Multiscale Edge-based Text Extraction from Complex Images

From: Xiaoqing Liu and Jagath Samarabandu; Multiscale Edge-based Text Region Extraction from Complex Images (https://pdfs.semanticscholar.org/b49b/3bfc48a6f9fa03889b219233f5fcc248e747.pdf), Proceedings of the IEEE International Conference on Mechatronics & Automation Niagara Falls, Canada, July 2005.

1. Introduction

Text that appears in images contains important and useful information. Detection and extraction of text in images have been used in many applications. In this mini project, we will try to implement a multiscale edge-based text extraction algorithm, which can automatically detect and extract text in complex images. The proposed method is a general-purpose text detection and extraction algorithm, which can deal not only with printed document images but also with scene text. It is robust with respect to the font size, style, color, orientation, and alignment of text and can be used in a large variety of application fields, such as mobile robot navigation, vehicle license detection and recognition, object identification, document retrieving, page segmentation, etc.

2. Methodology

2.1 Candidate Text Detection

This stage aims to build a feature map by using three important properties of edges: edge strength, density and variance of orientations. The feature map is a gray-scale image with the same size of the input image, where the pixel intensity represents the possibility of text.

2.1.1 Multi-scale edge detector

A convolution operation with a compass operator (as shown in Fig. 1) results in four oriented edge intensity images $E(\theta),(\theta \in \{0,45,90,135\})$, which contain all the properties of edges required in our proposed method.

-1	-1	-1
2	2	2
-1	-1	-1

0^{o} kernel

-1	-1	2
-1	2	-1
2	-1	-1

45^{o} kernel

-1	2	-1
-1	2	-1
-1	2	-1

90^{o} kernel

2	-1	-1
-1	2	-1
-1	-1	2

135^{o} kernel

Fig. 1. Compass operator

2.1.2 Feature map generation

Regions with text in them will have significantly higher values of average edge density, strength and variance of orientations than those of non-text regions. We exploit these three characteristics to generate a feature map which suppresses the false regions and enhances true candidate text regions.

The steps involved are:

- **Getting the strong and the weak edges:**
  Strong edges is obtained by OTSU thresholding of the edges in the vertical direction.
  Weak edges are obtained by the difference of dilated and closed morphology of the vertical edges.
- **Labelling of edges:**
  The edges are labelled in proportion to the area of the connected component they belong to.
- **Filtering out long vertical edges:**
  The labelled edges having very high area of connected component are filtered out. By employing this non linear weight filtering, the proposed method distinguishes text regions from texture-like regions, such as window frames, wall patterns, etc.

$$fmap(i,j) = \bigoplus_{s=0}^{n} \sum_{\theta} N\{ \sum_{x=-c}^{c} \sum_{y=-c}^{c} E(s,\theta,i+x,j+y) \times W(i,j)\}$$

