

Text Region Extraction from Punjabi News Videos using Feature Analysis

Pooja¹, Renu Dhir²

Dr. B. R. Ambedker National Institute of Technology, Jalandhar
poojachoudhary80@gmail.com¹, dhirr@nitj.ac.in²

Abstract: This paper presents a text region extraction approach based on feature analysis. This scheme consists of 5 main phases that are Frame Extraction, Edge Detection and Binarization, Fusion, Normalization and Feature Extraction. For Frame extraction MATLAB simulator is used where video reader is used to extract frames from the video according to the length of video. Edge detection and Binarization is processed on that extracted frames then both the detected edges and binarized image will be fused using PCA based fusion method. Fused image is then normalize and feature extraction is perform by using SIFT and SURF based algorithm. The results will show that extracted text region from the frames of the news videos.

Keywords: Text Region Extraction, Sobel, Otsu Threshold, PCA, SIFT, SURF.

1 Introduction

Automatic extraction of descriptive features that is relevant to the subject materials such as images, video, etc. are required for the content-based multimedia database indexing and retrieval tasks [Elagouni, 12]. Some of the typical low-level features that are extracted in the form of images and video include measures of shape, colour or texture.

Although these features can easily be obtained but they do not give a precise idea of the image content. Text, human faces and other descriptive features are the high level entities that needs to be extracted and has recently attracted research interest significantly [Saluja, 15]. Also text embedded in images and video, especially the captions provide important and brief content information, such as the title, location, date and time of an event, information about the speakers etc which can be a keyword resource as powerful as the information provided by speech recognizers. Besides, text-based search has been successfully applied in many applications. The robustness and computation cost of feature matching algorithms based on other high-level features is not efficient enough to be applied to large databases.

Text detection and recognition [Chen, 04] in images and video frames aims at integrating advanced optical character recognition and text-based searching technologies. It is now recognized as a main component in the development of advanced image and video annotation and retrieval systems.

The major part of any video or a picture is usually in the form of text. Complex or non complex picture portion give a reasonable outline of that picture.

The procedure for edge detection is explained below:

- a) Install text in a picture or by externally including a text in the video or the picture. But as a matter of fact, improvement or extraction of text is a perplexing procedure in dynamic videos.
- b) Hence now the complex dynamic video is subdivided further into few static frames.
- c) In case of edges that are colored, the frames are at first changed over to binary images. This is done to decrease the complex quality of the image .
- d) To expel the noisy substance from the image, several Gaussian filters are applied consistently to the whole frame set. Morphological operations are then applied to blend the broken areas.
- e) The procedure of edge detection is now finished. Bounding boxes are now produced utilizing the identified edges around both textual and non textual regions.
- f) The bounding boxes are classified into textual and non textual bounding boxes on the basis of aspect ratio, no of pixels in a bounding box etc.
- g) The bounding boxes that are retained are those that fulfil the textual content criteria.

Once the bounding boxes are identified, remaining region is flooded using the image in painting algorithm in such a way, that it recoups the lost or adulterated parts of an image so that the reproduced image looks natural.

There are 2 main applications of edge detection:

1. Content based image retrieval
2. Vehicle number plate detection

There are several loopholes in text extraction from Video content. These are:

- a) Erroneous frames are generated when rolling text in the videos.
- b) Frames that are generated sometimes include garbage data in the form of dots, lines, text, image(s) etc.
- c) Backgrounds in the video are either non uniform, textured, includes text or the background in the same or different colour comprises of rounded and vertical text.
- d) When mixing of colour and reverse video is done, non uniform background is created.
- e) Videos might contain handwritten text of any type.
- f) Fluctuating lines of text and low quality images sometimes do not easily bina-rized.
- g) Overlapping of image and textual content occurs.

1.1 Feature Extraction

Feature extraction is associated to degradation of dimensions. It is a process of transforming the input into downgrading of feature's set, when the input to be processed is too huge to an algorithm. Also, this process is used when there is suspicion of redundancy, like repeating of images shown as pixels, or similar measurements in meters and feet both.

Feature extraction begins with an initial set of measured data and creates features, i.e. derived values which are intended to be non-redundant and informative too, aiding the subsequent learning and generalization steps, and leading to improved interpretations by human [Jiang, 09]. It is expected that the extracted features comprises of the admissible information from the data, so that the desired function can be performed by applying this representation, rather than that of the entire initial data.

