iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** V **Month of publication:** May 2016

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

An Efficient Video Search Engine

Sachin S. Desai[1], Asst.Prof.Nayana Shenvi[2]
Goa College of Engineering

***Abstract*—Now a days with the advancement of internet and readily available tools for creating videos it is an important fact that users of internet more specifically users of social media must have some tools for retrieving information from the videos those are easily spread. Not only for the users of social media but also for video forensics is an efficient content based video search engine in high demand. In this paper an efficient video search engine is proposed based on SIFT(Scale Invariant Feature Transform) features.**

***Key words*—SIFT, Feature Extraction, Content Based**

I. INTRODUCTION

Now a days with the advancement of internet and readily available tools for creating videos it is an important fact that users of internet more specifically users of social media must have some tools for retrieving information from videos those are easily spread. Not only for the users of social media or internet but also for video forensics, such as to extract some information from a crime scene captured from some video camera installed anywhere.

In this context having an efficient content based video retrieval system is a must. In a content based video retrieval system some key frames from the video is to be extracted and within the video or within any database of other videos that key frame should be searched based on some robust feature extraction and feature matching techniques. Work has been done before for content based video retrieval . Yarmohammadi, H and others [1] uses the information theory based technique for content based video retrieval . The authors uses Shot Boundery Detection , Key Frame Extraction , and Video Indexing. Dyana, A and others [2] show that combining features for shape and motion trajectory of video objects works well for content based video retrieval . The authors also use CSS based shape representation and trajectory based motion representation. Previously Asha, S. and others [3] proposes SURF based content based video retrieval technique .B.V Patel [4] video retrieval of Near-Duplicates using K-NN retrieval of Spatio-Temporal descriptors describes a novel methodology for implementing video search functions such as retrieval of near-duplicate videos and recognition of actions in surveillance video.

An efficient video search-engine or content based video retrieval method is proposed based on the SIFT(**Scale Invariant Feature Transform**) features . This technique tries to extract a matched frame from an input video or from a database of images .Firstly the SIFT features are extracted from the selected input frame after that based on those SIFT features the matching is done to retrieve similar frames from another video or from a database of images.

The rest of the paper is organized like this ,the next section describes SIFT feature descriptor , describes the method for feature extraction . Feature matching technique is given in section 3 .Section 4 describes the system flow chart , Finally a conclusion is given .

II. SIFT

David Lowe proposed Scale Invariant Feature Transform which is a digital image descriptor for matching and recognition of digital image. The SIFT descriptors are of 128 descriptor vectors which are mainly used in point matching between different views of a 3-D scene and object recognition in computer vision. These descriptor are robust to rotation, scaling, translation transformation in the image and also robust to illumination variations. Therefore SIFT are useful for image matching and recognition under real world environment.

The SIFT descriptor consist of detecting interest point from the image which is grey level image for which local gradient direction of the image intensities are put together to give a brief description of the image structure in a local neighborhood around the key point, which should be used for matching corresponding key points between the different image. Then the SIFT descriptor is applied at the dense grids which are shown to cause the better performance for object categorization, texture classification, image alignment and biometrics. The SIFT descriptor can also be used for colour image and 2+ I-D spatio-temporal video. An overview of the algorithm is presented here. There are mainly four major stages of computation involved in SIFT algorithm.

A. Scaled-space Extrema Detection

©IJRASET: All Rights are Reserved

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Scaled-space Extrema Detection is the first step that searches over all scales and image locations. It uses a difference-of- Gaussian function for identification of potential interest points which are invariant to scale and orientation. Laplacian of Gaussian (LoG) is calculated for the image with various (J values which acts as a blob detector and detect blobs of different sizes with the change in (J. Gaussian kernel with low (J will give a high value for small corner in an image while Gaussian kernel with high (J fits well for broader corner. Thus, it will find the local maxima across the different scale and space which will again give a vector of (x,y,(J) values. These mean that at (J scale there is a potential keypoint at (x,y) which are shown in equation 1