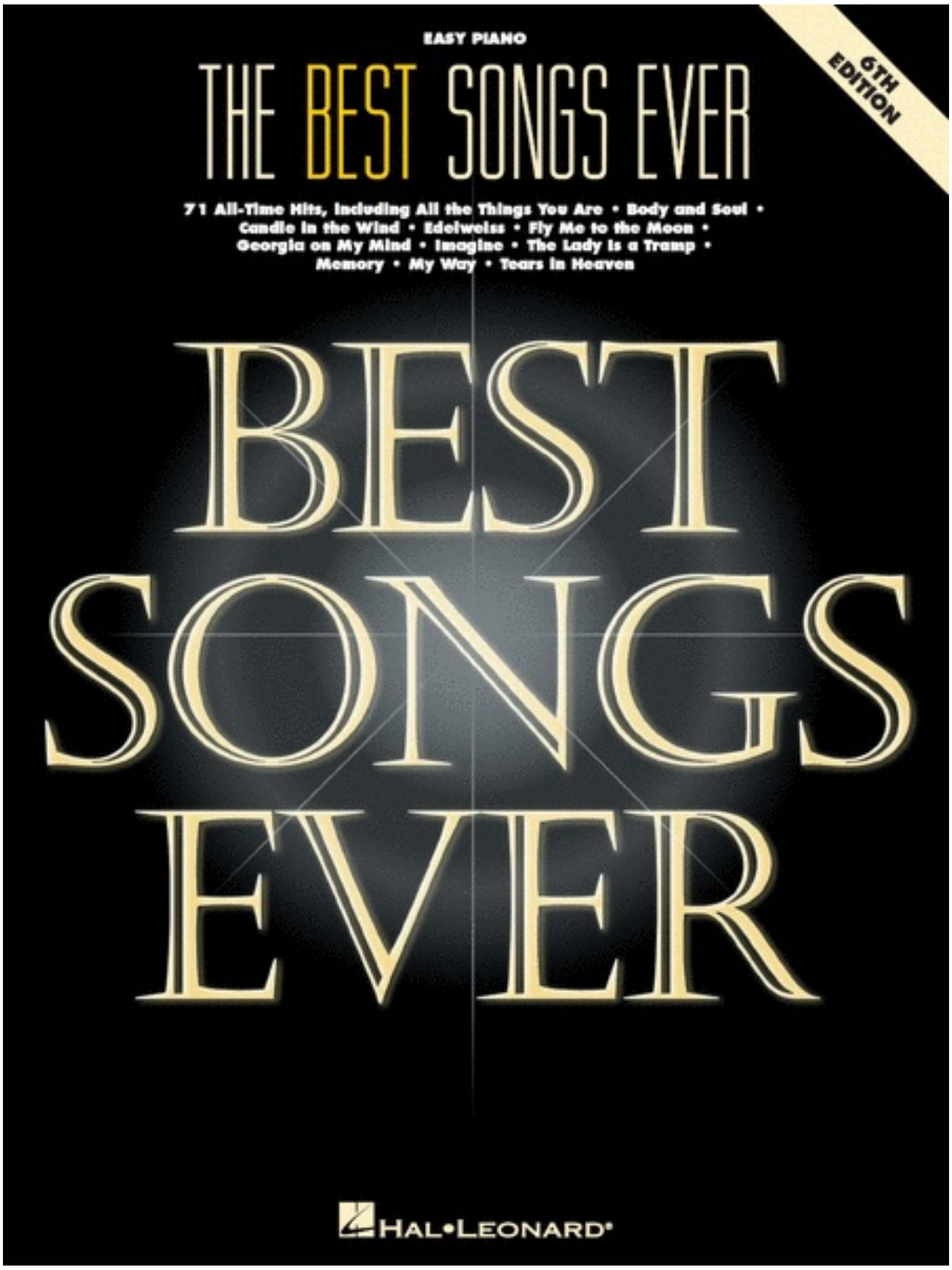
2.2 Text Region Localization

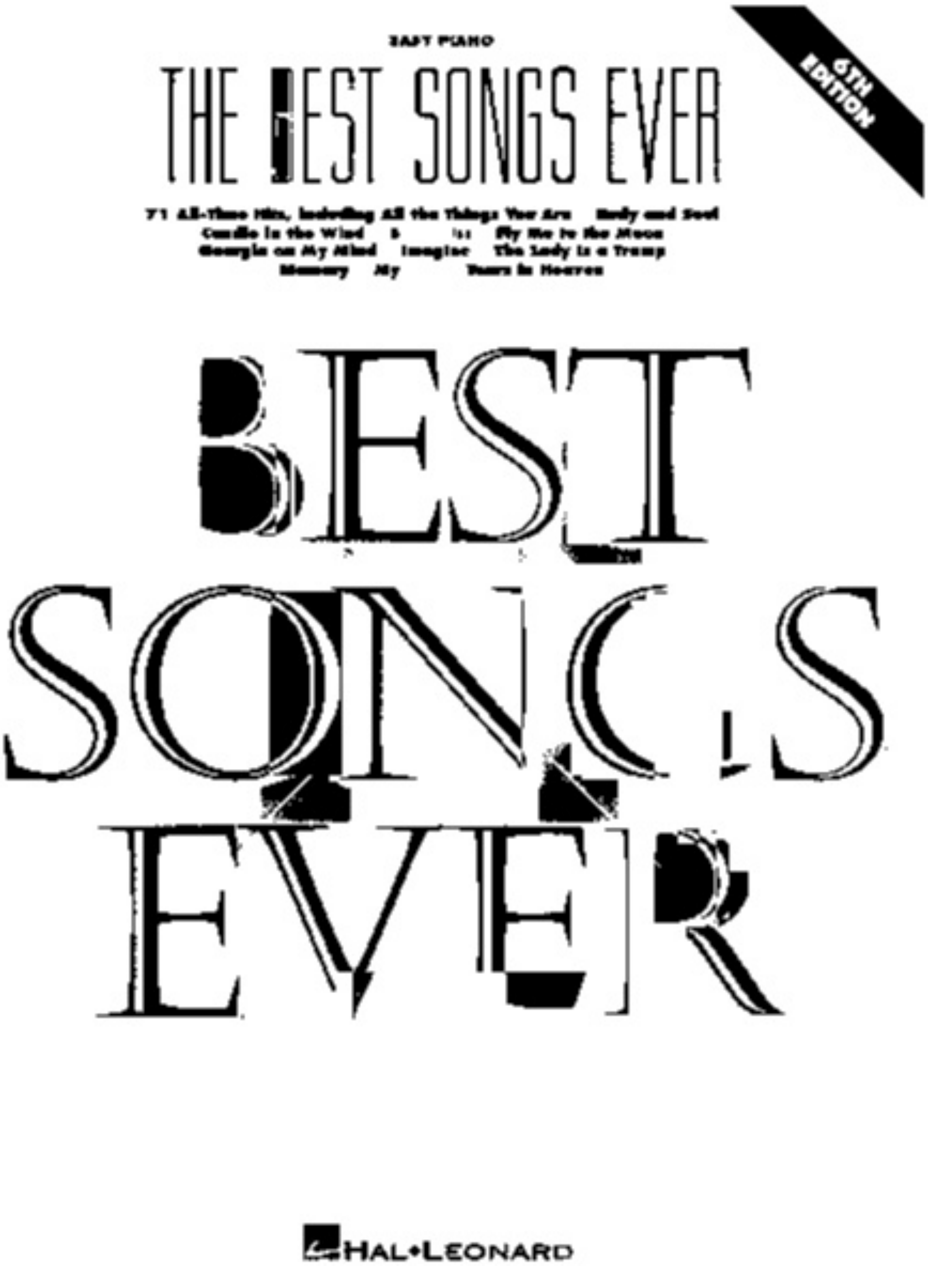
Since the intensity of the feature map represents the possibility of text, a simple global thresholding can be employed to highlight those with high text possibility regions resulting in a binary image. A morphological dilation operator can easily connect the very close regions together while leaving those whose position are far away to each other isolated.

