A THESIS REPORT

ON

"TEXT Extraction from Images/videos using Morphological operation"

Submitted in partial fulfillment of the requirement for the award of the degree

of

Master of Technology

In

Signal Processing & Digital Design

Submitted by

MeezanMd "Chand"

(2K13/SPD/12)

Under the Supervision of

Dr. S. Indu

(Associate Professor)

HOD(Head of Department)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

SHAHBAD DAULATPUR, DELHI -110042, INDIA

CERTIFICATE

This is to certify that the dissertation title "**TEXT Extraction from Images/videos using Morphological operation**" submitted by **MeezanMd "Chand"**, Roll. No. 2K13/SPD/12, in partial fulfillment for the award of degree of Master of Technology in Signal Processing & Digital Design at **Delhi Technological University, Delhi**, is a bona-fide record of student's own work carried out by him under my supervision and guidance in the academic session 2015-16. To the best of my belief and knowledge the matter embodied in dissertation has not been submitted for the award of any other degree or certificate in this or any other university or institute.

Date: _____

(Dr. S. Indu)

Associate Professor

Head of Department

ECE Department

Delhi Technological University

ACKNOWLEDGEMENT

First of all, Praise is to ALLAH Tala, the Almighty, on whom ultimately we depend forsustenance and guidance , Then I express my sincere gratitude & heart full thanks towards my Guide **Dr. S. Indu Mam, Associate Professor, Electronics and Communication Department, Delhi Technological University, New Delhi**for giving me the guidance and opportunity to work on above mentioned thesis work. Her constant help, encouragement and inspiration throughout the project work is resulted into the completion of my thesis work .

I would like to thank my guide for one more reason that, she never let me feel helpless when any help needed to me from department, as of being our **Head of Department ,** She helped me every bit for providing all the necessary facilities, which were indispensable in the completion of this project.

I take this opportunity to express my hearty thanks to all those who helped me in the completion of my project work. I am very grateful to the authors of various articles on the Internet, for helping me become aware of the research currently ongoing in this field.

I warmly thank and appreciate my parents and elder family members fortheir material and spiritual support in all aspects of my life.

I also would like to thank my all friends as they have provided assistance in numerous ways. Last, but not the least, I would like to thank my classmates for their valuable comments, suggestions and unconditional support

MeezanMd "Chand"

(2K13/SPD/12)
M.Tech. (SPDD)
Department of Electronics & Communication Engineering,
Delhi Technological University, Delhi-110042

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT :

Popularity of digital image & video is increasing rapidly, Text characters embedded in images and video sequences represents a rich source of information for content-based indexing and retrieval applications However, these text characters are difficult to be detected, recognized, extracted, due to their various sizes, gray scale values and complex backgrounds. This thesis investigates methods for building an efficient application system for text extraction of any gray scale values embedded in images and video sequences. Text embedded in multimedia data, as a well-defined model of concepts for humans' communication, contains much semantic information related to the content. This text information can provide a much truer form of content–based access to the image and video documents if it can be extracted efficiently. My work in this thesis is not focused on Text Recognition rather than on Text Extraction parts(Text Detection).

Text Extraction plays a major role in finding vital and valuable information. These text characters are difficult to be detected and recognized due to their deviation of size, font, style, orientation, alignment, contrast, complex colored, textured background. Due to rapid growth of available multimedia documents and growing requirement for information, identification, indexing and retrieval, many researches have been done on text extraction in images & videos. Several techniques have been developed for extracting the text from an image & video. The proposed methods were based on morphological operators, wavelet transform, artificial neural network, skeletonization operation, edge detection algorithm, histogram technique etc. All these techniques have their benefits and restrictions. This thesis discusses morphological operation based scheme (using extraction of region containing Text) for extracting the text from an image & video.

Text extraction is a challenging problem due to the presence of complex backgrounds and the variable length of the text Strings appearing on screen with variable heights. My work would follow in the sequence like, I converted frames into gray scale ,then localization of text frame, followed by ROI(Region of Extraction) steps &then binary conversion(Binarization) in themeanwhile morphological operation has been done on binary images, in last execution of text extraction/detection step taken place

CHAPTER 1
INTRODUCTION

Text extraction, recognition in images and videos is a research area which attempts to develop a computer system with the ability to automatically read from images and videos, the text con- tent visually embedded in complex backgrounds. As an example of the general object recognition issues, this computer system, shown in figure 1.1, should answer two typical question "Where & What": "where is a text string?" and "what does the text string say?" in an image or a video, In other words, using such a system, text embedded in any type of backgrounds can be automatically detected and each character or word can be recognized. But Our focus is only on the detection part(More precisely Extraction) ,which is discussed later on from this chapter to last one.

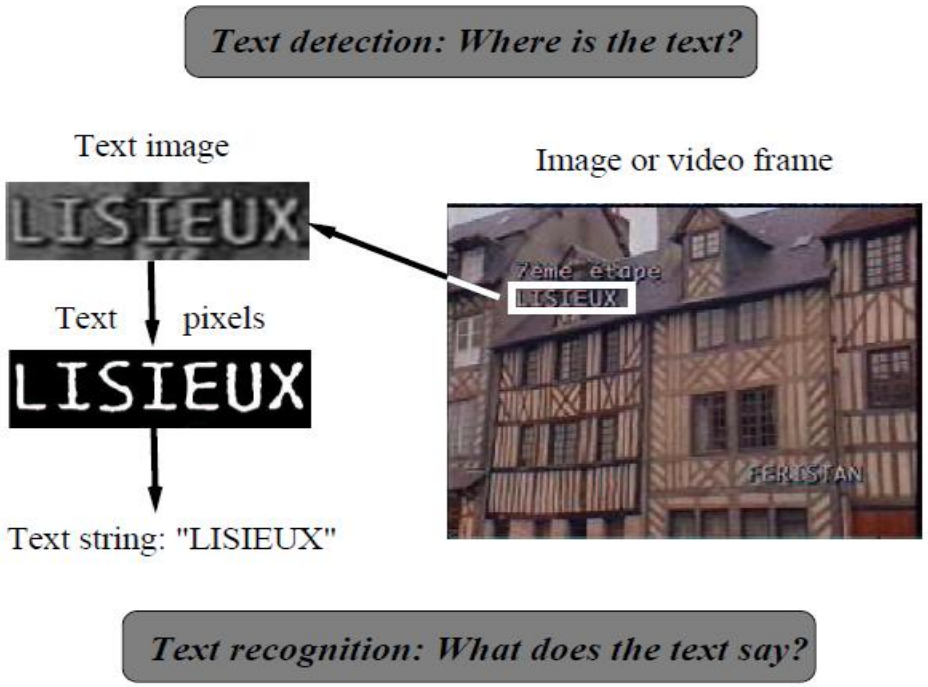


Figure 1.1 Text detection & Recognition in Images & Videos.

1.1 Motivation for Extracting Text from Image and Video Documents(Goals)

Image and video, as two most popular multi-media documents, are being produced on a daily basis by a wide variety of sources, including the long distance educational programs, medical diagnostic systems, business and surveillance applications, broadcast and entertainment industries, etc. Recently, with the increasing availability of low cost portable cameras and camcorders, the number of images and videos being captured are growing at an explosive rate. According to statistics provided by Flickr [1] and Youtube [2], the quantity of photos uploaded to Flickr has beenincreasing 20% year-over-year over the last 5 years and reached 6 billion in August 2011, and more than 13 million hours of video were uploaded to Youtube during 2010 and 48 hours of video were uploaded every minute. Given such vast quantities of image and video documents, it is quite probable that most images and videos we are interested in are available and can be accessed somewhere on the Internet.

The investigation of text detection and recognition in complex background is motivated by cutting edge applications of digital multimedia. Today more and more audio and visual information is captured, stored, delivered and managed in digital forms. The wide usage of digital media files provokes many new challenges in mobile information acquisition and large multimedia database management. Among the most prominent are:

1. Automatic broadcast annotation: creates a structured, searchable view of archives of the broadcast content
2. Digital media asset management: archives digital media files for efficient media management;
3. Video editing and cataloging: catalogs video databases on basis of content relevance;
4. Library digitizing: digitizes cover of journals, magazines and various videos using advanced image and video.
5. Mobile visual sign translation: extracts and translates visual signs or foreign languages for tourist usage, for example, a handhold translator that recognizes and translates Asia signs into English or French.

1.2 TEXT TYPE (SCENE TEXT , DOCUMENTS TEXT) :

Several differences exist between reading text in documents and in scene images. One example is shown in Figure 1.2 to know about scene text, The first primary difference is in the problem of locating the text to be Detected(later might be recognized). In a standard one or two column document, almost no detection must be done. Lines of text are easy to identify and simple heuristics are usually sufficient to prepare the input for recognition. In more complex documents such as newspapers and magazines, there is an added difficulty of distinguishing between text and image regions. However, these are often separate and aided by several intentional cues such as high contrast in the text and strong rectangular boundaries in imagesIn addition, text almost always appears in canonical horizontal or vertical orientations, and many techniques have been developed to estimate the global rotation necessary to horizontally "level" text lines in the image plane. Recent developments have also been made in the analysis of pages using camera-based acquisition techniques. More complex world transforms must be considered in these cases, but often an explicit model of the page can be used to rectify the text and return itto a planar appearance. Uneven lighting can be problematic in these situations, but the general binary nature of text on a page still makes local processing feasible for providing good binarizations.

The problem of finding text in an arbitrary image of a scene can be radically more complex. First, the contents of the input images(Frames) are generally more varied.the variety of potential image contents is vast, and it occupies all of the input images(Frames). Text regions in images need not be well-bounded the way they usually are in documents. Furthermore, text in scenes is often only a few words in one place. There are no large paragraphs or long lines to analyze. Although text is generally designed to be readable, there are often adverse effects of the imaging conditions that can make it difficult to identify. Distance from the camera can make text small and low resolution, without much detail. Specularitiescan mix text regions with a reflected image. Perspective distortion can produce text with a varying font size &orientation ,Unlike Text Documents, Small amounts of text can appear anywhere, at any size, with any world orientation, and on any surfaces.

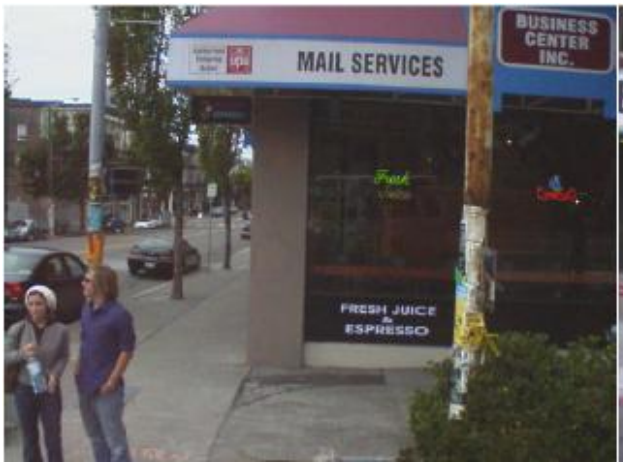
Figure 1.2 Example: Scene Text reading

Fortunately, it can be noted that there is a considerable amount of text objects in image and video documents. As a well-defined model of concepts for humans communication, text

embedded in multi-media data contains much semantic information related to the content. If this text information can be extracted and harnessed efficiently, it can provide a much truer form of content–based access to the image and video documents. Figure 1.3 presents some examples of text indicating the semantic contents of images (Figure 1.3-a to 1.3-d) and video frames (Figure 1.3-e to 1.3-h). Note that the two vehicles shown in Figure 1.3-a and 1.3-b have very high similarities in shape and color, but the text objects in the images can tell us the difference: One is a fire truck and the other is a vending truck.

(a)

(b)



(c)

(d)

(e)

(f)

(g) (h)

Figure 1.3 Examples of text indicating the semantic content of images.

Text in image and video can be classified into two categories: *caption text* and *scene text*. Caption text is artificially overlaid on the images or video frames at the time of editing. Although some types of caption text are more common than others, it is observed that caption text can have arbitrary font, size, color, orientation, and location.

For video document, a caption text event may remain stationary and rigid over time, or itmay translate, grow or shrink, and change color. A good text extraction system must beable to handle as wide a variety of these types of text as possible in order to support acontent-based indexing system. Text naturally occurring within the scene, or scene text, also occurs frequently in image and video. Examples include text appearing on vehicles (license plate,maker/model/dealer names, company names on commercial vehicles, bumper stickersetc.), text on signs and billboards, and text on T-shirts and other clothing. Technically, the extraction of scene text is a much tougher task due to varying size, position, orientation, lighting, and deformation. Contrast to a large amount of text extraction techniques for caption, only limited work is found in the literature that focuses on robustscene text extraction from images and videos.

Figure 1.4 shows the examples of caption text and scene text.

Figure 1.4 Examples of caption text and scene text.

Generally, most image-based text extraction approaches can be used for video documents as well, since video can be considered as a sequence of images (frames).However,compared with still images, video has some unique properties that may affect the text extraction. On one side, video usually has low resolution, low contrast, and color bleeding caused by compression, which are undesirable characteristics for text extraction; on the other side, text in video typically persists for at least several seconds to give human viewers the necessary time to read it. This temporal redundancy of video is very valuable for text verification and text tracking.

1.3 Problem Statement and Scope of This Dissertation

Many interchangeable terminologies are used in the literature to indicate textextraction process, such as text segmentation, text detection, and text localization. In thisdissertation, we follow the terminology which is easily defined & understandable. A system that can extract text objects in image and videodocuments is a text extraction system, which is composed of the following five stages:

(1) **Pre Processing steps**: This stage finds all frames in an image or videothat contain text objects or simply we are converting video into the various frame which can be further analyzed.

(2) **Text localization**: This stage is related to text regions of every frames which we have found in first steps, Findingthe region where all text objects is located;

(3) **ExtractingRegion of Interest(ROI)**: This stage is for video documents only. This step is for a text event that moves or changes over time in consecutive frames**,** the temporal and spatial locations of the text events is determined ,here we are extracting the exact portion of every frames where we have "portion of Text" only.

(4) **Text binarization**: In this stage, we converts the all frames into binary image found in previous step.

(5) **Text Detection**: This stage outputs the recognized text objects by detecting all text ,Where we find each and every character of text in bounding box.

The architecture of a text extraction system is illustrated in Figure 1.5

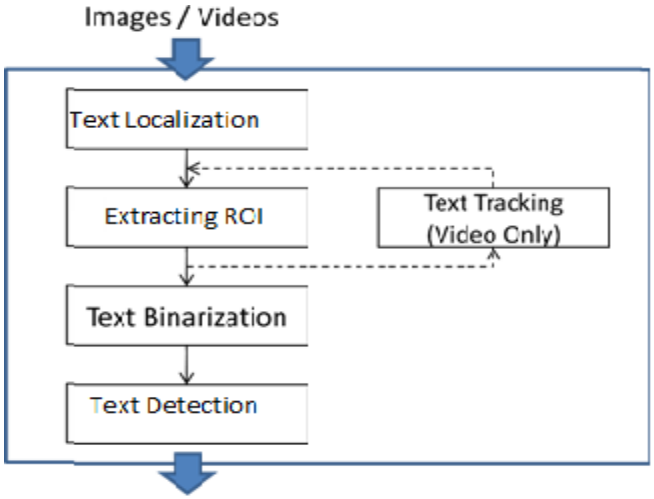


Figure 1.5 Architecture of text Extraction system.

Amongthefivestagesoftextextractionor detection process, most steps arecloselyrelatedto each other in few or many ways, andExtracting Exact ROI(Region of Interest) & then Implementing Morphological operation on them is morechallengingstepthatattractedthemaximum time among all utilized time on this project.Inthisdissertation,wediscussthetextlocalization ,Region Extraction & text Extraction or DetectionproblemsforTEXT found inimagesandvideosaswellastexttracking(Means Extraction ROI)problemforvideos,andpresentedtheresultsofwork.

InChapter2,wereviewthetextextractionapproachesreportedintheliterature,includingdetection

,localization,tracking,andrelatedapplications,anddiscussthestateofthearttextextractionresearchandlimitationsoftextextraction.

InChapter3,This chapter Contents maximum theoretical parts of my thesis, we discussed all related steps in this chapter in brief way & in more detail way too, where ever needed just the brief explanation I went through briefly & wherever needed (As per Work in this thesis) More detailing we have explained that term in more detailed way ,with the help of pictorial diagram too, Term explained in this chapter is Binarization ,Thresholding , Segmentation , Edge Detection & Finally Morphological Operation ,Main focused of this chapter is on Morphological operation & Their example with picture .

InChapter4,wediscussMain Method of my Thesis in this chapter, In Flow Diagram ,Then In This chapter, we have included Results with discussion ,Conclusion & Future Work too.

In Chapter 5, we have listed the all references which I have referred during our thesis completion work partially or fully .

CHAPTER2:

LITERATURE REVIEW

Hundreds& even more

numbersofapproacheshavebeenproposedfortextobjectextractioninimageandvideo documentssince1990s.Chenetal.[3]and Jungetal.[4]presentedcomprehensivesurveysoftextextractionapproachesproposedbefore1999and2003,respectively.Inthischapter,weconcentrateontheprogressmadeafter2003andreview recentlyproposedtextdetectionandlocalizationapproaches(Section2.1),texttrackingapproaches(Section2.2),andtextextractionbasedapplications(Section2.3).Basedontheliteraturereview, wediscussthestateoftheartandthelimitationsoftextextractionsystems(Section2.4).

2.1 DetectionandLocalizationApproaches

In[4],thetextdetectionandlocalizationapproachesaredividedintotwocategories:(i)Regionbased approachesthatusethedifferentpropertiesbetweentextregionandbackgroundregiontoseparatetextobjects.Thiscategorytypically worksinabottom-upwaybyseparatingtheimageintosmallregionsandthengroupingsmallcandidateregions into text regions;(ii)Texturebasedapproaches thatuse distinct texturepropertiesoftexttoextracttextfrombackground.Thiscategorytypicallyworksinatopdown waybyextractingtexturefeaturesofimageandthenlocatingtextregions.Inthischapter,wefollowthiscategorizationandfurthersubdividetheregion-basedapproachesbasedonthefeaturesusedintheapproaches.