Feature extraction generally comprises of reduction in the amount of resources needed to present large amount of data (set) [Vamvakas, 08]. Most of the time, a problem arise from the involvement of variable number, when operating analysis of complex data. Generally, an

ample amount of memory space and computing power is required while analysing with huge number of variables. Also, it might cause a classification algorithm to misfit to training samples and discern badly to new samples. Feature extraction is a broad word for practice of creating mixture of variables to sidestep these obstacles while still expressing the data with plentiful veracity.

In character recognition system, the pre-processing phase is followed by feature extraction. The basic concept of pattern recognition is to accept a pattern as input and assign it as one of the likely output classes. This method is divided into two steps:

- Feature selection
- Classification

Feature Selection is essential to the complete process because the classifier would be incapable of recognizing from inadequately selected features. According to Lippman, there is a certain criteria by which features are chosen, that is: “Features are those that comprises of information required to differentiate between classes, should not be sensitive to irrelevant variability in the input, and also be limited in number, to permit, efficient computation of discriminate functions and to restrict the amount of training data required”

Feature extraction is a vital stride in the development of any pattern classification and goes for the extraction of the pertinent data that describes every class. In this procedure significant features are separated from items/alphabets in order to shape feature vectors. These feature vectors are then utilized by classifiers to perceive the input section with target yield section. It becomes less demanding for the classifier to group between various classes by taking a gander at these features as it allows genuinely simple to recognize. Feature extraction is the procedure to recover the most essential information from the crude data. Feature extraction is finding the arrangement of parameter that characterizes the state of a character decisively and remarkably. In feature extraction stage, every character is presented by a feature vector, which turns into its identity. The significant objective of feature extraction is to separate an arrangement of features, which boosts the acknowledgment rate with minimal measure of elements and to produce comparative list of features for assortment of instance of the same symbol. The generally utilized feature extraction techniques are:

- Template coordinating
- Deformable layouts
- Unitary Image conversion
- Graph portrayal
- Projection Histograms
- Contour profiles
- Geometric moment invariants
- Gabor countenance
- Zoning
- Spline bend estimation
- Fourier descriptors
- Gradient features and
- Zernike Moments.

Importance of feature extraction:

At the point when the pre-processing and the desired level of division of a line, a word, a character or an image has been accomplished, some feature extraction approach is connected

to the fragments to get features which is trailed by utilization of grouping and post processing methods [Kumar, 14]. It is important to concentrate on the component extraction stage as it observably affects the proficiency of the acknowledgment framework. Feature selection of a feature extraction technique is the absolute most essential element in accomplishing high acknowledgment execution. "extraction of unprocessed data that would be the most suitable one for grouping purposes, while downsizing the inside of class pattern variability and improving the between class pattern variability"- Feature extraction could be defined like this. Along with this, choice of a relevant feature extraction system as per the data to be connected should be finished with most extreme consideration. Thinking about every one of these variables, it gets to be vital to take a gander at the different accessible procedures for feature extraction in a given space, covering unfathomable conceivable outcomes of cases. This work is to extract the text region using following steps which will be discussed in the next sections:

Step 1- Frame Extraction

Step 2- Edge Detection & Binarization

Step 3- Fusion

Step 4- Normalization

Step 5- Feature Extraction

2 Frame Extraction

As the quick advancement of system and interactive media innovation, there are ample of videos on the Internet and the sharing of sight and sound pictures (multimedia), audios and videos information has turned out to be progressively prevalent. Among all these online information streams, video data is the most entangled information. The issue of how to browse the huge volume of video is critical to be settled out. Key frame is the exceptionally illustrative frame in the sequel of video frames, which can be utilized to depict the key picture of a video shot; it mirrors the primary substance of a video shot or even an entire video well and rightly. The extraction of key frames has turned out to be so well known in video streams' investigation, recovery and enquiry that it draws in more consideration of analysts and researchers [Calic, 02].

Some great key frame extraction strategies are recorded as follows:

- a) **Based on Video shot Method:-** The video shot based technique is the easily accessible and fastest extraction strategy. Some enhanced exemplary techniques include frame average technique and histogram average technique. By techniques, Zhang et al. proposed to make utilization of color histogram to excerpt key frames. Yeung et al. juiced out key frames by method for figuring the utmost separation of the feature space. These strategies have the upside of basic and low calculation complexity, and the frames extracted have regular classical context. It is great at the camera of low substance movement or still. Nonetheless, in these strategies, the limitations are the unpredictability (complexity) of substance in the present video shot is not considered and the quantity of key frames is constrained as a fixed value, and the motion content in the video shot can't be productively portrayed.
- b) **Based on Content Analysis Method:-** This strategy is to concentrate key frames taking into account the change of colors, texture and other visual data of every frame, when this data changes essentially, the key frame is the current frame. This technique is extremely straightforward and can choose relating key frames as indicated by the change level of content in the video shot [Wattanasachothai, 15]. But its shortcoming is that the key frames extricated are not generally the most illustrative connotation one and can't show the progressions of motion information quantitatively, hence which will bring about unsteady key frame extraction, for instance, too much extraction of frames when meet continuous camera or mass of content movement.
- c) **Motion-based Analysis Method:-** One of the motion-based algorithms is Motion minimal algorithm. By investigating the optical stream, Wolf ascertains the movement of a shot, and chooses key frames from the local minimum in the movement [Liu, 10]. Wolf's motion based technique can choose the suitable number of key frames on the premise of the shot structure. The disservice is that the algorithm has huge calculation and don't have solid power since it relies on upon local data, and it don't give careful consideration to the context changes brought by the cumulative dynamic either.
- d) **Cluster-based Method:-** It is among the most effective algorithm to the principle shot content as the cluster examination strategy was enforced to the key frame

extraction. Zhang and Hanjalic endorsed cluster splitting algorithm with analysis of efficiency of cluster; the key frame is the one which is most nearly to the center of cluster. Joshi et al. implemented the fuzzy cluster technique in a short video sequel with just gradient video shot, the frames in the middle will be chosen as key frames for every video shot. The main hindrance is the difficulty to acquire general cluster parameters for cluster algorithm.

Key frame is the frame which can present the striking content and data of the shot [Liu, 12]. The foundation of video analysis and content based video recovery are Key frame extraction. It is a crucial part in study of video and management, giving a suitable video outline to video indexing, searching and recovery. The utilization of key frames downsizes the information required in video indexing and gives the structure to manage the video content. Lately, numerous calculations of key frame extraction are concentrated on authentic video stream. In this way it can present inefficient processing and complexity in computation when there is requirement of decompression prior to video processing. The key frames extricated must outline the attributes of the video, and the picture qualities of a video can be followed by all the key frames in sequence of time. Moreover, the video content can be perceived. The key frame extraction has an essential principle, that is, the key frame extraction would preferably be wrong than insufficient. So it is important to toss the frames with tedious or repetitive data amid the extraction.

In consideration of key frame extraction, assuming to be an essential part in video recovery and video indexing, a great deal of analysis and study has been done on the methods. The generally utilized key frame extraction procedures are as per the following:

- Key frame extraction in view of shot movement. Huang, along with Gresle processed the intra and reference histograms and afterward figured an activity indicator. The local minima are chosen as the key frames in view of the activity curve.
- Key frame extraction established on macro block statistical feature of MPEG video stream. Ebroul&Janko generate the frame contrast metrics by dissecting insights of the macro block features, which were extracted from the MPEG compacted stream. The key frame extraction technique is actualized by utilizing contrast measurements of curve simplification by algorithm of discrete contour evolution.
- Key frame extraction on the basis of motion analysis. Wolf processed the optical stream for every frame and after that utilized simple motion measurements to assess the adjustments in the optical stream along the grouping. Where the metric as a component of time has its local minima, in those spots key frames are found.

In this work, various sample videos from the Punjabi news channel has been collected and their frames has been extracted. Frames have been extracted using default video reader in MATLAB simulator. On the basis of duration of the video, video reader generates the frames. For example, A sample video of size 32.7 MB and duration of 4:53min was taken. The total number of frames extracted was 8785. Some of the frames are as shown in Fig. 1.



Fig. 1. Frames of the Sample Video (.avi)

3 Edge Detection & Binarization for Text Region

A numerical technique which distinguishes points in a computerized image at which the image brightness changes very sharply or has discontinuities is termed as Edge Detection. An arrangement of curved line fragments termed as edges are formed of the points at which image brightness changes sharply. Likewise we characterize step detection as discovering discontinuities in 1D signals and the issue of discovering signal discontinuities that occur over time is known as change detection. A basic tool in image processing, machine and computer vision is Edge Detection. It is especially utilized in the areas of feature detection and feature extraction.

We capture imperative occasions and changes in properties of the world for identifying sharp changes in image brightness. Under rather broad suspicions for an image formation model, it can be demonstrated that fundamental discontinuities in image brightness are liable to relate to depth discontinuities, surface orientation discontinuities, changes in material properties and scene illumination variations [Gao, 10].

The result of applying an edge detector to an image might prompt a set of associated curves that demonstrate the boundaries of objects, the limits of surface markings and additionally bends that relate to discontinuities in surface orientation. Therefore by applying an edge recognition algorithm technique to an image, it might essentially lessen the measure of data to be processed. It might therefore preserve the critical basic properties of an image while sifting through the data that might be viewed as less pertinent. The success of edge detection step prompts the resulting task of translating the information contents in the original image. But as a matter of fact, it is dependably impractical to get such perfect edges from real life images of moderate complexity.