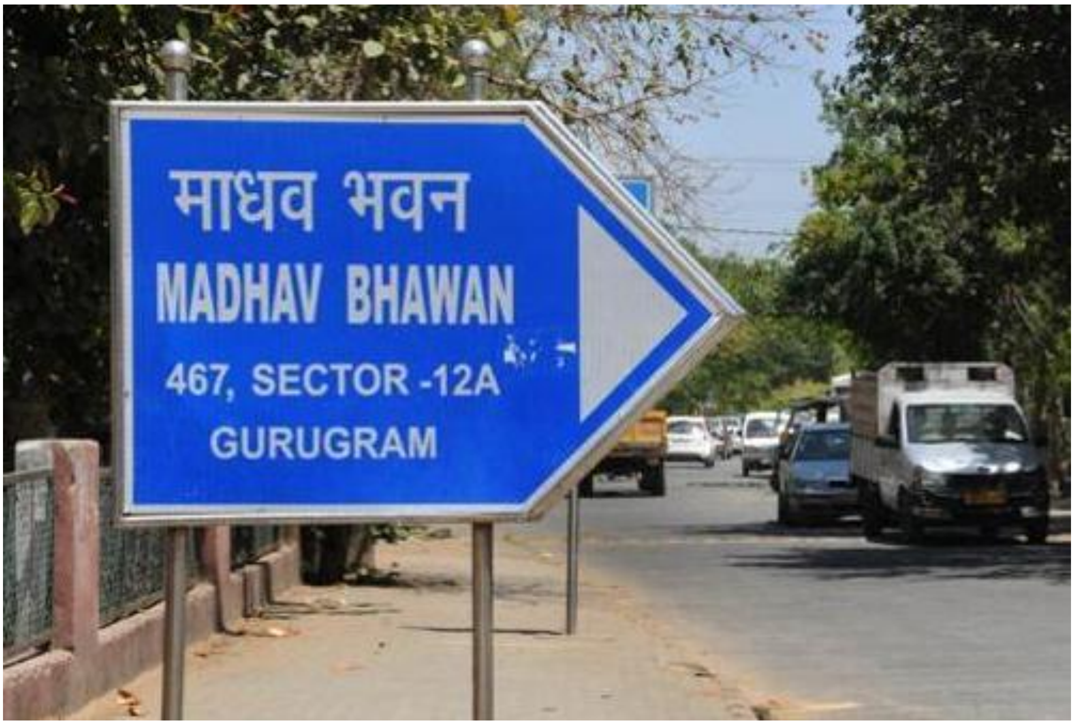
2.3 Character Extraction

Standard text processor such as Tesseract is operated on the feature map to get the characters.