2.1.1 Region-BasedApproaches

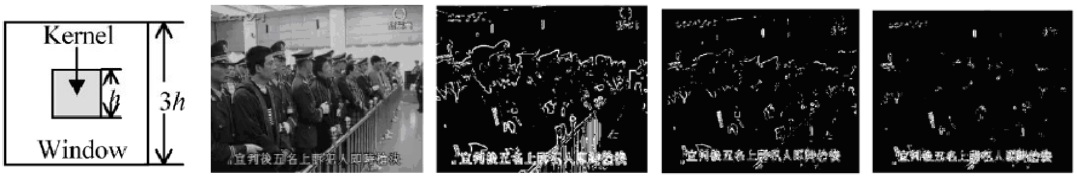
Weuseafeature-orientedstrategytogroupregion-basedapproachesbyfocusingonthedifferentpropertiesoftextobjects,sothatweareabletoknow:(i)whatfeaturescanbeusedtodescribethepropertiesoftextobjects;(ii)whythesefeaturescandistinguishtextobjectsfromotherobjects;and(iii)howthesefeaturesareusedtogeneratetextregionsintheapproaches.Thecategorizedgroupsincludes:gradientandedgebased,colorbased,cornerbased, strokebased,andmultiple-featurebasedapproaches.

(1) *GradientandEdgeBasedApproaches*

Itisobservedthattextobjecttypicallyhashighedgedensities,similaredgeheights,bigedgegradientmagnitudes,andlargeedgegradientdirection variations duetothefactthattextiscomposedofseveralalignedcharacterswithsimilarsizesandsharpcontrast tobackground.Gradientandedge-basedapproachesareimplementedusingthisobservation. Usually,morphologicaloperation and block-based dividing and mergingareusedtogeneratecandidatetextregionsbasedonlocaledgeinformation,andspatialand

geometricalconstraintsareusedtoremovefalsealarms.

Lyu*etal.*[5]proposedatextextractionapproachformultilingualvideos.Eightfrequentlyusedtext featuresaregroupedintotwocategories:(i)Language-independentfeatures:contrast,color,orientation,andstationarylocation;(ii)Language-dependent features:strokedensity,fontsize,aspectratio,andstrokestatistics.Textisdefinedasaregionthathast heunionofthesecharacteristics.Inordertodealwiththetextobjectswithvarioussizes,anewsequent ialmulti-resolutionparadigmispresented.Comparedwithparallelmulti-resolutionparadigmmethods[6][7]thatoftenspendunnecessarytimeondetectingthesametextstri ngindifferentresolutionlevels,thesequentialmulti-resolutionparadigmerasesatextedgeimmediatelyonceitisdetectedataresolutionlevelsothatnote xtedgescanappearseveraltimesatdifferentresolutionlevels.Inedgedetectionstage,twoconcentri csquaresshowninFigure2.1-aareusedtolocatetextedgesandsuppressbackgroundedgesbylocalthresholding.Thebackground complexityandtheedgestrengthhistograminsidethelargersquare“window”areanalyzedtodeter mine the localthresholdfor thesmallersquare“kernel”.Afterthat,edge densityandahysteresismaskareusedtoextractandrecovertextedges.Thedetectionprocessisshow ninFigure2.1-b. In textlocalizationstage,horizontalandverticalprojectionsoftheedgemaparecomputediterativelyto locatetheboundariesoftextobjectsinacoarse-to-fineway.Falsepositivesareeliminatedusingmulti-frameverification.EnglishandChinesearetestedinthepaper.



Originalimage Global LocalthresholdingText-likearearecovery

(a) Concentrickernel (b)Stepwiseresultsoftextdetection

Figure2.1Concentrickernelandstepwisedetectionresults(FromLyu*etal.*[5]).

Liu*etal.*[10]uselinefeaturestodetecttextobjectsinvideos.First,animprovedCannydetectorwith

twophasethresholding isperformedtocomputeedgesintheimage.Alledgelinesareexpressedaslistsusingan8-connectedcomponentalgorithmandthe non-textlinesareeliminatedbygeometricconstraints.Then,height,area,centerposition,andedgedensityareextractedfromtheboundingboxofeachlineinthelisttoformalinevectorgraph.By groupingalltheneighboringlinestogetherbasedonthe linevectorgraph usingthe 8-connectedcomponentalgorithmagain,the image isdividedintoseveralisolated regions withcloselydistributedlines.Afterremovingnon-textregionsbydistributionandnumberofedgesandregionaspectratio,thefinaltextregionsareextracted.Bycombiningthismethodwiththetemporalredundancyofvideo,Mi*etal.*[11]computeregionlocationsimilarityandedgemapsimilarityovermultipleframestorefinethecandidateboundariesoftextobjectsinvideo.

Basedonastructuralmodeloftext,Tran*etal.*[12]useridgesdetectedatseveralscalestoextracttextobjects.GiventheLaplacianimageofanimage,"ridge"includestwotypesofpoints,ridgepointsandvalleypoints,whicharedefinedasthepointsthattheirLaplacianvaluesarelocalmaximumsorminimumsinthedirectionofhighestcurvature.First,Gaussiankernelswithdifferentscalesofstandard deviationareemployedtosmooththeoriginalimage.Then,ridgepointsaredetectedbycomputing eigenvaluesandeigenvectorsofHessianmatrix[13]ateachscale.Finally,afterlinkingtheridgepointsateachscale,atextstringismodeledhierarchicallybasedonridgelengthconstraintsandspatialconstraints.Aridgeatacoarsescalerepresentsthecenterlineofthetextstringandnumerousshortridgesatasmallerscalerepresenttheskeletonsofthecharacters. This method is invariant tothesize andorientation of charactersandrobust todifferentwritingsystems.

Insteadofusingalldetectededges,Dubey[14]usesonlythevertical edgestofindtext regions basedonthe observationthat verticaledgescanenhance the characteristic ofthetextovertheotherpartsoftheimageandeliminatealargeportionofirrelevantinformation.Afterdetectingverticaledgesby Sobeloperator,regiongroupingisperformedinhorizontaldirectionbasedontheintensitiesandintervalsofedgepixels.Theregioncomposedofanumberofcontinuousverticallineswithsimilarlengthsismarkedasatextregion.Asnotedbytheauthors,however,thisapproachisunabletodetecttheverticallyalignedtextobjects.

Zhou*etal.*[15]presentacoarse-to-fineapproachbasedonedgeinformationandgeometricalconstraintstofindtextregionsinvideos.I

ftheedgedensityinaslidingwindowislargerthanapredefinedthreshold,itismarkedasacandidatet extregion.Thencandidateregionsarelabeledas*ConnectedComponents*(CCs)bymorphological closingoperationandfilteredbygeometricalconstraints.Textverificationandenhancementarei mplementedbyusingthetextpolarityconsistency,theoverlappingtextarea,andthestabilityofcha racterstrokesoftheconsecutiveframes.Edgedensityandtemporalredundancyarecombinedby Wang*etal.*[16]tolocalizetextinvideos.Sobeldetectorisappliedontwointegratedframeswhosepi xelvaluesaretheminimumormaximumintensitiesover30consecutiveframes.Afterdividingthei mageintoblocks,thetextregionsarelocalizedandmergedbasedontheedgedensitiesofblocks.

Liu*etal.*[17] use edgeinformationandHaarwaveletto localizetextobjects. Theedgemapisobtainedusinganeigenvaluebasedmethodandweakedgesareeliminatedbygradi entmagnitudethresholding.Candidatetextregionsaregeneratedandfilteredusingconnectedcom ponenttechnique,averagededgepixelgradientmagnitude,edgegradientdirectionvariance,ande dgepixelnumberbasedontheedgemap.Waveletenergythatisdefinedasthesummationof LH,HL ,andHHsub-bandsisusedtoverifycandidateregions.

Byusingverticaland horizontalprofiles of theedgemap, Park*etal.*[18]localizeKoreantextinoutdoorsignboardimages.Verticalprofilesandafuzzyc-meansmethodareadoptedtogetbinarizedregionsforindividualcharacters.Afternoiseremovalus inggeometricconstraints,theverticallyadjacentcomponentsaremergedtogeneratecharacterco mponentsbasedonthecharacteristicsofKoreancharacters.

(2) *Color-BasedApproaches*

Color-basedapproachesarebasedontheobservationthatthecoloroftextobjectishomogeneousanddiffer entfromthebackgroundcolors.Theapproachesusuallyfirstextracttheregionswithhomogenous colorsbasedoncolorsimilarityusingclusteringmethod,intensityhistogram,orbinarizationmeth od,thenthecandidatetextregionsarelocalizedbyspatialinformationandgeometricalconstraints. Fu*etal.*[19]proposeatextdetectionmethodincomplexbackgroundbasedonmultipleconstraints. PreliminarysegmentationisimplementedbyK-meansclusteringbasedonYCbCrcolorvectors.K=3or4dependingonthenumberofhumpsthatap pearinthehistogramofanimage.AfterobtainingCCsusingclusteringresults,fourconstraintsarea ppliedtoperformpost-processingtoeliminatebackgroundresidues.

(i)Colorconstraint,allCCs corresponding totextobjectsshouldhavehomogeneouscolors.

(ii) Edgemagnitudeconstraint,theboundariesoftextCCsshouldgowithstrongedges.

(iii) Thicknessconstraint,characterstrokesshouldhavepropersizeandCCswhoseheightorwidthexceedsthethresholdsareremoved.(iv)Componentsrelationconstraint,suchasdimensionalrange,combinationoftwocomponents,andcompactnessoftwocomponents.However, ifseveraltextobjectswithdiversecolorsappear ina videoframe simultaneously,theK-meansclustering(K=3or4)mayfailtofindalltextcolors.

Clavelli*etal.*[20]segmenttextobjectsincolorposterimagesbasedoncolorinformationandspatialrelationshipofcharacters.TheCCsarecreatedusingRGBcolorsimilarityofneighboringpixelsandtoosmallortoobigcomponentsareremoved.Then,eachpairofneighboringcomponentsisconsideredasseedstobuildtextlinesbasedonintervalandorientation.Thefinalsetoftextlinesislabeledastextregions.pixelsinRGBspaceforeveryrow.Ingroupingstep,thehorizontalchainsarelinkedvertically basedonspatialdistributionandthepropertiesofcharactertoformatwodimensionalspatialcolorcomponent,whichisacandidatetextregion.

Fu*etal.*[21]andLiu*etal.*[22]discriminatecharactersfrombackgroundbyassumingthatthreeneighboringcharacterscanberepresentedasaGaussianMixtureModel(GMM)basedonthedistances betweencentroidsofcharacters,theregionareasofcharacters,andtheregiondensities.Therefore,textregionscanbedistinguishedfromnon-textregionsbyaprobabilitycomputedusingthisassumptionandBayesrule.Themethodsfirstcompute CCsusingthebinarizedinputimage.TheobtainedCCsaremergedbymorphologicalclosingbasedondistanceandtheneighborhoodsofCCsaregeneratedbyVoronoiregions.Then,forallneighborsetwiththreeCCs,GMMmethodisusedtodeterminewhetherthethreeCCsarecharacters.Theparametersaredeterminedbymaximum-minimumsimilarity(MMS),whichcanmaximizethesimilaritiesofobservationsandmodelsfromthesameclasses,andminimizethesimilaritiesforthosefromdifferentclasses.

(3) *Corner-BasedApproaches*

Textobjectsarerichofcorners,whicharetypicallyuniformdistributedovertextregions.Basedonthisproperty,manycorner-basedapproachesareproposedtoseparatetextfromotherobjects.Block-basedcornerdensityandmorphologicaldilationbasedcornermergingareoftenemployedtogeneratecandidatetextregionsusingextractedcorners.

ThecornermapgeneratedbySUSANcornerdetector[23]isusedbyHua*etal.*[24]tolocalizetextobjectsinvideoframes.Candidatetextregionsareformed bycornerrefinement,merging,anddilation.Cornerdensity,edgedensity,theratioofverticaledge

densityandhorizontaledgedensity,andcenteroffsetratioofedgesextractedfromvertical,horizontal,andoverall edgemapsare computedto decomposed text regions intotextlinesandremovefalsealarms.

HarriscornersareusedbyZhao*etal*.[25]tolocalizetextandcaptioninvideodocumentaswell.Afterfindingthecornerpointsintheimagebyharrisdetector,morphologicaldilationisusedtomergeclosecornersintooneregion.Falsepositiveregionsarefilteredoutusingfivefeatures,theareaofaregion,thesaturationthatindicatestheproportionoftheforegroundintheboundingbox,theorientationofaregion,theaspectratioofboundingbox,andthepositionofaregionwithitscentroid.Formoving captions,thetext

detectionresultsarecombinedwiththemotionvectortodetectmovingtextregions.

Insteadofusingmorphologicaloperationstomergecornersasin[26],Sun*etal*.[27]useblock-basedcornerdensitytolocalizetextobjects.Theblockswithhighcornerdensitiesareconnectedtogeneratetext regions,whicharethenverified bycolordeviation.

(4) *Stroke-BasedApproaches*

Comparedwithedgeandcolorfeatures,strokecanbeconsideredasahigh-levelfeaturebecausestrokeistheprimeelementofcharacterandcontainsintrinsicpropertiesoftextobject.Oncestrokeinformationisextractedsuccessfully,textcanbelocalizedeasily.Themostobviousandwidelyusedpropertyofstrokeisthatthestrokesofacharactertypically

havesimilarwidthandcolor.However,theapproachesbasedonly

onstrokewidthmayfailwhentextobjectshavemorethanonedominatingstrokewidths.

Jung*etal*.[28]designastrokefilter tosegmenttextbasedonthe observationthatatextstrokehashomogenousintensityandaredifferentfromitslateralregions.AstrokefilterisshowninFigure2.3.Foreachpixel,threerectangularregionsarounditaremarkedasalocalregion.Thestrokefilterresponseiscomputedbasedonthedifferencesofthreemeansoftherectanglesoverfourorientationsandthreescales.Thepixelwithhigherresponseismorelikelybelonging tostroke-

likestructure.Grayintensityandconstantgradientvectorareusedasfeaturesforcandidatetextverification.

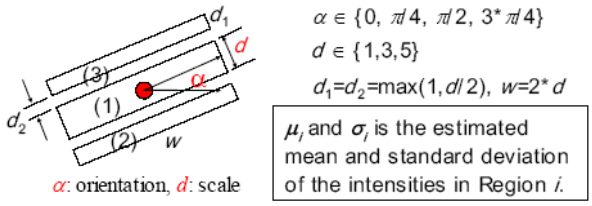


Figure2.2Localregionsforstrokefilter(FromJung*etal.*[28]).

Srivastav*etal.*[29]usestrokewidthandnearest-neighborconstraintstodetectscene text.Firstly,theedgemapisgeneratedbymergingtheedgesdetectedinquantized RGBchannelsandtheobtainededgesarelabeledusing8-connectedcomponent.Then,basedondimension,alignment,anddistance,CCsaregroupedusing nearest-neighborconstraints.Thestrokewidthsarecomputedinfourdirections(0,45,90,135degrees).Ifall strokewidthshavesimilarvalues,thisregionismarkedasatextregion. Thepropertythatthecorrespondingstrokepointshaveoppositegradientdirectionsisalsonoticedby Zhang*etal.*[30].Histogramoforientedgradient(HOG)with4binsisappliedtocapturethisproperty foreachdetectedclosedboundary.Ifbin1and3andbin2and4havesimilarnumbersofboundarypoints,theboundaryismarkedasaboundaryofcharacter.Spectralisusedtogroupcharacterstotextobjects.

(5) *Multiple-FeatureBasedApproaches*

Themultiple-featurebasedapproachescombinethetextinformationcapturedbydifferentfeaturestodetectandlocalizetextobjects.

Dinh*etal.* [31]localizetextusingedge cue andcolor cue. Edgesare extractedbySobeldetectorandfilteredbygeometricconstraints.Theedge-basedcandidatetextregionsaretheCCsobtainedusingtheremainingedges.Colorsegmentationis donebymeanshiftsegmentation.Thecolor-basedcandidatetextregionsaregeneratedusingthegeometricconstraints.Afterthat,twotypesofcandidatetextregionsarecombinedusing2Dtensorvotingtoyieldfinaldetectionresults.Kim *etal.*[32] usecolorconsistencyandedgeorientationconsistencyoftextobjectsinconsecutiveframestoextractstatictextregions.Morphologicalerosionanddilationareusedtoremovefalsepositives.

Liu*etal.*[33]useintensityandshapefilterstolocalizetextobjectinimage.

Intensityfilterisbuiltbasedontheobservationthattextobjectsusuallyhavebetter contrastwiththeiradjoiningbackgroundsthannon-textcomponents.Ablockismarkedasnon-textblockifitsintensityhistogramissimilartothatofitsadjoiningarea.Shapefilterisbuiltbasedontheassumptionthattextusually isconstitutedbydifferentcharacters.Theshapedistancecomputedusinginnerdistance[34]isutilizedtoremovenon-textregions.

2.1.2 Texture-BasedApproaches

Besidestheregionproperties,thedistincttexturalpropertiesbetweentextandbackgroundcanalso provideimportantinformationfortextextraction.Fortheapproachesinthiscategory,statisticalfeatures,frequencytransforms,Gaborfilters,andmachinelearningbasedmethodsareoftenusedtodescribeanddistinguishthetextureoftextandbackground.

Liu*etal*.[35]detecthorizontal,vertical,up-right,andup-leftedgemapsoftheinputimage.Mean,standarddeviation,energy,entropy,inertia,localhomogeneity,andcorrelationarecomputedforaslidingwindowoverthefouredgemaps.K-means(K=2)methodisemployedtoclustertheimageintotextandnon-textregionsbasedontheabovefeatures.Geometricconstraintsareusedtodiscardfalsepositives.

Shivakumara*etal*[36]proposeanalgorithmtodetectvideotextthroughtheclassificationoflowandhighcontrastimages.Afterdividinganimageintoblocks,thearithmeticfilterandmedianfilterareappliedoneachblocktogetoutputsAFandMF.ThedifferencebetweenAFandMFisdefinedasDiff.BasedontheirobservationsthatSobeldetectorusuallyyieldsmoreedgesthanCannydetectorforhighcontrastblocks,ablockisclassifiedasahighcontrastblockif:(i)thenumberofedgesdetectedbySobelofAFisbiggerthanthenumberofedgesdetectedbyCannyofDiff;(ii)thenumberofstrongedgesdetectedbySobleofMFisbiggerthanthatofDiff.Aframewithhighcontrastblocksismarkedashighcontrastframe.Then,similarto[35],statisticfeatures,suchasmean,standarddeviation,entropy,andK-meansmethod(K=2)areadoptedtoyieldcandidatetextclusterandbackgroundcluster.Thethresholdsofhighandlowcontrastframearecomputedseparatelybasedonthegradientedgemapsoftwotypesofframes.