Fragmentation is characterized as the procedure that hampers the edges extracted from non-insignificant images, implying that the edge curves are not associated. There are few missing edge segments and false edges that do not correspond to intriguing phenomena in the image. Therefore it confounds the task of translating the image information.

Edge detection accounts to be one of the fundamental steps in the phenomena like image analysis, image processing, image pattern recognition, and computer vision techniques etc.

Sobel Edge Detection:- Edge recognition strategies mostly deal with the supposition that the edge happens where there is brokenness in the function intensity or an extremely steep gradient intensity in the picture. Utilizing this suspicion, if the intensity derivative value is taken over the picture and discover focuses where the subsidiary is most extreme, then the edge could be found [Vincent, 09]. The slope is a vector, whose segments measure how fast values of pixels are changing with separation in the x and y heading. Along these lines, the segments of the angle might be discovered utilizing the accompanying estimate:

$$(\partial f(x,y))/\partial x = \Delta x = (f(x+dx,y) - f(x,y))/dx \quad (1)$$

$$(\partial f(x,y))/\partial y = \Delta y = (f(x,y+dy) - f(x,y))/dy \quad (2)$$

where dx and dy measure separation along the x and y bearings individually. In discrete pictures, one can consider dx and dy as far as quantities of pixel between two focuses. dx = dy = 1 (pixel dispersing) is the time when pixel organizes are (i, j) therefore,

$$\Delta x = f(i+1,j) - f(i,j) \quad (3)$$

$$\Delta y = f(i,j+1) - f(i,j) \quad (4)$$

Keeping in mind the end goal to distinguish the vicinity of an inclination irregularity, one could compute the adjustment in the angle at (i, j). This should be possible by finding the accompanying greatness measure

$$M = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad (5)$$

and the gradient direction θ is given by

$$\theta = \tan^{-1}(\Delta y / \Delta x) \quad (6)$$

Pseudo-codes for Sobel edge recognition strategy

- Data inserted: A Sample Image
- Yield: Detected Edges
- Step 1: Intake the data picture
- Step 2: Implying mask Gx,Gy to the data picture
- Step 3: implementing Sobel edge recognition algorithm and the angle
- Step 4: Masks control of Gx,Gy independently on the data picture
- Step 5: Results consolidated to locate the total size of the angle
- $|G| = \sqrt{(G_x^2 + G_y^2)}$ (7)
- Step 6: the total size is the yield edges

For edge detection, sobel edge detection method is used. Edge detection of the extracted frame is implemented using both DotNet and MATLAB simulator as shown in fig. 2 (left) and (right) respectively.



Fig. 2. Edge Detection using DotNet (left) and MATLAB (right)

Binarization:- Binarization is the procedure of changing over a pixel picture to a binary picture. A digital picture that has just two conceivable qualities for every pixel, is a binary picture. Commonly, the two hues utilized for a binary picture are high contrast; however any two hues can be utilized. The shading utilized for the object(s) in the picture is the frontal area shading while whatever remains of the picture is the foundation shading. Double pictures are additionally called bi-level or two-level [Zhi-Yong, 12]. This implies every pixel is put away as a solitary piece—i.e., a 0 or 1. The names, black-and-white, B&W, monochrome or monochromatic are regularly utilized for this idea, however might likewise assign any pictures that have one and only specimen for every pixel, for example, greyscale pictures. Binary pictures frequently emerge in advanced picture handling as veils or as the consequence of specific operations, for example, segmentation, thresholding, and dithering. A binary picture can be put away in memory as a bitmap, a stuffed cluster of bits. Around 37.5 KB of memory space is required for a 640×480 image. As a result of the little size of the picture records, fax machine and report administration arrangements for the most part utilize this configuration. Most parallel pictures additionally pack well with basic run-length pressure plans.

Otsu's strategy, is utilized to naturally perform clustering based picture thresholding, or, the attenuation of a gray-levelled picture to a binary picture. The expectation of the algorithm is that the image comprises of two classes of pixels taking after bi-modular histogram, that are foundation pixels and frontal are a pixels, it then ascertains the ideal threshold isolating the two classes so that their joined spread, namely, intra-class difference is insignificant, or proportionately on the grounds that the whole of pairwise squared distance is consistent, so that there is maximum inter class change [Han, 10].

