Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Video Annotation Methodology Based on Ontology for Transportation Domain

Khushboo Khurana*, Dr.M.B.Chandak
CSE Department of Computer Science Engg.
SRCOEM, Nagpur, India

Abstract— Increase in the importance of video data has led to flooding of video content over the internet and offline world. For such abundance of content the access of relevant video in reduced time is a matter of concern. So the need for a system which will assist in content based access of video arises. This paper introduces a novel approach for video annotation. The key frames are extracted from the video and are analyzed. Instead of complete video frames, only the key frames are analyzed to identify the objects present. The object detectors are trained for identification of the object. The detected objects are then added in the annotation file. The annotation is based on ontology which eases the semantic retrieval of videos. Ontology based video annotation greatly accelerates the performance of retrieval systems.

Keywords— Video Annotation, ontology based annotation, image annotation, SVM classification, SIFT features, key frame extraction

I. INTRODUCTION

Age of internet may cease to exist only with the end of human race. From copyright to freelance, internet is now the new barter system of the world. It takes your time and gives knowledge and information. Internet provides us with knowledge and information in the form of textual data, images, videos and multimedia content. With every increasing minute the content over internet is also increasing. Various educational video lectures, CCTV surveillance, transport and other types have increased the importance of multimedia video content. Videos tend to provide more informative content in a simple manner. This tremendous database of video needs proper approach to handle its access. Often these are not catalogued and are accessible only by the sequential scanning of the sequences. To make the use of large video databases more feasible, we need to be able to automatically index, search and retrieve relevant material [1]. Not only the web based information but also the increasing amount of offline data as well as personal data requires attention. The personal information present on computer systems need to be managed by the users which can be extensively time consuming and is a matter of concern. The search option available finds the image or video files based on the file name. For searching relevant video based on content, annotation can be added to ease the semantic retrieval. Content based video retrieval has many applications, but is a challenging problem. Difficulties lie in coming up with appropriate representation for visual content which reflects the semantics of the video. One solution is to recognize objects in the video [2]. Understanding semantic of video content is immediate for humans but far from immediate for a computer. This is referred as semantic gap [3]. To determine the video content, in this paper we have used Support Vector Machine (SVM) classifier as the detector for object and annotations are added to the video. Video annotation implies extraction of information and to attach such metadata to the video which will accelerate the retrieval speed, ease of access, analysis and categorization. It permits fast and better understanding of video content and enhances the performance of retrieval and reduces human time & efforts for better study of videos. Video annotation is imperative technique that assists in video access. In this paper we have proposed annotation based on ontology to have more complete semantic annotation. Ontology helps in establishing links to other concepts and disambiguating the results of classification. Amongst various video annotation techniques, ontology based techniques are fast emerging and give better results than other techniques [4]. In our approach we have combined machine learning and ontology. Considering that the videos have annotation, for its access we need to only consult the annotation file to determine whether the video is pertinent to the search, reducing overall access time. Also retrieved videos will be semantically correct. Though the system can be easily extended for other types of videos, in the paper we have used videos form transportation domain. Different object detectors are used for the objects- airplane, car, bus, bike and ship. Rest of the paper is laid out as follows. The next section presents the related work in the field. In section 3 we have proposed the methodology for video annotation which is based on ontology. Ontology based annotation helps in understanding the semantic concept related to the video. Section 4 describes the phases and the implementation details. Section 5 presents the experimental results and we finally conclude in section 6.

II. RELATED WORK

Video annotation is budding as the active field of research. First step towards video annotation is the analysis of video. Analysis can be of visual data, textual information, audio or combination of these; complete video can be analyzed for object detection or motion detection. Obtaining key frames from the video can be another approach. Text in the video

which may be captions or some other text appearing can be extracted using segmentation, output of which can be passed to optical character recognition (OCR) module to convert the text into ASCII [5]. An annotation approach that uses local, global and motion features is proposed in [6]. The Support Vector machine (SVM) classifier is trained using labelled videos and weights are assigned to unlabeled ones. Similarity between unlabeled video with all the relevant labelled videos is calculated and maximum similar video is found; the annotation for this video becomes the annotation for the unlabeled video. Assortments of other machine learning techniques also exist. Ontology based annotation approach are the recent innovation. In [7], automatic video annotation technique based on ontology is proposed. MPEG-7 visual descriptors are used as features which are mapped to semi concepts and finally to objects. Ontologies are exploited to hierarchically describe the contents among heterogeneous users and devices. They have applied the technique in the smart TV environment. Video concept detection approaches using semantic inference rules and SVM classifier for efficient video search, sharing and browsing is proposed. Bogdan Vrusias, Dimitrios Makris, et.al., have presented a framework for semantic annotation of CCTV video which combines computer vision algorithm that extract visual semantics, along with Natural Language Processing that builds the domain ontology from unstructured text annotations. The visual and text semantics are to be linked to provide video annotation [8]. Important frames can be extracted from the video and video annotation problem can be transformed into image annotation. Image annotation is the process of assigning a class or description to an unknown image. Human based annotations can be comprehensive but may be slow and costly. In [9], automatic annotation method for images using ontology is proposed. Low-level features and textual descriptions are used for ontology construction. Image annotation is implemented as retrieval process by comparing input image with representative images of all classes. In this paper we have combined the SVM for classification of objects and supplied transport domain ontology for the creation of annotation.

