

Article

Not peer-reviewed version

---

# DeepFlower: Archival and Retrieval of Videos

---

[Manjunath Aradhy](#)<sup>\*</sup>, Jyothi V K, [Sharath Kumar](#), Guru D S, Vinay Kumar

Posted Date: 29 November 2023

doi: 10.20944/preprints202311.1909.v1

Keywords: Keyframe selection; Dimensionality Reduction; ReliefF; Bi-normal separation; Indexing; KD-tree; retrieval of flower video; Deep Flower



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# DeepFlower: Archival and Retrieval of Videos

Jyothi V K <sup>1\*</sup>, V. N. Manjunath Aradhya <sup>2</sup>, Sharath Kumar Y H <sup>3</sup>, D S Guru <sup>1</sup> and N Vinay Kumar <sup>1</sup>

<sup>1</sup> Department of Studies in Computer Science, Manasagangothri, University of Mysore, Mysuru 570 006, Karnataka, India; dsg@compsci.uni-mysore.ac.in (D.S.G.); vinaykumar.natraj@gmail.com (N.V.K.)

<sup>2</sup> Department of Computer Applications, JSS Science and technology University, Mysuru, Karnataka, India; aradhya@sjce.ac.in

<sup>3</sup> Department of Information Science and Engineering, Maharaja Institute of Technology Mysore (MITM), Mandya, Karnataka, India; sharathyhk@gmail.com

\* Correspondence: jyothivk.mca@gmail.com; Tel.: +91-9148822089

**Abstract:** This paper presents a model for archival and retrieval of the videos of natural flowers. To design an efficient video retrieval system the stages namely, keyframe selection, feature extraction, feature dimensionality reduction and indexing are essential for fast browsing and accessing of videos. Three different keyframe selection approaches are proposed using clustering algorithms after segmenting flower regions from its background. Deep Convolutional Neural Network is used as a feature extractor. After keyframe selection, a video is represented with a set of keyframes. To reduce the feature dimension of a video, two feature selection methods are utilized. For an efficient archival and fast retrieval of flower videos an indexing method called KD-tree is recommended. For a given query video, similar videos are retrieved both in relative and absolute search modalities. An extensive experimentation conducted on a relatively large flower video dataset. The data set consists of 7788 videos of 30 different species of flowers. The videos are captured with three different devices in different resolutions. The comparative study reveals proposed keyframe selection approaches gives better results. It has also been observed that the videos retrieved in absolute approach with features selected from Binormal separation metric and indexing gives good results.

**Keywords:** keyframe selection; dimensionality reduction; relief; bi-normal separation; indexing; kd-tree; retrieval of flower video; deep flower

## 1. Introduction

Designing and developing a flower video retrieval system is a domain specific with many applications. Flower video retrieval system is an application in the field of floriculture for commercial trades. Floriculture is one of the important commercial trades in agriculture [48]. Day to day there is an increase in the demand for flowers. Floriculture involves nursery, flower trade, seed production from flowers [49]. Further, it is found useful in horticulture, interest in knowing the flower names for decoration, cosmetics and medicinal use etc., [47]. Due to the development of technology in business, floriculture traders can store a large volume of videos of flowers. Users can analyse the entire flower in a video before purchasing it and its seeds, instead of visiting the nurseries for flowers of their interest. Also, they can view different species of flowers in videos along with different variants available in each species. In such cases, it is essential to develop an automated system to search and retrieve users desired videos of flowers. Designing an automated system for the retrieval of flower videos is a challenging task, when the videos of flowers are captured in natural environment in all ecological conditions such as rainy, sunny and cloudy. The flowers in video pose a number of challenges such as occlusions, illumination, view point variation, scaling and multiple instances. To design an efficient video retrieval system, the phases such as feature extraction, keyframe selection, feature selection and indexing are essential.

Keyframe selection can be performed generally in two different methods namely, sequential and cluster-based [62]. In the sequential technique, video frames are compared sequentially using a threshold [4–6]. The following approaches are found in sequential keyframes selection. An aggregation method with thresholding [13], cumulative frame difference of frames [14], an

aggregation of depth, motion and spatial saliency feature maps using dynamic programming optimization algorithm and an object based sequential approach [16]. In cluster based keyframe selection, similar frames are grouped together to select one of the frames in that group as a keyframe [3,7] using clustering algorithms. Clustering algorithms such as fuzzy c-means [9], hierarchical [11], affinity propagation clustering [12], Gaussian model [3,18,26], K-Means [3], Hidden Markov Model (HMM) are found in the literature.

To extract keyframes from a video, features are extracted from each frame of the video. Features such as color, texture and entropy [3,11], the fusion of saliency features texture, curvature, multi scale contrast and motion [10], an energy feature consuming with potential energy and a kinetic energy [12], RGB correlation, the histogram of color and moments of inertia [13], DCT domain features are extracted. Keyframes can be selected using the scores of brightness, hue count, distribution of edge, contrast [8], Harries corner detector [18], Scale Invariant Feature Transformation [21], Speeded Up Robust Feature, Bag of Words [22], multi resolution histograms [23], Color and edge histogram [24], SURF descriptors [25], block wise intensity [5], color, motion and edge features [27,28] are extracted.

In literature, we found that bundle of hand crafted features are required to select keyframes. Features of the keyframes obtained from deep learning approach are represented as a feature vector. When the dimension of the feature vector of a video is high, then there may exists a small amount of redundant and unnecessary information. The feature dimensionality reduction is required to select most discriminant features to reduce the feature dimension and to eliminate the redundant features when the dimension of the feature is high [30]. Generally, there are two different categories in dimensionality reduction methods, namely, feature transformation and feature selection [69]. The following methods on feature transformation are found such as PCA [5,33,35,43], Linear Discriminant Analysis (LDA) [31], generalized eigen decomposition method [32], semi-supervised LDA [34,42], normalized discriminant analysis [40], canonical correlation analysis [42]. The feature selection methods are categorized into wrapper, filter and embedded based methods [53]. Examples of feature selection techniques are ReliefF algorithm [57,71], Bi-normal separation [58], Fisher Score, Chi-Squares, Information gain and Lasso [70]. After dimensionality reduction, the feature space reduces. Further it will be used for indexing and retrieval. Indexing is a technique used to speed up the accessing of videos more accurately. Due to the large volume of data manual organization cannot be done. To organize videos an efficient indexing mechanism is essential to retrieve the required and more relevant videos. Indexing methods such as Fast-Pattern-Index tree [19], Hierarchical affinity hybrid tree multidimensional indexing are used for video retrieval.

This paper presents a video retrieval system using Deep Convolutional Neural Network (DCNN) because of their capability of overcoming the drawback of traditional feature extraction algorithms of hand crafted features. Initially DCNN was designed for image classification [36]-[39]. To retrieve related text from image deep learning is applied [44]. Deep learning techniques are suitable for identification and retrieval of videos and images. In Computer vision and machine learning, retrieval of relevant videos is a challenging task because of interclass similarity and intraclass variability [45]. Videos can be retrieved based on query types such as an image, video, sketch or objects [46]. In the proposed work videos are retrieved based on a query video.

From the literature, it is clear, all the keyframe selection methods are traditional ways. Traditional methods require bundle of hand crafted features. This paper presents, a keyframe selection approach using deep learning method. Further, we have not found that a suitable feature selection method for feature dimension reduction for features extracted from the keyframes of the flower videos. Therefore, this paper presents suitable feature selection method. Fast accessing of the flower videos using indexing approach is the first of its kind in the literature. Flower video retrieval system using keyframe selection, feature selection and an indexing scheme, presents query by video framework for the retrieval of similar videos for a given query.

The overall contributions of the proposed work are:

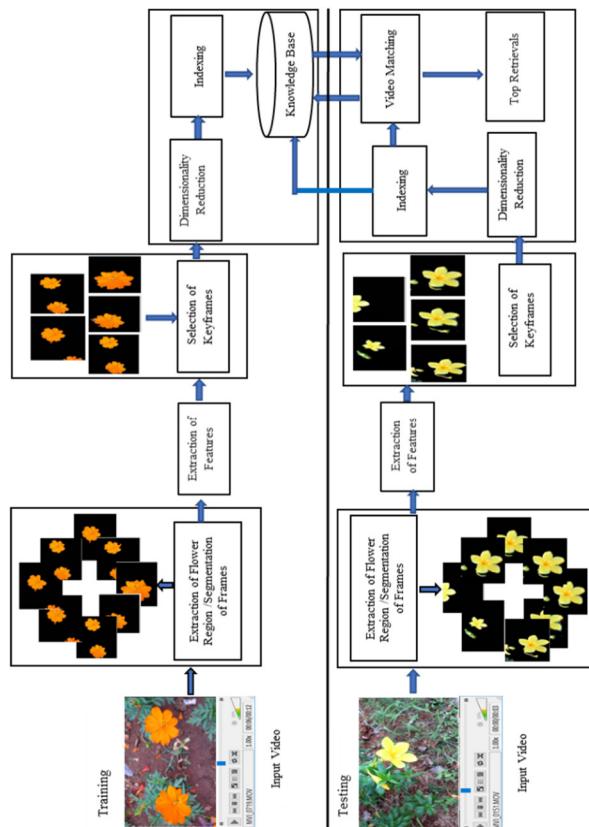
1. Proposal of algorithmic models for the selection of keyframes from natural flower videos with clustering techniques using Deep Convolutional Neural Network (DCNN) as a feature extractor.

2. Proposal of Dimensionality Reduction (DR) methods for the selection of essential features to reduce the feature dimension.
3. Proposal of an Indexing scheme for fast browsing and retrieval of flower videos.
4. Creation of reasonably a large dataset of flower videos which shall be made available public for research purpose.

The remaining part of the paper is organized as follows. The proposed work is explained in Section 2. Experimentation is described in Section 3. The comparative study between proposed model and the traditional model is discussed in Section 4. The work is concluded in Section 5.

## 2. Proposed Work

The retrieval of flower videos has majorly four sections, preprocessing, feature extraction, selection of keyframes and retrieval of videos of flowers as shown in Figure 1. The preprocessing is explained in Section 2.1. It includes conversion of a video to frames and extraction of flower region/segmentation of frames. Features extraction using deep learning technique is depicted in Section 2.2. The process of selection of keyframes is presented in Section 2.3. And finally, the retrieval stage, consists of dimensionality reduction, indexing and retrieval using indexing and are explained in Section 2.4.



**Figure 1.** Block diagram of the proposed flower video retrieval system.

### 2.1. Preprocessing

Preprocessing of raw videos is an essential task which is carried out before designing any video retrieval system. Initially, the system converts a video into frames in preprocessing. It generates 30 frames on an average per second and resizes the frames to 256 x 256 for further processes.

Let  $\mathbf{DB}_v$  be a collection of flower video database of 'N' number of videos

$$\text{i.e., } \mathbf{DB}_v = \{V_1, V_2, V_3, \dots, V_i, \dots, V_N\} \quad (1)$$

Then in general any video  $V_i$  in the database is a set of  $n$  number of frames and is given by

$$V_i = \{F_1, F_2, F_3, \dots, F_i, \dots, F_n\} \quad (2)$$

## Extraction of Flower Region/Segmentation of Frames

Segmentation is the process of selecting the region of interest and eliminating the background and unwanted region from the frame [63]. In the proposed work, segmentation is carried out to obtain the flower/s region/s of interest by removing the background region of the frames of the video. The video frames are segmented before keyframe selection. Image segmentation techniques can be divided into two types, one is an automatic and another one is semi-automatic [64]. In automatic segmentation there is no user interaction to segment the region of interest [50]. Semi-automatic techniques require user interaction [65]. In literature we can find several segmentation methods such as threshold based [72,73], graph based [74,75], Region based [76,77]. In this work, we have utilized automatic and region based segmentation algorithm. Due to the region merging algorithm combines regions, which are sets of pixels and they grow iteratively by combining smaller regions or pixels. Region merging techniques use statistical tests to decide the merging regions. It preserves the global properties for the perceptual units of the frame.