Celine*etal*.[37][38] propose a selectivemetric-basedclusteringmethodtoextracttextfromnaturalscene.First,basedontheirstudythatRGBspacecanhandlevariabilityofnaturalscenesbetterthanothercolorspaces,Euclideandistancebasedand CosinesimilaritybasedclusteringmethodsareappliedonGRBcolorspacecomplementarilytopa

rtitiontheoriginalimageintothreeclusters:textualforeground, textualbackground,andnoise.Thenthetextualforegroundisidentifiedbyfindingthelargestregularityofclusters.FinallyLog-Gaborfilterisemployedtocombinespatialinformationandfrequencyinformationtolocatecharactersanddetectcharacteredges.

Itisknownthataregionwithrichtextureinspatialdomainhashighfrequenciesinfrequencydomain.Takingadvantageofthisfact,manytexture-basedapproachesseparate text bylocalize highfrequencyregions in frequencydomain.*FourierTransform*(FT),*DiscreteCosineTransform*(DCT),and*DiscreteWaveletTransform*(DWT)aremostusedforthispurpose.

Shivakumara*etal.*[39]utilizeFourier-statisticalfeaturesinRGBspacetodetecttextobjectsinvideo.Firstly,aframeisdividedintoblocks, whichareclassifiedastextandnon-textblocksusing thefollowingcues:thesharpnessoftheedges,the straightnessandcursiveness ofthe edges, and theedge proximity. Theframe containing textblocks isdefinedastextframe.Then,Fouriertransform(FT)isappliedonRGBchannelsfortextframes.Basedonenergy,entropy,mean,inertia,andmomentfeatures,a K-meansmethod(K=2) is employed to separate the image into textandnon-textregions inFourierdomainbasedonthefactthatFTgiveslargervaluesfortextpixelsandlowervaluesfornon-textpixels.

ByusingaFourier-Laplacianfilter,Phan*etal.*[40]andShivakumara*etal.*[41]proposeamethodtodetectmulti-orientedtextobjectsinvideo.Aftertransformingaframetofrequencydomain,anideallow-passfilterandLaplacianfilterareusedtocomputemaximumdifference(MD)mapthatisdefinedasthedifferencebetweenthemaximumvalueandtheminimumvaluewithina1×Nwindowofthefilteredimage.AK-means(K=2) methodisadopted toclassifytheMD map into twoclustersbased onthe

magnitudesofMDvalues.TheregionswithlargeMDvaluesaremarkedascandidatetextregions,becausetextobjectshavehighcontrasttothebackground.Then,candidatetextregionsarelabeled asCCsanddividedintotwotypes,simpleand complex,basedonthenumberofintersectionsoftheskeletonsofCCs.Acomplexconnectedcomponentisseparatedintosimplecomponentsusingtheskeletonsegmentsobtainedbydeletingtheintersectionpointsoftheskeleton.Skeletonstraightnessandedgedensityarecomputedforeachsimplecomponenttoremovefalsepositives.

Leon*etal.*[42]useDWTtoextractioncaptiontext.Fortheregioncoveredbyaslidingwindow,ifthere arehighvaluesinatleastoneofLHandHL subbandsofaHarrtransform,thepixelsareclassifiedastextcandidates.Binarypartitiontreeareemployedtoverifythe candidateregionsbasedonoccupancy,aspectratio,height,area,andcompactnessfeaturesofgeneratedboundingboxes.

TheapproachesproposedbyGllavata*etal.*[43],Saoi*etal.*[44],andShivakumara*et al.* [45]alluse HL,LH,andHH sub-bands of DWT andK-meansmethodfortextdetectionandlocalization,butdifferentfeaturesareusedforclustering.Wavelet coefficients are selectedas the clusteringfeaturein [43]tolocalize text regionswithhighvaluedcoefficients. In[44],thethreesub-bandsimageareddividedinto8×8 blocksandthestandarddeviationsofthreesub-bandsareutilizedasfeaturesforclustering.Thestatisticfeatures,suchasmean, entropy,moments,etc.,extractedfromthreesub-bandsareusedasclusteringfeaturesin[45].Basedonthesefeatures,theapproachesclustertheinputimageintotextandnon-textregionsusingK-means(K=2)method.

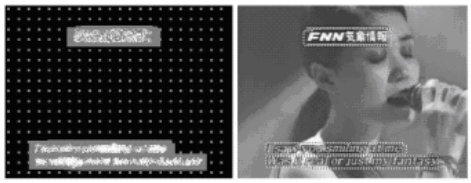
resultsofseveralclusteringalgorithmstogethertoimprovethequalityoftheindividualclustersandrobustness,isadoptedtofusemulti-frameinformation.Forasetofconsecutiveframes,thefeaturesareextractedfromeachframebyapplyingDWTandclusteredbyafuzzyC-meansmethod.ThenFCEisemployedtooutputanintegratedframewiththreeclusters,“text”,“background”and“complexbackground”,basedontheindividualclusteringresultofeachframe.Theclusterthathasthe smallestdistancetotheidealtextfeaturesarelabeledas“text”cluster.Acoarse-to-fineprojectionmethodyieldsthefinalboundaryboxesoftextobjects.

Withtherapiddevelopmentofclassifiertechniques,manysupervisedtextdetectionapproachesha

vebeenproposedbasedonArtificialNeuralNetwork(ANN),SupportVectorMachine(SVM),Bo ostingalgorithms,etc.Chen*etal.*[41]presentatwo-stepapproachfortextextraction.First,verticalandhorizontaledgesdetectedbyCannyoperatorare extendedinhorizontalandverticaldirectionbydilationoperatorstoextractcandidatetextregions. Boundingboxesofcandidateregionsaregeneratedbyabaselinealgorithm[46]andempiricalcons trains.Second,fourfeaturesareusedtoextracttextregions: (i)gray-scale spatial derivativesfeature;(ii)DCTcoefficients; (iii) distancemapfeature,whichonlyreliesonthepositionofstrongedgesintheimage,and(iv)constant gradientvariancefeature,whichnormalizesthecontrastatagivenpointusingthelocalcontrastvari ancecomputedinaneighborhoodofthispoint.Finally,multi-layerperceptrons(MLP)andSVMareusedtoverifythetextobjects.

InsteadofusingSVMtoclassifytext objects,Kim*etal.*[47]noticethatSVM canextractfeatureswithinitsownarchitectureefficientlybasedonkernelfunctionsandpresent anapproach usingSVMto analyze thetextural properties oftexts.The intensitiesoftherawpixelsintheoriginalimageareinputtotheSVMdirectlywithoutexternaltextu refeatureextractionmodule.Acontinuouslyadaptivemeanshiftalgorithm(CAMSHIFT)isempl oyedtogettheboundingboxesbymodifyingthesearchwindowsiterativelybasedonSVMoutputs .AnexampleoftextdetectionusingCAMSHIFTisshowninFigure2.4.Thelimitationisthattheapp roachcanonlydetecttextobjectsinhorizontalrectangleshapes.SimilarutilizationofSVMfortext extractionisalsoreportedbyCheng*etal.*[48]

(a) (b)



(c) (d)

Figure2.3ExampleoftextdetectionusingCAMSHIFT(FromKim*etal.*[47]).(a)inputimage(540 ×400),(b)initialwindowconfigurationforCAMSHIFTiteration(5×5-

sizedwindowslocatedatregularintervalsof(25,25)),(c)textureclassifiedregionmarkedaswhitea ndgraylevel(white:textregion,gray:non-textregion),and(d)finaldetectionresult.

WangandChen'smethod[49]usewavelettransformtoextracttextobjectsinvideodocuments.Int heedgedetectionstage,DWTwithsevenwaveletfunctionsisappliedonthevideosequenceandtwo differentcombinationofwaveletfunctionsareusedtodetectstaticanddynamictextedgesseparatel y.Intheclassificationstage,thetexturefeatures,suchasgray-levelco-occurrencematrix,maximumprobability,energy,andentropy,calculatedbasedonthecombinati onofDWTsub-bandsareinputtoaBayesclassifiertolocalizefinaltextregions.

Pan*etal.*[50][51]localizetextinsceneimagesusing*ConditionalRandomField*(CRF).Textregio nsfirstaredetectedbyHOGandWaldBoostalgorithm.Theconfidencemapshowingtheprobabili tyofaregioncontainingtextarecomputedusingtheWaldBoostoutputbasedonaboostedclassifier calibrationmethod.CCsareobtainedbyNiblack'sbinarizationalgorithm.Then,CRFisadoptedf orconnectedcomponentanalysis.Basedontheneighborhoodgraphof CCs,14unaryandbinaryfe aturesthatrepresentthepropertiesofCCsandcomponentneighboringrelationshipareusedtocons tructaCRFmodel.Textregionsarelabeledbyminimizingtheenergyfunctionofthegraph.Finally, minimumspanningtreeisemployedtogrouptextcomponentsintotextobjectsbasedon9distance metricfeaturesandtextlinefeatures.Inexperiments,5000textsampleswereusedtobuildWaldBo ostand400textcomponentsand13000non-textcomponentswereusedforCRFtraining.

2.1.3 OtherApproaches

Becausemanyvideosarestored,processed,andtransmittedinMPEGcompressedformat,somem ethodsusetheDCTcoefficientsprovidedbyMPEGtoextracttextincompresseddomaindirectly. Crandall*etal.*perform8×8block-wiseDCTonavideoframe.Foreachblock,19optimalcoefficientsthatbestcorrespondtotheprope rtiesoftextaredeterminedempirically.Thesumoftheabsolutevaluesofthesecoefficientsiscomp utedandregardedasameasureofthe"textenergy"ofthatblock.Ablockismarkedasatextblockifth e"textenergy"oftheblockislargerthanapre-definedthreshold.Similarly,Qian*etal.*[52]alsouseDCTtransformtoextracttextinMPEGformat. Thedifferenceisonlysevenof DCTcoefficients(threehorizontal,threeverticalandonediagonal)are selectedtocompute"textenergy".

Basedonthepartpropertyoftext,Markovmodelisusedfortextlocalizationbyregardingeachchara cterofthetextasapart.Zhang*etal.*[53]proposeaparts-basedapproach forscenetextdetectionusingahigh-order*MarkovRandomField*

(MRF)modelwithbeliefpropagation,whichcanovercomethelimitationofthepairwiseMRFthat spatialrelationshipofthreecharacterscannotbecaptured.Byutilizingtemporalinformation,Liu*et al.*[54]extend*DiscriminativeRandomField*(DRF)from2Dto3Dbyintegrating bothintra-frameneighbors andinter-frameneighbors,andbuild twographicalmodelsthatcombinetheDRFwiththe*HiddenMarkovModel*(HMM)todetecttexto bjectsinvideos.

SparserepresentationisadoptedbyPan*etal.*[55][56]tosegmenttextobjects fromcomplexbackground.Basedontheassumptionthatanimageiscomposedoftextandcomplex background, twodictionariesarechosen.One givesthe sparse representationtothetextforegroundandthenon-sparserepresentationtocomplexbackgroundwhiletheotheronedoestheopposite.Thenimageisd ecomposedandtext isextractedbythresholding.SparserepresentationisalsoadoptedbyZhao*etal.*[57].

Byexploringtheuniquepropertiesofcertainvideogenresandlanguages,somegenre-dependentandlanguage-dependenttextextractionapproachesareproposedbasedondomainknowledge.Theapproachesp roposedby Tang*etal.*[58]andChoudary*etal.*[59]are both designed toextract textininstructionalvideos.In [58]a unifiedapproach isusedtodetectvarioustextobjects,suchashandwritingonboard,handwrittenslides,electronicsli des,bookpages, andwebpages,basedonthepropertyofinstructionalvideothattherearemanycomparativelystatic periodsoftimewhichcorrespondtotheframeswithhighlightedtextforexplaining.Therefore,thel owmotionframesareextractedusingthedifferenceoftwoconsecutiveframes.ThenaSVMclassif ierisusedtoverifytextframesbyedgefeatures.[59]focusesontheinstructionalvideosofchalkboar dpresentationsandusethecontentfluctuationcurvebymeasuringthenumberofchalkpixelstodes cribethecontentineachframeofinstructionalvideos.Theframeswithenoughchalkpixelsareextra ctedaskeyframes.Hausdorffdistanceisusedinbothapproachestoremovetheredundanttextfram escausedbythesametextcontenthighlightedmultipletimes.

Wu*etal.*[60][61]presentanapproachtodetectandlocatetextonroadsignsinvideos.Thisstrategy iscomposedoftwostages:localizingroadsignsanddetectingtext.Byassumingthatthemotionofc ameraisalongitsopticalaxishorizontallyandscenetextliesonplanarsurfaces,candidateroadsignr egionsareselected.Then,edge-basedcoarsedetection,GaussianMixture Model(GMM)

basedcoloranalysis,
andtextalignmentanalysisareemployedtodetecttextondetectedroadsignregions.
BasedontheobservationthatDevanagariandBangla,twomostpopularIndianscripts,havecertain headlinewhichisanuniquecharacteristicofthescripts,Bhattacharya*etal.*[62]useCCsgeneratedb yOtsu'sbinarizationandmorphologicalopeningtoextractheadlines.Textregionsarelocalizedus ingtheobtainedheadlinesandneighboringrelationshipbetweencomponents.ThepropertiesofFa rsi/Arabictextisanalyzedandcomparedwithotherlanguagesby M.Moradi*etal.*[63].Cornerdetec tionisusedtolocalizecandidateregions,whichisverifiedusingDCT,statisticalfeatures,andSVM classifier.

2.2 TextTrackingApproaches

Text trackingstageutilizes thetemporal redundancyandspatial consistencypropertiesoftextobjectsinvideodocumentstospeedupdetectionandlocalizationofth esametextinconsecutiveframes.Moreover,Texttrackingalsoservesasatextverificationstagetor emovefalsealarms.Comparedtotextdetectionandlocalization,however,onlyafewworksarerep ortedfortexttrackinginrecent years.
Crandall*etal.*presenttwotrackingalgorithmstotrackrigidtextandchanging textin videos. ThemotionvectorsofMPEG-compressedvideosareusedinthefirst algorithm. Given a localized textregion in one frame, motion vectorsfromrelativelyfeaturelessmacroblocksandsmallnoisemotionvectorsarediscardedinord ertoselectreliablemotionvectors.Thenthemagnitudeanddirectionofremainingmotionvectorsa reclusteredandasinglemotionvectorfortheregionisyieldedbyaveragingvectorsinthecluster.
Atrajectory-predictionbased algorithmandanedge-pixel basedsearchalgorithmareadoptedtorefinethetrackingresults.Inthesecondalgorithmforchangin gtext,basedontheobservationthatthebasicshapesofthecharactersinthesametextobjectremainc onstant,thecontoursoftextobjectsintwoconsecutiveframesarematchedtodetermineiftheybelo ngtothesametextobjects.
Gallavata*etal.*[64]proposeasimilarmotion vectorbasedtexttrackingapproachforMEPGvideos.Amacroblockwhoseintersectionratioislar gerthanapredefinedthreshold isdefinedas atextmacroblock. Themode of the motion vector setischosenasthemotionvectorwhichpredictsthetextblockpositioninthenextframe.Compared with[**],onlythemotionvectorsintextmacro

blocksareextracted.Andinsteadofusingclusteringmethod,themodeofmotionvectorsisusedasreliablemotionvectors.

Basedonestimatedplanartransformationoverblocksofmultipleframes,Myers*etal.*[65]tracksceneetextundergoingscalechangesand3Dmotion.Theapproachassumesthatatextregionisplanarinthesceneandhasasufficientnumberofhightextured points, whichcanbecomputedbyLaplacian-of-Gaussian operator.These pointsaretrackedandlocalizedusing normalizedcorrelationofa

smallimagepatchcenteredatthecurrentpositionofthepoint.Afterreconstructingtheprojectingtransformationsforalltheframessimultaneouslybasedonthetrackedpoints,themotionoftextregionsisestimated.

Foradetectedtextblock,Tanaka*etal.*[66]findthebestmatchingblockintheotherframe bysearchingthe16×16blockscenteredattheblockonthecurrentimageforreductionofcalculationtime.Cumulativehistogramisusedtomeasurethesimilarityofblocks.Thelabelofaregionisdeterminedbythemostpopulatedlabelofblocksintheregion.

2.3 Applications

Clearly,themostimportantandwidelyusedapplicationofatextextractionsystemistext-basedimage/videoindexingandretrievalbyusingtherecognizedoutputsofthesystem.Besidesthat,manyotherapplicationsoftextextractionsystemhavebeendevelopedaswell,suchascamera-basedtextreadingsystemforvisuallyimpairedpeople,wearabletranslationrobot,wearabletext-trackingsystem,vehiclelicenseplaterecognition,androadsigntextdetectionformobiledevice. Inthissection,wegiveabriefintroductionoftheseapplications.

- C*amera-basedtextreadingsystemforblindpersons*:Figure2.5depictsthesystemdesignedEzaki *etal.*[67][68].Thetextcapturedbycameracanberecognizedandconvertedtovoicesignalbyasynthesizer.In[67],thetextextractionstageisimplementedusingedgeinformation,colorinformation,andmorphologyoperations.In[97],thetextblocksareextractedbyfindingbimodalhistogramusingFisher'sDiscriminantRateandOtsu'sbinarizationmethod.

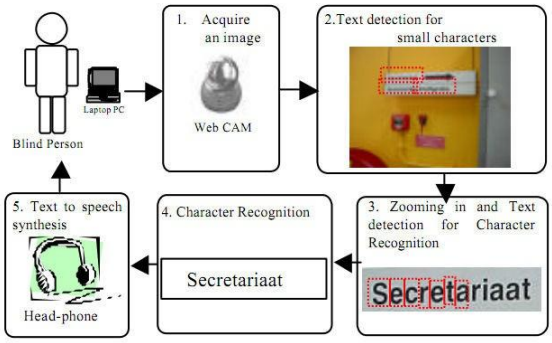


Figure2.4Camera-basedtextreadingsystem(FromEzaki*etal.*[67]).