III. THE PROPOSED METHODOLOGY

We present a novel approach for the annotation of video which is based on ontology. The proposed system consists of extraction of frames that contain most significant visual information. Hence the video is reduced to less number of images known as key frames. The classifiers for different objects are trained using the features for the respective objects. We have used scale invariant feature transform (SIFT) as the features for training and testing the classifiers. These features are invariant to scale, rotation, translation and are widely used for object recognition in images .When the features of key frames are given as input to different classifiers; they can detect whether the object is present in the frame or not. The proposed system uses Support Vector Machine (SVM) as the classifier. If the object is present then the frame names are given to the annotation module which will create an xml file corresponding to each key frame. The xml file contains the object / objects present in the image (video frame) along with object ontology. Figure 1 shows an overview of the proposed system. The input video after being reduced to key frames is sent to the feature extraction module. The SIFT features are found here. The detectors are trained with the help of these SIFT features. When images are to be classified; their SIFT features are passed to the classifiers. Once the objects in the frames are identified the annotation module provides the annotated frames for the video. These form the annotations for the video. The domain ontology is also constructed and provided to the annotation module.

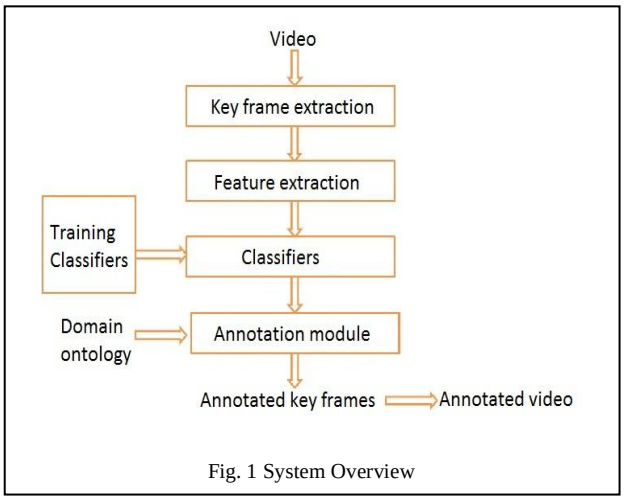


Fig. 1 System Overview

The proposed system can be broadly divided into following phases.

Phase 1: To obtain key frames from the video

Phase 2: (a) To extract SIFT features from all the key frames

(b) To train different object classifiers using SIFT features and to test the objects present in the key frames

Phase 3: To annotate the key frames based on ontology

IV. PHASES AND IMPLEMENTATION DETAILS

A. Key Frame Extraction

Video may consist of large amount of frames. People are unable to perceive millions of individual frames, but they can perceive episodes, scenes, and moving objects. A scene in a video is a sequence of frames that are considered to be

semantically consistent. Scene changes therefore demarcate changes in semantic context. Segmenting a video into its constituent scenes permits it to be accessed in terms of meaningful units. A set of frames that best represent the visual content of the scenes can then be extracted. These frames are called key frames and are used in the latter task of video annotation [1]. Key frame extraction algorithms select a subset of the most informative frames from videos. Key frame extraction finds applications in several broad areas of video processing research such as video summarization, creating chapter titles in DVDs, video indexing, and prints from video [10]. Analysis of video can hence be reduced to only analysis of the key frames. As per our previous work on key frame extraction from video, in this paper we have used the algorithm proposed in [11]. The algorithm uses edge difference to find the similarity between two frames. The algorithm is explained briefly by following steps.

Step 1: Find the edge difference between all the consecutive frames of video.
Step 2: Find threshold
Step 3: Edge differences which exceed the threshold are considered. Second frame out of the two consecutive frames is output as key frame.

Depending upon the change in the content of the video, key-frames are found. If there is less change in the content, fewer frames are key frames and if more change in the content then all the important frames are output. Refer [11] for the complete detailed algorithm for the key frame extraction.

B. Compute Features And Classification

The next step is the identification of objects present in the video. As we have reduced the video to key frames we need to find the objects present only in these frames. Object recognition is extensively important in various fields such as machine vision industry for the purposes of inspection, registration, and manipulation, in the field of surveillance, etc. Most systems for object recognition depend on correlation-based template matching. These can be effective for certain engineered images; but becomes computationally infeasible where object rotation, scale, illumination, and pose are allowed to differ, and even more when dealing with occlusion. An alternative to searching all image locations for matches is to extract features from the image that are invariant to image formation process and matching only to those features.