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \qquad (1)$$

Where L is the blurred image with

where L is the blurred image with σ amount of blur, $G = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}}$ is the Gaussian Blur operator , $I(x,y)$ is the pixel at row x and column y of the image I, and $*$ is the two dimensional convolution operator in x and y. If the amount of blur in a particular image is $k\sigma$, then the amount of blur in next level will be $2^{\frac{1}{\text{Number of blurred images}+1}}$, is a constant, which is given by . LoG is little

costly compared to Difference of Gaussians, therefore SIFT uses DoG that are the approximation of LoG. DoG are computed from the difference of Gaussian blurring of an image with two adjacent value of σ, let it be σ and kσ as shown in equation 2

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma)$$

where $L(x,y,k\sigma)$ and $L(x,y,\sigma)$ are the blurred image with blur amount $k\sigma$ and σ respectively

For different octaves of the image, this process is performed in Gaussian Pyramid as shown ion figure 1.

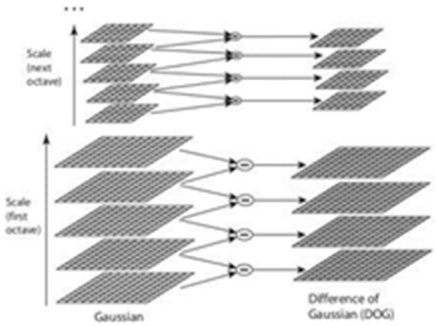


Fig 1: Difference of Gaussian (DOG

Then the local extrema over scale and space are searched on the image. For example one pixel value is compared to its 8 neighbours as well as 9 pixels in next scale and 9 pixels in previous scales and so on.. For every pixel with spatial

location x and y in image I', the pixel with spatial location

©IJRASET: All Rights are Reserved

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

$(x-1\,y-1),(x-1,y),(x-1,y+1),(x,y-1),(x,y+1),(x+1,y-1),(x+\ 1,\ y)$ and $(x+1,y+1)$

in the current image and consecutive scale images and I''' are the neighbouring pixels along with positions $(x.y)$ in I'' and I'''. If any one of the pixel value is the greatest of all the 26 neighbours, then the pixel is considered as the maxima point, and if any of the pixel value is the least of all the neighbours, the pixel is minima point. All these maxima and minima points are considered as candidate keys. Potential keypoint is identified if it is a local extrema which basically means that keypoint is best represented in that scale which is shown in figure 4: Normally, the value of different parameter that are used in this technique are given as number of different octaves = 4, number of scale levels value = 5, initial σ value =1.6, etc. as feasible values..

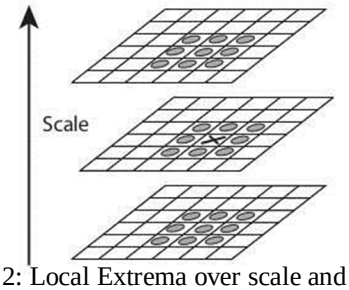


Fig 2: Local Extrema over scale and space

B. 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. After the potential keypoint location is calculated, it has to be refined to obtain more accurate results. For that Taylor series expansion of scale space is used to obtain more correct location of extrema. The keypoint is rejected if the intensity of this extrema ies less than a predefined threshold value (normally 0.003). Edges also need to be removed as DoG has higher response for edges which uses Harris corner detector is used. For computation of principle curvature, it uses a 2x2 Hessian matrix (H). For every candidate

key-point $p(i,j)$ at coordinate (i,j), the Hessian matrix is calculated as follows:

$$H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \qquad (3)$$

$h_{11}\ h_{12}\ h_{21}\ h_{22}$

$$h_{11} = p(i+1,j) + p(i-1,j) - 2*p(i,j)$$
$$h_{12} = h_{21} = p(i+1,j) + p(i,j-1) - 2*p(i,j)$$
$$h_{22} = (p(i+1,j+1) - p(i+1,j-1) - p(i-1,j+1) + p(i-1,j-1))/4$$

$$\frac{(h_{11}+h_{22})^2}{(h_{11}*h_{22})-(h_{12})^2} < \frac{(C_{edge}+1)^2}{C_{edge}}$$

, then retain the key-point,otherwise discard it. where C_{edge} is the ratio between the largest and non-zero smallest eigen-values in the block of the image. From Harris corner it is known that for edges, one value is larger than the other. The keypoint is rejected if this ratio is higher than a threshold. Usually it uses 10 as a threshold. This step eliminates the low contrast keypoint and edge keypoint and the only accurate keypoint is obtained.

