Ontology Based Video Annotation for Transportation Domain

N. Agwan .[1], K. Davange .[2], P. Kanade [3], R. Karpe.[4]

[1] Nayana N. Agwan, computer engineering department, SRES COE Kopargaon, Maharashtra, India
[2] Kajal P. Davange, computer engineering department, SRES COE Kopargaon, Maharashtra, India
[3] Poonam G. Kanade, computer engineering department, SRES COE Kopargaon, Maharashtra, India
[4] Rajashree E. Karpe, computer engineering department, SRES COE Kopargaon, Maharashtra, India

ABSTRACT

Video provides more information than other forms. By looking the popularity of video data on the internet tremendous video database is created with every increasing minute so we need a proper approach to handle such database and these problem solved by video annotation. Video contains sequence of frames instead of analyzing all the frames extract those frames which contain more information with the help of canny edge detector and for identification different classifier are trained for different objects and identified object represent in ontology structure.

Keyword: - *Video annotation, domain ontology, Key frame extraction, SIFT feature extraction, classifier.*

---

1. INTRODUCTION

Digital cameras and mobile phone cameras, internet become popular in daily life. It produce numerous videos in various domain like sports ,educational video lectures, CCTV surveillance and transport domain that increases the importance of multimedia video content. Video provides more information than the audio, textual data and images for ease of access need video annotation. Video annotation is the process of extracting and attaching data about data(metadata) which provides important information of video which is used for increasing retrieval speed, easy access, analysis and categorization[1]. Video annotation is imperative technique that assist in video access. Video annotation increases performance of retrieval information, better understanding of video content and reduces human time and efforts. The process of Video annotation starts from key frame extraction[6] . Features of object extracted using SIFT feature and detection of the object is done by training different classifier. In proposed system video annotation is based on ontology[3]. It avoids the ambiguity in the result of classification. Different object detectors are used for the objects like car, bus, airplane[8]. In the proposed system transportation domain videos are used and system can be extended to other type of domain videos.

2. PROPOSED SYSTEM

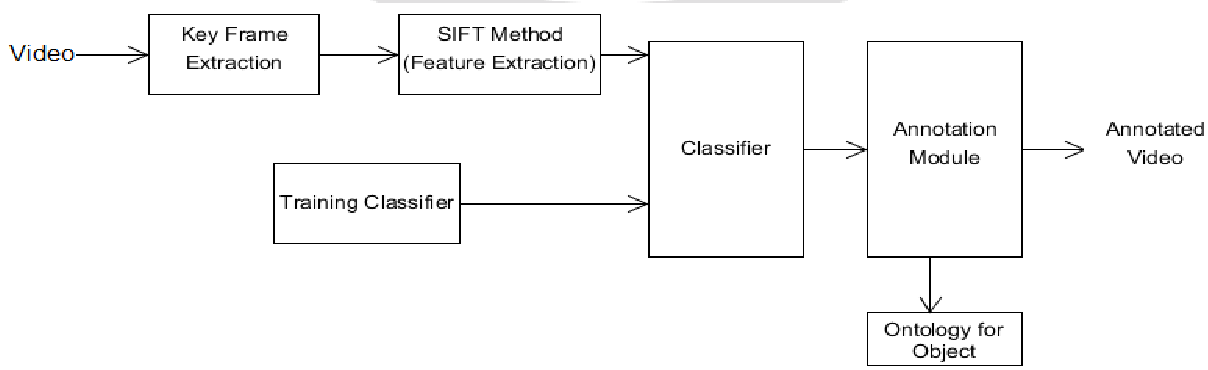


Fig -1: Proposed System for Ontology Based Video Annotation for Transportation Domain

The proposed system broadly classified into three main phases:-

2.1 To extract key frames from the video.

Video contains sequence of frames. All these frames do not contain important information of video, some of these frames best represents the visual scene of the video. Key frames are extracted by using Canny edge detector algorithm[2]:

Input to the Canny edge detector algorithm is, Video V, consisting of N frames.

Output of the Canny edge detector algorithm is, Key frames for input video.

Algorithm Key frame extraction

{

Step 1:

For each video frame k = 1 to N-1

{

1. Read frame V_k and V_{K+1}
2. Obtain gray level image for V_k and V_{K+1}

G_k = Gray image of V_k

G_{K+1} = Gray image of V_{K+1}

3. Find the edge difference between G_k and G_{K+1} using Canny edge detector.

Let diff (k) be their difference

$$\text{diff}(k) = \sum_i \sum_j (G_k - G_{K+1})$$

where i , j are raw and column index

}

Step 2:

Compute mean and standard deviation

$$\text{Mean} = M = \frac{\sum_{i=1}^{N-1} \text{diff(k)}}{N-1}$$

$$\text{Standard deviation, } S = \frac{\sqrt{\sum_{i=1}^{N-1} (\text{diff}-\text{M})^2}}{(\text{N}-1)}$$

Step 3:

Compute the threshold value

Threshold = M + a × S

Where a is constant

Step 4:

Find the Key frames

For K=1 to (N - 1)

{

If diff (k) > Threshold

{

Write frame V_{K+1} as the output key frame

}

}

}

Key frames are depends on changes in the content of video. If there is more changes in the video content then all the frames are consider as key frames and if less changes then less key frames are found.

2.2 Feature extraction, training classifier and testing the object present in key frames.

A] Compute SIFT feature[4] from all key frames.

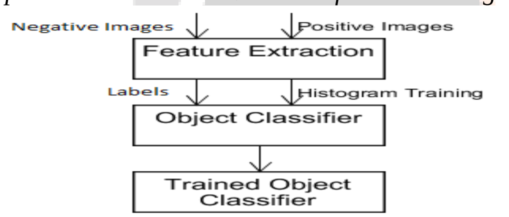
Scale Invariant Feature Transform:-

All SIFT algorithm proposed by Lowe [4] have the major stages of computation used to generate the set of image features:

1) Scale-space extrema detection: The first stage of computation searches over all scales image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.
2) Keypoint localization: At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.
3) Orientation assignment: One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.
4) Keypoint descriptor: The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination [5].