Algorithm:

- Calculate probabilities of every intensity level and histogram.
- Set up beginning as $w_i(0)$ and $\mu_i(0)$.
- Step through every single conceivable thresholds $t=1$, maximal intensity
 - Update w_i and μ_i .
 - Evaluate $\sigma_b^2(t)$.
- The desiring threshold compares to the maximum $\sigma_b^2(t)$.

For binarization, otsu based threshold method is used. Binarization of the extracted frame is implemented using both dotNet and MATLAB simulator as shown in fig. 3 (left) and (right) respectively.



Fig. 3. Binarization using Otsu method in DotNet (left) and MATLAB (right)

4 Fusion and Normalization

The procedure of consolidating significant data from two or more images into a solitary image is termed as Image Fusion. Out of the majority of the input images, the subsequent image of "Image Fusion" will be the most useful and informative. There are different techniques that exist to perform Image Fusion. The High pass filtering strategy is the fundamental one. Other techniques are Discrete Wavelet Transform, uniform rational filter bank, and Laplacian pyramid. Data Fusion has turned into a convention which demands more general and formal solutions for various application cases. Both high spatial and high spectral data in a solitary image is required in a several instances of image processing. This phenomenon is critical in various cases of remote sensing. In some cases, the instruments are

not fit for giving such information either due to configuration issues or because of observational constraints of the instruments. Data fusion is one conceivable solution for this.

The methods of Image fusion can be broadly classified into two groups -

1. Spatial domain fusion
2. Transform domain fusion.

The registration of the image used in image fusion should already be done. A major source of error in image fusion is misregistration.

Following are some well-known image fusion techniques:

1. High pass filtering technique
2. IHS transform based image fusion
3. PCA based image fusion[Chiang, 14]
4. Wavelet transform image fusion
5. Pair-wise spatial frequency matching

Techniques for fusion, for example, averaging, Trove method, principal component analysis (PCA) and IHS based strategies fall under spatial domain classification. The high pass filtering based technique is another imperative spatial domain fusion strategy. In this strategy, the high recurrence subtle elements are infused into unsampled variants of MS images. Spatial domain approaches produce spatial distortion in the fused image which thus is a defect of this methodology. While we go for further for processing, for example, for classification problem, spectral distortion turns into a negative element of this approach. However, Frequency domain approaches can handle this spatial distortion on image fusion. For analysing remote sensing images, the multiresolution analysis has turned into an extremely helpful tool. The discrete wavelet transform has turned into an exceptionally helpful tool for image fusion [Lili, 15]. Some other fusion techniques, for example, Laplacian pyramid based, curvelet transform based and so on are also some of the additional methods in image fusion. A better performance in spatial and spectral quality of the fused images is shown in these methods as compared to other spatial methods of fusion.

Principal Component Analysis (PCA):- It is a numerical tool which changes correlated variables into uncorrelated ones. Another set of axes are produced as a result which are orthogonal in nature. The first chief component is taken with the most extreme fluctuation. The second main component is constrained in the subspace opposite to the first component and it goes on so [Naidu, 08]. This algorithm replaces the spatial segment of the multi-spectral image with that of the panchromatic image. It permits the spatial subtle elements of the panchromatic picture to be consolidated into the multi-spectral image. The PCA based image fusion algorithm is shown in the flow diagram as shown in fig. 4.

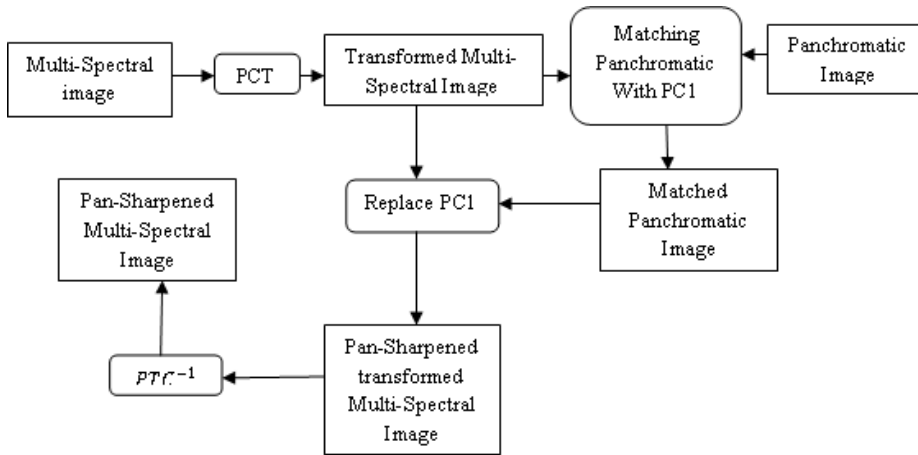


Fig. 4. PCA based image fusion [Metwalli, 09]