C. Orientation Assignment

Based on local gradient directions one or more orientations are assigned to each key point location . All the image operation that has been transformed are performed relative to the assigned orientation , scale , and location for each feature which provide invariance to these transforms . In order to achieve invariance to image rotation an orientation is assigned to each key point . Depending on the scale , a neighborhood point is chosen around the key point location . Then for that region the gradient magnitude and direction is calculated . After that it is the creation of orientation histogram with 36 bins covering 360 degree . It is usually weighted by gradient magnitude and Gaussian weighted circular window with σ equal to 1.5 times the scale of key point . Then in

©IJRASET: All Rights are Reserved

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

the histogram the highest pick is chosen and any peak above 80% of it is also taken for calculating the orientation which creates a key pint with same location and scale . They also contribute to the stability of matching.

D. Key Point Descriptor

At the selected scale, the local image gradients are measured in the region around each keypoint. They are converted into a representation that also works for significant levels of local shape distortion and illumination.

Keypoint descriptors are now obtained. A 16x16 neighborhood around the keypoint is selected and divided into 16 sub-blocks of 4x4 size block. For every sub-block, 8 bin orientation histogram is created which lead to a 128 bin value which also represents a vector forming keypoint descriptor. There are several techniques which are meant to obtain robustness to illumination, rotation, noise etc.

III. FEATURE MATCHING

By identifying their nearest neighbors, keypoint between similar images are matched. It can also be happen that the second closest-match may be very near to the first which may due to noise and some other factors. For this case, ration of closest-distance to second -closest distance is obtained. They are rejected if it is greater than 0.8. These steps discarded almost 90% of false matches and nearly only 5% correct matches are discarded ..

IV. THE PROPOSED METHOD

In the current scope of work the work flow goes like this , the user selects a frame of interest from the input online video stream by selecting a key from the keyboard . After that the selected frame is treated as an input image to the system, then SIFT features are extracted from the frame/image , and finally the features are matched with a database of images and after matching the best matched image is fetched from the database . Figure 3 describes the overall system flow chart .

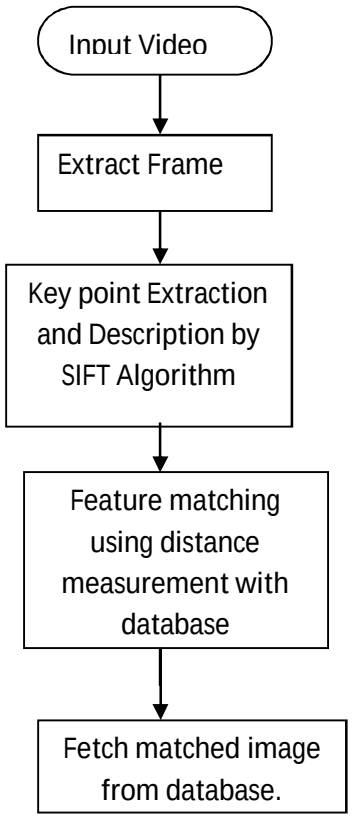


Fig 3: The overall system flowchart

V. CONCLUSION

A fast and simple method of content based video retrieval is presented . In this method the robust image/frame features are extracted

©IJRASET: All Rights are Reserved

using the robust algorithm named as SIFT . Then the SIFT feature descriptors are used for matching the input image/video frame with a database of videos.

