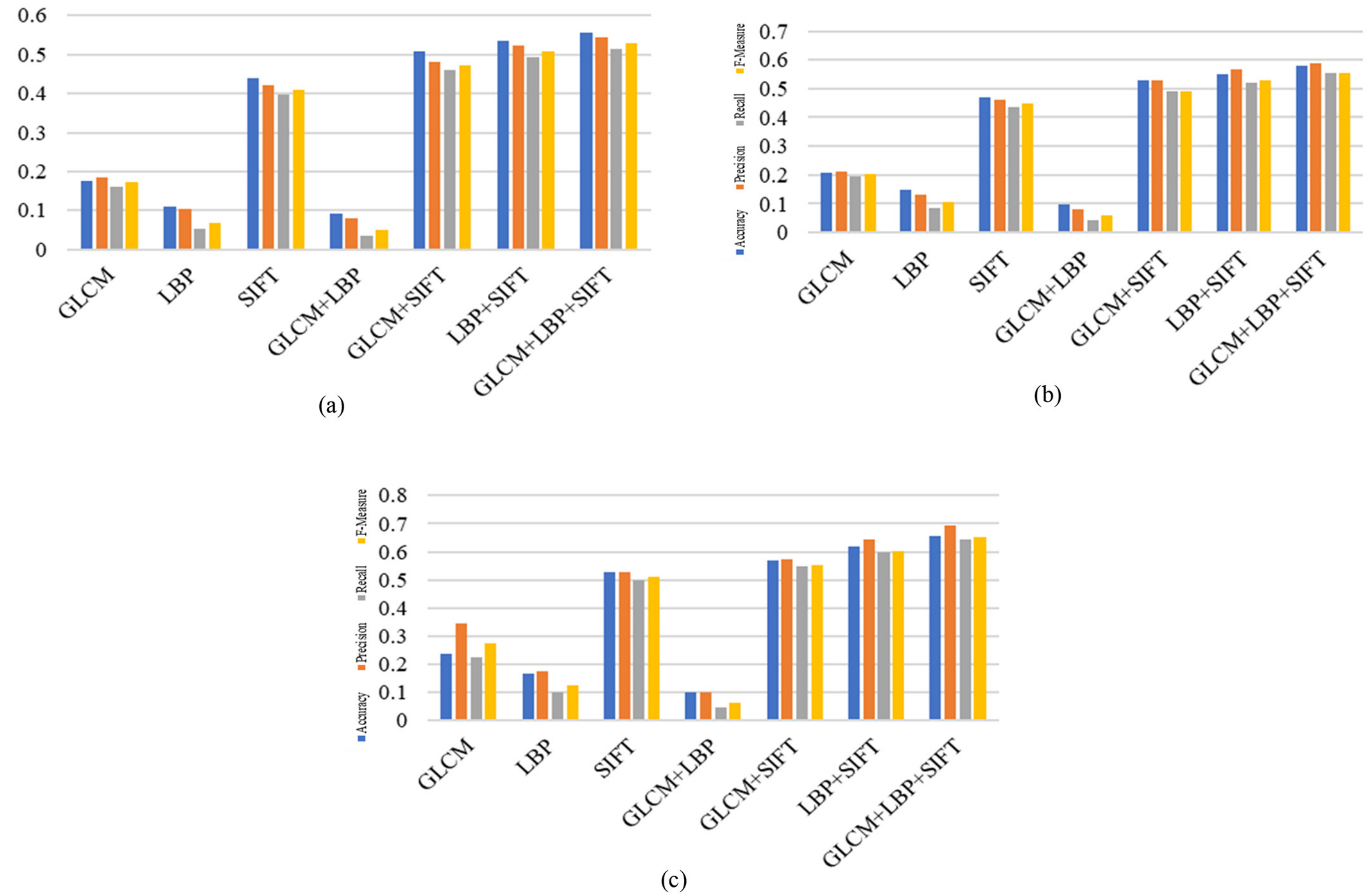


Fig. 14. Features extracted from entire keyframe for Canon dataset: (a) 30% Train – 70% Test, (b) 50% Train – 50% Test, (c) 70% Train – 30% Test.