For recognition of objects present in the image we have trained different classifiers for different objects. For training the classifiers and then to determine the objects we need to provide object features (SIFT features). SIFT approach transforms the image into an assortment of local feature vectors; these vectors are invariant to translation, scaling, rotation and partially invariant to illumination changes [12].They are highly distinctive and easy to extract and allows for correct object identification with low probability of mismatch. Extrema detection produces many keypoint candidates, out of which stable ones which have low contrast or are poorly localized along an edge are neglected. After assigning an image location, scale, and orientation to each keypoint, a descriptor for the local image region that is highly distinctive yet is as invariant as possible to remaining variations, such as change in illumination is found [13]. In our approach we find keypoints and descriptors from the images; then histogram of visual words is constructed. These features are given as input to the object classifiers for training. We have adopted the technique of machine learning i.e., first training the object detectors and then testing. For actual classification, features are calculated for the test images which are given as input to the trained classifier. The object is then recognized. Support Vector Machine (SVM) is used as the classifier. It is a supervised technique which classifies into two classes. In our system it classifies as positive or negative i.e., object present or absent. SVM is trained as per the following algorithm:

Algorithm TrainClassifier
{
Step 1: Compute SIFT keypoints and descriptors for positive and negative training images.
Step 2: Compute the histogram using features of positive and negative images. It is the histogram of visual words.
Step 3: Labels for images are computed. +1 is assigned for positive examples and -1 for negative examples.
Step 4: Train the liner SVM classifier with histogram and labels as input.
}

For training we need to make the classifier learn positive images i.e. when the object is present and with the images where the object is absent. SIFT features of both are found, using which histogram is constructed. This histogram along with the labels is given as input to the SVM; resulting in a trained object classifier. Different classifiers are used for different object recognition. Figure 2 illustrates the training procedure for single object classifier.

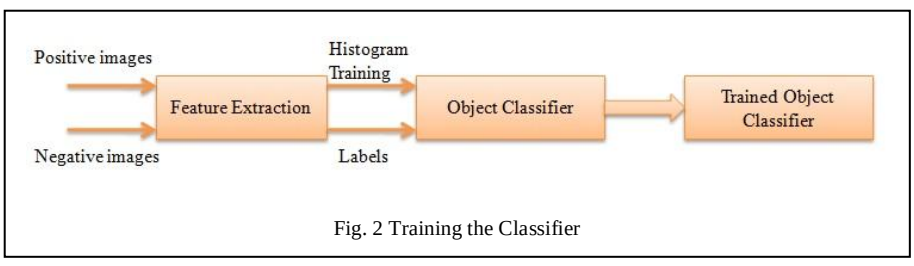


Fig. 2 Training the Classifier

After training is complete; the classifier is ready for object identification. The SIFT features of the test images (video key frames found in section 4.1) are given as input to all the trained object classifiers one by one. Each SVM classifier returns a score value for the test frame. This score may be negative or positive. The algorithm ***DetectObjectInKeyFrame*** is proposed for detection of objects. Figure 3 shows the flowchart representation of the algorithm.

Algorithm DetectObjectInKeyFrame

{

Step 1: Compute SIFT features and its histogram for input test images / video key frames.

Step 2: Find labels for input images

Step 3: Provide above two as input to the classifier. The SVM classifier returns score for each input image; let these be score(i)

Where i – number of images given as input.

Step 4: if(score(i) is positive)

{

Object is present, display image(i) and send it to annotation module.

}

else

{

Object is absent.

}

}

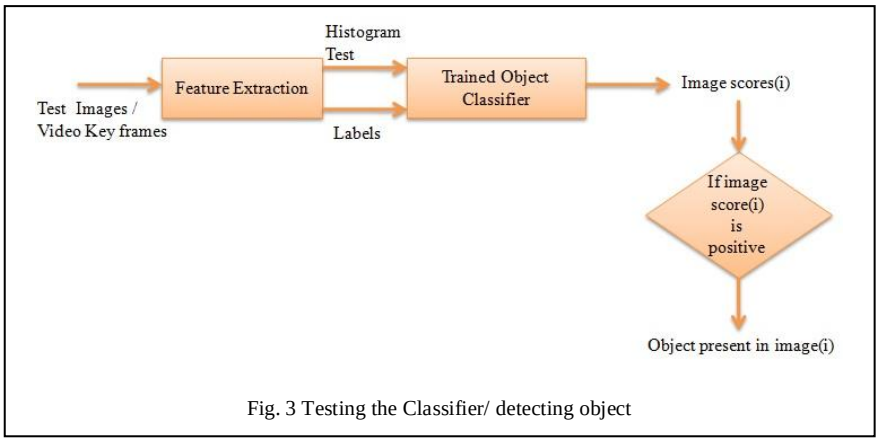


Fig. 3 Testing the Classifier/ detecting object

The SIFT features are computed for the video key frames and histogram is constructed. Histogram along with test labels is given to the trained object classifier. The result obtained is **scores(i)** corresponding to i number of input images/ key frames. The score for test image may be positive or negative. If the score is positive it indicates the presence of the object; whereas negative scores correspond to nonexistence of the object. Only if the object is present, the image is passed to the annotation module for adding the object annotation. Consider for example a bus classifier is trained to identify bus in an image. Figure 4 shows the input images given to the Trained Bus Classifier (TBC). As shown in figure, first three images are of bus and remaining images do not contain bus. Hence, when these images are given as input to the TBC it should classify first three images as "Containing Bus". The TBC produces the scores for the images as shown in figure 4. The scores **2.0058, 2.5664, 2.5940, -2.3256, -1.4303, -0.5336** correspond to the input images. As per our algorithm ***DetectObjectInKeyFrame*** positive score images (images having score: 2.0058, 2.5664 and 2.5940) are declared as consisting object Bus. Images with negative score (images having score: -2.3256, -1.4303 and -0.5336) are reported as Bus absent. After identification of the images that contain "bus" they are given to the annotation module for creation of xml file. We have trained 5 different linear SVM classifiers for identification of five different objects. These objects are airplane, bus, bike, car and ship. Figure 5 depicts in detail the training and testing for object identification. For example consider the airplane classifier, the classifier is trained with positive images containing airplane and with negative images, which do not contain airplane.