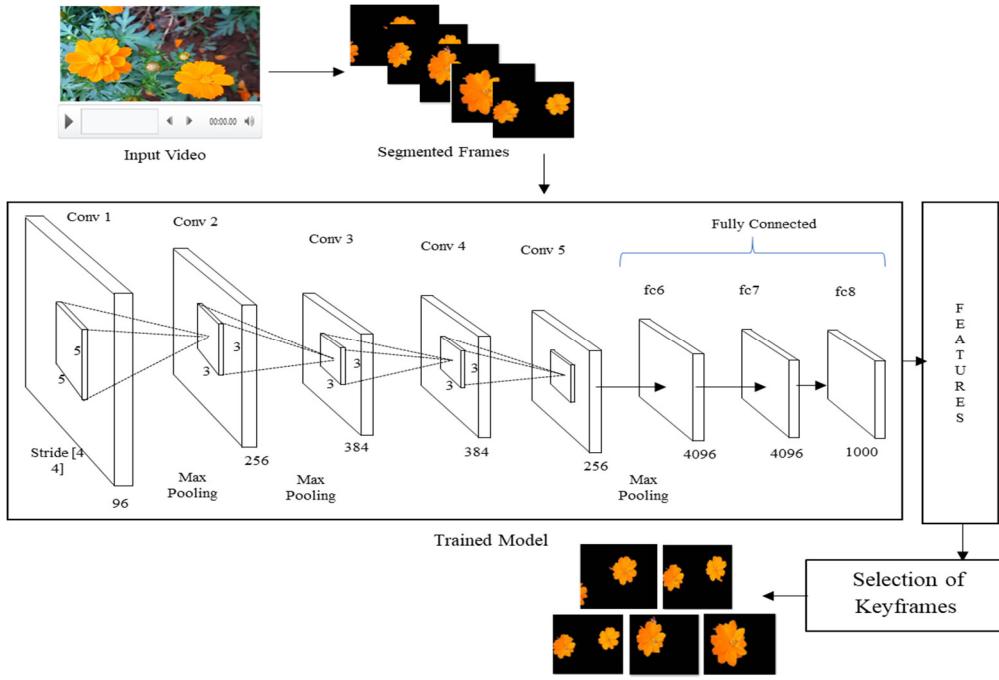
After converting a video  $V_i$  into frames, the flower region/s of interest of each frame of the video  $V_i$  is extracted using statistical Region Merging (SRM) algorithm [50]. After segmentation, the video  $V_i$  is named as  $SV_i$  and is given by

$$SV_i = \{SF_1, SF_2, SF_3, \dots, SF_i, \dots, SF_{sn}\} \quad (3)$$

it consists of  $sn$  number of segmented frames. After segmentation features are extracted using deep convolutional neural network.

### 2.2. Extraction of Features

In this Section we present the extraction of features using Convolutional Neural Network architecture. Deep learning technique, extracts features in hierarchy wise, starts from lower level hierarchy from pixel wise then abstracts the extracted features in the next hierarchy. Lower level layer extracts low level features such as edges and corners. The higher layers extract high level patterns such as parts and object. The final hierarchy is the abstract level of all the hierarchies. Deep Convolutional Neural Network learns features directly from data without depending on human designed traditional features (Yoshua, 2009). The ability to automatically learn features from raw input data is becoming an important for the applications to computer vision and machine learning. In literature we found AlexNet, GoogleNet, VGGNet, ResNet deep learning architectures. We recommended AlexNet because it can be trained on small size images i.e  $227 \times 227$ , which is more appropriate. VGG and GoogleNet were trained on large size images approximately  $(1000 \times 1000)$  (Pang et al., 2017). In the proposed work, for the extraction of deep features, the architecture of AlexNet (Krizhevsky et al., 2012) is used after segmentation of frames of the video. It is a layered architecture in which the layers perform convolution, pooling and ReLU operations of the data (Guo et al., 2016). Convolution puts the input data of the video through a set of convolutional filters, each of which activates features. The output of the first layer is an input to the second layer and it continues same for all layers. Pooling simplifies the output to reduce the number of parameters that the network needs to learn by performing non-linear down sampling. In Rectified Linear Unit (ReLU) negative values are mapped to zero for faster and more effective training it maintains positive values. The proposed architecture is composed of eight layers as shown in Figure 2. The first five, conv1 to conv5 are convolutional layers and the next three namely fc6 to fc8 are fully connected layers. An image of size  $227 \times 227 \times 3$  is input to the first convolution layer, it filters with 96 kernels. The second layer filters first layer output with 256 kernels. The third layer takes the input from the second layer and filters it with 384 kernels. The fourth layer has 384 kernels. The fully connected layers fc6 and fc7 have 4096 neurons. The fc8 layer has 1000 neurons, i.e., softmax layer. The features obtained from fully connected eighth layer output are used to design the proposed model.



**Figure 2.** Structure of AlexNet ConvNet architecture used for Flower video retrieval system.

### 2.3. Keyframe Selection

Keyframe selection is the fundamental process in video content analysis to design a video retrieval system [2,3]. It is essential to maximally reduce the redundant frames to minimize the computational burden. Keyframes contain salient information of the video. Most of the traditional keyframe selection approaches select keyframes based on hand crafted conventional features. Still they lack in achieving good performance. In this paper we propose keyframe selection approaches using Deep Convolutional Neural Network (DCNN) as a feature extractor with clustering approaches.

The segmented frames of a video  $\mathbf{SV}_i$  in Equation (3), contains the redundant frames, which can be eliminated by representing the video with essential frames, this process is called keyframe selection of a video of flower/s. Now video  $\mathbf{V}_i$  in Equation (3) can be redefined as

$$\mathbf{SKV}_i = \{\mathbf{SKF}_1, \mathbf{SKF}_2, \dots, \mathbf{SKF}_m\} \quad (4)$$

it consists of ' $m$ ' number of keyframes among ' $n$ ' number of frames, which are obtained by the proposed deep learning cluster based keyframe selection methods. The features of keyframes are represented as follows.

$$\mathbf{SKF}_1 = [f_{11}, f_{12}, f_{13}, \dots, f_{1r}]$$

$$\mathbf{SKF}_2 = [f_{21}, f_{22}, f_{23}, \dots, f_{2r}]$$

$$\mathbf{SKF}_3 = [f_{31}, f_{32}, f_{33}, \dots, f_{3r}]$$

$$\mathbf{SKF}_m = [f_{m1}, f_{m2}, f_{m3}, \dots, f_{mr}]$$

Therefore, the feature vector of the video  $\mathbf{V}_i$  in Equation (1) can be represented as follows,

$$\mathbf{FMV}_i = \{[f_{11}, f_{12}, f_{13}, \dots, f_{1r}], [f_{21}, f_{22}, f_{23}, \dots, f_{2r}], [f_{31}, f_{32}, f_{33}, \dots, f_{3r}], \dots, [f_{m1}, f_{m2}, f_{m3}, \dots, f_{mr}]\} \quad (5)$$

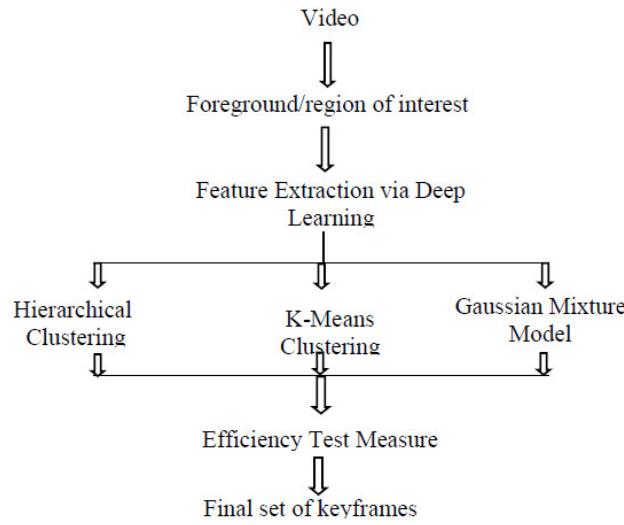
The  $\mathbf{FMV}_i$  can be obtained from the following summarized contributions of keyframe selection using DCNN with clustering approaches.

1. Keyframe selection using Hierarchical clustering
2. Keyframe selection using K-means clustering
3. Keyframe selection using Gaussian Mixture Model

The keyframe selection of the proposed model is as shown in the following Figure 3. Each of the three methods consists of three major phases.

- a. Preprocessing - Segmentation

- b. Feature Extraction - DCNN
- c. Clustering and selection of final set of keyframes



**Figure 3.** Proposed model of the process of keyframe selection.

#### 2.3.1. Keyframe Selection with Hierarchical clustering using DCNN

In this approach, initially the video frames are segmented, the features are extracted from each segmented frame using ConvNet [51]. The proposed method selects keyframes using Hierarchical clustering algorithm. Hierarchical clustering algorithm employs a bottom-up approach for grouping similar frames to form clusters. Initially it assumes each frame itself as a cluster. For a frame it finds minimum distanced frame, then it merges these two frames to form a hierarchy and creates a cluster. Similarly, it repeats for all the frames of the video. Finally, it creates a single hierarchy for all the patterns. The proposed model slices the hierarchy, to obtain K-number of clusters. Then it finds the centroid of each clusters generated from Hierarchical clustering. The frame near to each centroid are selected as keyframes. Once the set of keyframes are obtained for a video then the model applies fidelity measure to obtain the best set of keyframes by comparing with other two proposed clustering methods. The Algorithm 1 shows the proposed keyframe selection approach with Hierarchical clustering.

**Algorithm 1:** Hierarchical\_keyframes\_selection ( $V_i$ )

**Input:** Frames ( $F_n$ ) of Video  $V_i$

**Output:**  $K$ -centroids, keyframes ( $SKV_i$ ),  $Kfdb$  = keyframes database

```

for i=1 to n frames of  $V_i$ 
  extract DCNN features
  if( min dist( $F_i$  ,  $F_{i+1}$ ))
    C=Merge( $F_i$  ,  $F_{i+1}$ )
  for end
  If C= single cluster      //single hierarchy
  
```

```

Split C into K number of hierarchies           // to obtain K-clusters

Kfdb=Find K-centroids           // frame nearest to centroid

If end

return(Kfdb)

```

### 2.3.2. Keyframe Selection Using K-Means Clustering

The proposed method extracts DCNN features from each segmented frame of the video then clusters the similar frames with K-means clustering algorithm [55]. It creates 'K' number of clusters based on the similarity of the segmented frames and it selects frame near to the centroid of each cluster as a keyframe. Once a set of keyframes are obtained then the model applies fidelity measure [54]. The Algorithm 2 represents the proposed keyframe selection process using K-means clustering [55].

#### Algorithm 2: KMeans\_keyframes\_selection ( $V_i$ )

**Input:** Frames ( $F_n$ ) of Video  $V_i$  , K - number of clusters and  $\mu_1, \mu_2, \mu_3, \dots, \mu_K$  are the means of each initial clusters

**Output:** K-centroids, keyframes ( $SKV_i$ ),  $Kfdb$  = keyframes database

```

for i=1 to n frames of  $V_i$ 

    extract DCNN features

    select  $\mu_1, \mu_2, \mu_3, \dots, \mu_K$  are the means of each initial clusters

    find  $S_i$  number of nearest frames to  $\mu_i$ 

    Recalculate  $\mu_i$ 

    Until there is no change in  $\mu_i$ 

    Return  $\mu_1, \mu_2, \mu_3, \dots, \mu_K$ 

    for end

    for i=1 to K //K number of clusters

         $K_i$ =find frame near to  $\mu_i$ 

         $Kfdb=K_i$            // keyframes

        If end

        return(Kfdb)

```

The above algorithm illustrates that it takes n number of frames of a video and the number of required clusters K and  $\mu_1, \mu_2, \mu_3, \dots, \mu_K$  are the means of each initial clusters, then calculates the nearest

frame to each cluster mean using Euclidean distance as a proximity measure. It recalculates the mean value to create new clusters having similar frames, then cluster centroids are chosen as keyframes to represent a video.

### 2.3.3. Keyframe Selection Using Gaussian Mixture Model (GMM)

This method selects keyframes using GMM [67,68] clustering algorithm. Initially the video frames are segmented, the features are extracted from each segmented frame using ConvNet [51]. Gaussian Mixture Model is a probabilistic distribution model, it groups segmented keyframes based on probability of the similar frames. From GMM, K-number of clusters are obtained. Then the centroid of the GMM clusters are selected as keyframes. Finally, fidelity measure has been applied to obtain a best set of keyframes. The algorithm 3 signifies the proposed keyframe selection process using GMM.

**Algorithm 3:** GMM\_keyframes\_selection ( $V_i$ )

**Input:** Frames ( $F_n$ ) of Video  $V_i$ , K - number of clusters

**Output:** K-centroids, keyframes ( $SKV_i$ ),  $Kfdb$  = keyframes database

```

for i=1 to n frames of  $V_i$ 
    extract DCNN features
    estimate maximum likelihood expectation using GMM distribution
    group the similar frames to from K number of clusters
for end
for i=1 to K //K number of clusters
     $K_i$ =find frame near to centroid of each cluster
     $Kfdb=K_i$       // keyframes
    If end
return( $Kfdb$ )

```

Among all the above three approaches the method which generates highest fidelity value has been used to represent a video for further processes, such as Feature Dimensionality Reduction, Indexing and Retrieval.