- *Wearabletranslationrobot*:Shi*etal.*[69]developedawearabletranslationrobotwhichc anautomaticallyextracttextobjectsfromthevideocapturedbythecameraandtranslate multiplelanguagesinreal-time.Therobotismadeupofahead-mountedcamera,wearablecomputer,andahead-mounteddisplay.Theimplementation ofthetext detection is based on the difference ofgraylevel betweentext andbackground,which is computed usingthe difference curve that indicates thechangesofgraylevel.
- *Autonomoustextcapturingandtrackingrobot*:TherobotdesignedbyTanaka*etal.*[70]is equipedwithahead-mountedcamera,NTSC-DVconverter,andtwolaptops.Therobotcanlookaroundtheenvironmentusinganonboa rdactivevideocamera.Thetechnique supportingthe robotisDCTcoefficientbased textdetectionandblock

basedtrackingmethods.Similardevicesarealsousedinthetext-trackingwearablecamerasystemdevelopedbyTanakaandGoto[71][72].

- *Roadsignandtextdetection*:Basedontheobservationthatsignregionhashomogeneousc olorandholesthatarethecharacterslocatedinthesign,Bounman*etal.*[##]proposealowc omplexitysignandtextlocalizationmethodformobileapplicationsusingadividingand mergingmethodologyandaregiongrowingtechnique.Wu*etal.*[85][86]detectandlocat etextonroadsignsinvideosbyassumingthemotionofcameraisalongitsopticalaxishoriz ontallyandscenetextliesonplanarsurfaces.Signtextislocalizedbyedge,color,andalign mentanalysis.
- *Vehiclelicenseplaterecognition*:Thecharacterscontainedinthevehiclelicenseplatecan providemuchuseful information forlicense platedetection

andrecognition.Forexample,Wazalwar*etal.*[73]computeEulernumberofacandidatep lateregiontoverifylicenseplatebecausealicenseplatecontainsmanyholescausedbythe characters.Huang*etal.*[74]useHarriscornerdetectortoextractcornersoftextincandidat eplateregiontolocalizelicenseplate.

2.4 StateoftheArtandLimitations

Textextractioninimage andvideodocuments,as animportantresearchbranchofcontent-based information retrievaland indexing,continues tobeatopicof much interesttoresearchers.Within-depthinvestigationofthepropertiesoftextobjectsandthecomprehensiveunderstandingoftextextr actionproblem,alargenumberofrecentlyproposedapproachesintheliteraturehavecontributedtoa nimpressiveprogress in this research area. In this section, we summarize the progress made in recent years based onthe approaches reviewed in the previous sections and discuss future research directions.For text detection and localization, prior to 2003, very little work was done onscene text which usually has varying size, position, orientation, lighting, andtransformation [4]. Now, scene text extraction has received extensive study and manyapproaches were presented in recent years. For example, [12] uses ridges detected atseveral scales of a Laplacian image to extract scene text objects, some author uses the text colorextracted from the focus of mobile camera as the seed color to localize scene text, andone author[79] uses the stroke features computed based on gradient directions of edge points todetect scene text. Meanwhile, temporal nature of video documents, which were

considered by only a few approaches before 2003 [82], has been utilized by almost all recent video-based text extraction approaches, such as the minimum and maximum intensities based integrated frames used in [23], the edge and color similarities of consecutive frames used in [22][48], and the multi-frame based fuzzy clustering ensemble used in [63], and so on. Moreover, many techniques that were proposed for other research fields have been successfully adopted and applied for text extraction recently, including the mean shift algorithm initially proposed for face detection and tracking , discriminative random fields initially proposed for man-made building detection , Sparse representation initially proposed for the research on the receptive fields of simple cells , histogram of oriented gradient initially proposed for human detection , inner distance initially proposed for shape classification [50]. Besides the study of text regions, some recent approaches

focus on the background of text and use background information to localize text. For instance, [80] captures transient color between text and background to find captions in videos. [81] first find the uniform background with holes as the text background and then localize text in the marked regions. Although much progress has been achieved in the past few years, more work still remains to further improve the performance of text extraction systems. For example, the extraction of single characters and touching characters has not been solved because of the lack of important spatial information between neighboring characters; Transparent text, text in low resolution, text with varying lighting, and text with color bleeding cannot be localized accurately due to unreliable, insufficient, or incomplete character information; How to combine several complementary character features or extraction algorithm outputs efficiently still need more investigation. Although text tracking is an indispensable step for text extraction in videos, only a few text tracking approaches have been reported in recent years when compared with text detection and localization approaches. More effort is needed to focus on tracking, not only for static and scrolling text, but also for dynamic (growing, shrinking, and rotating) text objects.

CHAPTER 3

There are few basic principle which I have used exclusively during our project completion or during extraction of TEXT from image &video ,The main operation is MORPHOLOGICAL OPERATION which I have discussed in this chapter in more detailed way & rest steps involved discussed in neither very briefly nor in more detailed ,without going into much detail of project I would like to list all those steps which have been used in project here as followings

1. **SEGMENTATION**
2. **THRESHOLDING**
3. **BINARIZATION**
4. **MORPHOLOGICAL OPERATION**

all above mentioned four steps is detailed in section 4.1,4.2,4.3&4.4 of this chapter **,**as of followingafter a short description of these four steps, these four steps are directly involved in many times or at least once in our project ,but besides all above four steps ,there is one more term/step i.e EDGE DETECTION (which is incorporated partially or fully in almost all type of detection & extraction process)**,**so EDGE DETECTION is also the focused topic to discuss in the last of this chapter in more detailed way

3.1 SEGMENTATION

In Image processing, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s). When applied to a stack of images

Example of Segmentations:: Simple Scenes Segmentations of simple gray-level images can provide useful information about the surfaces in the scene.

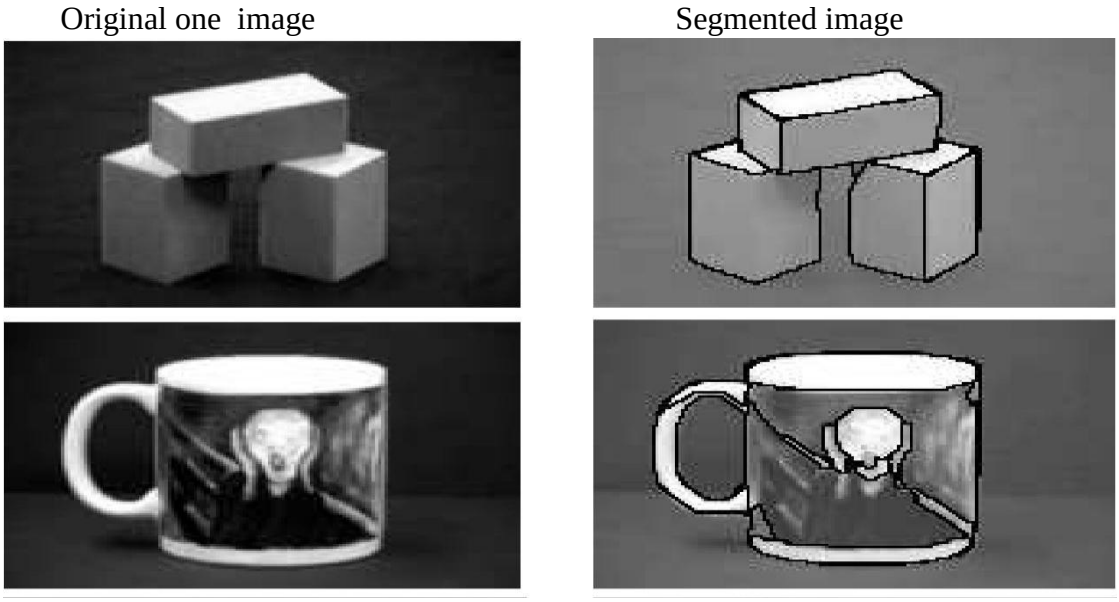


Figure 3.1 Example of Segmentation

Applications ::

Some of the practical applications of image segmentation are:

>>Content-based image retrieval

>>Machine vision

>>Medical imaging

- Locate tumors and other pathologies
- Measure tissue volumes
- Diagnosis, study of anatomical structure
- Surgery planning
- Virtual surgery simulation
- Intra-surgery navigation

>>Object detection

- Pedestrian detection
- Face detection
- Brake light detection
- Locate objects in satellite images (roads, forests, crops, etc.)

>>Recognition Tasks

- Face recognition
- Fingerprint recognition
- Iris recognition

>>Traffic control systems

>>Video surveillance

Several general-purpose algorithms and techniques have been developed for image segmentation. To be useful, these techniques must typically be combined with a domain's specific knowledge in order to effectively solve the domain's segmentation problems.

Importance of Segmentation ::

- Segmentation is generally the first stage in any attempt to analyze or interpret an image automatically.
- Segmentation bridges the gap between low-level image processing and high-level image processing.
- Some kinds of segmentation technique will be found in any application involving the detection, recognition, and measurement of objects in images.
- The role of segmentation is crucial in most tasks requiring image analysis. The success or failure of the task is often a direct consequence of the success or failure of segmentation.
- However, a reliable and accurate segmentation of an image is, in general, very difficult to achieve by purely automatic means

Descriptive Definition of Segmentation ::

• Segmentation subdivides an image into its constituent regions or objects. That is, it partitions an image into distinct regions that are meant to correlate strongly with objects or features of interest in the image.

• Segmentation can also be regarded as a process of grouping together pixels that have similar attributes.

• The level to which the subdivision is carried depends on the problem being solved. That is, Segmentation should stop when the objects of interest in an application have been isolated. There is no point in carrying segmentation past the level of detail required to identified those elements .

• It is the process that partitions the image pixels into non-overlapping regions such that:

1. Each region is homogeneous (i.e., uniform in terms of the pixel attributes such asintensity, color, range, or texture, and etc.) and connected.

2. The union of adjacent regions is not homogeneous.

Image segmentation Steps (in Algorithmic Form) ::

- All pixels must be assigned to regions.
- Each pixel must belong to a single region only.
- Each region must be uniform.
- Any merged pair of adjacent regions must be non-uniform.
- Each region must be a connected set of pixels.

Image Segmentation Strategies OR Base of Categorization of Segmentation ::

- Image segmentation algorithms generally are based on one of two basic properties of intensity values: discontinuity and similarity.

- Discontinuity based approach: Partition an image based on abrupt changes in intensity.

- Similarity based approach: Partition an image based on regions that are similar according to a set of predefined criteria. (a) Thresholding (b) Region growing (c) Region splitting & merging

Techniques based on discontinuity attempt to partition the image by detecting abrupt changes in gray level. Point, line, and edge detectors.

Techniques based on similarity attempt to create the uniform regions by grouping together connected pixels that satisfy predefined similarity criteria. Therefore, the results of segmentation may depend critically on these criteria and on the definition of connectivity.

The approaches based on discontinuity and similarity mirror one another in the sense that completion of a boundary is equivalent to breaking one region into two.

Thresholding :: The simplest method of image segmentation is called the thresholding method. This method is based on a clip-level (or a threshold value) to turn a gray-scale image into a binary image. [More Detail explanation about Thresholding is discussed in later section 4.2]

Clustering methods :: The K-means algorithm is an iterative technique that is used to partition an image into K clusters.

The basic algorithm is

1. Pick K cluster centers, either randomly or based on some heuristic method, for example K-means++
2. Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center
3. Re-compute the cluster centers by averaging all of the pixels in the cluster
4. Repeat steps 2 and 3 until convergence is attained (i.e. no pixels change clusters)

In this case, distance is the squared or absolute difference between a pixel and a cluster center. The difference is typically based on pixel color, intensity, texture, and location, or a weighted combination of these factors. K can be selected manually, randomly, or by a

heuristic. This algorithm is guaranteed to converge, but it may not return the optimal solution. The quality of the solution depends on the initial set of clusters and the value of K.

Compression-based methods ::

Compression based methods postulate that the optimal segmentation is the one that minimizes, over all possible segmentations, the coding length of the data. The connection between these two concepts is that segmentation tries to find patterns in an image and any regularity in the image can be used to compress it. The method describes each segment by its texture and boundary shape. Each of these components is modeled by a probability distribution function and its coding length is computed as follows:

1. The boundary encoding leverages the fact that regions in natural images tend to have a smooth contour. This prior is used by Huffman coding to encode the difference chain code of the contours in an image. Thus, the smoother a boundary is, the shorter coding length it attains.
2. Texture is encoded by lossy compression in a way similar to minimum description length (MDL) principle, but here the length of the data given the model is approximated by the number of samples times the entropy of the model. The texture in each region is modeled by a multivariate normal distribution whose entropy has a closed form expression. An interesting property of this model is that the estimated entropy bounds the true entropy of the data from above. This is because among all distributions with a given mean and covariance, normal distribution has the largest entropy. Thus, the true coding length cannot be more than what the algorithm tries to minimize.

For any given segmentation of an image, this scheme yields the number of bits required to encode that image based on the given segmentation. Thus, among all possible segmentations of an image, the goal is to find the segmentation which produces the shortest coding length. This can be achieved by a simple agglomerative clustering method. The distortion in the lossy compression determines the coarseness of the segmentation and its optimal value may differ for each image. This parameter can be estimated heuristically from the contrast of textures in an image. For example, when the textures in an image are

similar, such as in camouflage images, stronger sensitivity and thus lower quantization is required.

Histogram-based methods ::

Histogram-based methods are very efficient compared to other image segmentation methods because they typically require only one pass through the pixels. In this technique, a histogram is computed from all of the pixels in the image, and the peaks and valleys in the histogram are used to locate the clusters in the image. Color or intensity can be used as the measure.

A refinement of this technique is to recursively apply the histogram-seeking method to clusters in the image in order to divide them into smaller clusters. This operation is repeated with smaller and smaller clusters until no more clusters are formed.

One disadvantage of the histogram-seeking method is that it may be difficult to identify significant peaks and valleys in the image.

Histogram-based approaches can also be quickly adapted to apply to multiple frames, while maintaining their single pass efficiency. The histogram can be done in multiple fashions when multiple frames are considered. The same approach that is taken with one frame can be applied to multiple, and after the results are merged, peaks and valleys that were previously difficult to identify are more likely to be distinguishable. The histogram can also be applied on a per-pixel basis where the resulting information is used to determine the most frequent color for the pixel location. This approach segments based on active objects and a static environment, resulting in a different type of segmentation useful in video tracking.

Edge detection ::

Edge detection is a well-developed field on its own within image processing. Region boundaries and edges are closely related, since there is often a sharp adjustment in intensity at the region boundaries. Edge detection techniques have therefore been used as the base of another segmentation technique.

The edges identified by edge detection are often disconnected. To segment an object from an image however, one needs closed region boundaries. The desired edges are the boundaries between such objects or spatial-taxons.

Spatial-taxons are information granules, consisting of a crisp pixel region, stationed at abstraction levels within a hierarchical nested scene architecture. They are similar to the Gestalt psychological designation of figure-ground, but are extended to include foreground, object groups, objects and salient object parts. Edge detection methods can be applied to the spatial-taxon region, in the same manner they would be applied to a silhouette. This method is particularly useful when the disconnected edge is part of an illusory contour

Segmentation methods can also be applied to edges obtained from edge detectors. Lindeberg and Li developed an integrated method that segments edges into straight and curved edge segments for parts-based object recognition, based on a minimum description length (MDL) criterion that was optimized by a split-and-merge-like method with candidate breakpoints obtained from complementary junction cues to obtain more likely points at which to consider partitions into different segments, This EDGE DETECTION is discussed later on in more detailed way.

Region growing method ::

Actually region growing, clustering & Split and Merge all three comes into the category of **"Region Segmentation"** ,I have already discussed about "clustering" in previous subsection of same section 4.1 ,Now here is "Region Growing" & "Split and merge" discussion as follows

Region-growing methods rely mainly on the assumption that the neighboring pixels within one region have similar values. The common procedure is to compare one pixel with its neighbors. If a similarity criterion is satisfied, the pixel can be set to belong to the cluster as one or more of its neighbors. The selection of the similarity criterion is significant and the results are influenced by noise in all instances.

The method of Statistical Region Merging(SRM) starts by building the graph of pixels using 4-connectedness with edges weighted by the absolute value of the intensity

difference. Initially each pixel forms a single pixel region. SRM then sorts those edges in a priority queue and decide whether or not to merge the current regions belonging to the edge pixels using a statistical predicate.

One region-growing method is the seeded region growing method. This method takes a set of seeds as input along with the image. The seeds mark each of the objects to be segmented. The regions are iteratively grown by comparison of all unallocated neighboring pixels to the regions. The difference between a pixel's intensity value and the region's mean, δ , is used as a measure of similarity.The pixel with the smallest difference measured in this way is assigned to the respective region. This process continues until all pixels are assigned to a region. Because seeded region growing requires seeds as additional input, the segmentation results are dependent on the choice of seeds, and noise in the image can cause the seeds to be poorly placed.

Another region-growing method is the unseeded region growing method. It is a modified algorithm that does not require explicit seeds. It starts with a singleregion A{1}—the pixel chosen here does not markedly influence the final segmentation. At each iteration it considers the neighboring pixels in the same way as seeded region growing. It differs from seeded region growing in that if the minimum 'δ' is less than a predefined threshold T then it is added to the respective region A{j}. If not, then the pixel is considered different from all current regions A{i} and a new region A{n+1} is created with this pixel.

One variant of this technique, proposed byHaralick and Shapiro (1985), is based on pixel intensities. The mean and scatter of the region and the intensity of the candidate pixel are used to compute a test statistic. If the test statistic is sufficiently small, the pixel is added to the region, and the region's mean and scatter are recomputed. Otherwise, the pixel is rejected, and is used to form a new region.

A special region-growing method is called λ segmentation (see also lambda-connectedness). It is based on pixel intensities and neighborhood-linking paths. A degree of connectivity (connectedness) is calculated based on a path that is formed by pixels. For a

certain value of λ , two pixels are called λ connected if there is a path linking those two pixels and the connectedness of this path is at least λ. λ connectedness is an equivalence relation.