The process under image fusion by PCA is explained here. The principal component changes the resample paint bands of the multi-spectral image to the same resolution as the panchromatic image. The histogram that is matched to the first principal component accounts to be the panchromatic image formed [Pradesh, 15]. This is done in order to remove the spectral differences between the two images, which happened because of different types of sensors or distinctive procurement of dates and edges. The histogram matched panchromatic imagery replaces the primary essential component of the multi-spectral image. By computing the converse of principal component transformation, the new consolidated multi-spectral imagery is acquired. Fig. 5 shows the PCA based fused image.

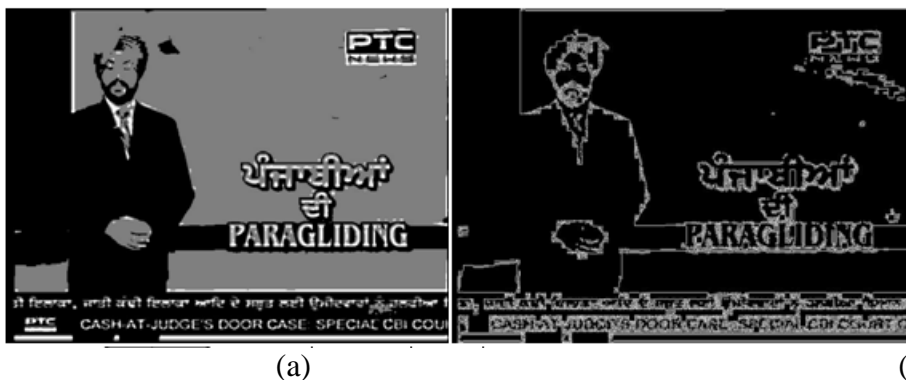


Fig. 5. PCA based fusion using DotNet (a) and MATLAB (b)

4.1 Normalization

Normalization is a procedure that progresses the range of intensity of pixel value. For instance, applications incorporate photos with insufficient contrast because of glare. In this normalization of the fused image is on the basis of an assumption that the Punjabi text in the video is majorly on the bottom part of the frame, so cropping is done on the bottom part of the frame as shown in figure 6.

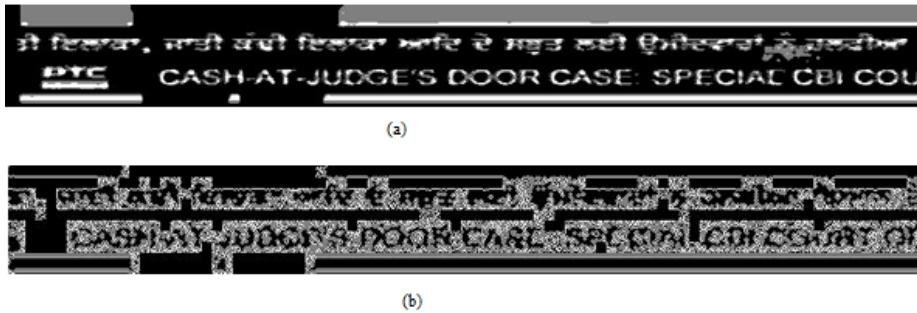


Fig. 6. Cropped Image using Dotnet (a) MATLAB (b)

5 Text Region Extraction using SURF & SIFT

The main focus of this work is to extract text region from the frames. So, two feature extraction methods are used: SURF and SIFT [Sakai, 15].

5.1 SURF based Feature Extraction

Speeded up Robust Features (SURF) is a feature identifier and descriptor that can be utilized for errands, for example, object acknowledgment or enlistment or order or 3D reproduction. It is incompletely enlightened by the scale-invariant component change (SIFT) descriptor [Khan, 15]. The standard variant of SURF is a few times speedier than SIFT and guaranteed by its creators to be more vigorous against various picture changes than SIFT.