## 2.4. Retrieval

After the selection of keyframes, each video of  $DB_v$  in Equation (1) is represented as keyframes  $SKV_i$  as shown in Equation (4). Then, features of keyframes of a video  $SKV_i$  is represented as a feature vector as shown in Equation (5). The features are extracted using DCNN, as it extracts features from layer wise the dimension of the feature vector is high. There may be sparse, redundant and unwanted features in the feature vector  $FMV_i$ . To select most essential features, feature selection methods are required and are explained in Section 2.4.1.

### 2.4.1. Dimensionality Reduction (DR)

Features having large dimension increases retrieval error and increases the computational complexity drastically. Selecting a subset of most necessary features from the original by eliminating irrelevant, noisy, redundant features is a challenging task. In the proposed model, features are extracted using DCNN from a set of keyframes for the retrieval of flower videos. When the length of the feature dimension is very high then the filter based feature selection methods are suitable for the feature dimension reduction. Therefore, in this work, we have proposed two filter based feature selection methods namely ReliefF algorithm and Bi-normal separation metric explained in Section 2.4.1.1 and 2.4.1.2 respectively.

#### 2.4.1.1. ReliefF Algorithm

To select necessary features, in the proposed work, the ReliefF is used as a supervised feature selector. ReliefF is initially proposed by Kira and Randell as a filter-based feature selector. It computes ranks and weights for the predictors. The feature weights ranges from -1 to +1. The important attributes are those large positive weighted attributes. Rank indicates that the importance of the attribute. It assigns ranks based on K nearest neighbors.

#### 2.4.1.2. Bi-Normal Separation (BNS)

Bi-normal separation feature selection metric is the inverse cumulative probability function of the difference between true positive ratio and false positive ratio of the standard normal distribution [58]. Some value is substituted to zero to avoid undefined value. If the feature is more prevalent in the positive class, then its threshold is further from the tail of the distribution than that of the negative class. The BNS metric measures the separation between these two thresholds. It is defined as follows,

$$BNS(dfv_i, FC_j) = \left| F^{-1}(P(dfv_i | C_j)) - F^{-1}(P(dfv_i | \bar{C}_j)) \right|$$

Where  $dfv_i$  is the discriminating feature of the video  $v_i$  with respect to the category of the flower video i.e  $FC_j$

After eliminating the irrelevant features from the feature vector of the video in question (5), the reduced feature vector is defined as follows.

$$DRV_i = [drf_1, drf_2, drf_3, \dots, drf_p] \quad (6)$$

#### 2.4.2. Indexing

For the retrieval of videos, the proposed model uses reduced features obtained from Equation (6). Along with reduced features an indexing scheme called KD-tree, explained in Section 2.4.2.1 is used for easy and fast accessing of relevant videos. Therefore, after indexing the training videos of the database can be defined as follows.

$$DB_v = \{DRV_1, DRV_2, DRV_3, \dots, DRV_i, \dots, DRV_n\} \quad (7)$$

For any query video  $TQ_j$  denoted as  $TQ_j = \{QSF_1, QSF_2, \dots, QSF_n\}$ , be a set of segmented frames of the query video  $TQ_j$ . Then  $KQ_j = \{QSKF_1, QSKF_2, \dots, QSKF_n\}$ , be a set of keyframes of the query video. Then the features of  $KQ_j$  are denoted as follows

$$QSKF_1 = [qf_{11}, qf_{12}, \dots, qf_{1m}]$$

$$QSKF_2 = [qf_{21}, qf_{22}, \dots, qf_{2m}]$$

$$QSKF_n = [qf_{k1}, qf_{k2}, \dots, qf_{km}]$$

The reduced dimension of a query video is denoted as follows,

$$drQ_j = \{drqf_1, drqf_2, \dots, drqf_q\} \quad (8)$$

Finally, after indexing the testing video is defined as  $TQ_j = IdrQ_j$

To find the similarity between a query video 'TQ<sub>j</sub>' to the database videos of 'DB<sub>v</sub>', nearest neighbor approach of KD-tree indexing is used. Figure 13 shows the sample query video, selected keyframes and retrieved videos from the proposed video retrieval system.

#### 2.4.2.1. KD-Tree Indexing

The process of retrieval of videos takes long time when the size of the database is extensively large. To reduce the retrieval time and to fast the retrieval process, designing an indexing scheme is an essential task. We can find indexing methods namely, B-tree [79], B<sup>+</sup> tree [80], R tree, KD-tree [59]. Due to the feature dimension of flower video is multi-dimensional, this paper proposes a mechanism for indexing using KD-tree indexing method. KD-tree is a K dimensional feature vector indexing tree structure. Feature vector of a video 'V' is multi-dimension as shown in Equation (5). Each individual video of the database is indexed using KD-tree as shown in Equation (6). The KD-tree indexing model is the efficient data structure for multidimensional features for organizing K-dimensional data and it is also useful for searching based on a multidimensional key. To construct the KD-tree, at the root, we split the set of features into two subsets called left sub-tree and right-subtree. It is an efficient data structure for retrieval of natural flower videos. The KD-tree indexing mechanism in [59]-is used in the proposed work.

### 3. Experiments

#### 3.1. Creation of Very Large Dataset

Due to the advanced technology of widely using devices such as mobiles and cameras, capturing and storage of flower videos is easy, when the size of the database increases the difficulty lies in searching flower videos of user's interest. Standard dataset of flower videos is not publicly available. We have collected different species of flower videos with three different devices namely, Samsung Galaxy Grand Prime (SGGP) mobile of 8 mega pixels, which consists of 2611 videos. Sony Cyber shot camera of 14.1 mega pixels, consists of 2521 videos. Canon camera of 16 mega pixels, which consists of 2656 videos. The duration of the captured videos ranges from 4 to 60 seconds. There exists a small inter class and large intra class variations. Videos captured in the real environment with varied ecological conditions such as summer, rainy and winter seasons. We created datasets, in South Karnataka during 2016 to 2018. Videos captured in flower shows namely Mysore palace flower show, Mysore Dasara flower show, Lalbhagh flower show at Bangalore, nurseries, parks, flowers in house gardens etc. Videos in flower shows, flower pots/plants are organized by floriculturists and horticulturists for visitors, Figure 4. shows that the video frames/images captured in flower shows. The challenges considered such as viewpoint variations, cluttered background, large illumination, partial occlusion and multiple instances of the flowers. The large intra-class variations of flower video samples from the dataset we captured are shown in Figure 14.



**Figure 4.** Video frames/ Images captured in flower shows.

### 3.2. Experimentation on Keyframes Selection

In this Section we present the experimental results and the comparative study between proposed approaches with conventional cluster based approaches for the selection of keyframes.

To examine the proposed keyframe selection approaches the performance measure namely, Fidelity measure is used. It finds the maximum of minimal distance between the set of keyframes or semi-Hausdorff distance. Fidelity measure is used to compare the selected keyframe with remaining frames of the video [54]. In the proposed work, the selected keyframes are compared with frames of the sequence of a video. The proposed deep learning cluster based keyframes selection methods are compared with existing conventional cluster based keyframes selection methods [3,11]. Table 1 describes the average fidelity values for 30 classes of flower videos of the proposed methods.

We have proposed three cluster based methods and conducted experiments for the selection of 2 keyframes to 5 keyframes by creating 2 clusters to 5 clusters respectively. Our experimentation is limited to 5 keyframes since to create more than 5 keyframes no convergence is achieved. The average fidelity results for 2 to 5 keyframes on SGGP, Sonycyber Shot and Canon datasets are shown in Table 1. The fidelity values are normalized between 0 to 100.

**Table 1.** Average fidelity values of proposed deep learning keyframe selection approaches.

Dataset	Hierarchical				K-Means				GMM			
	2kfs	3kfs	4kfs	5kfs	2kfs	3kfs	4kfs	5kfs	2kfs	3kfs	4kfs	5kfs
SGGP	70.32	67.49	64.65	64.24	70.52	68.55	69.59	65.35	71.76	73.17	78.58	<b>78.76</b>
Sony Cyber Shot	61.02	59.91	58.99	57.74	57.83	58.89	60.02	59.83	68.55	71.57	63.88	<b>79.01</b>
Canon	68.01	65.55	65.59	65.78	67.44	66.16	67.29	69.32	70.55	78.78	72.47	<b>82.56</b>

#### 3.2.1. Comparative Study between Conventional V/S Deep Learning Approaches for Keyframes Selection

The proposed deep learning cluster based keyframes selection approaches are compared with conventional cluster based keyframes selection approaches [3,11]. Both conventional and deep learning methods generates good results for 5 keyframes. Therefore, we have shown comparative study for 5 keyframes. Table 2 describes the average fidelity values for 30 classes of flower videos of the proposed deep learning methods and conventional methods for 5 keyframes.

**Table 2.** Comparison of average fidelity values of proposed deep learning and conventional keyframe selection approaches.

Dataset	Conventional approaches			Proposed DCNN approaches		
	Hierarchical [11]	K-Means [3]	GMM [3]	Hierarchical	K-Means	GMM
SGGP	54.36	55.00	55.50	64.24	65.35	<b>78.76</b>
Sony Cyber Shot	52.10	50.77	64.73	57.74	59.83	<b>79.01</b>
Canon	54.72	53.18	55.66	65.78	69.32	<b>82.56</b>

In this work, we presented three keyframes selection approaches using DCNN as a feature extractor. Among the three different approaches, the keyframe selection method using deep learning features with GMM clustering approach gives good fidelity results. These keyframes are used to represent a video and are used for the retrieval of videos. The proposed approaches give better results than the existing methods. The sample keyframes are shown in Figure 15.

### 3.3. Result Analysis of Retrieval of Flower Videos

For retrieval, each video of the database is represented as a set of keyframes selected from proposed keyframe selection using clustering approach namely Gaussian Mixture Model. For a given query video of a flower/s, the similar videos are retrieved via two different modalities, namely, absolute modality and relative modality. In absolute mode, top 5 to top 25 videos which are absolute

to the query video are retrieved. In relative mode, top 5% and top 10% videos which are relative to the query video are retrieved. The two different ways of retrieval of videos are as follows,

**Absolute Modality:**

- i. With Dimensionality Reduction (DR) and with KD-tree indexing
  - a. ReliefF feature selection algorithm and with indexing.
  - b. Bi-normal separation feature selection metric and with indexing
- ii. Without dimensionality reduction and with KD-tree indexing
- iii. Data base search (without DR and without KD-tree indexing)

**Relative Modality:**

- i. With dimensionality reduction and with KD-tree indexing
  - a. ReliefF feature selection algorithm and with indexing.
  - b. Bi-normal separation feature selection metric and with indexing
- ii. Without dimensionality reduction and with KD-tree indexing
- iii. Data base search (without DR and without KD-tree indexing)

To examine the proposed retrieval system the performance measures namely, precision, recall and F-measure are used and are given below.

$$Precision = \frac{\text{Total number of videos retrieved are relevant}}{\text{Total number of videos retrieved}} \quad (9)$$

$$Recall = \frac{\text{Total number of videos retrieved are relevant}}{\text{Total number of similar videos in the database}} \quad (10)$$

$$F - Measure = \frac{2 * Precision * Recall}{(Precision + Recall)} \quad (11)$$

Result analysis of video retrieval of SGGP, Sony Cyber shot and Canon devices dataset with above mentioned modalities are shown below.

### 3.3.1. Absolute Modality

#### 3.3.1.1. with DR and with KD-Tree Indexing

In this section, we present the retrieval results obtained with reducing feature dimension and using indexing mechanism in absolute modality. To reduce the dimension of the features, the proposed system is tested up to 1500 features. The Figures 5 and 6 show that the features required for an efficient retrieval in absolute modality for ReliefF and BNS algorithms. Features selected from 100 to 1500 with varying 100 features.

**a. ReliefF feature selection and with KD-tree indexing.**

Table 3a-c show the precision, recall and F-Measure for top 5 to top 25 videos retrieved with varying the training and testing samples of SGGP, Sony Cybershot and Canon datasets in absolute mode. To select the most discriminating features, the feature selection method such as ReliefF algorithm is used for feature dimensionality reduction and KD-tree indexing mechanism is applied to retrieve top 5 to 25 similar videos from the database for a given query video. Feature selection method and an indexing scheme reduces the computation time and cost and it increases the efficiency in terms of precision.