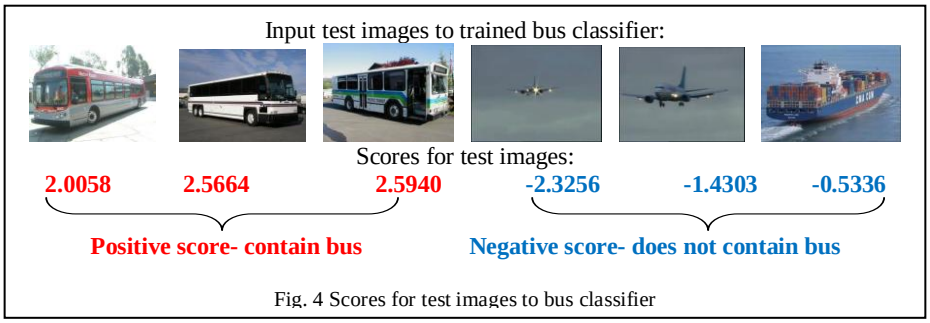


Fig. 4 Scores for test images to bus classifier

The bus classifier is trained with images containing bus (positive images) and with negative images, which do not contain bus. Similarly, all the other object identifiers are trained. For classification of key frames amongst these five categories, key frame features are given as input to different trained classifiers. First the frames are given to airplane classifier, if there is any frame containing airplane it is given to the annotation module. Then frames are given to car classifier which authorizes only the images of car and passes to the annotation module. Same is repeated for bus, bike and ship classifiers. If any frame contains more than one object, all objects will be present in the annotation file.

C. Ontology Based Annotation

Ontology consists of entities and their relationships, which are organized hierarchically. It may be in the form of classes and subclasses where each class may consist of one or more instances. Ontology can be defined as an explicit specification of a conceptualization [14]. For example, "cat" is a subclass of class "animal". Ontology is simply a knowledge representation method which provides the way to consolidate the information in a structured manner [15]. It helps to associate semantic to any object or its image and provides a better understanding in proper context.

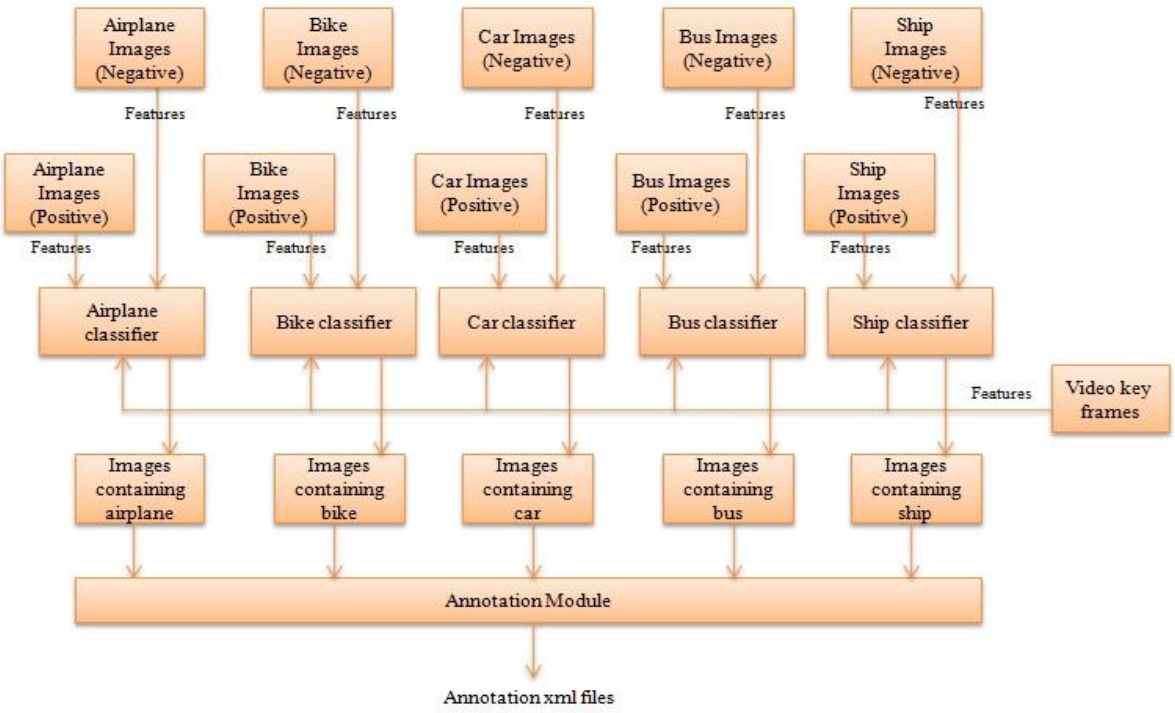


Fig. 5 Classification procedure using 5 classifiers- Airplane, Bike, Bus, Car, Ship.

