Transcript Extraction from Document Metaphors With Intricate Background

S.M.Krishna Ganesh #1, Dr.S.Venkatesan Jeyakumar #2

#1 St.Joseph University in Tanzania, P.O.Box – 11 007, Dar Es Salaam, Tanzania
#2 St.Joseph University in Tanzania, P.O.Box – 11 007, Dar Es Salaam, Tanzania

ABSTRACT

The automatic detection of Region of Interests (ROI) is an active research area in the design of machine vision systems. Text embedded in images contains large quantities of useful semantic information which can be used to fully understand Images. Moreover, text consist of words that are well defined models for humans communication. The proposed method is to apply a variation of Contourlet transform on images to decompose it into set of directional sub bands with texture details capture in different orientations at various scales. Contourlets not only possess the main features of wavelets (namely, multi scale and time-frequency localization), but also offer a high degree of directionality and anisotropy. As the contourlet transform is not shift-invariant, non-sub sampled contourlet transform NSCT is used. Here, the idea is to capture Multi-oriented Texture details at high frequency component so as to produce text regions whereas at low frequency gives rise to non-text region, instead of recognizing the edges in only horizontal, vertical & diagonal directions. Variations of text due to differences in size, style, orientation, alignment; low image contrast and presence of complex background makes the problem of automatic text extraction extremely challenging. If the text can be extracted from these images, we can apply it to many areas such as robot navigation, helping the visual impaired and improving robot intelligence and also for guiding foreigners.

Key words: Directional filter bank, Laplacian Pyramid, Non-Subsampled Contourlet transform

Corresponding Author: S.M.Krishna Ganesh and Dr.S.Venkatesan Jeyakumar

INTRODUCTION

The recent advances in digital technology, more and more databases are multimedia in nature, containing images and video in addition to the textual information. Text extraction from images is an active research area as it is contributing more to Content based image indexing and Text & database retrieval. Text embedded / inserted in images is used to describe the contents of an image & also it can be easily extracted compared to other semantic contents and it facilitates applications such as keyword-based image search, automated processing & reading of documents, text-based image indexing,. Variation in Font style, size, Orientation, alignment & complexity of background makes the text extraction as a challenging task. Text in images/video images are classified into Caption text and Scene text. Caption text is the one which is inserted text & otherwise called as superimposed/artificial text. Natural images are called as Scene text/graphics.
text. Researchers have addressed three different approaches to extract the text from images namely Bottom-up, Top-down & Hybrid approaches. Bottom up approach starts with the identification of sub-structures, such as CCs or edges, and then merging these sub-structures to mark bounding boxes for text. Top-down approach looks for global information in the page & splits the page from column level to word level. Edge based & Connected component (CC) based methods are categorized under Bottom-up methods & Textured based methods under Top-down approach. Proposed method employs Non sub sampled Contourlet transform as a texture based method to extract text from document images.

MATERIALS AND METHODS

Kim et al [3] proposed “Scene text extraction in natural scene images using hierarchical feature combining and verification.”. Three local image features are used to find local variations of intensity, color and color continuity of the text area. High-level text feature is used to verify the text candidate regions by examining stroke compositions. Strokes are combined to form a character. Multi-resolution wavelet transform is used for obtaining 8-directional stroke features. Local area is classified as text or non-text with SVM. The method has difficulties in finding text where characters are too big, or illumination changes are so severe. Basavaraj et al [1] proposed “Video Text Extraction from Images for Character Recognition”. Proposed text extraction method use morphological operations for edge detection and sobel operator for edge enhancement. Connected component labeling is used for labeling text contour region. Uneven reflections result in incomplete character segmentation. Anil. K. Jain [4] et al proposed “Automatic Text Location in Images and Video Frames”. Input image is a binary image, a synthetic web image, a color image or video frame. It is decomposed into multiple foreground images based on color reduction and multivalued image decomposition. Connected component analysis is implemented to individual foreground images. Finally, the outputs of all the channels are fused to locate text in the input image. Method tries to extract only horizontal text of relatively large size. You-Long et al [2] proposed “Gabor Filter Based Text Extraction from Digital Document Images”. Image filtering is done based on Gabor filter. Gabor filtered images suggest that: Text regions are quite rich in middle and high frequency components. Region corresponding to the text field has an approximate rectangular boundary. Potential text region partition is done based on binarization, mending operation and boundary tracing algorithm. To make text regions separable from non-text regions we employ two measure parameters: Standard Rate (SR), High Frequency Content (HFC).

PROPOSED METHOD

TEXT REGION DETECTION

The contourlet transform is an extension of the wavelet transform which uses multi scale and directional filter banks. Here images are oriented at various directions in multiple scales, with flexible aspect ratios. The contourlet transform effectively captures smooth contours images that are the dominant feature in natural images. The main difference between contourlets and other multi scale directional systems is that the contourlet transform allows for different and flexible number of directions at each scale, while achieving nearly critical sampling. In addition,
the contourlet transform uses iterated filter banks, which makes it computationally efficient; specifically, it requires $O(N)$ operations for an $N$ pixel image. The contourlet transform [9] is a multidirectional and multi scale transform that is constructed by combining the Laplacian pyramid [11] with the directional filter bank (DFB) proposed in [14]. Due to downsamplers and up samplers present in both the Laplacian pyramid and the DFB, the contourlet transform is not shift invariant. An over complete transform the non sub sampled contourlet transform (NSCT) has been proposed & NSCT has been applied in our proposed system. The NSCT is a fully shift-invariant, multi scale, and multidirection expansion that has a fast implementation. Here filters are designed with better frequency selectivity thereby achieving better sub band decomposition. The structure consists of a bank of filters that splits the 2-Dfrequency plane into sub bands. This transform can thus be divided into two shift-invariant parts: 1) a non sub sampled pyramid structure that ensures the multi scale property and 2) a non sub sampled DFB structure that gives directionality.