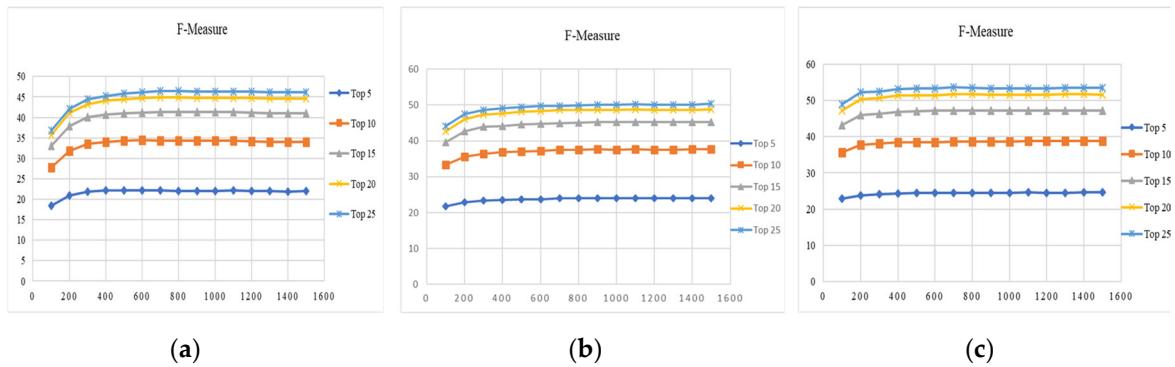


Figure 5. Feature selection graph in absolute mode using ReliefF algorithm for datasets (a) SGGP (b) Sonycyber Shot (c) Canon.

**Table 3 (a). SGGP dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results**

Precision						Recall						F-Measure					
Train-Test in %	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10
30–70	72.87	66.42	61.2	56.23	51.6	13.25	23.86	32.56	39.35	44.64	21.96	34.06	41.06	44.65	46.13		
40–60	75.58	69.14	64.73	60.75	57.11	10.61	19.14	26.55	32.81	38.18	18.28	29.17	36.44	41.12	44.11		
50–50	77.39	71.75	67.5	64.07	60.69	8.82	16.22	22.63	28.32	33.18	15.6	25.8	32.85	37.94	41.36		
60–40	80.73	75.27	71.39	67.78	64.89	7.55	14.04	19.9	25	29.67	13.64	23.17	30.28	35.4	39.37		
70–30	82.73	77.59	73.65	70.2	67.26	6.68	12.47	17.68	22.36	26.61	12.22	21.06	27.78	32.9	36.88		
80–20	84.91	80.7	76.38	73.06	69.94	5.99	11.32	15.98	20.3	24.18	11.07	19.51	25.81	30.89	34.82		

**Table 3 (b). SonyCybershot dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results**

Precision						Recall						F-Measure					
Train-Test in %	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25		
30–70	75.74	70.15	64.86	59.39	54.56	14.65	26.88	36.69	43.98	49.81	24.01	37.66	45.28	48.81	50.33		
40–60	80.34	74.75	69.96	65.75	61.5	11.67	21.49	29.93	37.05	42.65	20	32.48	40.59	45.8	48.66		
50–50	82.53	77.34	73.19	69.3	65.8	9.67	17.95	25.29	31.68	37.27	17.03	28.42	36.47	42.07	45.98		
60–40	84.44	79.72	75.91	72.4	69.12	8.25	15.45	21.84	27.57	32.66	14.81	25.29	32.98	38.7	42.9		
70–30	85.1	81.09	77.57	74.33	71.49	7.09	13.38	19.08	24.21	28.93	12.92	22.5	29.84	35.46	39.89		
80–20	87.15	83.04	79.76	76.99	73.81	6.38	12.08	17.28	22.11	26.31	11.75	20.69	27.7	33.38	37.59		

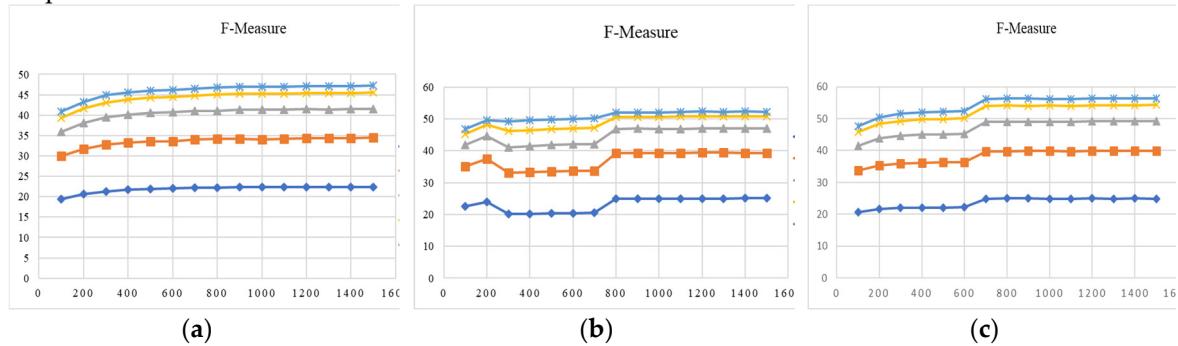
Table 3 (c). Canon dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results

Precision						Recall						F-Measure					
Train- Test in %	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25		
30–70	82.4	76.83	71.26	65.71	60.52	14.86	27.35	37.43	45.37	51.59	24.6	39.04	47.33	51.7	53.63		
40–60	84.74	80.22	75.89	71.56	67.1	11.63	21.73	30.46	37.82	43.78	20.06	33.2	41.99	47.67	50.99		
50–50	85.73	81.73	77.68	74	70.44	9.53	17.99	25.43	31.92	37.62	16.87	28.69	37.06	42.99	47.17		
60–40	87.89	83.74	80.34	76.83	73.54	8.2	15.53	22.16	27.99	33.13	14.77	25.55	33.66	39.61	43.99		
70–30	91.26	87.32	84.11	81.04	77.6	7.26	13.85	19.9	25.43	30.19	13.27	23.38	31.27	37.44	41.93		

80-20	92.71	89.61	86.82	84.29	81.47	6.43	12.38	17.88	23.06	27.65	11.89	21.32	28.89	35.12	39.91
-------	-------	-------	-------	-------	-------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------

### b. Bi-normal separation feature selection metric and with KD-tree indexing

Table 4a-c show the precision, recall and F-Measure for top 5 to top 25 videos retrieved with varying the training and testing samples of SGGP, Sony Cyber shot and Canon datasets in absolute mode. To select the most discriminating features, the feature selection method such as Bi-normal separation metric is used. KD-tree indexing mechanism is applied to retrieve top 5 to 25 similar videos from the database. Bi-normal separation feature selection metric and KD-tree indexing scheme reduces much more computation time and it increases the efficiency in terms of precision when compared with the ReliefF.



**Figure 6.** Feature selection graph in absolute mode using BNS algorithm for datasets (a) SGGP (b) Sonycyber Shot (c) Canon.

**Table 4 (a). SGGP dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results**

Train-Test in %	Precision					Recall					F-Measure				
	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25
30-70	74.07	67.15	61.77	57.13	52.77	13.52	24.17	32.98	40.1	45.83	22.4	34.5	41.57	45.48	47.31
40-60	75.77	70.11	65.3	61.23	57.66	10.63	19.45	26.9	33.27	38.78	18.32	29.64	36.89	41.63	44.72
50-50	77.11	72.28	68.14	64.44	61.18	8.77	16.37	23	28.75	33.76	15.51	26.04	33.35	38.42	41.97
60-40	80.13	75.63	72.16	68.78	65.53	7.51	14.13	20.17	25.53	30.17	13.57	23.32	30.69	36.1	39.96
70-30	82.13	77.51	74.28	71.19	68.4	6.63	12.48	17.89	22.77	27.22	12.13	21.08	28.11	33.49	37.68
80-20	85.14	80.51	77.09	73.94	71.19	5.98	11.27	16.15	20.58	24.68	11.07	19.44	26.1	31.33	35.55

**Table 4 (b). SonyCybershot dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results**

Train-Test in %	Precision					Recall					F-Measure				
	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25
30-70	78.35	72.7	66.98	61.47	56.37	15.31	28.13	38.22	45.97	51.92	25.06	39.34	47.07	50.84	52.27
40-60	81.95	77.28	72.63	67.9	63.64	12.05	22.49	31.37	38.59	44.54	20.62	33.9	42.44	47.58	50.65
50-50	84.27	79.4	75.78	71.9	68.18	10	18.68	26.49	33.2	38.96	17.58	29.5	38.1	43.97	47.94
60-40	85.79	81.51	78.34	75.01	71.95	8.431	15.93	22.81	28.92	34.37	15.13	26.05	34.35	40.46	45.01
70-30	87.34	83.25	80	77.57	74.51	7.34	13.9	19.94	25.61	30.51	13.38	23.34	31.11	37.38	41.95
80-20	88.69	85.22	82.12	79.53	76.98	6.58	12.53	18.04	23.2	27.9	12.12	21.43	28.85	34.9	39.69

Table 4 (c). Canon dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results

Train-Test in %	Precision					Recall					F-Measure				
	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25
30–70	83.54	78.43	73.81	68.76	63.45	24.93	39.93	49.14	54.26	56.36	24.93	39.93	49.14	54.26	56.36
40–60	86.4	82.11	78.04	74.29	70.18	11.85	22.25	31.37	39.31	45.89	20.44	34.02	43.27	49.58	53.46
50–50	87.73	83.69	80.01	76.48	73.16	9.763	18.43	26.19	33.08	39.14	17.27	29.42	38.21	44.55	49.1
60–40	89.6	85.75	82.62	79.23	76.28	8.343	15.84	22.73	28.86	34.44	15.03	26.1	34.6	40.88	45.73
70–30	92.22	89.31	86.17	83.32	80.44	7.32	14.08	20.28	26.05	31.24	13.4	23.81	31.92	38.43	43.44
80–20	93.29	91.07	88.86	86.37	83.99	6.46	12.48	18.2	23.48	28.45	11.95	21.52	29.45	35.83	41.12

### 3.3.1.2. Without Dimensionality Reduction and with KD-Tree Indexing

Table 5a–c show the precision, recall and F-Measure for top 5 to top 25 videos retrieved with varying the training and testing samples of SGGP, Sony Cybershot and Canon datasets in absolute mode. Retrieval without dimensionality reduction and with KD-tree indexing takes very long time and it gives less result than with dimensionality reduction method of absolute mode.

Table 5 (a). SGGP dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results: Without DR – with Indexing

Train-Test in %	Precision					Recall					F-Measure				
	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25
30–70	67.71	60.82	55.66	50.85	46.58	12.16	21.64	29.37	35.36	40.18	20.2	30.98	37.16	40.25	41.59
40–60	69.74	63.4	58.63	54.8	51.3	9.66	17.31	23.75	29.29	33.97	16.67	26.46	32.73	36.85	39.41
50–50	71.54	66.18	61.7	57.88	54.68	8.06	14.77	20.43	25.3	29.58	14.26	23.55	29.76	34.02	37.02
60–40	75.2	69.8	65.57	61.84	58.5	6.99	12.91	18.03	22.48	26.39	12.62	21.33	27.52	31.96	35.17
70–30	77.26	71.98	68.11	64.51	61.31	6.19	11.46	16.16	20.23	23.88	11.33	19.38	25.45	29.88	33.25
80–20	79.56	74.77	70.7	66.98	63.81	5.56	10.38	14.61	18.33	21.68	10.29	17.9	23.65	27.99	31.38

Table 5 (b). SonyCybershot dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results: Without DR – with Indexing

Train-Test in %	Precision					Recall					F-Measure				
	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25
30–70	72.91	66.96	61.49	55.55	50.86	14.13	25.74	34.86	41.3	46.67	23.15	36.03	42.97	45.73	47
40–60	77.11	71.67	66.64	62	57.65	11.22	20.65	28.51	34.93	40.05	19.23	31.18	38.65	43.17	45.62
50–50	78.76	73.45	69.28	65.1	61.42	9.23	17.08	23.96	29.79	34.8	16.27	27.03	34.55	39.55	42.91
60–40	81.72	76.27	72.04	68.3	64.71	7.96	14.72	20.7	25.94	30.55	14.31	24.12	31.27	36.44	40.14
70–30	82.29	77.71	74.09	70.49	67.35	6.84	12.79	18.18	22.88	27.12	12.47	21.52	28.45	33.54	37.45
80–20	84.43	79.41	76.03	72.79	69.53	6.15	11.5	16.41	20.82	24.68	11.33	19.69	26.33	31.47	35.3

Table 5 (c). Canon dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results: Without DR – with Indexing

Train-Test in %	Precision					Recall					F-Measure				
	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25
30–70	79.26	73.15	67.4	61.65	56.53	14.23	25.9	35.31	42.49	48.18	23.58	37.03	44.7	48.45	50.07
40–60	82.01	76.68	72.04	67.46	63.13	11.19	20.71	28.78	35.49	41.04	19.31	31.67	39.73	44.82	47.87
50–50	83.18	78.18	73.96	69.92	66.27	9.21	17.14	24.11	30.07	35.24	16.3	27.37	35.18	40.55	44.28
60–40	85.66	80.42	76.71	73.06	69.49	7.96	14.82	21	26.47	31.16	14.36	24.42	31.97	37.53	41.44
70–30	89.5	84.59	80.45	77.08	73.69	7.08	13.34	18.93	24.05	28.54	12.97	22.55	29.79	35.48	39.71
80–20	91.14	87.36	83.8	80.6	77.5	6.29	11.98	17.16	21.84	26.1	11.63	20.66	27.77	33.36	37.78

### 3.3.1.3. Data Base Search (without DR and without KD-Tree Indexing)

The following tables illustrates database search results of datasets SGGP, Sony Cyber shot and Canon for top 5 to top 25 retrievals in absolute way without using dimensionality reduction and indexing techniques but the features are extracted using DCNN. For any given query video, it searches similar videos from the database using Euclidean distance as a proximity measure. Tables 6a–c show the precision, recall and F-Measure of top 5 to top 25 videos retrieved with varying the training and testing samples of SGGP, Sony Cyber shot and Canon datasets in absolute mode.