Figure 6 shows the transport domain ontology which used for annotation. The main domain is transport hence the root node is transport. We have considered three transports namely – air, water and road. The leaf nodes are the objects which are recognized by the classifiers. Annotation file is created in the xml format which contains the object present in the image (leaf of the ontology) as well as all the parent nodes present in the ontology. For example, if the object identified in any frame is airplane, the corresponding annotation xml file will contain airplane as well all its parent- air transport and root- transport. For bike, the annotation file will have bike, two-wheeler, light vehicles, road, and transport.

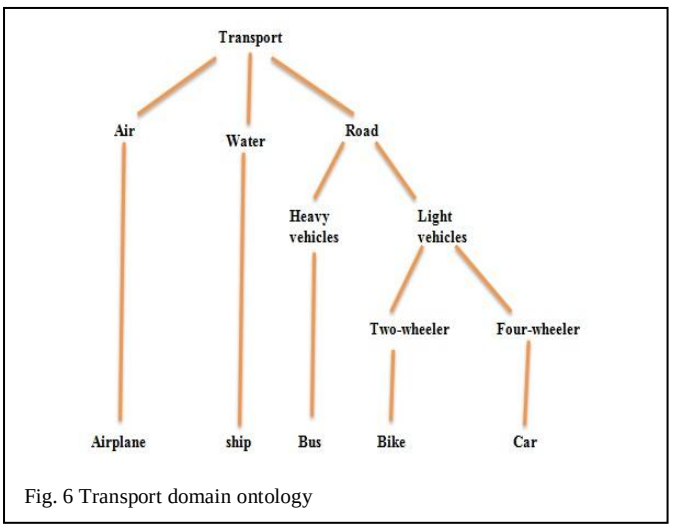


Fig. 6 Transport domain ontology

Figure 7 (a) shows the xml annotation created for one of the key frame containing airplane. *ImageName* is the name of the key frame for the corresponding xml file. *VideoName* is the video to which the frame belongs. *KeywordAnnotation* consist of the ontology based description of the object present in the key frame. So, even if the search is for air transport the *airplane1* video will be retrieved. Ontology based annotation files will help in the content based retrieval of the videos. Figure 7(b) shows the xml file that will be created for the frames/ images containing car. Annotation file contains the complete ontology structure.

```
<?xml version="1.0"?>
- <annotation>
    <ImageName> img10 </ImageName>
    <VideoName> airplane1 </VideoName>
  - <KeywordAnnotation>
      <Root> Transport </Root>
      <Domain> Air transport </Domain>
      <Subdomain> </Subdomain>
      <Object> Airplane </Object>
    </KeywordAnnotation>
</annotation>
```

(a)

```
<?xml version="1.0"?>
- <annotation>
    <ImageName> img12 </ImageName>
    <VideoName> car7 </VideoName>
  - <KeywordAnnotation>
      <Root> Transport </Root>
      <Domain> Road transport </Domain>
      <Subdomain> Light vehical </Subdomain>
      <Subdomain1> Four Wheeler </Subdomain1>
      <Object> Car </Object>
    </KeywordAnnotation>
</annotation>
```

(b)

Fig. 7 (a) Annotation file for airplane images. (b) Annotation file for car images

V. EXPERIMENTAL RESULTS

We have implemented the proposed video annotation system using the transport domain ontology in Matlab-R2012 in Windows-7 machine. For feature extraction module we have used the VL Fleat Matlab toolbox. The VLFeat open source library implements popular computer vision algorithms including HOG, SIFT, MSER, k-means, hierarchical k-means, agglomerative information bottleneck, SLIC superpixels and quick shift. It is written in C for efficiency and compatibility, with interfaces in MATLAB for ease of use, and detailed documentation throughout. It can be downloaded from the download section of [16]. Only the visual characteristics of video are considered. The input given to the key frame extraction module is a video in avi format. The evaluation of the system is performed on the transportation domain videos. The key frames for each video are stored in separate folders. The system can take as input these key frame images or folder of images for object identification and annotation. The GUI for the system is shown in figure 8. The main focus of the paper is to find key frames and then all the key frames of one video are given at a time to the system and it finds the objects in each frame and creates an xml file corresponding to each frame. The results obtained are shown in section V-A. We have also evaluated the system to perform object identification and annotation for image database. An image database consisting of images of different objects from the transport domain is constructed. Object name and the number of images to be extracted from the collection of images can be specified. There are 1739 images in the collection. The results are shown in section V-B.