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 (Gyamfi et al., 2018; Wang et al., 2016) 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,

$$DR(FMV_i) = \{ f_1, f_2, f_3, \ldots, f_{dr} \} \tag{17}$$

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 (Kumar and Gopal, 2011, and Khan et al., 2012). 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 (Vapnik (1998) and Duda et al., 1997).

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, Sonycyber Shot camera and Canon camera.

Table 5
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

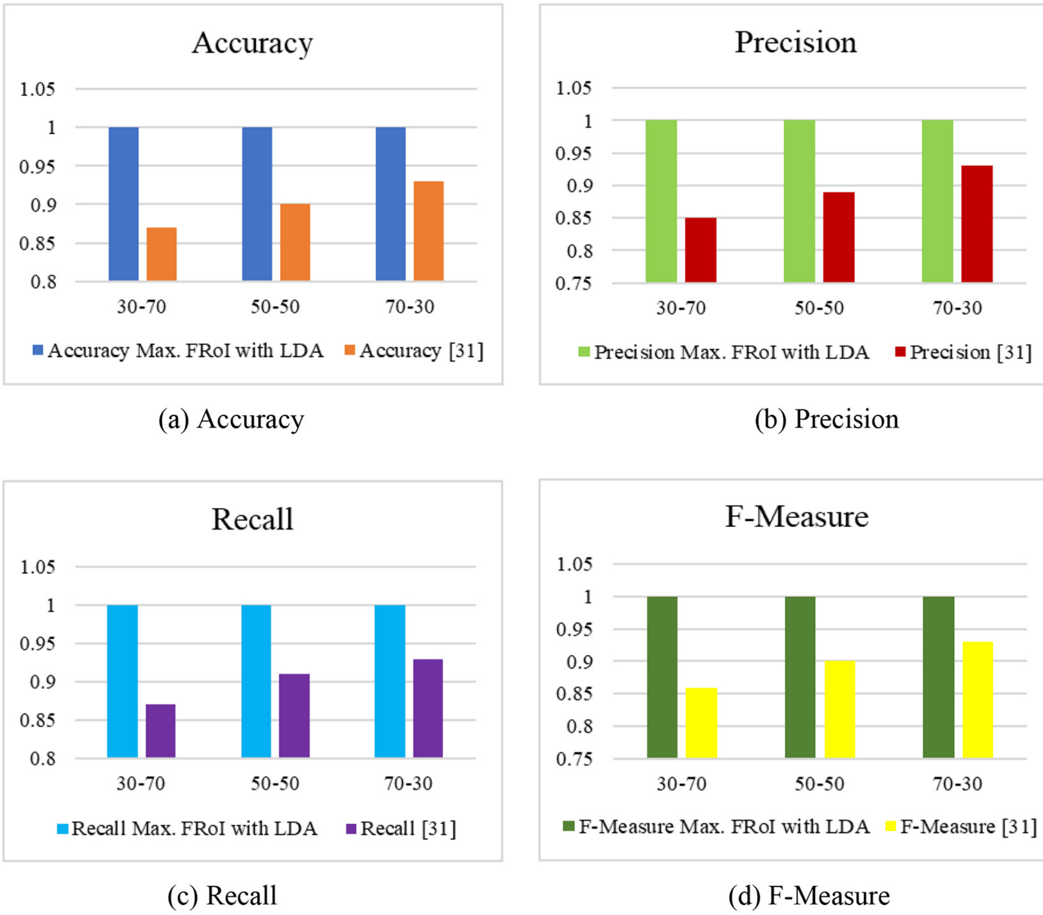


Fig. 15. Comparative study between proposed work and deep learning model (Jyothi et al., 2018) for SGGP dataset.

SGGP dataset consists of 2611 videos of 8 M pixels. Sonycyber Shot camera dataset consists of 2521videos of 14.1 M pixels. And Canon camera cosists of 2656 videos of 16 M pixels. Videos captured with the duration ranges from 4 to 60 s. 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. 4.

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. 5.