Table 6 (a). SGGP dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results

Train-Test in %	Precision					Recall					F-Measure				
	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25
30–70	67.72	60.86	55.71	50.89	46.62	12.2	21.7	29.41	35.41	40.23	20.2	31	37.2	40.29	41.63
40–60	69.74	63.44	58.67	54.82	51.34	9.66	17.3	23.77	29.31	34.01	16.67	26.5	32.76	36.88	39.44
50–50	71.54	66.18	61.7	57.88	54.68	8.06	14.8	20.43	25.3	29.58	14.26	23.6	29.76	34.02	37.02
60–40	75.22	69.82	65.65	61.9	58.56	6.99	12.9	18.06	22.52	26.44	12.63	21.3	27.57	32.01	35.23
70–30	75.22	69.82	65.65	61.9	58.56	6.19	11.5	16.17	20.24	23.89	11.33	19.4	25.46	29.89	33.27
80–20	79.56	74.77	70.7	67	63.82	5.56	10.4	14.61	18.34	21.69	10.29	17.9	23.65	28.00	31.38

Table 6 (b). SonyCybershot dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results

Train-Test in %	Precision					Recall					F-Measure				
	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25
30–70	72.92	66.98	61.54	55.6	50.89	14.1	25.7	34.88	41.32	46.69	23.15	36	43	45.76	47.02
40–60	77.14	71.72	66.69	62.04	57.7	11.2	20.7	28.52	34.95	40.07	19.23	31.2	38.67	43.19	45.66
50–50	78.81	73.64	69.38	65.23	61.5	9.24	17.1	23.98	29.83	34.83	16.27	27.1	34.59	39.6	42.96
60–40	81.8	76.37	72.09	68.37	64.78	7.97	14.7	20.71	25.96	30.57	14.32	24.1	31.28	36.47	40.17
70–30	82.26	77.73	74.12	70.53	67.4	6.84	12.8	18.19	22.89	27.13	12.46	21.5	28.46	33.55	37.47
80–20	84.43	79.45	76.08	72.86	69.6	6.15	11.5	16.42	20.83	24.69	11.33	19.7	26.34	31.48	35.33

Table 6 (c). Canon dataset – Precision, Recall, F-Measure of top 5 to 25 retrieval results

Train-Test in %	Precision					Recall					F-Measure				
	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25	Top 5	Top 10	Top 15	Top 20	Top 25
30–70	79.47	73.23	67.44	61.71	56.59	14.3	25.9	35.32	42.52	48.22	23.63	37.1	44.72	48.49	50.12
40–60	82.01	76.68	72.03	67.46	63.12	11.2	20.7	28.77	35.49	41.04	19.3	31.7	39.73	44.82	47.87
50–50	83.17	78.24	74.05	70.01	66.35	9.21	17.2	24.14	30.11	35.29	16.3	27.4	35.23	40.61	44.34
60–40	85.71	80.51	76.77	73.14	69.57	7.97	14.8	21.01	26.51	31.2	14.36	24.4	31.99	37.59	41.5
70–30	89.5	84.58	80.46	77.14	73.75	7.08	13.3	18.93	24.07	28.57	12.97	22.5	29.79	35.51	39.74
80–20	91.17	87.37	83.9	80.62	77.61	6.29	12	17.18	21.85	26.14	11.63	20.7	27.8	33.36	37.84

### 3.3.2. Relative Modality

#### 3.3.2.1. With DR and with KD-Tree Indexing

**a. ReliefF feature selection algorithm and with KD-tree indexing** Retrieval of videos in relative video search modality using ReliefF dimensionality reduction algorithm and KD-tree Indexing retrieves top 5% and 10% of videos from database. Precision, Recall, F-Measure and time are show in Table 7a-c for SGGP, Sony Cyber shot and Canon datasets respectively.

To reduce the dimension of the features, the proposed system is tested up to 1500 features. The Figures 7 and 8 show the features required for an efficient retrieval in relative modality. Features selected from 100 to 1500 with varying 100 features.

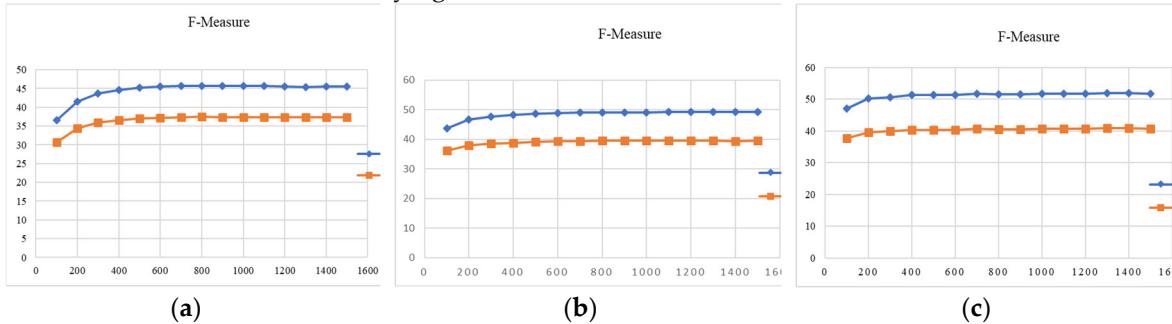


Figure 7. Feature selection graph in relative mode using ReliefF algorithm for datasets (a) SGGP (b) Sony Cyber Shot (c) Canon.

Table 7 (a). SGGP dataset - Top 5% and 10% retrieval results

Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	36.75	24.14	49.94	64.12	40.83	34.01	1.1	1.08
40–60	36.5	24.25	49.81	64.05	40.57	34.08	2.16	2.16
50–50	36.6	24.32	50.29	64.72	40.76	34.23	3.5	3.81
60–40	36.97	24.44	50.53	65.14	41.11	34.44	3.6	3.91
70–30	37.15	24.47	50.78	65.24	41.29	34.48	5.76	6.07
80–20	37.44	24.67	50.95	65.59	41.55	34.74	6.32	6.75

Table 7 (b). Sony Cyber shot dataset - Top 5% and 10% retrieval result

Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	36.75	24.14	49.94	64.12	40.83	34.01	1.1	1.08

30–70	44.13	27.81	59.99	74.61	49.26	39.53	1.48	1.46
40–60	44.61	28.21	60.73	74.95	49.83	39.98	1.47	1.57
50–50	44.73	28.24	60.5	75.19	49.8	40.04	5.79	6.06
60–40	44.74	28.38	60.5	75.05	49.82	40.16	6.14	6.49
70–30	44.66	28.42	60.33	75.14	49.71	40.22	8.03	10
80–20	44.43	28.3	59.63	74.54	49.28	39.99	8.04	8.66

Table 7 (c). Canon dataset - Top 5% and 10% retrieval results

Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	46.62	29.17	63.48	76.02	51.78	40.79	2.29	4.9
40–60	46.9	29.47	63.23	76.19	51.79	41.06	4	2.01
50–50	46.33	29.28	62.49	75.65	51.09	40.73	3.61	3.74
60–40	46.62	29.23	62.62	76.19	51.29	40.75	3.21	6.81
70–30	46.71	29.48	63.14	76.36	51.51	41	3.05	3.27
80–20	47.68	29.96	63.84	77.34	52.37	41.64	3.49	3.83

### b. Bi-normal feature selection metric and KD-tree indexing

Relativel of videos in relative search modality using Bi-normal separation feature selection metric and KD-tree Indexing retrieves top 5% and 10% of videos from the database. Precision, recall and F-Measure and time taken are show in Table 8a–c for SGGP, Sony Cybershot and Canon datasets. Retrieval in relative mode using Bi-normal feature selection metric gives better results than ReliefF feature selection algorithm.

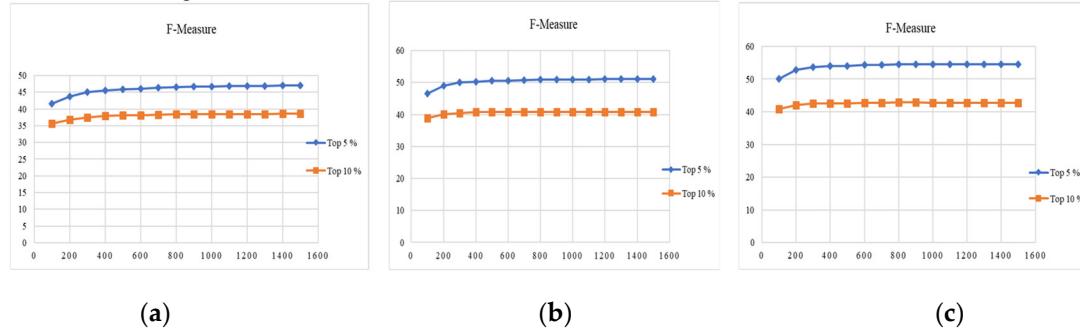


Figure 8. Feature selection graph in relative mode using BNS algorithm for datasets (a)SGGP (b) Sonycyber Shot (c) Canon.

Table 8 (a). SGGP dataset - Top 5% and 10% retrieval results

Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	42.28	27.4	57.55	72.73	47.01	38.59	1.8	1.77
40–60	41.99	27.52	57.69	72.73	46.82	38.68	1.99	2.08
50–50	41.59	27.25	57.57	72.52	46.48	38.35	2.33	2.47
60–40	42.34	27.63	58.25	73.63	47.23	38.92	2.33	2.47
70–30	42.42	27.75	58.4	73.99	47.31	39.1	2.25	2.48
80–20	43.03	27.98	59.07	74.41	47.96	39.41	2.49	2.79

Table 8 (b). Sonycyber Shot dataset - Top 5% and 10% retrieval results

Train – Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	45.57	28.66	62.34	76.93	51.03	40.75	1.75	1.74
40–60	46.11	29.08	63.23	77.36	51.68	41.23	1.79	1.87
50–50	46.42	29.22	63.25	77.9	51.86	41.44	2.45	2.59
60–40	46.52	29.39	63.48	77.87	52	41.61	3.56	3.81
70–30	46.63	29.46	63.63	78.11	52.12	41.72	3.73	3.97
80–20	<b>46.69</b>	<b>29.42</b>	<b>63.4</b>	<b>77.89</b>	<b>52.03</b>	<b>41.62</b>	<b>3.87</b>	<b>4.19</b>

Table 8 (c). Canon dataset - Top 5% and 10% retrieval results

Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	49.09	30.59	66.81	79.49	54.54	42.75	1.33	1.31
40–60	49.62	30.9	66.86	79.72	54.83	43.04	1.41	1.51
50–50	49.11	30.79	66.18	79.34	54.17	42.81	1.99	2.14
60–40	49.15	30.72	65.92	79.62	54.06	42.77	2.14	2.36
70–30	49.36	31.01	66.53	79.92	54.39	43.08	2.55	2.78
80–20	<b>50.45</b>	<b>31.49</b>	<b>67.33</b>	<b>80.91</b>	<b>55.36</b>	<b>43.72</b>	<b>2.94</b>	<b>3.27</b>

### 3.3.2.2. Without DR and with KD-Tree Indexing

Table 9a–c show the precision, recall, F-Measure and time taken for top 5 % and top 10 % videos retrieved with varying the training and testing samples of SGGP, Sony Cybershot, Canon datasets respectively in relative modality.