ENERGY COMPUTATION AND BOOSTING OF SUB BANDS

The original image is decomposed into eight directional subband outputs using the NSCT at three different scales and the energy of each sub band can be obtained from the decomposed image. The energy of the image block associated with sub band is defined as

$$E_i = \sum_{x=1}^{n} \sum_{y=1}^{m} | I(x,y)^2 |$$

Here $I(x,y)$ denote the image intensity corresponding to subband. Normalized energy value is used instead of energy value to avoid threshold inaccuracies due to spatial intensity variations across the image. The Sub bands are categorized as Strong & weak based on the value of the Computed Energy. To extract dominant directional energy, it is necessary to select a threshold. Now, difference between each sub band with the maximum energy band is determined. Sub bands are arranged in ascending order based on difference in the energy. The sub bands occupying first few places from the beginning of the list will have minimum difference & will become candidate for Strong bands. Then proper threshold is applied to separate this sorted list into two sets as Strong & Weak subbands. Energy levels of identified weak sub bands are boosted so as to bring out the proper edges in edge detection stage.

EDGE DETECTION

The Sobel operator is used as an edge detector because of following reasons. It is isotropic, so that strokes of all directions are equally treated. It generates double edges, which makes text areas denser than non-text areas. Edges have been detected for the multidirectional sub bands.

TEXT REGION LOCALIZATION

The process of localization involves further enhancing the text regions by eliminating non-text regions. This stage localizes text regions through following steps: morphological dilation, heuristic filtering.
MORPHOLOGICAL DILATION AND HEURISTIC FILTERING

Here Detected edges of Strong & boosted weak sub bands are dilated. The basic effect of the dilation operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels (i.e. white pixels, typically) by adding pixels to the boundaries of the objects in an image. Thus areas of foreground pixels grow in size while holes within those regions become smaller. 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. Various structuring elements have been experimented to find the suitable one & dilated images are produced using a disk shaped structuring element of 6 pixels radius. Here Dilation is performed to enlarge or group the identified text regions. The geometric ratio between the width and the height of the text characters is considered to eliminate possible non-text regions. This ratio value will be defined after experimenting on different kinds of images to get an average value. In this project, regions with minor to major axis ratio less than 10 are considered as candidate text regions for further processing.

TEXT EXTRACTION

The Binarization is applied to extract the text from the identified text region. It will enable the extracted text to be passed & recognized by the Common OCR systems. Binarization is a technique by which the gray scale images are converted to binary images. The most common method is to select a proper threshold for the image and then convert all the intensity values above the threshold intensity to one intensity value representing either “black” or “white” value. All intensity values below a threshold are converted to one intensity level and intensities higher than this threshold are converted to the other chosen intensity. It segments an image into foreground and background. The foreground contains interested characters & this process generates an output image with white text against a black background.

EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed method 20 test images are used.
Type of images taken:- jpg
Orientation of text: - Horizontal, vertical, multiple directions
Font Size: - Ranging from 7 to 30

Fig.1 Input Image with different font size, color, Orientational alignment
CONCLUSION

In this paper, we have presented a method to extract text from document images with complex background by using the variation of Contourlet transform. The gradients of the contours are used to detect the text by combining strong & boosted weak subbands. Proposed method is very effective and efficient in localizing and extracting text-based features from document images with complex background. We have applied our algorithm on several images with complex backgrounds and obtained encouraging results & experimental results show that proposed method out performs the edge based method & Connected component method. The results indicate that our methodology using Non sub sampled contourlet transform has the efficacy to discriminate between text & non text for different kinds of images.

REFERENCE


THE AUTHORS

Mr. S.M.Krishna Ganesh has completed Masters of Technology Degree in Computer Science and Engineering at Kalasalingam University in the year 2009 and completed M.B.A (DDE) from Madurai Kamaraj University in the year 2009, Tamil Nadu, India. He is currently working as Lecturer in St. Joseph College of Engineering and Technology, Dar Es Salaam, Tanzania, East Africa. He has more than 10 publications in various international journals to his credit. He has guided more than 20 projects to final year B.E/B.Tech students with good industry and teaching experience. His areas of interests are Image Processing, Computer Networks, Neural networks and Bioinformatics.
Dr. S. Venkatesan Jayakumar received his B.Sc degree in chemistry from the Bharathidasan University and Ph.D (Organic Chemistry) from Madras University in Chennai, Tamil Nadu, India. He worked as Research Scientist at RECON R&D, BPRL, Karnataka and was offered a position as a Senior Research Scientist at ADVINUS THERAPEUTICS PVT.LTD., Bangalore, Karnataka. He worked for four years developing new lead molecules for Leishmaniasis – African sleeping sickness. He is currently working as Assistant Professor in St.JOSEPH UNIVERSITY, TANZANIA with a focus on Computational Chemistry, Computer Applied Organic Synthesis. His area of interest includes Bio informatics, Green Chemistry and Solid Phase Synthesis.