Principle/Idea of Region Growing :

- Region growing is the simplest region based segmentation that groups pixels or sub-regions into larger regions based on pre-defined criteria
- The pixel aggregation starts with a set of "seed" points in a way that the corresponding regions grow by appending to each seed points those neighboring pixels that have similar properties(such as gray level,texture,color,shape)
- Region growing based techniques are better than the other techniques in noisy images where edges are difficult to detect .
- The seed point can be selected either by a human or automatically by avoiding areas of high contrast (large gradient) .

Figure 3.2 (Explaining Region Growing Segmentation) .

Split-and-merge segmentation is based on a quad tree partition of an image. It is sometimes called quad tree segmentation. This method starts at the root of the tree that represents the whole image. If it is found non-uniform (not homogeneous), then it is split into four child squares (the splitting process), and so on. If, in contrast, four child squares are homogeneous, they are merged as several connected components (the merging process). The node in the tree is a segmented node. This process continues recursively until no further splits or merges are possible When a special data structure is involved in the

implementation of the algorithm of the method, its time complexity can reach**O(*n*log*n*)**, an optimal algorithm of the method

The basic idea of region Split & Merging is to break the image into a set of disjoint regions which are coherent within themselves:

- Initially take the image as a whole to be the area of interest.
- Look at the area of interest and decide if all pixels contained in the region satisfy some similarity constraint.
- If TRUE then the area of interest corresponds to a region in the image.
- If FALSE split the area of interest (usually into four equal sub-areas) and consider each of the sub-areas as the area of interest in turn.
- This process continues until no further splitting occurs. In the worst case this happens when the areas are just one pixel in size.
- This is a divide and conquer or top down method.

If only a splitting schedule is used then the final segmentation would probably contain many neighboring regions that have identical or similar properties.

Thus, a merging process is used after each split which compares adjacent regions and merges them if necessary. Algorithms of this nature are called **split and merge** algorithms.

To illustrate the basic principle of these methods let us consider an imaginary image.

- Let I denote the whole image shown in Fig 3.3 (a).
- Not all the pixels in I are similar so the region is split as in Fig 3.3 (b).
- Assume that all pixels within regionsI_1 , I_2 andI_3 respectively are similar but those in are not.
- Therefore I_4 is split next as in Fig 3.3 (c).
- Now assume that all pixels within each region are similar with respect to that region, and that after comparing the split regions, regions I_{43} and I_{44} are found to be identical.
- These are thus merged together as in Fig 3.3 (d).

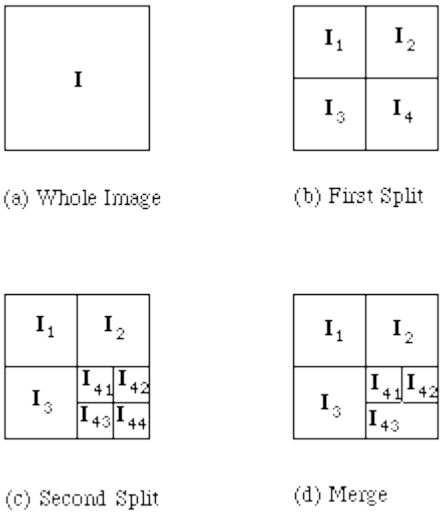


Figure 3.3 Example of region splitting and merging

We can describe the splitting of the image using a tree structure, using a modified quadtree. Each non-terminal node in the tree has at most four descendants, although it may have less due to merging.

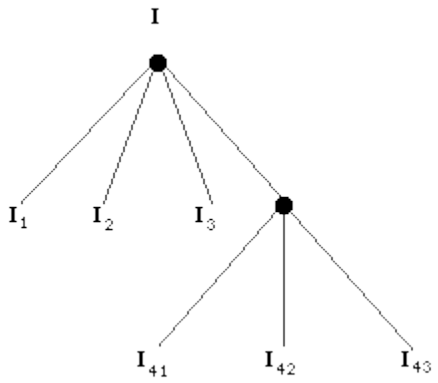


Figure 3.4 Region splitting and merging tree

Watershed Transformation ::

The watershed transformation considers the gradient magnitude of an image as a topographic surface. Pixels having the highest gradient magnitude intensities (GMIs) correspond to watershed lines, which represent the region boundaries. Water placed on any pixel enclosed by a common watershed line flows downhill to a common local intensity

minimum (LIM). Pixels draining to a common minimum form a catch basin, which represents a segment.

There are different technical definitions of a watershed. In graphs, watershed lines may be defined on the nodes, on the edges, or hybrid lines on both nodes and edges. Watersheds may also be defined in the continuous domain. There are also many different algorithms to compute watersheds. Watershed algorithm is used in image processing primarily for segmentation purposes

General Definition :A drainage basin or watershed is an extent or an area of land where surface water from rain melting snow or ice converges to a single point at a lower elevation, usually the exit of the basin, where the waters join another waterbody, such as a river, lake, wetland, sea, or ocean .

Figure 3.5 Example to understand Watershed

The Watershed transformation is a powerful tool for image segmentation, it uses the region-based approach and searches for pixel and region similarities.

We will represent a gray-tone image by a function: : $\mathbb{Z}^2 \rightarrow \mathbb{Z}$,$f(x)$ is the gray value of the image at point x

A section of f at level i is a set $X_i(f)$ defined as:

$$X_i(f) = \{x \in \mathbb{Z}^2 : f(x) \geq i\}$$

And in the same way we define $Z_i(f)$ as:

$$Z_i(f) = \{x \in \mathbb{Z}^2 : f(x) \leq i\}$$

$$\Leftrightarrow X_i(f) = Z^c{}_{i+1}(f)$$

If we look at the image f as a topographic surface, imagine that we pierce each $M_i(f)$ of the topographic surface S and then we plunge this surface into a lake, the water entering through the holes floods the surface and if two or more floods coming from different minima attempt to merge, we avoid this event by building a dam on the points of the surface where the floods would merge.

At the end of the process only these dams will emerge and this is what define the watershed of the function f

Image gradient :An image gradient is a directional change in the intensity or color in an image. Image gradients may be used to extract information from images. And understanding image gradient is very much necessary to understand about watershed

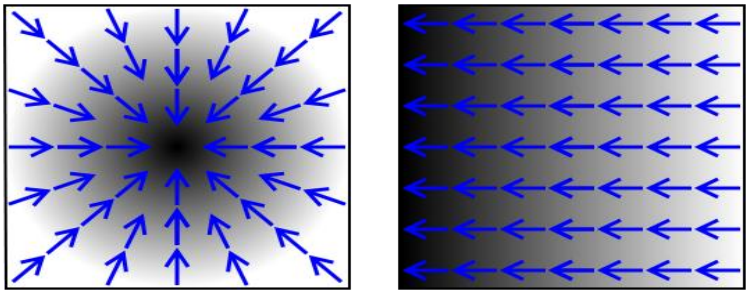


Figure 3.6 Image gradient

Inter-pixel watershed :

S. Beucher and F. Meyer introduced an algorithmic inter-pixel implementation of the watershed method, given the following procedure:

1. Label each minimum with a distinct label. Initialize a set S with the labeled nodes.

2. Extract from S a node x of minimal altitude F, that is to say $F(x) = \min\{F(y)|y \in S\}$. Attribute the label of x to each non-labeled node y adjacent to x, and insert y in S.

3. Repeat Step 2 until S is empty

Topological watershed :

The topological watershed was introduced in 1997, and beneficiate of the following fundamental property. A function W is a watershed of a function F if and only if $W \leq F$ and W preserves the contrast between the regional minima of F; where the contrast between two regional minima M1 and M2 is defined as the minimal altitude to which one must climb in order to go from M1 to M2

Algorithms :

Different approaches may be employed to use the watershed principle for image segmentation.

- Local minima of the gradient of the image may be chosen as markers, in this case an over-segmentation is produced and a second step involves region merging.
- Marker based watershed transformation make use of specific marker positions which have been either explicitly defined by the user or determined automatically with morphological operators or other ways.

Meyer's flooding algorithm :

One of the most common watershed algorithms was introduced by F. Meyer in the early 90's.

The algorithm works on a gray scale image. During the successive flooding of the grey value relief, watersheds with adjacent catchment basins are constructed. This flooding process is performed on the gradient image, i.e. the basins should emerge along the edges. Normally this will lead to an over-segmentation of the image, especially for noisy image material, e.g. medical CT data. Either the image must be pre-processed or the regions must be merged on the basis of a similarity criterion afterwards.

- A set of markers, pixels where the flooding shall start, are chosen. Each is given a different label.
- The neighboring pixels of each marked area are inserted into a priority queue with a priority level corresponding to the gradient magnitude of the pixel.
- The pixel with the lowest priority level is extracted from the priority queue. If the neighbors of the extracted pixel that have already been labeled all have the same label, then the pixel is labeled with their label. All non-marked neighbors that are not yet in the priority queue are put into the priority queue.
- Redo step 3 until the priority queue is empty.

3.2 THRESHOLDING

Thresholding is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images

Definition :: The simplest thresholding methods replace in an image with a black pixel if the image intensity I{i,j} is less than some fixed constant T (that is,I{i,j}<T} or a white pixel if the image intensity is greater than that constant. In the example image(Shown below), this results in the dark tree becoming completely black, and the white snow becoming completely white

Original one image Threshold effect created image

Figure 3.7Thresholding example

As we have seen in Segmentation,Segmentation involves separating an image into regions (or their contours) corresponding to objects. We usually try to segment regions by identifying common properties. Or, similarly, we identify contours by identifying differences between regions (edges).

The simplest property that pixels in a region can share is intensity. So, a natural way to segment such regions is through thresholding, the separation of light and dark regions. Thresholding creates binary images from grey-level ones by turning all pixels below some threshold to zero and all pixels about that threshold to one. (What you want to do with pixels at the threshold doesn't matter, as long as you're consistent.)

Problems of Thresholding :: The major problem with thresholding is that we consider only the intensity, not any relationships between the pixels. There is no guarantee that the pixels identified by the thresholding process are contiguous. We can easily include extraneous pixels that aren't part of the desired region, and we can just as easily miss isolated pixels within the region (especially near the boundaries of the region). These effects get worse as the noise gets worse, simply because it's more likely that a pixels intensity doesn't represent the normal intensity in the region.

When we use thresholding, we typically have to play with it, sometimes losing too much of the region and sometimes getting too many extraneous background pixels. (Shadows of objects in the image are also a real pain—not just where they fall across another object but where they mistakenly get included as part of a dark object on a light background.)

The another major problem of thresholding is how to select the optimal value of threshold T. Usually in order to get the optimal value of T, we need to statically analyze the so-called "histogram" (or "intensity histogram") of the input gray image.

Type of Thresholding& Briefly Explanation of Each Type ::

Optimal Global Thresholding :

•A threshold is said to be globally optimal if the number of misclassified pixels is minimum

– Histogram is bimodal (object and background)

– Ground truth is known OR the histograms of the object and the background are known

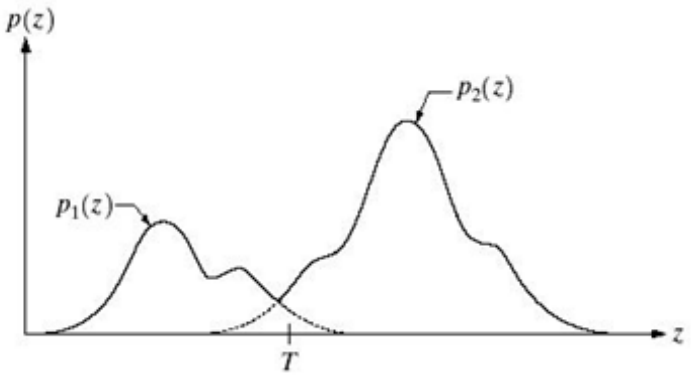


Fig 3.8This pic is explaining about gray level probability density function of two regions in an image.

• If the individual (normalized) histograms are known as p1(z) and p2(z)

– The (normalized) histogram of the overall image is P1p1(z)+P2p2(z) where P1=N1/(N1+N2) and P2=N2/(N1+N2) Notice that P1+P2=1

[Probability of Error]

>> Probability of erroneously classifying a class 1 pixel to

class 2 is E2 (T) = integrating p1(z) from T to Infinity

>> Probability of erroneously classifying a class 2 pixel to class 1 is

E1(T) = Integrating p2(z) from –(infinity) to T

>> Overall probability of error is then E(T)=P2E1(T)+P1E2(T)

Minimum Error Thresholding :

Assume object pixels are distributed (histogram) according to pO (x) and the background pixels by pB (x) . The error is misclassifying object pixels is

$$\int_{-\infty}^{t} p_O(x)\,dx$$

And misclassifying background pixels as object pixels is

$$\int_{t}^{\infty} p_B(x)dx$$

Let θ be the fraction of pixels in the object. Then the total error is …

$$E(t) = \theta \int_{-\infty}^{t} p_O(x)dx + (1-\theta)\int_{t}^{\infty} p_B(x)dx$$

To find the minimum $E(t)$ we differentiate …

$$0 = \partial E / dt = \theta p_O(t) - (1-\theta)\, p_B(t)$$

OR

$$\theta\, p\, O\, (t) = (1-\theta)\, pB\, (t)$$

Local Thresholding :

Another problem with global thresholding is that changes in illumination across the scene may cause some parts to be brighter (in the light) and some parts darker (in shadow) in ways that have nothing to do with the objects in the image.

We can deal, at least in part, with such uneven illumination by determining thresholds locally. That is, instead of having a single global threshold, we allow the threshold itself to smoothly vary across the image.

Another way of dealing with illumination is to consider

$$\Rightarrow \quad g(x, y) = r(x, y) * i(x, y)$$

$$\Rightarrow \quad \ln g(x, y) = \ln r(x, y) + \ln i(x, y)$$

Hysteresis Thresholding :

Take into account neighbors.

1. Choose two thresholds T_{high} and T_{low}
2. If $T > T_{high}$ the pixel is in the body. If $T < T_{low}$ the pixel is in the background.
3. If $T_{low} < T < T_{high}$ the pixel is in the body only if a neighbor is already in the body
4. Iterate

Automated Methods for Finding Thresholds ::

To set a global threshold or to adapt a local threshold to an area, we usually look at the histogram to see if we can find two or more distinct modes—one for the foreground and one for the background.

Recall that a histogram is a probability distribution:

$$p(g)=ng\ /n$$

That is, the number of pixels *ng* having grayscale intensity *g* as a fraction of the total number of pixels *n*. Here are five different ways to look at the problem:

(a) Known Distribution
(b) Clustering (K-Means Variation)
(c) Clustering (The Otsu Method)
(d) Mixture Modeling
(e) Pick & valley Distribution

Multispectral Thresholding ::

We are interested in a technique for segmenting images with multiple components (color images, Landsat images, or MRI images with T1, T2, and proton- density bands). It works by estimating the optimal threshold in one channel and then segmenting the overall image based on that threshold. We then subdivide each of these regions independently using properties of the second channel. We repeat it again for the third channel, and so on, running through all channels repeatedly until each region in the image exhibits a distribution indicative of a coherent region (a single mode).

e.g.

1. Compute histogram for each channel separately.
2. Find the peak in each histogram,Select two thresholds corresponding to some valley on each side of these peaks.Segment image into two regions. One between these thresholds and one outside.
3. Project into multi-spectral representation

Thresholding Along Boundaries ::

If we want our thresholding method to give stay fairly true to the boundaries of the object, we can first apply some boundary-finding method (such as edge detection techniques) and then sample the pixels only where the boundary probability is high.

Thus, our threshold method based on pixels near boundaries will cause separations of the pixels in ways that tend to preserve the boundaries. Other scattered distributions within the object or the background are of no relevance.

However, if the characteristics change along the boundary, we're still in trouble. And, of course, there's still no guarantee that we'll not have extraneous pixels or holes.

Adaptive Thresholding ::

• Uneven illumination

– An image can be modeled as the product of a reflectance component r(x,y) and an illumination component i(x,y) as f (x, y) = i(x, y)r(x, y) where

$0 < i(x, y) < \infty$ and $0 < r(x, y) < 1$

– i(x,y): the amount of source illumination incident on the scene being viewed

– r(x,y): the amount of illumination reflected by the object

Uneven illumination makes an originally perfectly segmentable image into an image that can not be segmented satisfactorily using a single threshold

• One way to overcome the uneven illumination problem is to first estimate the uneven illumination and then correct it accordingly (rectification)

– Upon correction, global thresholding can be employed

• Another way is to use adaptive thresholding by partition the original image into several subimages and utilize global thresholding techniques for each subimage

– Key issues are how to partition the image and how to estimate the threshold for each sub-image

Major procedures used for this example ::

– Dividing the image into sub-images

– Testing for bimodality for each sub-image

– Apply Optimal Global Thresholding for each identified image with bimodal histogram

Multi-Spectral thresholding ::

Many practical segmentation problems need more information than is contained in one spectral band. Color images are natural example, in which information are coded in three spectral band for example, red, green, &blue .multi spectral remote sensing images or satellite images may have even more spectral bands. One segmentation approach determines thresholds independently in each spectral band &Combined them into single segmented image

Algorithm of Multi-spectral thresholding:

1. Initialize the whole image as a single region.
2. Compute a smoothed histogram for each spectral band . Find the most significant peak in each histogram & determine two thresholds as local minima of either side of this maximum. Segment each region in each spectral band in sub-region according to these thresholds. Each segmentation in each spectral band is projected into multi spectral segmentation . Regions for the next processing steps are those in the multi-spectral image.
3. Repeat step 2 for each region of image until each region's histogram contains only one significant peak.

3.3 BINARIZATION

Principal stage of the any image analysis procedureis the binarization, according to which the pixels are classified into textand background. It is a crucial stage that can affect further stages includingthe final character recognition in last stage, image binarization (or thresholding) is the process that segments the gray-scale or color image into text and background by removing any existing degradations (such as bleed-through, large ink stains, non-uniform illumination and faint characters). It is an important pre-processing step of the image processing and analysis pipeline that affects further stages as well as the final Character Recognition stage.

Image binarization converts an image of up to 256 gray levels to a black and white image.Or bi-level (black & white) images. Frequently, binarization is used as a pre-processor before Final Stages .