Table 9 (a). SGGP dataset - Top 5% and 10% retrieval results

Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	37	24.4	50.2	64.6	41.1	34.3	11.5	11.7
40–60	36.8	24.5	50.2	64.5	40.9	34.4	12.6	13.1
50–50	36.7	24.4	50.4	64.8	40.9	34.3	18.5	19.4
60–40	37.2	24.6	50.8	65.4	41.4	34.7	23.5	25.0
70–30	37.4	24.7	51.1	65.7	41.6	34.7	24.1	26.5
80–20	<b>37.7</b>	<b>24.8</b>	<b>51.3</b>	<b>66.0</b>	<b>41.8</b>	<b>35.0</b>	<b>25.8</b>	<b>29.0</b>

Table 9 (b). Sony Cyber shot dataset - Top 5% and 10% retrieval result

-Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	41.2	26.3	56.2	70.5	46.0	37.4	11.4	11.6
40–60	41.4	26.5	56.5	70.4	46.3	37.5	11.8	12.3
50–50	41.4	26.4	56.2	70.4	46.2	37.4	17.1	18.1
60–40	41.4	26.6	56.2	70.2	46.2	37.6	23.3	24.7
70–30	41.3	26.5	55.9	70.1	46.0	37.5	23.8	25.9
80–20	<b>41.1</b>	<b>26.4</b>	<b>55.2</b>	<b>69.5</b>	<b>45.5</b>	<b>37.3</b>	<b>24.5</b>	<b>27.4</b>

Table 9 (c). Canon dataset - Top 5% and 10% retrieval results

Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	43.9	28	59.9	73.2	48.8	39.2	11.4	11.6
40–60	44.3	28.3	59.8	73.4	48.9	39.4	13	13.6
50–50	43.7	28	59.1	72.8	48.2	39	18.8	19.8
60–40	43.5	27.8	58.5	72.5	47.8	38.7	23.6	25.2
70–30	43.5	27.9	59	72.7	48	38.9	24	26.4
80–20	<b>44.5</b>	<b>28.4</b>	<b>59.6</b>	<b>73.5</b>	<b>48.9</b>	<b>39.5</b>	<b>26.7</b>	<b>30.0</b>

### 3.3.2.3. Data Base Search (without DR and without KD-Tree Indexing)

The following tables illustrates database search results of datasets SGGP, Sony Cyber shot and Canon for top 5% and top 10% retrievals in relative way without using dimensionality reduction and without using indexing technique. For any given query video, it searches similar videos in the database using Euclidean distance as a proximity measure. Table 10a-c show that the precision, recall, F-Measure and time of top 5% and top 10% videos retrieved with varying the training and testing samples of SGGP, Sony Cyber shot and for Canon datasets respectively in relative mode.

Table 10 (a). SGGP dataset - Top 5% and 10% retrieval results

Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	32.62	21.94	44.51	58.58	36.29	30.96	38.89	38.89
40–60	32.62	22.07	44.65	58.5	36.3	31.04	69.69	69.69
50–50	32.47	21.98	44.79	58.83	36.22	30.99	108.2	108.2
60–40	33.1	22.24	45.45	59.65	36.87	31.38	157.8	157.8
70–30	33.07	22.26	45.39	59.72	36.81	31.4	209.7	209.7
80–20	33.19	22.31	45.33	59.62	36.87	31.45	275.7	275.7

Table 10 (b). Sony Cyber shot dataset - Top 5% and 10% retrieval result

Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	36.1	23.59	49.27	63.55	40.31	33.54	40.4	40.4
40–60	36.5	23.83	50.02	63.82	40.84	33.82	70.98	70.98
50–50	36.6	23.76	49.81	63.74	40.81	33.72	109.9	109.9
60–40	36.36	23.76	49.52	63.37	40.56	33.68	157.8	157.8
70–30	35.91	23.63	48.86	62.95	40.02	33.47	203.6	203.6
80–20	35.9	23.58	48.61	62.62	39.89	33.35	251.2	251.2

Table 10 (c). Canon dataset - Top 5% and 10% retrieval results

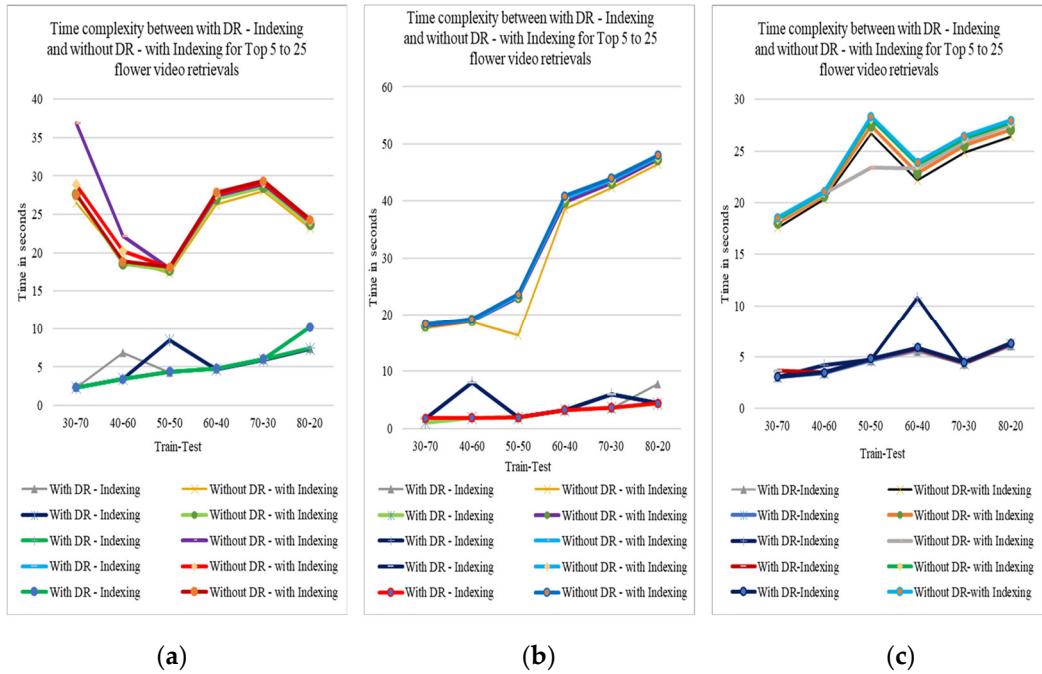
Train - Test in %	Precision		Recall		F-Measure		Time	
	5%	10%	5%	10%	5%	10%	5%	10%
30–70	39.78	25.82	54.15	67.38	44.12	36.1	42.11	42.11
40–60	39.97	25.93	53.8	67.13	44.06	36.13	76.94	76.94
50–50	39.62	25.71	53.34	66.6	43.61	35.78	121.1	121.1
60–40	39.48	25.45	53.04	66.52	43.39	35.49	174.9	174.9
70–30	39.45	25.59	53.43	66.75	43.5	35.67	233.9	233.9

80-20	40.35	26.02	54.09	67.54	44.28	36.2	312.3	312.3
-------	-------	-------	-------	-------	-------	------	-------	-------

### 3.3.4. Comparative Study between Retrieval Time of Flower Videos in Absolute Mode

#### 3.3.4.1. Retrieval Time between ReliefF DR with Indexing and without DR with Indexing

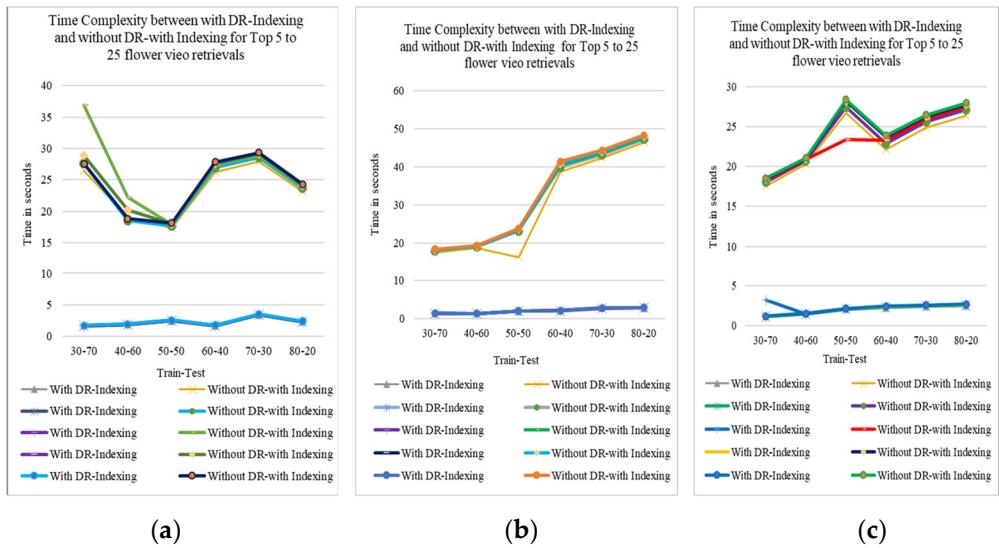
The following Figure 9 show the retrieval time difference between with ReliefF feature selection algorithm for feature dimensionality reduction and without dimensionality reduction. But both using Indexing approaches for the retrieval of top 5 to top 25 videos in absolute mode for SGGP, Sony Cybershot and Canon datasets respectively. From the figures it can be observed that without dimensionality reduction takes long time for retrieval of flower videos than with dimensionality reduction.



**Figure 9.** Retrieval time between ReliefF DR with Indexing and without DR with Indexing.

#### 3.3.4.2. Retrieval Time between BNS DR with KD tree Indexing and without DR with Indexing

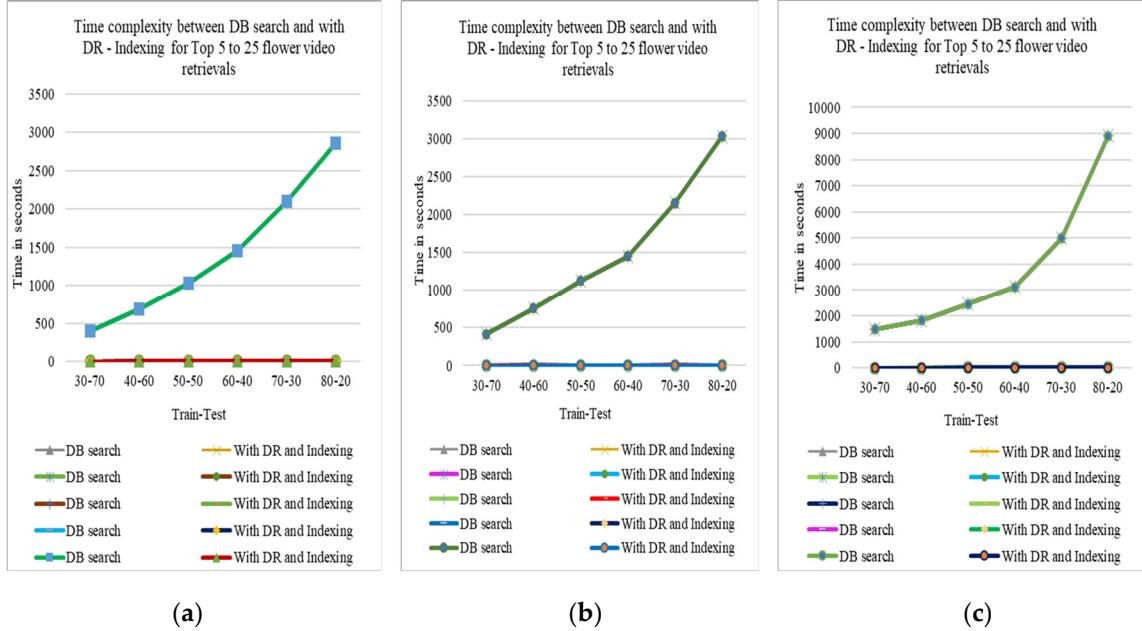
The following Figure 10 show the retrieval time difference between with Binormal separation feature selection metric for feature dimensionality reduction and without dimensionality reduction. But both using Indexing approaches for the retrieval of top 5 to top 25 videos in absolute mode for SGGP, Sony Cybershot and Canon datasets respectively. From the figures it can be observed that without dimensionality reduction takes long time for retrieval of flower videos than with dimensionality reduction. And it can be observed that videos retrieved from this modality takes less time than ReliefF. Therefore, it is an efficient way for the retrieval of flower videos.