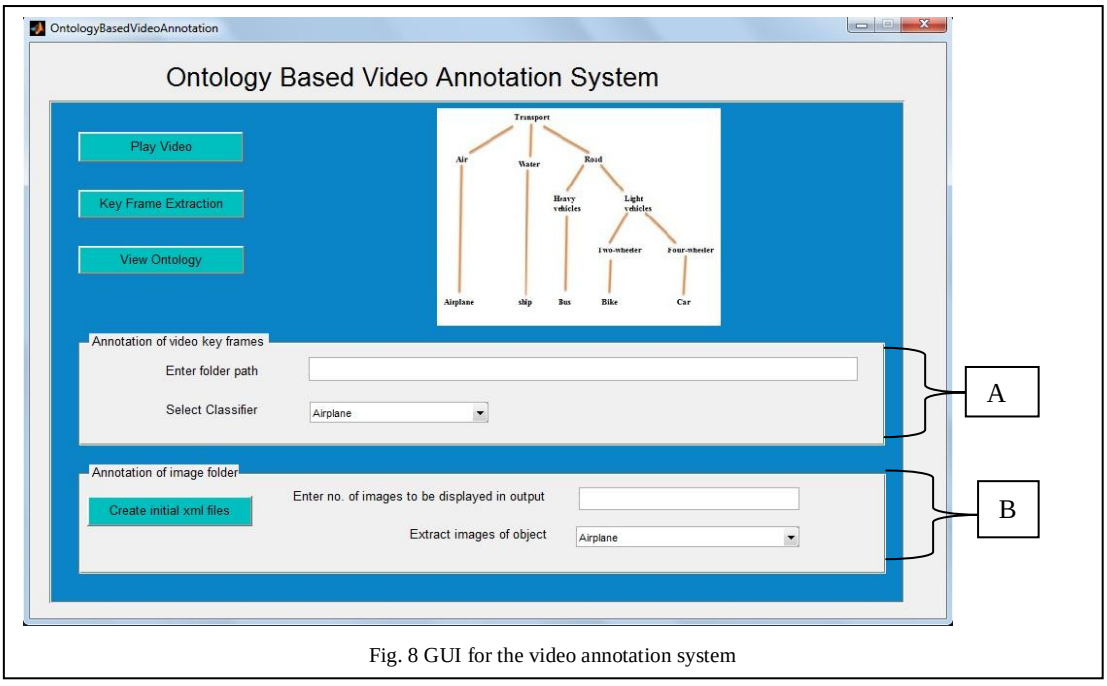


Fig. 8 GUI for the video annotation system

A. Key frame extracted from the video and annotation.

Key frames are still images from video which describe the videos important information. The "a1.avi" video is selected as the input for key frame extraction module as shown in figure 9. The key frames extracted are also shown. Set

of these key frames is given as input to the A part of the system (classification and annotation module) shown in figure 8. These images are passed through all the 5 classifiers one by one. Consider that all the extracted key frames from the "*a1.avi" video* are passed to all the five classifiers one by one; figure 9 shows the results obtained by all classifiers. The scores allotted to all the key frames by the airplane classifier are positive. Since the scores are positive for all the three frames; all the three frames are recognized as "containing airplane". Also corresponding to each frame an xml annotation file is created. As desired the other classifiers have shown absence of other objects since the scores reported are negative by other classifiers. The annotation file created for all the three key frames have an ontology structure as shown in figure 7(a).

Input video for key frame extraction:

Extracted key frames:



Scores by airplane classifier:	**2.6909**	**0.1462**	**1.3971**
Airplane classifier:	Creates annotation xml file for all the frames.		
Scores by car classifier:	**-1.6738**	**-0.4691**	**-0.4019**
Car classifier:	No image of car found		
Scores by bus classifier:	**-0.8927**	**-1.4737**	**-0.4779**
Bus classifier:	No image of bus found		
Scores by bike classifier:	**-0.7036**	**-1.7208**	**-2.2264**
Bike classifier:	No image of bike found		
Scores by ship classifier:	**-0.1515**	**-0.4542**	**-0.6161**
Ship classifier:	No image of bike found		

Fig. 9 Results of different classifiers and annotation when key frames of a video are input

We also rename the video depending upon the object detected. If the detected object is airplane as in above example; the name of the video is changed to "Airplane&" where &- is a random numeric value. An xml file corresponding to the video is also created. The retrieval module can use the xml files corresponding to the key frames for accessing the video or can directly judge whether the video is as required by accessing video name or xml corresponding to the video can be accessed. Depending upon the level of detail; retrieval module can be designed. If more than one object is present in the image then object detectors identify all the objects. Figure 10 shows the input image given to the system and the corresponding xml annotation file. The input image contains three objects bus, car and airplane. When given input to car classifier car is identified and ontology based annotation is added. Then it is input to bus classifier, bus is detected and added as second object, airplane is also identified.

(a)

```
<?xml version="1.0"?>
- <annotation>
    <ImageName> test7 </ImageName>
    <VideoName> Test_3multiobject </VideoName>
  - <KeywordAnnotation>
      <Root> Transport </Root>
      <Domain> Road transport </Domain>
      <Subdomain> Light vehical </Subdomain>
      <Subdomain1> Four Wheeler </Subdomain1>
      <Object> Car </Object>
    </KeywordAnnotation>
  - <KeywordAnnotationObject2>
      <Domain> Road transport </Domain>
      <Subdomain> Heavy vehical </Subdomain>
      <Object> Bus </Object>
    </KeywordAnnotationObject2>
  - <KeywordAnnotationObject3>
      <Domain> Air transport </Domain>
      <Subdomain> </Subdomain>
      <Object> Airplane </Object>
    </KeywordAnnotationObject3>
</annotation>
```

(b)

Fig. 10 (a) Input image containing car, bus and airplane (b) Annotation xml file for the input image

B. Object image extraction and annotation from collection of images

Apart from video key frame annotation, we have also gauged the system performance when large number of images are to be annotated. Figure 8 (B) depicts the system. Images of various objects are stored in a folder; these objects are mainly from the transport domain. The problem of image annotation is transformed into an image retrieval one. We need to specify the name of the object whose images are to be retrieved and number of images to be extracted. Annotation files are created for the corresponding extracted images. As shown in the tables 1, when the input to the system is to extract 5 images of the object bus, all 5 images of bus are retrieved. All the 5 images will be annotated as bus. When we try to extract 30 images of bus from the image database, 27 images out of 30 extracted are of bus. This indicates that 3 images are misclassified as bus. The remaining results for extracting bus are shown in the table 1. The column heads state the number of images to be extracted for the corresponding object. The entries in the table state the number of correctly retrieved images for the object. All the images are extracted in the decreasing order of scores which are found by the SVM classifier.