The simplest way to use image binarization is to choose a threshold value, and classify all pixels with values above this threshold as white, and all other pixels as black. The problem then is how to select the correct threshold. In many cases, finding one threshold compatible to the entire image is very difficult, and in many cases even impossible. Therefore, adaptive image binarization is needed where an optimal threshold is chosen for each image area.Binarization is the process of converting a pixel image to a binary image

USES::In the old days binarization was important for sending faxes. These days its still important for things like digitalising text or segmentation or for different form of image analysis.

PROCESS::At first the image is converted into gray-scale,Then a threshold gets applied,The threshold can either be set fixed or adaptive using a clustering algorithm, Binarisation is the basis of segmentation....

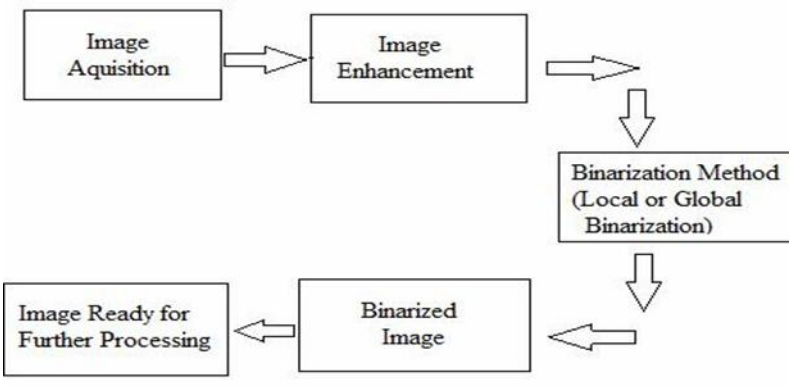


Figure 3.9 BLOCK DIAGRAM OF BINARIZATION

In A simple Fixed Thresholdingbinarization method fixed threshold value is used to assign 0's and 1's for all pixel positions in a given image. The basic idea for fixed binarization method is described as under $g(x,y) = 1$ if $f(x,y) >= T$; $g(x,y) = 0$,else

T shows global threshold value. For various threshold values in a simple fixed binarization methods the results are illustrated below.

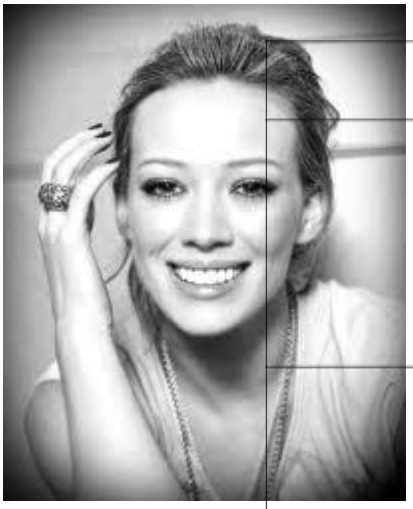


Figure 3.10 ORIGINAL IMAGE

Now Binarization Process is implemented on above image(Original Image) , Out put of Binarization on above Pic is shown as following Pic.

THRESHOLD VALUE	OUTPUT
T=108	

T=128

T=148

T=160

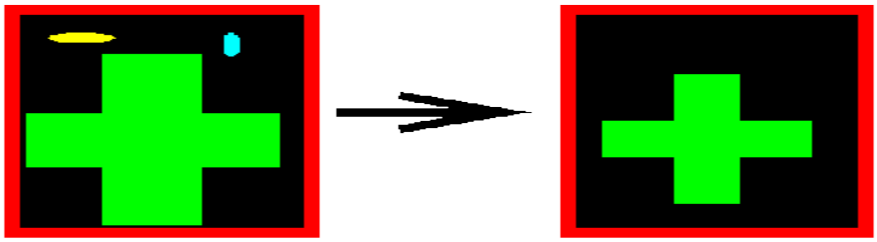
TABLE 3.1

As shown in table 3.1 outputs vary as given an input threshold value. From the above table it can be noticed that output at threshold T=160 is better for this image. But T=160 is not optimal threshold value for all the images. So it is very difficult to decide an optimal threshold value for a current input image. To overcome this difficulty we should go for more binarization techniques in which optimal threshold value can be computed according to input image.

3.4 MORPHOLOGY & MORPHOLOGICAL OPERATION

Morphology, is a word from the Greek and have meaning "study of shape" … & There is a term Mathematical morphology too, means a theoretical model based on lattice theory, used

for digital image processing ,In any of the field Morphology does mean always “study of shape” somehow or completely …



Binary images may contain numerous imperfections. In particular, the binary regions produced by simple thresholding are distorted by noise and texture. Morphological image processing pursues the goals of removing these imperfections by accounting for the form and structure of the image. These techniques can be extended to greyscale images

BASIC CONCEPTS OF MORPHOLOGY & ITS OPERATION ::

Morphological image processing is a collection of non-linear operations related to the shape or morphology of features in an image. According to sources which I referred, morphological operations rely only on the relative ordering of pixel values, not on their numerical values, and therefore are especially suited to the processing of binary images. Morphological operations can also be applied to greyscale images such that their light transfer functions are unknown and therefore their absolute pixel values are of no or minor interest.

Morphological techniques probe an image with a small shape or template called a structuring element. The structuring element is positioned at all possible locations in the image and it is compared with the corresponding neighborhood of pixels. Some operations test whether the element "fits" within the neighborhood, while others test whether it "hits" or intersects the neighborhood

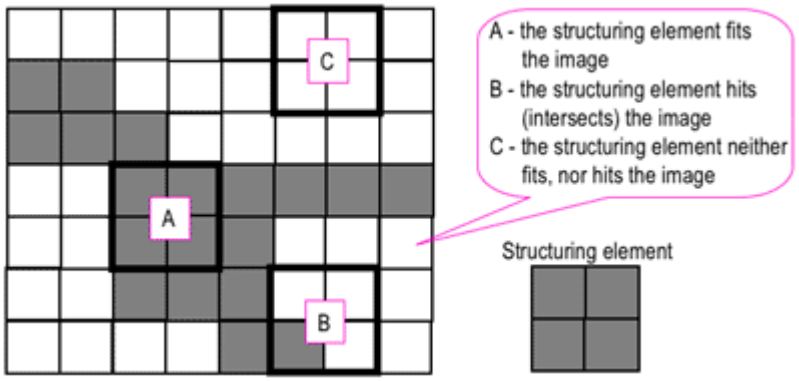


Figure 3.11 Probing of an image with a structuring element

(white and grey pixels have zero and non-zero values, respectively)

A morphological operation on a binary image creates a new binary image in which the pixel has a non-zero value only if the test is successful at that location in the input image.

The **structuring element** is a small binary image, i.e. a small matrix of pixels, each with a value of zero or one:

- The matrix dimensions specify the size of the structuring element.
- The pattern of ones and zeros specifies the shape of the structuring element.
- An origin of the structuring element is usually one of its pixels, although generally the origin can be outside the structuring element.

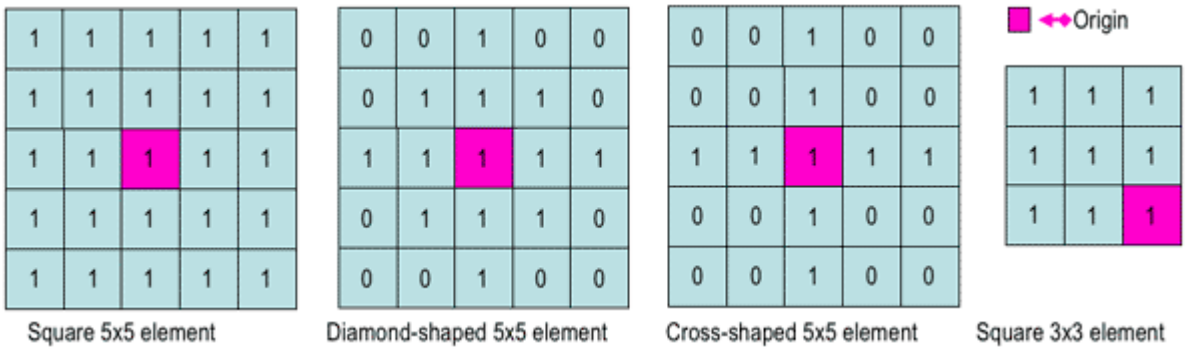


Figure 3.12 Examples of simple structuring elements.

A common practice is to have odd dimensions of the structuring matrix and the origin defined as the Centre of the matrix. Structuring elements play in morphological image processing the same role as convolution kernels in linear image filtering.

When a structuring element is placed in a binary image, each of its pixels is associated with the corresponding pixel of the neighborhood under the structuring element. The structuring element is said to fit the image if, for each of its pixels set to 1, the corresponding image pixel is also 1. Similarly, a structuring element is said to hit, or intersect, an image if, at least for one of its pixels set to 1 the corresponding image pixel is also 1.

		A	B	C
fit	s_1	yes	no	no
	s_2	yes	yes	no
hit	s_1	yes	yes	yes
	s_2	yes	yes	no

Figure 3.13 Fitting and hitting of a binary image with structuring elements s1 and s2.

Zero-valued pixels of the structuring element are ignored, i.e. indicate points where the corresponding image value is irrelevant.

FUNDAMENTAL MORPHOLOGICAL OPERATIONS ::

DIALATION [Common Name Dialate/Grow/Expand)]:: Dilation is one of the two basic operators in the area of mathematical morphology, the other being erosion. It is typically applied to binary images, but there are versions that work on grayscale images too. The basic effect of the operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels (i.e. white pixels, typically). Thus areas of foreground pixels grow in size while holes within those regions become smaller.

HOW IT WORKS ::

The dilation operator takes two pieces of data as inputs. The first is the image which is to be dilated. The second is a (usually small) set of coordinate points known as a structuring element (also known as a kernel). It is this structuring element that determines the precise effect of the dilation on the input image.

The mathematical definition of dilation for binary images is as follows:

- Suppose that X is the set of Euclidean coordinates corresponding to the input binary image, and that K is the set of coordinates for the structuring element.

- Let Kx denote the translation of K so that its origin is at x.
- Then the dilation of X by K is simply the set of all points x such that the intersection of Kx with X is non-empty.

The mathematical definition of grayscale dilation is identical except for the way in which the set of coordinates associated with the input image is derived. In addition, these coordinates are 3-D rather than 2-D.

As an example of binary dilation, suppose that the structuring element is a 3×3 square, with the origin at its center, as shown in Figure 4.14. Note that in this and subsequent diagrams, foreground pixels are represented by 1's and background pixels by 0's.

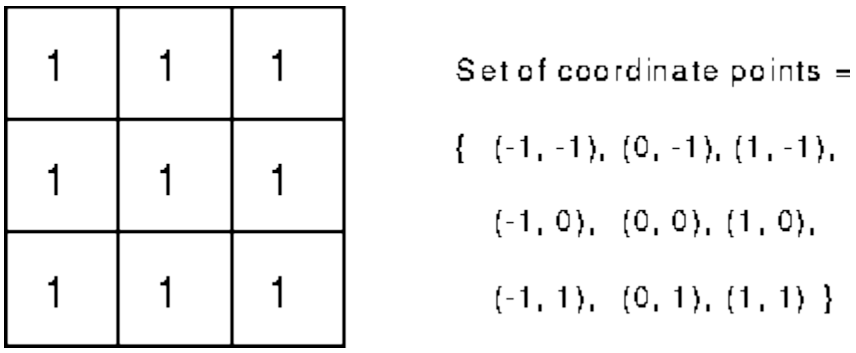


Figure 3.14 A 3*3 Square Structuring Element

For our example 3×3 structuring element, the effect of this operation is to set to the foreground color any background pixels that have a neighboring foreground pixel (assuming 8-connectedness). Such pixels must lie at the edges of white regions, and so the practical upshot is that foreground regions grow (and holes inside a region shrink).

Dilation is the dual of erosion i.e. dilating foreground pixels is equivalent to eroding the background pixels.

The dilation of an image f by a structuring element s (denoted f⊕s) produces a new binary image g=f⊕s with ones in all locations (x,y) of a structuring element's origin at which that structuring element s hits the the input image f, i.e. g(x,y) = 1 if s hits f and 0 otherwise,

repeating for all pixel coordinates (x,y). Dilation has the opposite effect to erosion -- it adds a layer of pixels to both the inner and outer boundaries of region

Results of dilation or erosion are influenced both by the size and shape of a structuring element. Dilation and erosion are dual operations in that they have opposite effects. Let f^c denote the complement of an image *f*, i.e., the image produced by replacing 1 with 0 and vice versa. Formally, the duality is written as

$$f \oplus s = f^c \ominus s_{rot}$$

where s_{rot} is the structuring element *s* rotated by 180°. If a structuring element is symmetrical with respect to rotation, then s_{rot} does not differ from *s*. If a binary image is considered to be a collection of connected regions of pixels set to 1 on a background of pixels set to 0, then erosion is the fitting of a structuring element to these regions and dilation is the fitting of a structuring element (rotated if necessary) into the background, followed by inversion of the result.

Binary Image DIALATION: by 2*2 Structuring element

Figure 3.15

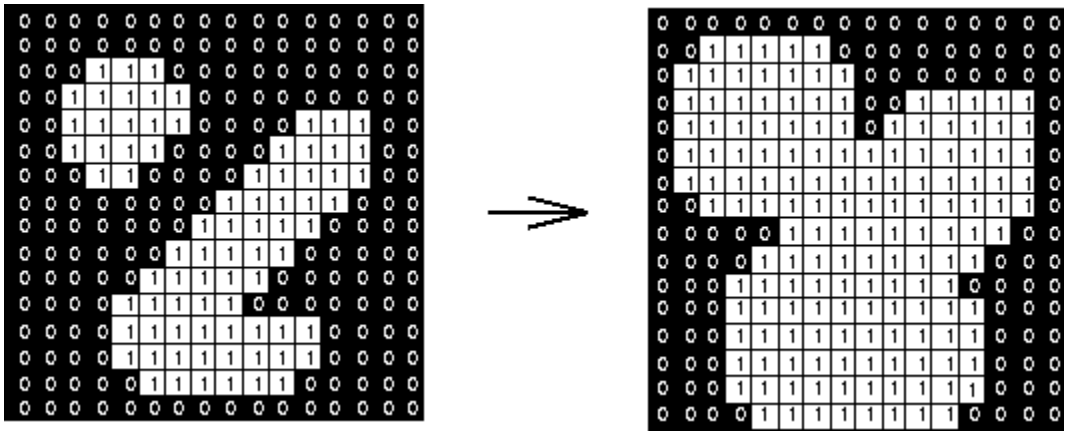


Figure 3.16 Effect of dilation using a 3×3 square structuring element

To compute the dilation of a binary input image by this structuring element, we consider each of the background pixels in the input image in turn. For each background pixel (which we will call the input pixel) we superimpose the structuring element on top of the input image so that the origin of the structuring element coincides with the input pixel position. If at least one pixel in the structuring element coincides with a foreground pixel in the image underneath, then the input pixel is set to the foreground value. If all the corresponding pixels in the image are background, however, the input pixel is left at the background value.

The 3×3 square is probably the most common structuring element used in dilation operations, but others can be used. A larger structuring element produces a more extreme dilation effect, although usually very similar effects can be achieved by repeated dilations using a smaller but similarly shaped structuring element. With larger structuring elements, it is quite common to use an approximately disk shaped structuring element, as opposed to a square one.

EROSION :: [Common Names: Erode, Shrink, Reduce]

Erosion is one of the two basic operators in the area of mathematical morphology, the other being dilation. It is typically applied to binary images, but there are versions that work on grayscale images. The basic effect of the operator on a binary image is to erode away the boundaries of regions of foreground pixels (i.e. white pixels, typically). Thus areas of foreground pixels shrink in size, and holes within those areas become larger.

How It Works ::

The erosion operator takes two pieces of data as inputs. The first is the image which is to be eroded. The second is a (usually small) set of coordinate points known as a structuring element (also known as a kernel). It is this structuring element that determines the precise effect of the erosion on the input image.

The mathematical definition of erosion for binary images is as follows:

- Suppose that X is the set of Euclidean coordinates corresponding to the input binary image, and that K is the set of coordinates for the structuring element.

- Let Kx denote the translation of K so that its origin is at x.
- Then the erosion of X by K is simply the set of all points x such that Kx is a subset of X.

The mathematical definition for grayscale erosion is identical except in the way in which the set of coordinates associated with the input image is derived. In addition, these coordinates are 3-D rather than 2-D.

As an example of binary erosion, suppose that the structuring element is a 3×3 square, with the origin at its center as shown in Figure 1. Note that in this and subsequent diagrams, foreground pixels are represented by 1's and background pixels by 0's.

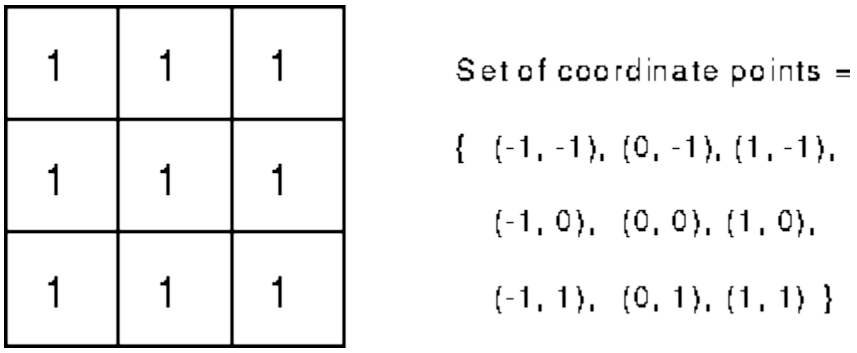


Figure 3.17: A 3×3 square structuring element

To compute the erosion of a binary input image by this structuring element, we consider each of the foreground pixels in the input image in turn. For each foreground pixel (which we will call the input pixel) we superimpose the structuring element on top of the input image so that the origin of the structuring element coincides with the input pixel coordinates. If for every pixel in the structuring element, the corresponding pixel in the image underneath is a foreground pixel, then the input pixel is left as it is. If any of the corresponding pixels in the image are background, however, the input pixel is also set to background value.

For our example 3×3 structuring element, the effect of this operation is to remove any foreground pixel that is not completely surrounded by other white pixels (assuming 8-

connectedness). Such pixels must lie at the edges of white regions, and so the practical upshot is that foreground regions shrink (and holes inside a region grow).