**Figure 10.** Retrieval time between BNS DR with KD tree Indexing and without DR with Indexing.

#### 3.3.4.3. Retrieval Time between ReliefF DR with Indexing and Data Base Search

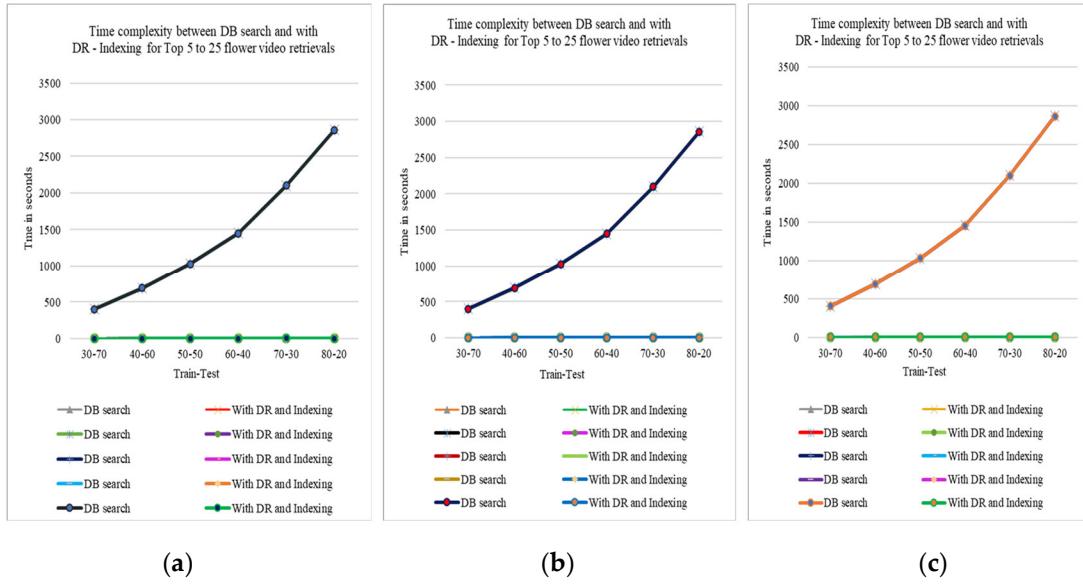
The Figure 11 show the retrieval time difference between with ReliefF feature selection algorithm with KD-tree indexing and database searching to retrieve top 5 to top 25 videos in absolute mode for SGGP, Sony Cybershot and Canon datasets respectively. And we can observe that DB search approach takes long time for retrieval of videos than ReliefF DR with KD-tree Indexing approach.



**Figure 11.** Retrieval time between ReliefF DR with Indexing and data base search.

#### 3.3.4.4. Retrieval Time between BNS DR with Indexing and Data Base Search

The Figure 12 show the retrieval time difference between with BNS DR feature selection metric for feature dimension reduction and KD-tree indexing with database searching to retrieve top 5 to top 25 videos in absolute mode for SGGP, Sony Cybershot and Canon datasets respectively. We can observe that DB search approach takes long time for retrieval of videos than with Bi-normal separation feature selection metric and KD-tree Indexing approach. And we can observe that this modality of retrieval is much more efficient than the ReliefF algorithm in the point of view of retrieval time.



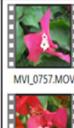
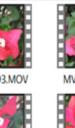
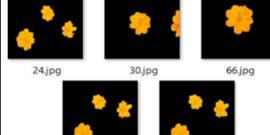
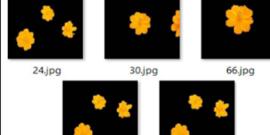
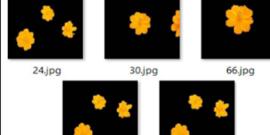
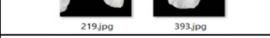
**Figure 12.** Retrieval time between BNS DR with Indexing and data base search.

#### 4. Comparative Study between Proposed and Traditional Retrieval Model

In the previous work the features such as Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP) and Scale Invariant Feature Transform (SIFT) are extracted. With the fusion of all these features the model achieved good performance. The retrieval accuracy of previous work achieved 53.83%, 60.18% and 65.73% for SGGP, Sony cyber Shot and Canon datasets respectively for 80% training and 20% testing. In the proposed work, to further improve the retrieval performance, deep features are extracted and dimensionality reduction methods are utilized in different modalities as shown results in Section 3. From the proposed model the results are improved and good results achieved via absolute modality for top 5 retrievals using Bi-normal feature selection dimensionality reduction method and KD-tree indexing. The retrieval accuracy in this modality obtained 85.14%, 88.69% and 93.29% for SGGP, Sony cyber Shot and Canon datasets respectively for 80% training and 20% testing.

#### 5. Conclusions

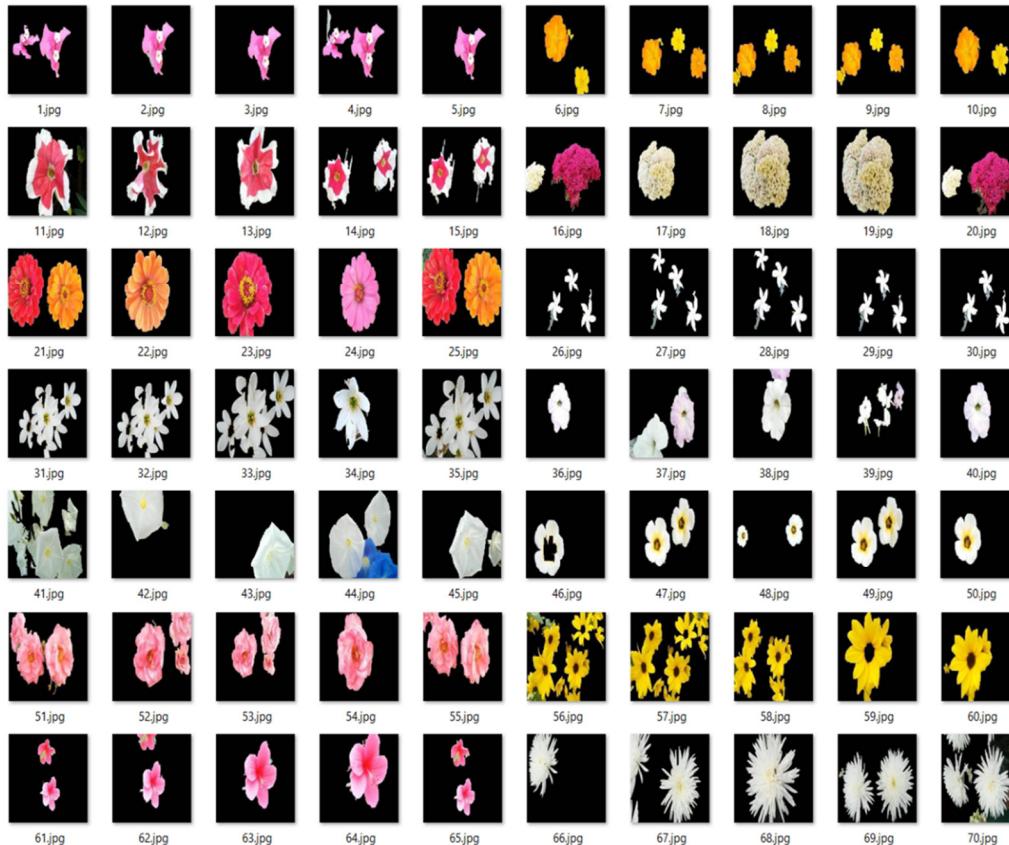
In this paper, for video summarization three different keyframe selection methods are presented. Among the three different methods, GMM clustering approach gives good results. These keyframes are used to represent a video. To design an efficient video retrieval system dimensionality reduction and KD-tree Indexing are used. Proposed two modalities to retrieve similar videos from the database for a given query video, namely, absolute and relative. Good results achieved via absolute modality for top 5 to top 25 retrievals. Retrieval of flower videos using Bi-normal feature selection dimensionality reduction and KD-tree indexing in absolute modality is efficient in terms of time and precision results.

Query Video	Selected Keyframes	Top 10 Retrieved Videos
 MVI_0742.MOV	     	 MVI_0757.MOV  MVI_0771.MOV  MVI_0782.MOV  MVI_0793.MOV  MVI_0797.MOV   MVI_0799.MOV  MVI_0837.MOV  MVI_0840.MOV  MVI_0863.MOV  MVI_0870.MOV
 MVI_0719.MOV	     	 MVI_0724.MOV  MVI_0729.MOV  MVI_2785.MOV  MVI_2794.MOV  MVI_8780.MOV   MVI_8782.MOV  MVI_8783.MOV  MVI_8980.MOV  MVI_8981.MOV  MVI_9096.MOV
 MVI_0103(20).MOV	     	 MVI_0103(20).MOV  MVI_0103(20).MOV  MVI_0103(20).MOV  MVI_0103(20).MOV  MVI_0103(20).MOV   MVI_0103(20).MOV  MVI_0103(20).MOV  MVI_0103(20).MOV  MVI_0103(20).MOV  MVI_0103(20).MOV
 MVI_0060.MOV	     	 MVI_0062.MOV  MVI_0073.MOV  MVI_8927.MOV  MVI_8943.MOV  MVI_8968.MOV   MVI_9388.MOV  MVI_9393.MOV  MVI_9394.MOV  MVI_9395.MOV  MVI_9398.MOV
 MVI_9187.MOV	     	 MVI_9220.MOV  MVI_9244.MOV  MVI_9245.MOV  MVI_9248.MOV  MVI_9249.MOV   MVI_9299.MOV  MVI_9301.MOV  MVI_9302.MOV  MVI_9303.MOV  MVI_9514.MOV

**Figure 13.** Examples of sample Queries, selected keyframes and retrieved top 10 videos by the proposed flower video retrieval system.



**Figure 14.** Samples of flower videos with large intraclass variation.



**Figure 15.** Samples of keyframes selected from proposed GMM clustering approach.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