SURF utilizes a number estimation of the determinant of Hessian blob finder to distinguish interest focuses, which can be figured with 3 integral operations utilizing a prior computed fundamental picture. Its component descriptor depends on the aggregate of the Haar wavelet reaction around the point of interest. It could likewise be figured with the guide of the indispensable picture. SURF descriptors can be utilized to find and perceive questions, individuals or countenances, to make 3D scenes, to track objects and to concentrate purposes of hobby. The SURF algorithm depends on the same standards and ventures as SIFT; be that as it may, points of interest in every stride are distinctive. The algorithm has three primary parts:

- Matching.
- Interest point location
- Local neighbourhood depiction

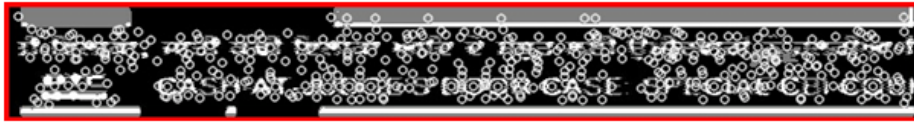


Fig. 7. SURF

Features

The above figure (Fig. 7) shows the extracted SURF features which covers the part of the text region of the frame.

5.2 SIFT based Feature Extraction

Scale-invariant element change (or SIFT) is an algorithm in PC vision to distinguish and portray local features in pictures. For any item in a picture, fascinating focuses on the article can be extricated to give an "element portrayal" of the item. This depiction, removed from a preparation picture, can then be utilized to recognize the article when endeavouring to find the item in a test picture containing numerous different articles [Kalamkar, 15]. To perform reliable acknowledgment, it is imperative that the components separated from the preparation picture be perceivable even under changes in picture scale, noise and brightening. Such focuses normally lie on high-differentiate locales of the picture, for example, object edges.

Another essential normal for these features is that the relative positions between them in the first scene shouldn't change starting with one picture then onto the next. For instance, if just the four corners of an entryway were utilized as components, they would work paying little heed to the entryway's position; however in the event that focuses in the edge were likewise utilized, the acknowledgment would come up short if the entryway is opened or shut. So also, includes situated in explained or adaptable items would normally not work if any adjustment in their inner geometry happens between two pictures in the set being handled. Be that as it may, by and by SIFT recognizes and utilizes a much bigger number of components from the pictures, which decreases the commitment of the mistakes brought about by these nearby varieties in the normal blunder of all element coordinating blunders.

Filter can vigorously recognize protests even among jumble and under fractional impediment, on the grounds that the SIFT feature descriptor is invariant to uniform scaling, introduction, and incompletely invariant to relative contortion and brightening changes.

Filter keypoints of articles are initially removed from an arrangement of reference images and put away in a database. An article is perceived in another picture by separately looking at every component from the new picture to this database and discovering hopeful coordinating elements in view of Euclidean separation of their element vectors. From the full arrangement of matches, subsets of keypoints that concur on the item and its area, scale, and introduction in the new picture are recognized to sift through great matches. The determination of steady

groups is performed quickly by utilizing a proficient hash table execution of the summed up Hough change. Every bunch of 3 or more elements that concede to an article and its stance is then subject to promote point by point model confirmation and along these lines anomalies are disposed of. At long last the likelihood that a specific arrangement of elements demonstrates the vicinity of an article is processed, given the precision of fit and number of plausible false matches. Object coordinates that breeze through every one of these tests can be distinguished as right with high certainty.

- Scale-space Extrema Detection
- Orientation Assignment
- Keypoint Localization
- Keypoint Matching
- Keypoint Descriptor

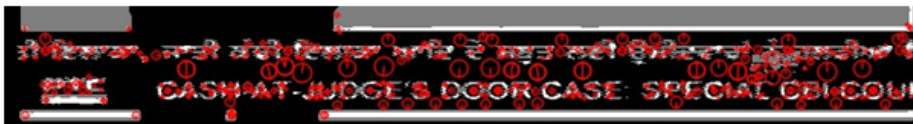


Fig. 8. SIFT

Features

The above fig shows the extracted SIFT features which also cover the part of the text region of the frame.

6 Conclusion

With the advancements in the field of digital world, the demands of the digital information are also increases. In this work, text region extraction from the Punjabi news videos has been done where firstly frames of the videos are extracted. These frames are further processed where edge detection is done by using sobel edge detection method and Binarization is done by using Otsu threshold method. These two outputs of a single frame are then fuse by using PCA based fusion method. Fusion is used to merge two images so that frame will produce more information. In this normalization of the fused image is on the basis of an assumption that the Punjabi text in the video is majorly on the bottom part of the frame, so cropping is done on the bottom part of the frame. Feature extraction in this work is done by using SURF and SIFT feature extraction methods on the cropped part of the frame. These features extract the text region from the frames.

References

1. K. Elagouni, C. Garcia, and F. Mamalet: Text Recognition in Videos using a Recurrent Connectionist Approach, In Proc. *Int. Conf. on Artificial Neural Networks*, 2012, 172–179.