Erosion is the dual of dilation, i.e. eroding foreground pixels is equivalent to dilating the background pixels.

Most implementations of this operator will expect the input image to be binary, usually with foreground pixels at intensity value 255, and background pixels at intensity value 0. Such an image can often be produced from a grayscale image using thresholding. It is important to check that the polarity of the input image is set up correctly for the erosion implementation being used.

The structuring element may have to be supplied as a small binary image, or in a special matrix format, or it may simply be hardwired into the implementation, and not require specifying at all. In this latter case, a 3×3 square structuring element is normally assumed which gives the shrinking effect described above. The effect of an erosion using this structuring element on a binary image is shown in Figure 3.18

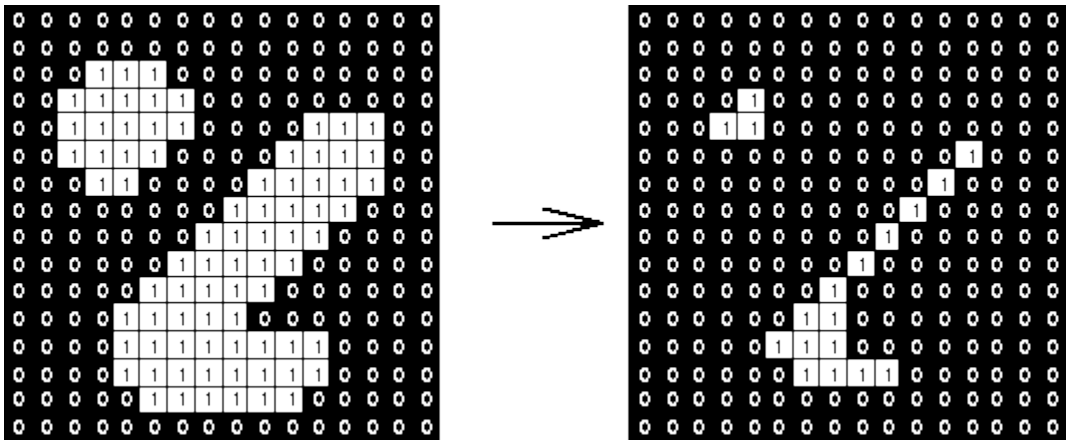


Figure 3.18 : Effect of erosion using a 3×3 square structuring element

The 3×3 square is probably the most common structuring element used in erosion operations, but others can be used. A larger structuring element produces a more extreme erosion effect, although usually very similar effects can be achieved by repeated erosions using a smaller similarly shaped structuring element. With larger structuring elements, it is

quite common to use an approximately disk shaped structuring element, as opposed to a square one.

OPENING :: Opening and closing are two important operators from mathematical morphology. They are both derived from the fundamental operations of erosion and dilation. Like those operators they are normally applied to binary images, although there are also graylevel versions. The basic effect of an opening is somewhat like erosion in that it tends to remove some of the foreground (bright) pixels from the edges of regions of foreground pixels. However it is less destructive than erosion in general. As with other morphological operators, the exact operation is determined by a structuring element. The effect of the operator is to preserve foreground regions that have a similar shape to this structuring element, or that can completely contain the structuring element, while eliminating all other regions of foreground pixels.

How It Works ::

Very simply, an opening is defined as an erosion followed by a dilation using the same structuring element for both operations. The opening operator therefore requires two inputs: an image to be opened, and a structuring element.

Graylevel opening consists simply of a graylevel erosion followed by a graylevel dilation.

Opening is the dual of closing, i.e. opening the foreground pixels with a particular structuring element is equivalent to closing the background pixels with the same element.

CLOSING ::

Closing is an important operator from the field of mathematical morphology. Like its dual operator opening, it can be derived from the fundamental operations of erosion and dilation. Like those operators it is normally applied to binary images, although there are graylevel versions. Closing is similar in some ways to dilation in that it tends to enlarge the boundaries of foreground (bright) regions in an image (and shrink background color holes in such regions), but it is less destructive of the original boundary shape. As with other morphological operators, the exact operation is determined by a structuring element. The effect of the operator is to preserve background regions that have a similar shape to this

structuring element, or that can completely contain the structuring element, while eliminating all other regions of background pixels.

HOW IT WORKS ::

Closing is opening performed in reverse. It is defined simply as a dilation followed by an erosion using the same structuring element for both operations The closing operator therefore requires two inputs: an image to be closed and a structuring element.

Graylevel closing consists straightforwardly of a graylevel dilation followed by a graylevel erosion.

Closing is the dual of opening, i.e. closing the foreground pixels with a particular structuring element, is equivalent to closing the background with the same element.

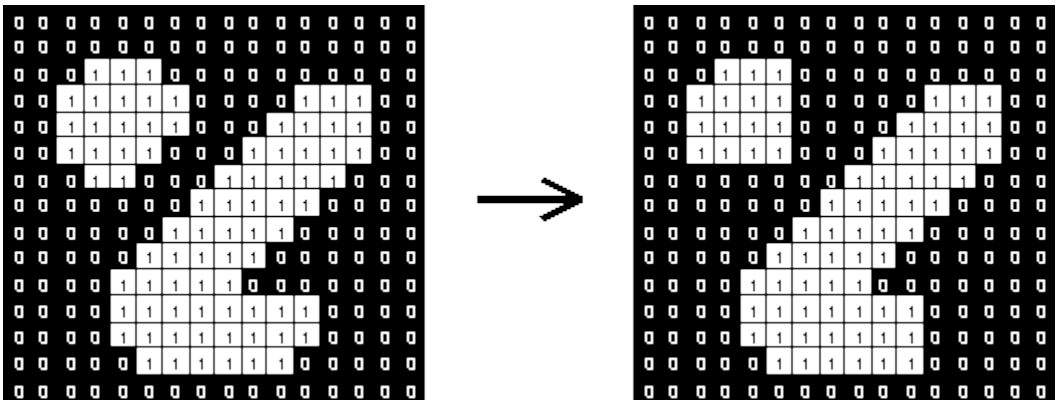


Figure 3.19: Effect of opening using a 3×3 square structuring element

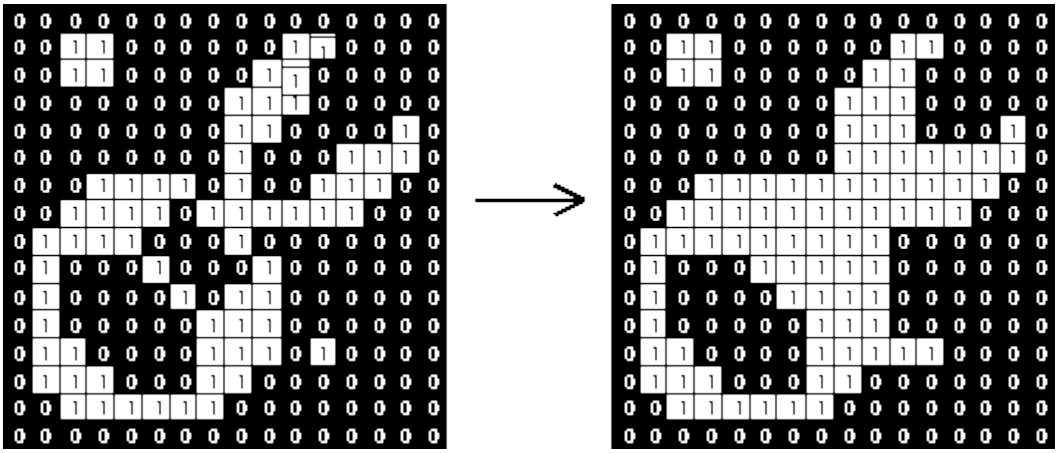


Figure 3.20: Effect of closing using a 3×3 square structuring element

COMPOUND OPERATIONS ::

Many morphological operations are represented as combinations of erosion, dilation, and simple set-theoretic operations such as the **complement** of a binary image:

$$f^c(x,y) = 1 \text{ if } f(x,y) = 0, \text{ and } f^c(x,y) = 0 \text{ if } f(x,y) = 1,$$

the **intersection** $h = f \cap g$ of two binary images f and g:

$$h(x,y) = 1 \text{ if } f(x,y) = 1 \text{ and } g(x,y) = 1, \text{ and } h(x,y) = 0 \text{ otherwise,}$$

and the **union** $h = f \cup g$ of two binary images f and g:

$$h(x,y) = 1 \text{ if } f(x,y) = 1 \text{ or } g(x,y) = 1, \text{ and } h(x,y) = 0 \text{ otherwise:}$$

The **opening** of an image f by a structuring element s (denoted by $f \circ s$) is an erosion followed by a dilation:

$$f \circ s = (f \ominus s) \oplus s$$

Opening is so called because it can open up a gap between objects connected by a thin bridge of pixels. Any regions that have survived the erosion are restored to their original size by the dilation:

Opening is an **idempotent** operation: once an image has been opened, subsequent openings with the same structuring element have no further effect on that image:

$$(f \circ s) \circ s) = f \circ s.$$

The **closing** of an image f by a structuring element s (denoted by $f \bullet s$) is a dilation followed by an erosion:

$$f \bullet s = (f \oplus s_{rot}) \ominus s_{rot}$$

3.5 EDGE DETECTION

Brief Explanation :

Definition of edges

1 Edges are significant local changes of intensity in an image.

1 Edges typically occur on the boundary between two different regions in an image.

• **Goal of edge detection**

2 Produce a line drawing of a scene from an image of that scene.

2 Important features can be extracted from the edges of an image (e.g., corners, lines, curves). These features are used by higher-level computer vision algorithms (e.g., recognition)

• **What causes intensity changes?**

- Various physical events cause intensity changes.
- Geometric events
    * object boundary (discontinuity in depth and/or surface color and texture)
    * surface boundary (discontinuity in surface orientation and/or surface color and texture)

- Non-geometric events
    * specularity (direct reflection of light, such as a mirror)
    * shadows (from other objects or from the same object)
    * inter-reflections

• **Edge descriptors**

Edge normal: unit vector in the direction of maximum intensity change. Edge direction: unit vector to perpendicular to the edge normal.

Edge position or center: the image position at which the edge is located. Edge strength: related to the local image contrast along the normal.

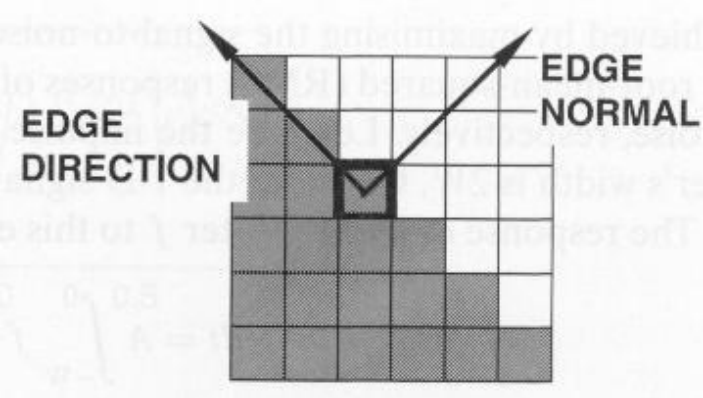


Figure 3.21Showing various parameters of Edge

• Modeling intensity changes

- Edges can be modeled according to their intensity profiles.

Step edge: the image intensity abruptly changes from one value to one side of the discontinuity to a different value on the opposite side.

Ramp edge: a step edge where the intensity change is not instantaneous but occur over a fi nite distance.

Ridge edge: the image intensity abruptly changes value but then returns to the starting value within some short distance (generated usually by lines).

Roof edge: a ridge edge where the intensity change is not instantaneous but occur over a fi nite distance (generated usually by the intersection of surfaces).

• The four steps of edge detection

(1) **Smoothing**: suppress as much noise as possible, without destroying the true edges.
(2) **Enhancement**: apply a filter to enhance the quality of the edges in the image (sharpening).
(3) **Detection**: determine which edge pixels should be discarded as noise and which should be retained (usually, thresholding provides the criterion used for detection).
(4) **Localization**: determine the exact location of an edge (sub-pixel resolution might be required for some applications, that is, estimate the location of an edge to better than the spacing between pixels). Edge thinning and linking are usually required in this step.

EDGE DETECTION METHOD ::

- Edge Detection by Derivatives
- Edge Detection Using Gradient
- The Prewitt Edge Detector
- The Sobel Edge Detector
- The Roberts Edge Detector

CHAPTER 4

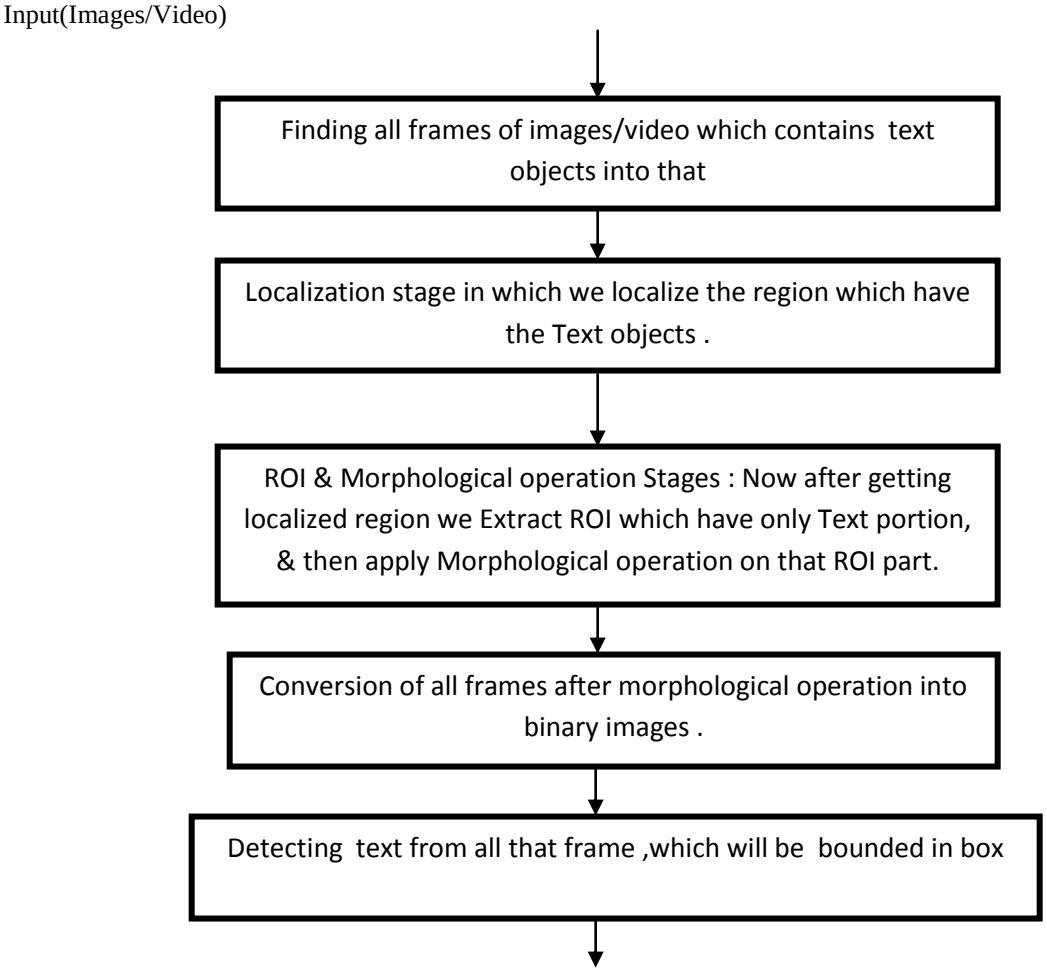
RESULTS&DISCUSSIONS

This Chapter is dedicated to Main Methodology(Extraction based on Morphological Operation) which I have applied in our project then we have included Future Work or Scope of my project & Results, Conclusion section of my Thesis ,as I have discussed in Introduction chapter too about the content of this chapter

The Sequence which is being followed in this chapter is

4.1 **Main Method 4.2 Discussions 4.3 Results 4.4 Conclusions 4.5 Future Work**

4.1 Main Method :: My main method which I have implemented to achieve the results is shown by following flow diagram



4.2DISCUSSIONS

DISCUSSIONS::

It is always tuff to prove that, your proposed algorithm is giving better output, and this is more tuff task to show that in terms of percentage of accuracy ,The localization and detection problems involve finding tight bounding boxesaround any text in a given frame , These are significantly harder than the corresponding problems in document analysis.But I think I have done it convincingly ,which is explained later on in this section & then in later section results in pictorial form is shown

I have taken three different videos of three different resolution parameters(Bits Rate, No of Frames ,Duration etc) which are listed in table 1 ,table 2 & table 3 & I named those videos as video1 , video2 & video3 and I have used a formula of percentage of accuracy for calculating accuracy of our results for few frames of these three videos , then I have found that accuracy percentage of every frame(which I have taken into consideration of analysis) is more than 90% .

Accuracy is defined here in terms of Number of character which my proposed algorithm is detecting & Number of character which that Input Frame (considered frame of Analysis) was actually having

Accuracy(A) = {Numbers of Character detected in any particular frame in OutPut}/{Total Numbers of Character which that Frame was actually }

Percentage of Accuracy = [**A***100]

Table1 ::

Video1(For Data Input)

VIDEO-1 Parameters	
Parameter	Value
Duration	631.18 sec
Bits Per Pixel	24
Frame Rate	25
Height	320
Width	480
Number of Frames	15779

Table2 ::

Video2(For Data Input)

VIDEO-2 Parameters	
Parameter	Value
Duration	432.14 sec
Bits Per Pixel	24
Frame Rate	30
Height	360
Width	638
Number of Frames	12964

Table3 ::

Video3(For Data Input)

VIDEO-3 Parameters	
Parameter	Value
Duration	766.97 sec
Bits Per Pixel	24
Frame Rate	25
Height	360
Width	640
Number of Frames	19174

4.3 RESULTS

4.3 RESULTS:: Resultsare shown in this section in the form of frame(screen shot) which have detected text of frame containing text in input video, three different videos is used as input data during my implementation of this method so I have taken screen shots of two different frames of all 3 inputs videos&hence here I am showing all 6 outputs , which is as of following table in next pages …

.

.

.