1. G.G.Lakshmi P, S. Dominic, Shot based keyframe extraction for ecological video indexing and retrieval. *Ecol. Inform.* **2013**, *23*, 107–117.
2. K.S.Thakre, A.M.Rajurkar, R.R.Manthalkar, Video Partitioning and Secured Keyframe Extraction of MPEG Video, ICISP2015. *Procedia Computer Science* **2016**, *78*, 790–798
3. D.S. Guru, Jyothi V. K and Y.H. Sharath Kumar, Cluster based approaches for key frame selection in natural flower videos, 17<sup>th</sup> International Conference on Intelligent Systems Design and Applications, ISDA 2017, Springer AISC, Volume 736, pp. 474–484.
4. S.L.Zhai, B.Luo, J.Tang, C-Y. Zhang, Video abstraction based on relational graphs, Fourth International Conference on Image and Graphics IEEE 2007; pp. 827–832.
5. M.K.Geetha, S.Palanivel, V.Ramalingam, A novel block intensity comparison code for video classification and retrieval. *Expert Syst. Appl.* **2009**, *36*, 6415–6420.
6. C.V.Sheena, N.K. Narayanan, Key-frame extraction by analysis of histograms of video frames using statistical methods. *Procedia Comput. Sci.* **2015**, *70*, 36–40.
7. J-L. Lai, Y. Yi, Key frame extraction based on visual attention model. *J. Vis. Commun. Image R.* **2012**, *23*, 114–125.
8. M.Srinivas, M.M.M. Pai, R.M. Pai, An Improved Algorithm for Video Summarization-A Rank Based Approach. *Procedia Comput. Sci.* **2016**, *89*, 812–819.
9. H.Zhou, A.H. Sadka, M.R.Swash, J.Azizi, U.A. Sadiq, Feature extraction and clustering for dynamic video summarization. *Neurocomputing* **2010**, *73*, 1718–1729.
10. K.Muhammad, M.Sajjid, M.Y.lee, S.W.Baik, Efficient visual attention driven framework for keyframes extraction from hysteroscopy videos. *Biomed. Signal Process. Control* **2017**, *33*, 161–168.
11. H. Liu, H. Hao, Key frame extraction based on improved hierarchical clustering algorithm, 11<sup>th</sup> International Conference on FSKD. *IEEE* 2014; pp. 793–797.
12. W.Yongxiong, S.Shuxin, D.Xueming, A self-adaptive weighted affinity propagation clustering for key frames extraction on human action recognition. *J. Vis. Commun. Image R.* **2015**, *33*, 193–202.
13. N.Ejaz, T.B. Tariq, S.W.Baik, Adaptive key frame extraction for video summarization using an aggregation mechanism. *J. Vis. Commun. Image R.* **2012**, *23*, 1031–1040.
14. C.Gianluigi, S.Raimondo, An innovative algorithm for key frame extraction in video summarization, Springer-Verlag. *J Real-Time Image Proc.* **2006**, 69–88.
15. L.Ferreira, L.A.d.S.Cruz, P.Assuncao, A generic framework for optimal 2D/3D key-frame extraction driven by aggregated saliency maps. *Signal Process. : Image Commun.* **2015**, *39*, 98–110.
16. W.Barhoumi, E.Zagrouba, On-the-fly extraction of key frames for efficient video summarization. *AASRI Procedia* **2013**, *4*, 78–84.
17. M.Chatzigiorgaki, A.N.Skodras, Real time keyframe extraction towards video content identification. *IEEE*: 2009, ISSN: 978-1-4244-3298.
18. L. Honghua, Y.Xuan, P.Jihong, Key frame extraction based on multi-scale phase based local features, ICSP2008 proceedings. *IEEE*: 2008; pp. 1031–1034.
19. J.H.Su, Y-T.Hung, H-H.Yeh, V.S.Tseng, Effective content-based video retrieval using pattern-indexing and matching techniques. *Expert Systems with Applications* **2010**, *37*, pp.5068–5085.
20. D.Liu, T.Chen, Video retrieval based on object discovery. *Comput. Vis. Image Underst.* **2009**, *113*, 397–404.
21. Y.Zhu, X.Huang, Q.Huang, Q.Tian, Large-scale video copy retrieval with temporal-concentration SIFT. *Neurocomputing* **2016**, *187*, 83–91.
22. J.Han, X.Ji, X.Hu, J.Han, T.Liu, Clustering and retrieval of video shots based on natural stimulus fMRI. *Neurocomputing* **2014**, *144*, 128–237.
23. C-C.Yu, F-D.Jou, C-C.Lee, K-C.Fan, T.C.Chuang, Efficient multi-resolution histogram matching for fast image/video retrieval. *Pattern Recognit. Lett.* **2008**, *29*, 1858–1867.
24. B.Liang, W.Xiao, X.Liu, Design of video retrieval system using MPEG-7 descriptors. *Procedia Eng.* **2012**, *29*, 2578–2582.
25. S.Asha, Sreeraj.M, Content based video retrieval using SURF descriptor. IEEE Computer Society, 2013, 212–215.
26. L.Deng, L-Z.Jin, A video retrieval algorithm based on ensemble similarity. *IEEE*: 2010; pp. 638–642.
27. N.A.Lili, Hidden Markov Model for content based video retrieval. IEEE Computer Society: pp. 353–358.
28. G.S.Nagaraja, S.R.Murthy, T.S.Deepak, Content based video retrieval using support vector machine classification. *IEEE*: 2015; pp. 821–827.
29. K.Chatterjee, S-C.Chen, Hierarchical affinity hybrid tree: A multidimensional index structure to organize videos and support content based retrievals. *IEEE IRI*: 2008; pp. 435–440.
30. J.Zhao, Asymptotic convergence of dimension reduction based boosting in classification. *J. Stat. Plan. Inference* **2013**, *143*, 651–662.

31. H.Xie, J.Li, Q.Zhang, Y.Wang, Comparison among dimensionality reduction techniques based on random projection for cancer classification. *Computational Biology and Chemistry* 2016, 65, pp.165–172.
32. X.Wang, W.Liu, J.Li, X.Gao, A novel dimensionality reduction method with discriminative generalized eigen-decomposition. *Neurocomputing* 2016, 173, 163–171.
33. U.S.Pacharaney, P.S.Salankar, S.Mandalalpu, Dimensionality reduction for fast and accurate video search and retrieval in a large scale database; IEEE: 2013, ISSN.978-1-4799-0727-4.
34. H.Gan, A noise-robust semi-supervised dimensionality reduction method for face recognition. *Optik* 2018, 157, 858–865.
35. L-A.Gottlieb, A.Kontorovich, R.Krauthamer, Adaptive metric dimensionality reduction. *Theor. Comput. Sci.* 2016, 620, 105–118.
36. C.Affonso, A.LD.Rossi, F.H.A.Vieira, A.C.P.D.L.F.D.Carvalho, Deep learning for biological image classification. *Expert Syst. Appl.* 2017, 85, 114–122.
37. X.Luo, R.Shen J.Hu, J.Deng, L.Hu, Q.Guan, A deep convolution neural network mode for vehicle recognition and face recognition. *Procedia Comoputer Sci.* 2017, 107, 715–720.
38. X.W.Gao, R.Hui, Z.Tian, Classification of CT brain images based on deep learning networks., Computer methods and programs in biomedicine 2017, 138, 49–56.
39. N.Najva, K.E.Bijoy, SIFT and Tensor based object detection and classification in videos using deep neural networks. *Procedia Comput. Sci.* 2016, 93, 351–358.
40. Z.Liang, S.Xia, Y.Zhou, Normalized discriminant analysis for dimensionality reduction. *Neurocomputing* 2013, 110, 153–159.
41. S.Wang, J.Lu, X.Gu, H.Du, J.Yang, Semi-supervised linear discriminant analysis for dimension reduction and classification. *Pattern Recognit.* 2016, 57, 179–189.
42. X.Chen, H.Yin, F.Jiang, L.Wang, Multi-view dimensionality reduction based on universum learning, *Neurocomputing* 2018, 275, 2279–2286.
43. F.S.Tsai, Dimensionality reduction for computer facial animation. *Expert Syst. Appl.* 2012, 39, 4965–4971.
44. B.Jiang, J.Yang, Z.Lv, K.Tian, Q.Meng, Y.Yan, Internet cross-media retrieval based on deep learning. *J. Vis. Commun. Image R.* 2017, 48, 356–366.
45. Y. Hongpeng, J. Xuguo, Yi C and F. Bin, Scene Classification based on single-layer SAE and SVM. *Expert Syst. Appl.* 2015, 42, 3368–3380.
46. H. Weiming, X. Nainhua, Li X Zeng, M. Stephen, A survey on visual content based video indexing and retrieval. IEEE transaction on systems, MAN and Cybernetics-Part C, Applications and reviews 2011, 41, 797–819.
47. M. Das, R. Manmatha, E M Riseman, Indexing flower patent images using domain knowledge. *IEEE Intelligent Systems* 1999, 14, 24–33.
48. D.S. Guru, Y. H. Sharath, S. Manjunath, Texture Features and KNN in classification of Flower Images, IJCA special Issue on Recent Trends in Image Processing and Pattern Recognition, RTIPPR, 2010, 21–29.
49. D.S. Guru, Y. H. Sharath, S. Manjunath, Texture Features in flower classification. *Math. Comput. Model.* 2011, 54, 1030–1036.
50. R. Nock and F. Nielsen, Statistical Region Merging, IEEE transactions on Pattern Analysis and Machine Intelligence, 2004, 26, 1–7.
51. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Proceedings of 25<sup>th</sup> International conference on Neural Information Processing Systems: 1, 2012; pp. 1097–1105.
52. Y. Guo, Y. Liu, A. Oerlemans, S.Lao, S. Wu, M.S.Lew, Deep learning for visual understanding: A review. *Neurocomputing* 2016, 187, 27–48.
53. A. Zafra, M. Penchenizkiy, S.Ventura, ReliefF-MI: An extension of ReliefF to multiple instance learning. *Neurocomputing* 2012, 75, 210–218.
54. C. Gianluigi and S. Raimondo, An innovative algorithm for key frame extraction in video summarization. *J Real-Time Image Proc.* 2006, 1, po.69–88.
55. R.O.Duda, P.E.Hart, D.G.Stork, Pattern classification, Second edition.
56. D. K Vandana, D. Singh, P. Raj, M. Swathi, P. Gupta, kd-tree based fingerprint identification system, 2<sup>nd</sup> international conference on Anti-counterfeiting, Security and Identification, IEEE xplore, 2008, pp.5–10.
57. K. Kira, L. Rendell, A practical approach to feature selection, Machine Learning, proceedings of the ninth conference, 1992, 249–256.
58. G. Forman, An extensive empirical study of feature selection paradigm for text classification. *J. Mach. Learn. Res.* 2003, 3, 1289–1305.
59. K. B. Nagasundara, D S Guru, S Manjunath, KD-tree based palmprint Indexing, ICCVR-2011, Bhubaneswar, India, pp.51–55.
60. K B Nagasundara, Theses: Biometric Indexing, 2012, Appendix-A, Kd-tree – A complete illustration with a example. pp. 146–152.

61. Y H K Sharath, D S Guru, Sketch based Flower Retrieval: Kd-tree based approach, Elsevier Procedia Computer Science, vol. 46, ICICT 2015, 1577 – 1584.
62. Manjunath S, 2012, Video Archival and Retrieval System (VARS), thesis.
63. Gonzales R C, Woods R E and Eddins, 2008, Digital Image Processing, Third Edition.
64. Emad Mabrouk, Ahmed Ayman, Yara Raslan, Abdel-Rahman Hedar, 2019, Immune system programming for medical image segmentation, Journal of Computational Science, Volume 31, February 2019, pp. 111–125.
65. S. A. Hojjatoleslami and J. Kittler. Region Growing: A New Approach. *IEEE Trans. Image Process.* **1998**, *7*, 1079–1084.
66. P A Vijaya, M Narasimha Murthy, D K Subramanian, Efficient bottom-up hybrid hierarchical clustering techniques for protein sequence classification. *Pattern Recognit.* **2006**, *39*, 2344–2355.
67. Turgay Celik, 2011, Bayesian change detection based on spatial sampling and Gaussian mixture model. *Pattern Recognition Letters* **2011**, *32*, 1635–1642.
68. Michael Reiter, Paolo Rota, Florian Kleber, Markus Diem, Stefanie Groeneveld-Krentz, Michael Dworzak, Clustering of cell populations in flow cytometry data using a combination of Gaussian mixtures. *Pattern Recognit.* **2016**, *60*, 1029–1040.
69. Jiliang Tang, Salem Aleyani and Huan Liu, 2014, Feature selection for classification: A Review, Data classification: Algorithms and Applications, Book chapter, pp.37–64.
70. Jun Chin Ang, Andri Mirzal, Habibollah Haron and Haza Nuzly Abdull Hamed. *ACM Transactions on Computational Biology and Bioinformatics* **2016**, *13*, 971–989.
71. Ryan J. Urbanowicz, Randal S. Olson, Peter Schmitt, Melissa Meeker, Jason H. Moore. Benchmarking Relief-Based Feature Selection Methods for Bioinformatics Data Mining. *J. Biomed. Inform.* **2018**.
72. Guru D S, Sharath K Y H and Manjunath S, 2010, Texture Features and KNN in Classification of Flower Images, IJCA Special Issue on 'Recent Trends in Image Processing and Pattern Recognition, pp.21–29.
73. Na Zhao, Shi-Kai Sui, Ping Kuang, 2015, Research on image segmentation method based on weighted threshold algorithm, IEEE ISBN: 978–1-4673–8266–3, 307–310.
74. Stuart Golodetz, Irina Voiculescu, Stephen Cameron, Simpler editing of graph-based segmentation hierarchies using zipping algorithms. *Pattern Recognit.* **2017**, *70*, 44–59.
75. Xiaomin Yu, Weibin Liu, Weiwei Xing, Behavioral segmentation for human motion capture data based on graph cut method. *J. Vis. Lang. Comput.* **2017**, *43*, 50–59.
76. Jifeng Ning, Lei Zhang, David Zhang, Chengke Wu, Interactive image segmentation by maximal similarity based region merging. *Pattern Recognit.* **2010**, *43*, 445–456.
77. Fanjie Meng, Dalong Shan, Ruixia Shi, Yang Song, Baolong Guo, Weidong Cai, Merged region based image retrieval. *J. Vis. Commun. Image Represent.* **2018**, *55*, 572–585.
78. Shuchao Pang, Juan Jose del Coz, Zhezhou Yu, Oscar Luaces, Jorge Diez, 2017, Deep learning to frame objects for visual target tracking. *Eng. Appl. Artif. Intell.* **2017**, *65*, 406–420.
79. P. Punitha, D. S. Guru, 2008 Symbolic image indexing and retrieval by spatial similarity: An approach based on B-tree. *Pattern Recognit.* **2008**, *41*, 2068–2085.
80. Kimia Razaei-Kalantari, A M Eftekhari-Moghadam, 2010, Symbolic image indexing and retrieval by spatial similarity: A new approach based on multi-dimensional B<sup>+</sup> tree, IEEE Xplore, ISBN:978–89–88678.
81. Guru D S, Jyothi V K, Kumar Y H S. Features fusion for retrieval of flower videos. *Springer Nature. Lecture Notes in Networks and Systems*; 2019; Volume 43, pp. 221–233.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.