2. S. Saluja and D. Rana,: Non text eradication from degraded and non degraded Videos and Images, In Proc. *Int. Conf. on Advances in Computer Engineering and Application (ICACEA)*, 2015, 768–772.
3. D. Chen and J. Odobez: Text detection and recognition in images and video frames, *Pattern Recognit.*, vol. 37, 2004, 595–608.
4. X. Jiang: Feature Extraction for Image Recognition and Computer Vision, In Proc. *IEEE Int. Conf. on Computer Science and Information Technology.*, 2009, pp. 1–15.
5. G. Vamvakas, B. Gatos, N. Stamatopoulos, and S. J. Perantonis: A Complete Optical Character Recognition Methodology for Historical Documents, 2008, 525–532.
6. G. Kumar: A Detailed Review of Feature Extraction in Image Processing Systems, In Proc. *Int. Conf. on Advanced Computing & Communication Technologies.*, 2014, 5–12.
7. J. Calic and E. Izquierdo: Efficient Key-Frame Extraction and Video Analysis, In Proc. *Int. Conf. on Information Technology*, 2002.
8. W. Wattanarachothai and K. Patanukhom: Key Frame Extraction for Text Based Video Retrieval Using Maximally Stable Extremal Regions,” In Proc. *Int. Conf. on Industrial Network and Intelligent System(INISCom)*, 2015, 29–37.
9. G. Liu: Key Frame Extraction from MPEG Video Stream, In Proc. *Int. Symp. on Information Processing (ISIP)*, 2010, 423 – 427.
10. H. Liu, L. Pan, and W. Meng: Key Frame Extraction from Online Video Based on Improved Frame Difference Optimization, In Proc. *Int. Conf. on Communication Technology (ICCT)*, 2012, 940 – 944.
11. W. Gao and X. Zhang: An Improved Sobel Edge Detection, In Proc. *Int. Conf. on Computer Science and Information Technology (ICCSIT)*, 2010, 67–71.
12. O. R. Vincent: A Descriptive Algorithm for Sobel Image Edge Detection Clausthal University of, In Proc. *Informing Science & IT Education Conference (InSITE)*, 2009, 97–107.
13. Zhang Zhi-Yong, Song Yang: The License Plate Image Binarization Based on Otsu Algorithm and MATLAB Realize, In Proc. *Int. Conf. on Industrial Control and Electronics Engineering (ICICEE)*, 2012, 1657–1659.
14. D. Han, J. Teng, Z. Yang, Y. Pang, and M. Wang: 2D Barcode Image Binarization Based on Wavelet Analysis and Otsu’s Method, In Proc. *Int. Conf. on Computer Application and System Modeling (ICCA SM)*, 2010, no. Iccasm, 30–33.
15. J. Chiang: Knowledge-based Principal Component Analysis for Image Fusion, *Int. J. Appl. Math. Inf. Sci.*, vol. 230, no. 1, 204, 223–230.



16. W. Lili, S. Yue, C. Deyun, and Y. Xiaoyang: ECT Image Fusion Based on PCA Transform and Wavelet Transform, In Proc. *Int. Conf. on Instrumentation and Measurement, Computer, Communication and Control*, 2015, 1188–1192.
17. V. P. S. Naidu and J. R. Raol: Pixel-level Image Fusion using Wavelets and Principal Component Analysis, *Defence Science Journal*, vol. 58, no. 3, 2008, 338–352.
18. M. R. Metwalli, A. H. Nasr, O. S. F. Allah: Image Fusion Based on Principal Component Analysis and High-Pass Filter, In Proc. *Int. Conf. on Computer Engineering & Systems, ICCES*, 2009, 63–70.
19. A. Pradesh: Edge Preserving Satellite Image Enhancement Using DWT-PCA Based Fusion and Morphological Gradient, In Proc. *IEEE Int. Conf. on Electrical, Computer and Communication Technologies (ICECCT)*, 2015, 1–5.
20. Y. Sakai, T. Oda, M. Ikeda, and L. Barolli: An Object Tracking System Based on SIFT and SURF Feature Extraction Methods, In Proc. *18th Int. Conf. on Network-Based Information Systems (NBIS)*, 2015, 0–4.
21. M. A. Khan and V. Fodor: Characterization of SURF and BRISK Interest Point Distribution for Distributed Feature Extraction in Visual Sensor Networks, *IEEE Trans. Multimed.*, vol. 17, no. 5, 2015, 591–602.
22. K. D. Kalamkar and P. S. Mohod: Feature Extraction of Heterogeneous Face Images Using SIFT and MLBP Algorithm, In Proc. *Int. Conf. on Communications and Signal Processing (ICCSP)*, 2015, 1286–1290.