4.4 CONCLUSION

Text objects occurring in image and video documents can provide much usefulinformation for content based information retrieval and for other applications, because they contain much deep information related to the documents contents. However, extracting text from images and videos is a very difficult task due to the varying font, size, color, orientation, and deformation of text objects. Although a large number of text extraction approaches have been reported in the literature review chapter, no specific designed text model and character features are presented to capture the unique properties and structure of characters and text objects.

In this dissertation, we proposed a method to detect and localize the text objects occurring in image and video documents based on morphological operation done on extracted region content text of all frame, and this text tracking method to track text event in image/video documents(frame).

First when we see the characters(Or Text),then characters typically have big contrast to their background and are located within certain distance interval, we first initializes the whole process by preprocessing steps in which we have found those parts of every frame in Image/video, Then we localize the text region where all text object is located ,& Next is ROI extraction followed by detecting text in output.

Based on the detection and found results of three different video as shown in early part of this Chapter, we can see that the proposed text detection method can detect the text objects with various fonts, sizes, and orientations efficiently. However, there are still some missed text objects and you can say fault of this method. We discuss the few limitations of the proposed text detection method and the future work to solve them in next section.

4.5FUTURE WORK

The work presented in this thesis is, hopefully, comprehended within the defined scope, but research never ends, therefore, future research is expected to explore horizons beyond the scope of this thesis. It is hoped that the limitations of this work would be considered as the beginning for the research in the future in many directions. The effectiveness and efficiency of the proposed system can be improved and enhanced in various way.

However Few Future works of this proposed method of Text Extraction could be in any of following type :

1. **LanguageVersatility ::** In current stage, we only evaluated the proposed method on English language-based datasets (i.e images/videos). In further work, we will test our method on other languages whose text objects are not in English ,for Example some East Asian languages, such asJapanese, Korean, and Chinese, their characters are different than English in many aspects ,But still proposed method would work on that too.
2. **Transparent characters ::** If a character is transparent, the edge of the character mayconnect with other background edges and are not closed. In this case, transparent characters are removed due to unreliable edges. Due to Mixed up of "Text Character" with background edges, few characters will not be detected properly in output, Sometimes Few Character of "Input Text" doesn't come in output due to unclosed boundary, So these problem could be removed by making unreliable edges to reliable by some other technique than proposed method would be implemented on that, These limitation can be resolved by having some more color information of Text (Input Text) .
3. **Recognition Step ::**Recognition of all individual Text/Character could be the novel research after extraction part from my proposed method .
4. **Reduction of Processing Time ::** In current proposed method we are having too much processing time , means execution time per frame to reach on the output stage from input, So we can reduce the execution time or in other word we can say that we are trying to reach towards output from input in real time execution time by removing some glitch due to which processing time is too much.

CHAPTER 5
REFERENCES

[1] http://blog.flickr.net/en/2011/08/04/6000000000/
[2] http://www.youtube.com/t/press_statistics
[3] Chen D., J. Luettin, K. Shearer, "A Survey of Text Detection and Recognition in Images and Videos", *Institute DalleMolled' Intelligence Perceptive (IDIAP) Research Report*, IDIAP-RR 00-38, 2000
[4] Jung K., K.I. Kim, and A.K. Jain, "Text Information Extraction in Images and Video: A Survey", *Pattern Recognition*, pp. 977-997, 2004
[5] Lyu M.R., J. Song, M. Cai, "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 15, pp. 243-255, 2005.
[6] Lienhart R. and A. Wernicke, "Localizing and Segmenting Text in Images and Videos", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 12, No. 4, pp. 256-268, 2002.
[7] Kim K.I., K. Jung and J.H. Kim, "Texture-based Approach for Text Detection in Image Using Support Vector Machine and Continuously Adaptive Mean Shift Algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, pp. 1631-1638, 2003.
[10] Liu Y., H. Lu, X. Xue, and Y.P. Tan, "Effective Video Text Detection Using Line Features", *Proceedings of the 8th International Conference on Control, Automation, Robotics and Vision*, pp. 1528-1532, 2004.
[11] Mi, Y. Xu, H. Lu, and X. Xue, "A Novel Video Text Extraction Approach Based on Multiple Frames", *International Conference on Information, Communications and Signal Processing*, pp. 678-682, 2005.
[12] Tran H., A lux, H.L. Nguyen T. and A. Boucher, "A Novel Approach for Text Detection in Images Using Structural Features", *Proceedings of the 3rd International Conference on Advances in Pattern Recognition*, pp. 627-635, 2005.
[13] Eberly D., "Ridges in Image and Data Analysis", *Kluwer Academic Publishers*, 1996.
[14] Dubey P., "Edge Based Text Detection for Multi-purpose Application", *Proceedings of International Conference Signal Processing*, 2006
[15] Zhou J., "A Robust System for Text Extraction in Video", *Proceedings of International Conference on Machine Vision*, pp. 119-124, 2007.
[16] Wang R., W. Jin, and L. Wu, "A Novel Video Caption Detection Approach Using Multi-Frame Integration", *Proceedings of International Conference on Document Analysis and Recognition*, 2005.
109
[17] Liu Y., S. Goto, and T. Ikenaga, "A Robust Algorithm for Text Detection in Color Image", *Proceedings of International Conference on Document Analysis and Recognition*, 2005.
[18] Park J., G. Lee, E. Kim, J. Lim, S. Kim, and H. Yang, "Automatic Detection and Recognition of Korean Text in Outdoor Signboard Images", *Pattern Recognition Letters*, Vol. 31, 2010.
[19] Fu L., W Wang, Y. Zhan, "A Robust Text Segmentation Approach in Complex Background Based on Multiple Constraints", *Proceedings of the 8th Pacific Rim conference on Advances in multimedia information processing*, 2005.
[20] Clavelli A. and D. Karatzas, "Text Segmentation in Colour Posters from the Spanish Civil War Era", *Proceedings of International Conference on Document Analysis and*

Recognition, 2009.
[21] Fu H., X. Liu, Y. Jia, and H. Deng, "Gaussian Mixture Modeling of Neighbor Characters for Multilingual Text Extraction in Images", *Proceedings of International Conference on Image Processing,* pp.3321-3324, 2006.
[22] Liu X., H. Fu, and Y. Jia, "Gaussian Mixture Modeling and Learning of Neighboring Characters for Multilingual Text Extraction in Images", *Pattern Recognition,* Vol. 41, pp. 484-493, 2008.
[23] Smith S.M. and J.M. Brady, "SUSAN-A New Approach to Low Level Image Processing", *International Journal on Computer Vision,* Vol. 23, No. 1, pp 45-78, 1997.
[24] Hua X.S., W. Liu, and H.J. Zhang. "An Automatic Performance Evaluation Protocol for Video Text Detection Algorithms", *IEEE Transactions on Circuits and Systems for Video Technology,* 14(4), pp. 498–507, 2004.
[25] Bai H., J. Sun, S. Naoi, Y. Katsuyama, Y. Hotta, and K. Fujimoto, "Video Caption Duration Extraction", *Proceedings of International Conference on Pattern Recognition,* 2008.
[26] Zhao X., K.H. Lin, Y. Fu, Y. Hu, Y. Liu and T.S. Huang, "Text From Corners: A Novel Approach to Detect Text and Caption in Videos", *IEEE Transaction on Image Processing,* Vol. 20, No. 3, pp.790-799, 2011.
[27] Sun L., G.Z. Liu, X. Qian and D.P. Guo, "A Novel Text Detection and Localization Method Based on Corner Response", *Proceedings of IEEE international conference on Multimedia and Expo,* 2009.
[28] Jung C., Q. Liu, and J. Kim, "A Stroke Filter and Its Application to Text Localization", *Pattern Recognition Letters,* Vol. 30, pp. 114-122,2009.
[29] Srivastav A. and J. Kumar, "Text Detection in Scene Images Using Stroke Width and Nearest-Neighbor Constraints", *IEEE region 10 conference,* 2008.
[30] Zhang J. and R. Kasturi, "Text Detection Using Edge Gradient and Graph Spectrum", *Proceedings of International Conference on Pattern Recognition,* 2010.
111
[31] Dinh, T.N., Park, J., and Lee, G. "Text Localization Using Image Cues and Text Line Information", *Proceedings of International Conference on Image Processing,* pp. 2261-2264, 2010.
[32] Kim D. and K. Sohn, "Static Text Region Detection in Video Sequences Using Color and Orientation Consistencies", *Proceedings of International Conference on Pattern Recognition,* 2008.
[33] Liu Z. and S. Sarkar, "Robust Outdoor Text Detection Using Text Intensity and Shape Features", *Proceedings of International Conference on Pattern Recognition,* 2008.
[34] Ling H. and D.W. Jacobs, "Shape Classification Using the Inner-Distance", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 29(2), pp. 286-299, 2007.
[35] Liu C., C. Wang, and R Dai, "Text Detection in Images Based on Unsupervised Classification of Edge-based Features", *Proceedings of International Conference on Document Analysis and Recognition,* 2005.
[36] Shivakumara P., W. Huang, T. Phan and C.L. Tan, "Accurate Video Text Detection Through Classification of Low and High Contrast Images", *Pattern Recognition,* Vol. 43, Issue 6, pp. 2165-2185, 2010.
[37] Mancas-Thilou C., B. Gosselin, "Spatial and Color Spaces Combination for Natural

Scene Text Extraction", *IEEE international Conference on Image Processing*, pp. 985-988, 2006.
[38] Mancas-Thillou C., B Gosselin, "Color Text Extraction with Selective Metric-based Clustering", *Computer Vision and Image Understanding*, pp. 97-107, 2007.
[39] Shivakumara P., T.Q. Phan and C.L. Tan, "New Fourier-Statistical Features in RGB Space for Video Text Detection", *IEEE Transactions on Circuits and Systems for Video Technology*, 2010.
[40] Phan T.Q., P. Shivakumara and C.L. Tan, "A Laplacian Method for Video Text Detection", *Proceedings of International Conference on Document Analysis and Recognition,* 2009.
[41] Shivakumara P., T. Q. Phan, and C. L. Tan, "A Laplacian approach to multi-oriented text detection in video", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 2, 2011.
112
[42] Leon M., V. Vilaplana, A. Gasull, and F. Marques, "Caption Text Extraction for Indexing Purposes Using a Hierarchical Region-based Image Model", *Proceedings of International Conference on Image Processing* 2009.
[43] Gllavata J., R. Ewerth, and B. Freisleben, "Text Detection in Images Based on Unsupervised Classification of High Frequency Wavelet Coefficients", *Proceedings of International Conference on Pattern Recognition,* 2004.
[44] Saoi T., H. Goto, and H. Kobayashi, "Text Detection in Color Scene Images Based on Unsupervised Clustering of Multi-channel Wavelet Features", *Proceedings of International Conference on Document Analysis and Recognition,* 2005.
[45] Shivakumara P., T.Q. Phan, and C.L. Tan, "A Robust Wavelet Transform based Technique for Video Text Detection", *Proceedings of International Conference on Document Analysis and Recognition* 2009.
[46] Chen D., J.M. Odobez and H. Bourlard, "Text Detection and Recognition in Images and Video Frames", *Pattern Recognition*, pp. 595-608, 2004.
[47] Kim K.I., K. Jung and J.H. Kim, "Texture-based Approach for Text Detection in Image Using Support Vector Machine and Continuously Adaptive Mean Shift Algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, pp. 1631-1638, 2003.
113
[48] Cheng Z., Y.C. Liu, "Caption Location and Extraction in Digital Video Based on SVM", *Proceedings of International Conference on Machine Learning and Cybernetics*, Vol. 6, pp 26-29, 2004.
[49] Wang Y.K. and J.M. Chen, "Detection Video Texts Using Spatial-temporal Wavelet Transform", *Proceedings of International Conference on Pattern Recognition,* 2006.
[50] Pan Y., C.L. Liu, and X. Hou, "Fast Scene Text Localization by Learning-based Filtering and Verification", *Proceedings of International Conference on Pattern Recognition,* 2010.
[51] Pan Y., X. Hou, and C.L. Liu, "Text Localization in Natural Scene Image Based on Conditional Random Field", *Proceedings of International Conference on Document Analysis and Recognition,* 2009.
[52] Qian X. and G. Liu, "Text Detection, Localization, and Segmentation in Compressed Videos", *Proceedings of IEEE international conference on acoustics, speech and signal processing,* Vol. 2, pp. 385-388, 2006.
[53] Zhang D. Q. and S. F. Chang, Learning to Detect Scene "Text Using a Higher-order

MRF with Belief Propagation", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 2004.
[54] Liu D. and T. Chen, "Object Detection in Video with Graphical Models", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, pp 14-19, 2006.
[55] Pan W. M., T. D. Bui, and C. Y. Suen, "Text Segmentation from Complex Background Using Sparse Representations", pp. 412-416, *Proceedings of International Conference on Document Analysis and Recognition*, 2007.
[56] Pan W.M., T.D. Bui, and C.Y. Suen, "Text Detection from Scene Images Using Sparse Representation", *Proceedings of International Conference on Pattern Recognition*, 2008.
114
[57] M. Zhao and S. Li, "Sparse Representation Classification for Image Text Detection", *the Second International Symposium on Computational Intelligence and Design*, pp76-79, 2009.
[58] Tang L. and J.R. Kender, "A Unified Text Extraction Method for Instructional Videos", *IEEE international conference on image processing*, Vol. 3, pp11-14, 2005.
[59] Choudary C., and T. Liu, "Summarization of Visual Content in Instruction videos", *IEEE Transactions on Multimedia*, Vol. 9, pp. 1443-1455, 2007.
[60] Wu W., X. Chen and J. Yang, "Detection of Text on Road Signs from Video", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 6, pp. 378-390, 2005.
[61] Wu W., X. Chen and J. Yang, "Incremental Detection of Text on Road Signs from Video with Application to a Driving Assistant System", *Proceedings of ACM Multimedia*, 2004.
[62] Bhattacharya U., S.K. Parui, and S. Mondal, Devangari and Bangla "Text Extraction from Natural Scene Images", *Proceedings of International Conference on Document Analysis and Recognition*, 2009.
[63] Moradi M., S. Mozaffari, and A.A. Orouji, "Farsi/Arabic Text Extraction from Video Images", P*roceedings of The 9th Iranian conference on electrical engineering*, page 1-6, 2011.
[64] Gllavata J., R. Ewerth and B. Freisleben, "Tracking Text in MPEG Videos", *Proceedings of the 12th annual ACM international conference on Multimedia*, pp 240-243, 2004.
[65] Myers G. and B. Burns, "A Robust Method for Tracking Scene Text in Video Imagery", pp. 30-35, *Proceedings of the Workshop on Camera Based Document Analysis and Recognition*, 2005.
[66] Tanaka M. and H. Goto, "Autonomous Text Capturing Robot Using Improved DCT Feature and Text Tracking", *Proceedings of International Conference on Document Analysis and Recognition*, pp. 1178-1182, 2007.
115
[67] Ezaki N., K. Kiyota, B.T. Minh, M. Bulacu, and L. Schomaker, "Improved Text-Detection Methods for a Camera-based Text Reading System for Blind Persons", *Proceedings of International Conference on Document Analysis and Recognition*, pp. 257-261, 2005.
[68] Ezaki N., M. Bulacu, L. Schomaker, "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons", *Proceedings of International Conference on Pattern Recognition*, 2004.
[69] Shi X. and Y. Xu, "A Wearable Translation Robot", *International Conference on*

Robotics and Automation", pp. 4400-4405, 2005.
[70] Tanaka M. and H. Goto, "Autonomous Text Capturing Robot Using Improved DCT Feature and Text Tracking", *Proceedings of International Conference on Document Analysis and Recognition,* pp. 1178-1182, 2007.
[71] Tanaka M. and H. Goto, "Text-Tracking Wearable Camera System for Visually – Impaired People", *Proceedings of International Conference on Pattern Recognition,* 2008.
[72] Goto H., M. Tanaka, "Text-tracking Wearable Camera System for the Blind", *Proceedings of International Conference on Document Analysis and Recognition,* pp. 141-145, 2009.
[73] Wazalwar D., E. Oruklu, and J. Saniie, "Design Flow for Robust License Plate Localization", *IEEE International Conference on Electro/Information Technology,* 2011.
[74] Huang H., G. Ma, and Y. Zhuang, "Vehicle License Plate Location Based on Harris Corner Detection", *IEEE International Joint Conference on Neural Network,* 2008.
[75] Bradski G.R., Real Time "Face and Object Tracking as a Component of a Perceptual User Interface", *Proceedings of IEEE Workshop Applications of Computer Vision,* pp. 214-219, 1998.
[76] Kumar S. and M. Hebert, "Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification", *IEEE International Conference on Computer Vision,* Vol. 2, pp. 1150-1157, 2003.
116
[77] Olshaus B. and D. Field, "Emergence of Simple-Cell Receptive Field Properties by Learning A Sparse Code for Natural Images", *Nature*, 381:607–609, 1996.
[78] DalalN.and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *Proceedings of IEEE Conference on Computer Vision and Pattern Recogntion,* pp. 886-893, 2005.
[79] Epshtein B., E. Ofek, and Y. Wexler, "Detecting Text in Natural Scenes with StrokeWidth Transform", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,2010.
[80]Bouman K.L., G. Abdollahian, M. Boutin, and E.J. Delp, "A Low Complexity Sign Detection and Text Localization Method for Mobile Applications", IEEE Transactions on Multimedia, Vol. 13, No. 5, pp. 922-934, 2011. 110
[81]Jafri S.A.R., M. Boutin, and E.J. Delp, "Automatic text area segmentation in natural images", Proceedings of International Conference on Image Processing, pp.3196-3199, 2008. [82]Crandall D., S. Antani, R. Kasturi, "Extraction of Special Effects Caption TextEvents from Digital Video", International Journal of Document Analysis and Recognition, Vol. 5, pp. 138-157, 2003 ,108 [83]PavlidisT.,"Limitations of CBIR", *Plenary talk, the Nineteenth InternationalConference on Pattern Recognition*, http://www.theopavlidis.com/technology/CBIR/PaperB/icpr08.pdf, 2008 .

[84]AnubhavKumar.,"An Efficient Algorithm for Text Localization & Extraction in Complex Video Text Images" , 2013 2nd international conference on information management in the knowledge Economy