Table 1 Number of correct images extracted for "Bus" object from a common database of images.

Object	Extract 5 images	Extract 10 images	Extract 30 images	Extract 50 images
Bus	5	10	27	45

C. Classifier Performance

The classifier performance is measured with the help of precision-recall curve. Precision is the fraction of positive predictions which are correct. Recall is the fraction of positive labels that have been correctly classified i.e. recalled. To find precision and recall, we need to compute:

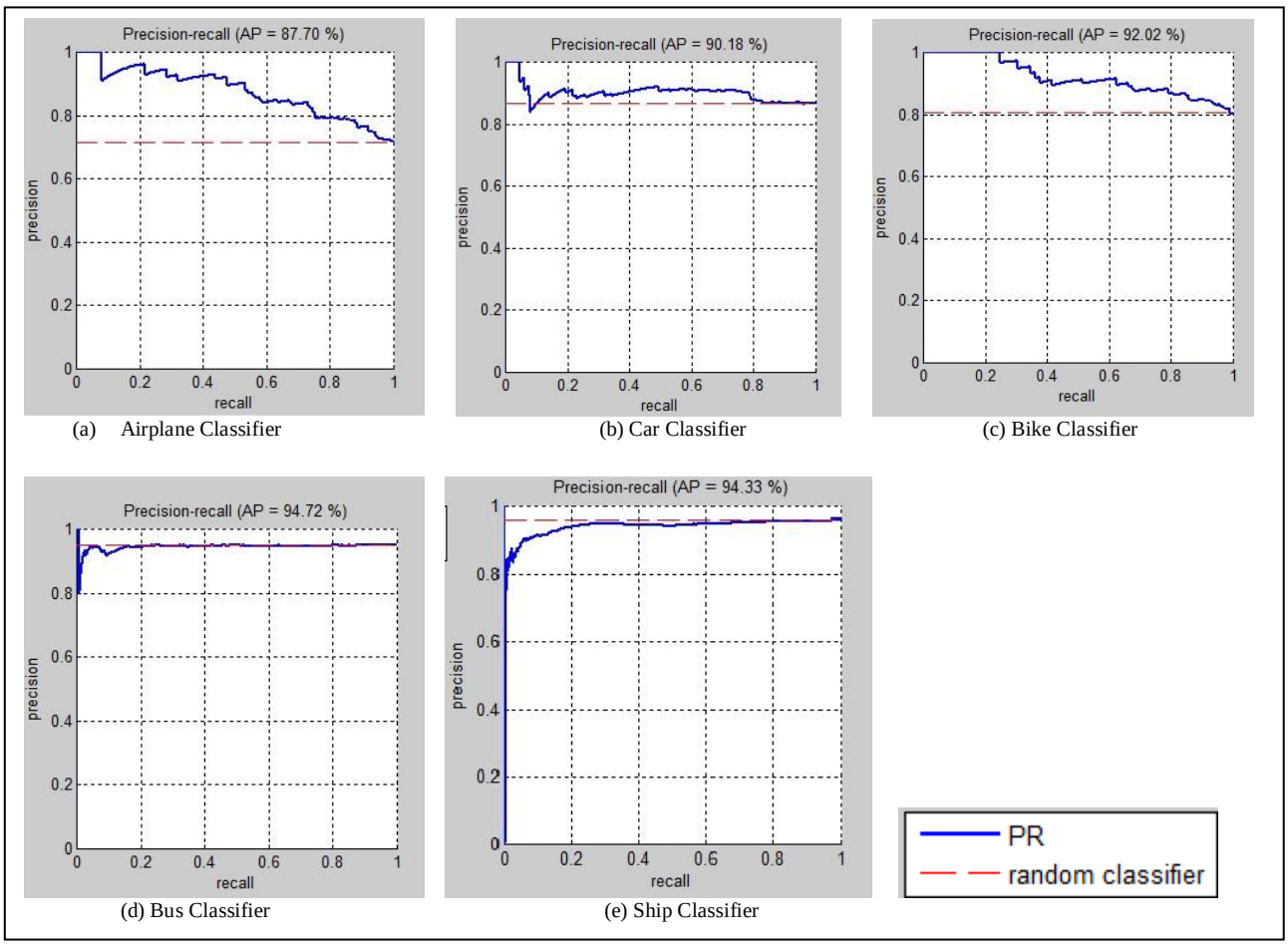
- True positives (TP) – Positive results that are declared as positive
- False positives (FP) – Negative results declared as positive
- Total positives (P) – Total number of positives

Precision and recall can be computed using the following equations:

Precision = TP / (TP+ FP)