B] Training of different object classifier using SIFT features.

Video being reduce to key frames and extraction of SIFT feature from the key frames then next step is to train the different object classifier for different objects.
In the proposed system we use the method of machine learning that is first training the classifier for different objects then testing. For actual classification feature are calculated for the test images which are given as input to the trained classifier. Support Vector Machine can be used for the classification. Here we consider SVM [7] as classifier, it is supervised technique which classifies into two classes. In our proposed system, SVM classifies as positive and negative, i.e. object is present or absent. Support Vector Machine is trained using following algorithm [9]:
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



.

Fig. 2 Training classifier

Fig. 2 shows training procedure for single object classifier. The SIFT feature are extracted for both positive and negative image from those feature histogram is constructed. Constructed histogram and labels are given as input to the SVM [7],[9] which results into the train object classifier. We need to train different classifiers for different objects.
C] Test the object present in the key frame.
Once training is completed classifier is ready for object identification [8]. Following algorithm is used for detection of object in the key frames.
Algorithm DetectObject
{
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 shows the flowchart representation of algorithm.

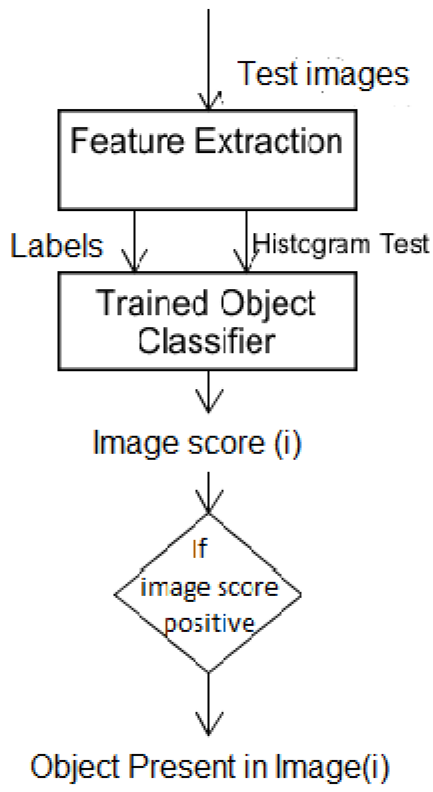


Fig.3 Testing the classifier or Detecting the object.

For the test image SIFT features [3] are extracted and histogram is constructed. Both histogram and labels are given to the train object classifier which generates the score (i) corresponding to i number of input images. If the score is positive then object is present else object is absent.

Only if the object is present the image is pass to the annotation module for adding the object annotation.

2.3 Annotation of key frames based on ontology.

Ontology consist of entities and their relationships, which are organized hierarchically [3]. It may be in the form of classes and subclasses where each class consist of one or more instances.

Following Fig.4 shows the transportation domain ontology.

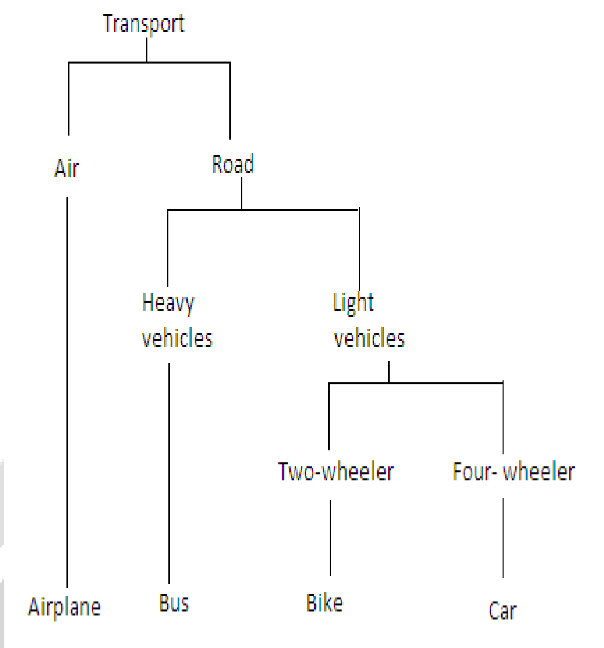


Fig.4 Transport domain ontology.

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 detected object is car then ontology for the car is given as main domain transport, sub domain is road, class is light vehicles, subclass is four-wheeler and object is car.

3. CONCLUSIONS

The above proposed system input is given as transportation video then key frames are extracted from video this frames are given to the feature extraction module and feature will be extracted by SIFT and classify Using SVM trained classifier. Annotation file is created for each object in video which help us for content based retrieval.

REFERENCES

[1]. Khushboo Khurana , Dr.M.B.Chandak "Video Annotation Methodology Based on Ontology for Transportation Domain ", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, June 2013.

[2]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.

[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] 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.

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

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

[7] Durgesh K. Srivastava , Lekha Bhambhu "Data Classifiaction Using Support Vector Machine ", Journal of Theoretical and Applied Information Technology.

[8]Archana Nagendran1, Naveena Dheivasenathipathy2,Ritika V. Nair3 and Varsha Sharma4 "Recognition And Tracking Moving Objects Using Moving Camera In Complex Scenes", International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.4, No.2, April 2014.

9] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin,"A Practical Guide to Support Vector Classification",Department of Computer Science National Taiwan University, Taipei 106, Taiwan http://www.csie.ntu.edu.tw/~cjlinInitial version: 2003 Last updated: April 15, 2010