REFERENCES

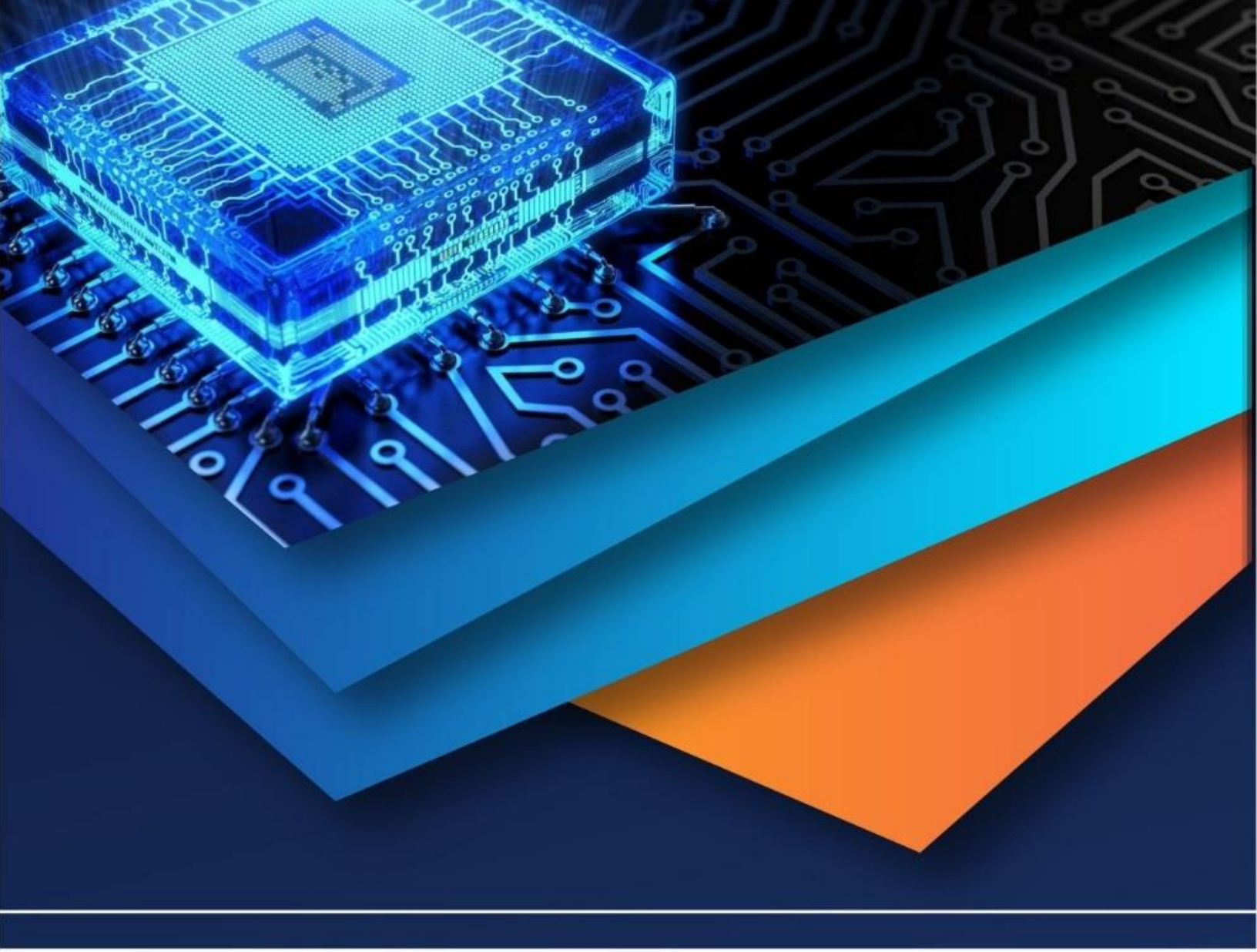
[1] Yarmohammadi, H.; Rahmati, M.; Khadivi, S., "Content based video retrieval using information theory," Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on , vol., no., pp.214,218, 10 - 12 Sept. 2013

[2] Dyana, A.; Subramanian, M.P.; Das, S., "Combining Features for Shape and Motion Trajectory of Video Objects for Efficient Content Based Video Retrieval," Advances in Pattern Recognition, 2009. ICAPR '09. Seventh International Conference on , vol., no., pp.113,116, 4-6 Feb. 2009

[3] Asha, S.; Sreeraj, M., "Content Based Video Retrieval Using SURF Descriptor," Advances in Computing and Communications (ICACC), 2013 Third International Conference on , vol., no., pp.212,215, 29-31 Aug. 2013

[4] B V Patel and B B Meshram "Content Based Video Retrieval Systems " , International Journal Of UbiComp(IJU) , Vol.3 , No.2 , April2012

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)