Recall = TP / P

The performance for five object classifiers is shown in the fig.11



(a) Airplane Classifier (b) Car Classifier (c) Bike Classifier

(d) Bus Classifier (e) Ship Classifier

VI. CONCLUSION AND FUTURE SCOPE

As per the results, if number of frames is less there is less misclassification and almost all the images containing the particular object are identified. The number of key-frames is generally less; hence the performance of video key frame

annotation is promising. But when the analysis is on large number of images the performance of the system can be improved by increasing the number of positive and negative images. Experimentation shows that bus classifier mostly gets confused with car images, bike classifier mostly gets confused with cycle images, ship images are confused with airplane images since water and sky has similarity. This paper mainly focuses on the video key frame annotation which is performed using the SIFT features and SVM as the classifier. Furthermore, the annotation file created for all the frames are based on ontology which can greatly assist retrieval and browsing. The system is implemented with five object detectors; it can easily be expanded by increasing the detectors. System can also be extended for annotation of other domain videos. Performance of the system can be made error free by increasing the positive and negative images for training. The system can be used for web based or offline searches. Currently the offline search on the computer is based on the file name. Our system can help the content based search of the videos or images within the computer. Let us consider that there is a folder of large number of videos or images on the computer system and we need to find the video of our interest based on its contents, the content based search option can be added. Another application can be in CCTV surveillance videos. If CCTV footage is semantically annotated then the system can respond to semantic queries and retrieve footage based on content. If the footage required is based on the query- "a red car entered the parking area". This may result in many videos, which may be restricted by adding the month or range of days for the action. Such semantic queries can be easily handled with the help of video annotation. The events and color of objects can also be added to the annotation. The possible extension and application areas of the system and video annotation are boundless; with innovation in the related domain, video access time and relevance can be greatly enhanced.

REFERENCES

[1] J. Calic and E. Izquierdo, ", "Efficient Key-frame Extraction and Video Analysis", International Symposium on Information Technology, April 2002, IEEE.

[2] V. Lavrenko, S. L. Feng , R. Manmatha, "Statistical Models for Automatic Video Annotation and Retrieval", In Acoustics, Speech, and Signal Processing. Proceedings (ICASSP'04). IEEE International Conference on, Vol. 3, pp. iii-1044, May- 2004.

[3] L. Ballan, M. Bertini, A. Bimbo, G. Serra, "Video Annotation and Retrieval Using Ontologies and Rule Learning", IEEE MultiMedia, Vol. 17, issue 4, 2010.

[4] K.Khurana, M.B.Chandak, "Key Frame Extraction Methodology for Video Annotation," International Journal of Computer Engineering and Technology, Volume 4, issue 2, pp.221-228, March- April 2013.

[5] R. Lienhart, W. Effelsberg, "Automatic Text Segmentation and Text Recognition for Video Indexing", Multimedia systems, Vol. 8, Issue No. 1, pp. 69 – 81, 2000.

[6] J. Lu, Y Tian, Y. Li, Y. Zhang, Z. Lu, "A Framework for Video Event Detection Using Weighted SVM Classifiers", In Artificial Intelligence and Computational Intelligence, AICI'09. International Conference on, Vol. 4, pp. 255-259, November- 2009.

[7] J. W. Jeong, H. K. Hong, D. H. Lee, "Ontology-Based Automatic Video Annotation Technique in Smart TV Environment", IEEE Transactions on Consumer Electronics, Vol. 57, No. 4, pp. 1830-1836, 2011.

[8] B. Vrusias, D. Makris, J. P. Renno, N. Newbold, K. Ahmad, G. Jones, "A Framework for Ontology Enriched Semantic Annotation Of CCTV Video", In Image Analysis for Multimedia Interactive Services, WIAMIS'07, Eighth International Workshop. IEEE, June 2007.

[9] P. Koletsis, E. G. Petrakis, "SIA: Semantic Image Annotation Using Ontologies And Image Content Analysis", In Image Analysis and Recognition, Springer Berlin Heidelberg, pp. 374-383, 2010. http://dx.doi.org/10.1007/978-3-642-13772-3_38

[10] C. T. Dang, M. Kumar, H. Radha, "Key Frame Extraction from Consumer Videos Using Epitome", In Image Processing (ICIP), 19th IEEE International Conference on. pp. 93-96, September 2012.

[11] K. Khurana, M. B. Chandak, "Key Frame Extraction Methodology for Video Annotation", International Journal of Computer Engineering and Technology, Vol. 4, issue 2, pp.221-228, 2013.

[12] D. G. Lowe, "Object recognition from local scale-invariant features", In Computer vision. Proceedings of the International Conference on Computer Vision. 2, pp. 1150–1157, 1999.

[13] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Vol. 60, issue 2, pp. 91-110, 2004.

[14] T. R.Gruber, "A Translation Approach To Portable Ontology Specifications. Knowledge Acquisition", Vol. 5, no. 2, pp.199-220, 1993.

[15] N. Magesh, P. Thangaraj, "Semantic Image Retrieval Based on Ontology and SPARQL Query", In proceedings of International Journal of Computer Applications (IJCA) – ICACT, Number 1, pp.12-16, August 2011.

[16] VL Fleat: http://www.vlfeat.org/