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 (Guru et al., 2018a, 2018b) are shown in section 4.5. The dataset we created is used to conduct experiments. In order to evaluate the system, metrices such as accuracy, precision, recall and F-measure are used and are given below. The results are tabulated with varying training and testing videos.

$$Accuracy = \frac{Sum\ of\ videos\ retrieved\ correctly}{Total\ number\ of\ query\ videos} \tag{18}$$

$$Precision = \frac{Total\ number\ of\ videos\ retrieved\ are\ relevant}{Total\ number\ of\ videos\ retrieved} \tag{19}$$

$$Recall = \frac{Total\ number\ of\ videos\ retrieved\ are\ relevant}{Total\ number\ of\ similar\ videos\ in\ the\ database} \tag{20}$$

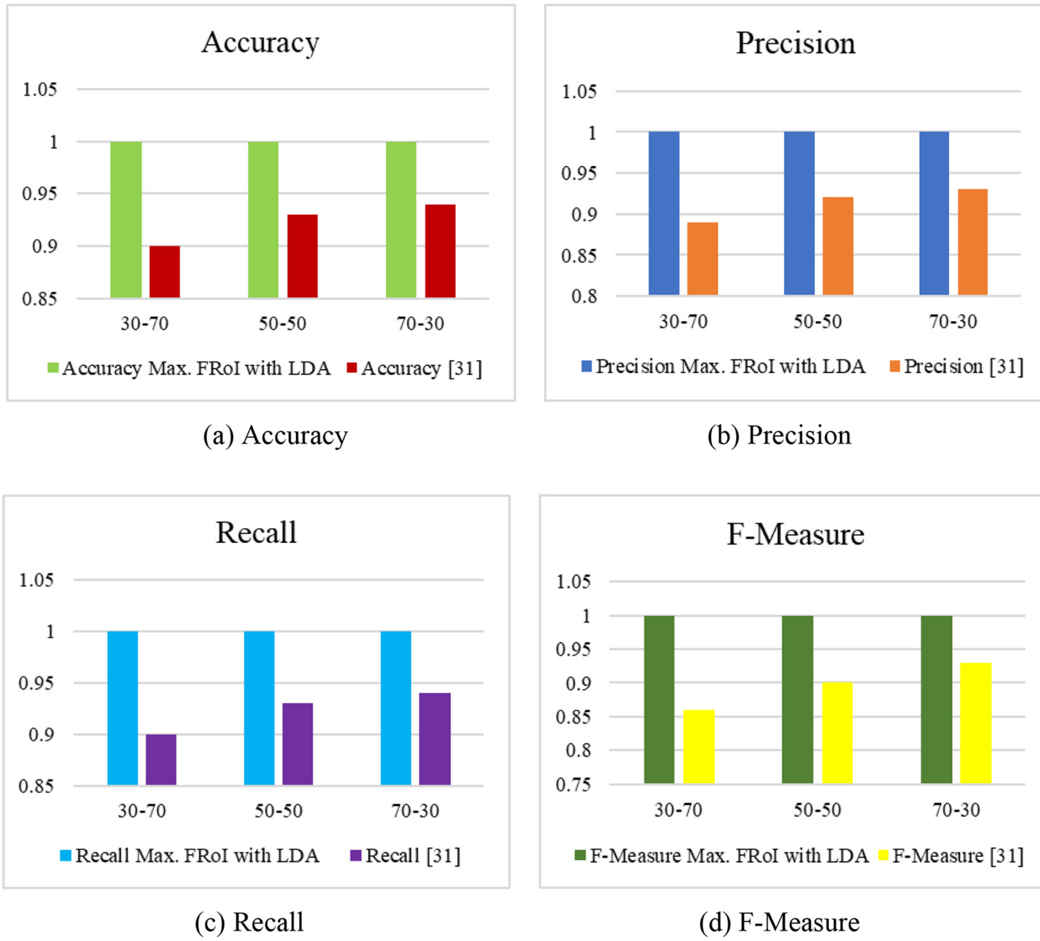
$$F{-}Measure = \frac{2{*}Precision{*}Recall}{(Precision + Recall)} \tag{21}$$

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. 6, Fig. 7 and Fig. 8 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.

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. 9, Fig. 10 and Fig. 11 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



(a) Accuracy (b) Precision

(c) Recall (d) F-Measure

Fig. 16. Comparative study between proposed work and deep learning model (Jyothi et al., 2018) for Sonycyber Shot dataset.

can observe that the Max. FRoI give improved results than all FRoI's for all the three datasets.