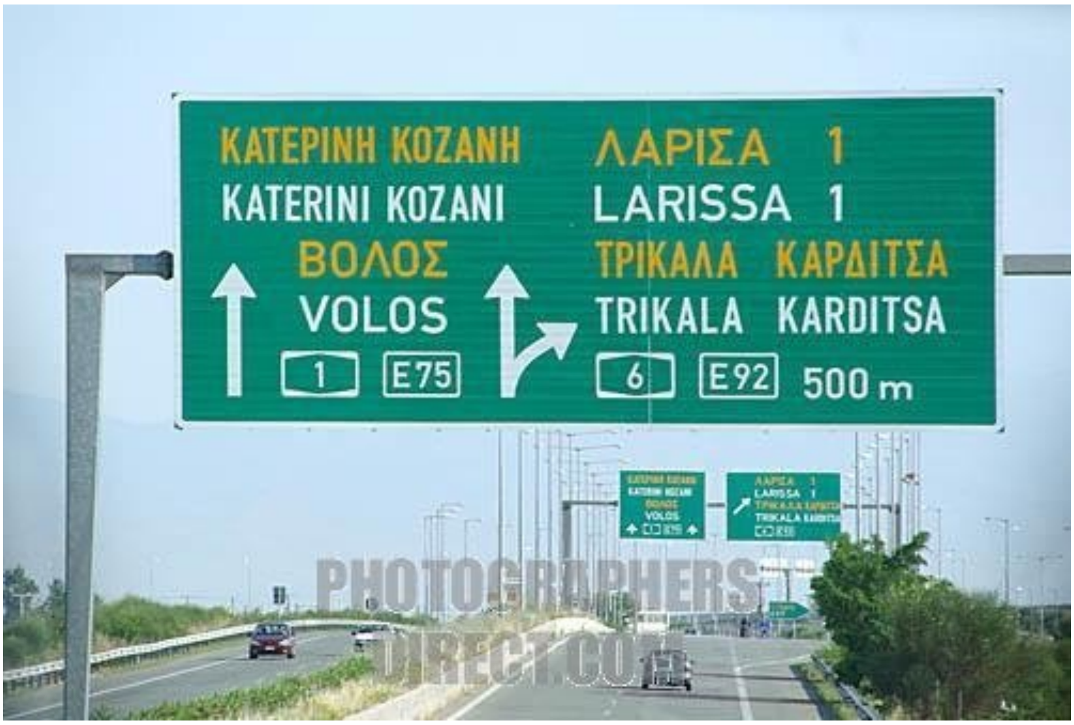
Results

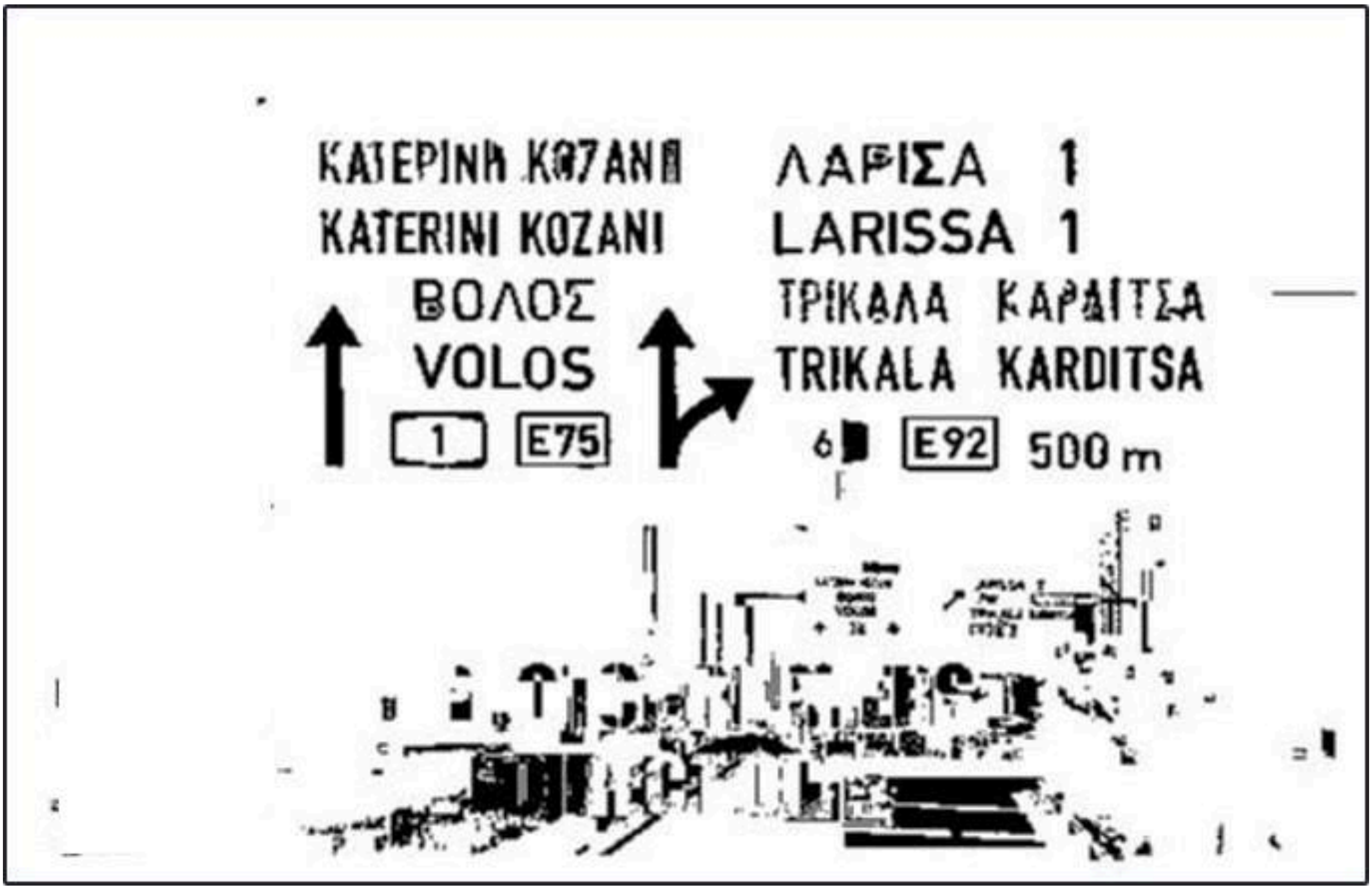



EASY PIANO
THE BEST SONGS EVER
6TH EDITION
71 All-Time Hits, Including All the Things You Are
Body and Soul
Candle in the Wind
Fly Me to the Moon
Georgia on My Mind
Imagine
The Lady Is a Tramp
Memory
Tears in Heaven
BEST SONGS EVER
HAL•LEONARD


माधव भवन
MADHAV BHAWAN
467, SECTOR -12A
GURUGRAM


धव भवन
ADHAV
467, SECTOR -12A
GURUGRA


ΚΑΤΕΡΙΝΗ ΚΟΖΑΝΗ
KATERINI KOZANI
ΒΟΛΟΣ
VOLOS
1
E75
ΛΑΡΙΣΑ 1
LARISSA 1
ΤΡΙΚΑΛΑ ΚΑΡΔΙΤΣΑ
TRIKALA KARDITSA
6
E92
500 m
PHOTOGRAPHERS
DIRECT.COM


KATERINI KOZANI
ΒΟΛΟΣ
VOLOS
1
E75
ΛΑΡΙΣΑ 1
LARISSA 1
TRIKALA KARDITSA
E92
500 m

Conclusion

We evaluate the performance of our proposed method and observe that the performance across the set of images is good overall.

Discussion

Aditya Narayan

- The performace of the proposed method is evaluated to be better in cases where there does not exist a lot of background noise. To ensure better results, the text localization process can be made adaptive to the size of the forground to ensure independence of scale.
- The performance was increased by switching to binary morphological operators instead of ones that operate over flating point images. To achieve this, OTSU thresholding was applied at each stage of feature map generation.
- The Gaussian pyramid generated was limited when no weak vertical edges were detected in the octave. This optimisation ensured that the number of octaves for each images was reduced to as low as four.

Adarsh Kumar Kosta

- The Multi-scale edge based detection proves to be a good strategy for text detection given the fact that the text regions inherently consist of multiple edges with strong edge density and varying orientations.
- The Gaussian pyramid approach makes the algorithm scalable and applicable to images of any resolution.
- The Feature map created defines the probabilty of the pixel belonging to a tet region.
- The Text recognition although is able to segragate text regions from textured regions generally, but regions with high edge density such as tree branches etc are also classified as text regions.