4.4. Max. FRoI with LDA

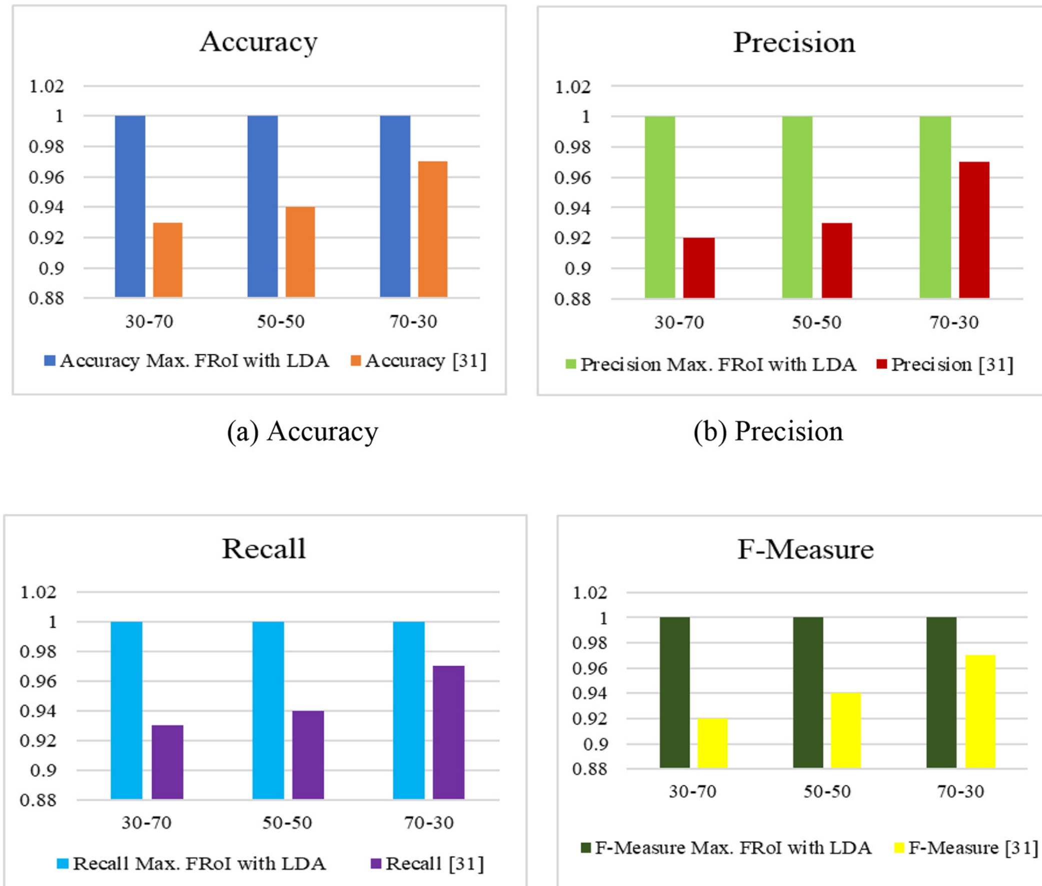
In this section we obtain discriminant features from Max. FRoI using LDA are passing to the model. It improves the retrieval performance by identifying the class of the query video. Table 2(a) to Table 2(c), Table 3 (a) to Table 3(c) and Table 4(a) to Table 4(c) show the result analysis of proposed retrieval system trained with the reduced features extracted from Max. FRoI is as shown in Eq. (14) 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.

4.5. Comparative study between proposed work and previous work

In the previous work (Guru et al., 2018a, 2018b) the features such as Gray Level Co-occurrence Matrix (GLCM) (Haralick et al., 1973), Local Binary Pattern (LBP) (Ojala et al., 2002) and Scale Invariant Feature Transform (SIFT) Lowe (2004) are extracted from entire keyframe. With the fusion of these features the model achieved good performance. The retrieval accuracy of previous work (Guru et al., 2018a, 2018b) achieved 53.83%, 60.18% and 65.73% are shown in Fig. 12, Fig. 13 and Fig. 14 for SGGP, Sonycyber Shot and Canon datasets respectively. In the proposed work, to further improve the retrieval performance, GLCM (Haralick et al., 1973), LBP (Ojala et al., 2002) and SIFT Lowe (2004) 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 (Guru et al., 2018a, 2018b). 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 5 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. Features extracted from entire keyframes of a video provide good results with the fusion of the features GLCM+LBP + SIFT as shown in Fig. 12 to Fig. 14. 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. 6 to Fig. 8 for SGGP, Sonycyber Shot and Canon datasets respectively. Max. FRoI's approach generates good results for the combination of features GLCM+LBP + SIFT as shown in Fig. 9 to Fig. 11 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. 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 2 and Table 3 show the combination of features



(a) Accuracy (b) Precision

(c) Recall (d) F-Measure

Fig. 17. Comparative study between proposed work and deep learning model (Jyothi et al., 2018) for Canon dataset.

LBP + SIFT achieves good performance for SGGP and Sonycyber Shot datasets. Table 4 show that the combinations of features LBP + SIFT and GLCM+LBP + SIFT achieves good performance for Canon dataset.

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. 5 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 (Jyothi et al., 2018), authors have proposed a flower video retrieval system using deep leaning approach, here the similar videos for a given query video are retrieved using Multiclass Support Vector Machine. For the extraction of features in (Jyothi et al., 2018), 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 (Jyothi et al., 2018), 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 (Jyothi et al., 2018) 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 (Jyothi et al., 2018) are compared with present work and the results are shown in Fig. 15, Fig. 16 and Fig. 17 for SGGP, Sonycyber Shot and Canon datasets respectively.

6. Future work

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

a) An attempt on shot boundary or class boundary detection when a video consists of multiple species of flowers can be further explored.

Explanation: In the present work, when a video consists of one class of flowers in some duration and then other classes in next duration. In such case, we manually split (cut) the video into shots. To overcome this drawback a video shot/class boundary detection is essential.

Further, the research can be automized for shot boundary detection instead of manually split the video shots.

b) The current research work limits the species of flowers to 30.

Explanation: The current research work limits the class size of flowers to 30. There is scope for extending the class size and explore different methodologies to retrieve flower videos in real time.

7. 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. Keyframes are segmented using statistical region merging algorithm. From the segmented keyframes features are extracted in three different modalities namely, all regions of flowers in the keyframe, maximum flower region in the keyframe and finally, features of maximum flower region with a set of discriminating features generated by LDA. The presented system is compared against our previous work (Guru et al., 2018a, 2018b) and the deep learning retrieval system (Jyothi et al., 2018), which reveals that the proposed system with Max. FRoI with LDA is superior to the existing models in terms of retrieval results. The proposed system retrieves similar videos when the query video contains multiple classes of flowers in a video. In proposed work, when video consists of one class of flowers in some duration and then other classes in next duration. In such case, we manually split (cut) the video into shots. Further, the research can be extended by automizing the shot boundary detection instead of manually split the video shots. And also, the current research work limits the species of flowers to 30, further the class size can be extended and can explore different methodologies to retrieve flower videos in real time.

Declaration of Competing Interest

We don't have any conflict of interest.

References

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1999. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. 19, 711–720.

Chaitra, K.N., Jyothi, V.K., Chandrajit, M., Guru, D.S., 2021. Flower classification in videos: a hog-PCA-NN method. 2nd International Conference on Artificial Intelligence: Advances and Applications, ICAIAA 2021.

Chen, W., Tian, Y., Wang, Y., Huang, T., 2015. Fixed-point gaussian mixture model for analysis friendly surveillance video coding. Comput. Vis. Image Underst. 142, 65–79.

Cui, M., Cui, J., Li, H., 2016. Dimensionality reduction for histogram features: a distance-adaptive approach. Neurocomputing 173, 181–195.

Das, M., Manmatha, R., Riseman, E., 1999a. Indexing flower patent images using domain knowledge. IEEE Intell. Syst. 14 (5), 24–33.

Das, M., Manmatha, R., Riseman, E.M., 1999b. Indexing flower patent images using domain knowledge. IEEE Intell. Syst. 14 (5), 24–33.

Duda, R.O., Hart, P.E., Stork, D.G., 1997. Pattern Classification. Second edition. .

Gao, X., Xuelong, L., Jun, F., Dacheng, T., 2009. Shot-based video retrieval with optical flow tensor and hmms. Pattern Recogn. Lett. 30, 140–147.

Geetha, M.K., Palanivel, S., Ramalingam, V., 2009. A novel block intensity comparison code for video classification and retrieval. Expert Syst. Appl. 36, 6415–6420.

Guru, D.S., Sharath, Y.H., Manjunath, S., 2010. Texture features and knn in classification of flower images. IJCA special Issue on Recent Trends in Image Processing and Pattern Recognition, RTIPPR, pp. 21–29.

Guru, D.S., Sharath, Y.H., Manjunath, S., 2011. Texture features in flower classification. Math. Comput. Model. 54, 1030–1036.

Guru, D.S., Jyothi, V.K., Kumar, Y.H.S., 2018a. Features fusion for retrieval of flower videos. Proceedings of DAL 2018, LNNS 43, pp. 221–234.

Guru, D.S., Jyothi, V.K., Kumar, Y.H.S., 2018b. Cluster based approaches for keyframe selection in natural flower videos. Springer AISC 736, ISDA 2018, pp. 474–484.

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 Syst. Appl. 91, 252–262.

Han, J., Ji, X., Hu, X., Han, J., Liu, T., 2014. Clustering and retrieval of video shots based on natural stimulus FMRI. Neurocomputing 144, 128–137.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernet. SMC-3 (6), 610–621.

Hong, R., Pan, J., Hao, S., Wang, M., Xue, F., Wu, X., 2014. Image quality assessment based on matching pursuit. Inf. Sci. 273, 196–211.

Hu, W., Xie, N., Li, L., Xianglin, Z., Maybank, S., 2011. A survey on visual content-based video indexing and retrieval. IEEE Trans. Syst. Man Cybernet. 41 (6), 797–819.

Jyothi, V.K., Guru, D.S., Sharath Kumar, Y.H., 2018. Deep learning for retrieval of natural flower videos. Elsevier Proc. Comput. Sci. 132, 1533–1542.

Khan, N.M., Ksantini, R., Ahmad, I.S., Boufama, B., 2012. A novel SVM+NDA model for classification with an application to face recognition. Pattern Recogn. 45, 66–79.

Kumar, M.A., Gopal, M., 2011. A hybrid SVM based decision tree. Pattern Recogn. 43, 3977–3987.

Li, J., Li, X., Yang, B., Sun, X., 2015. Segmentation-based image copy move forgery detection scheme. IEEE Trans. Inf. Foren. Secur. 10 (3), 507–518.

Liang, B., Xiao, W., Liu, X., 2012. Design of video retrieval system using MPEG-7 descriptors. Proc. Eng. 29, 2578–2582.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoint. Int. J. Comput. Vis. 60, 91–110.

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 Process. Image Commun. 25, 450–465.

Mufti, M., Shamim, K.M., McGinnity, Martin, Hussain, Amir, 2021. Deep learning in mining biological data. Cogn. Comput. 13, 1–33.

Nock, R., Nielsen, F., 2004. Statistical region merging. IEEE Trans. Pattern Anal. Mach. Intell. 26 (11), 1–7.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24 (7), 971–987.

Priya, G.G.L., Domnic, S., 2014. Shot based keyframe extraction for ecological video indexing and retrieval. Ecol. Inform. 23, 107–117.

Shekar, B.H., Uma, K.P., Holla, K.R., 2016. Video clip retrieval based on lbp variance. Proc. Comput. Sci. 89, 828–835.

Shen, X.-J., Mu, L., Li, Z., Wu, H.-X., Gou, J.-P., Chen, X., 2016. Large-scale support vector machine classification with redundant data reduction. Neurocomputing 172, 189–197.

Stauffer, C., Grimson, W.E.L., 1999. Adaptive background mixture models for real-time tracking. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.

Vapnik, V.N., 1998. Statistical Learning Theory. John wiley and sons, New York, USA.

Wäldchen, Jana, Rzanny, Michael, Seeland, Marco, Mäder, Patrick, 2018. Automated plant species identification—Trends and future directions. https://doi.org/10.1371/journal.pcbi.1005993.

Wang, S., Lu, J., Gu, X., Du, H., Yang, J., 2016. Semi-supervised linear discriminant analysis for dimension reduction and classification. Pattern Recogn. 57, 179–189.

Zhu, Y., Huang, X., Huang, Q., Tian, Q., 2016. Large-scale video copy retrieval with temporal concentration sift. Neurocomputing 187, 83–91.