Paper ID-475: Local Gradient Difference Features for Classification of 2D-3D Natural Scene Text Images

Lokesh NANDANWAR, Palaiahnakote Shivakumara, Ramachandra Raghavendra, Tong Lu, Umapada Pal and Daniel Lopresti and Nor Badrul Anuar

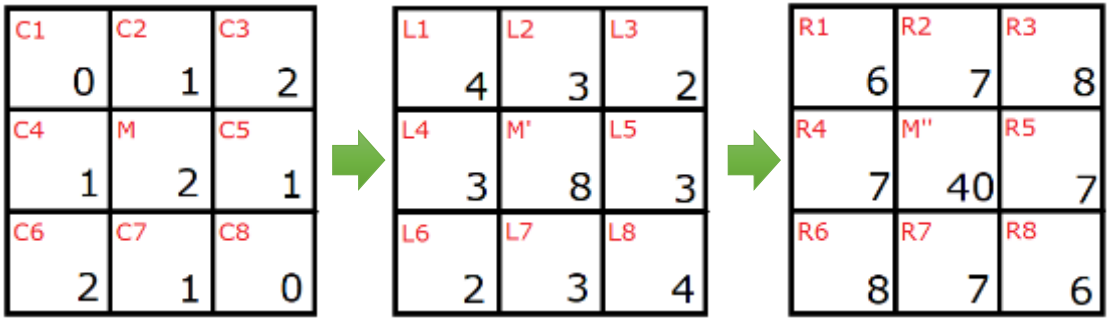
Motivation: Text detection performance before and after classification of 2D and 3D images

(a) Text detection by existing PSENet method in 2D and 3D text images before classification.

(b) Text detection by PSENet method in 2D and 3D text images after classification.

The Proposed Method

- Local Gradient Difference for Candidate Pixel Detection



$$LGD(M) = M'$$
$$= |M - C1| + |M - C2| + |M - C3| + |M - C4| + |M - C5| + |M - C6| + |M - C7| + |M - C8| = 8$$

$$LGR(M') = M''$$
$$= |M' - L1| + |M' - L2| + |M' - L3| + |M' - L4| + |M' - L5| + |M' - L6| + |M' - L7| + |M' - L8| = 40$$

Illustration for LGD and LGR for 3×3 Gradient window

The Proposed Method

- Local Gradient Difference for Candidate Pixel Detection

(a) Absolute of gradient images for 2D and 3D text images

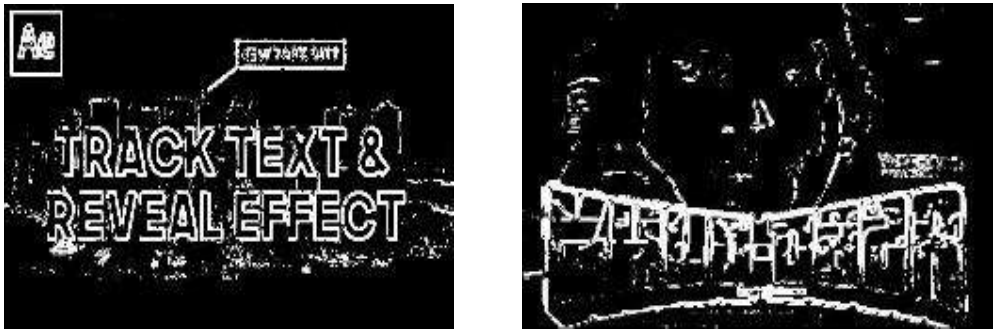
(b) Dominant pixels detection by the Max–Min clustering

(c) Local Gradient Difference (LGD) images for 2D and 3D images

The Proposed Method

- Local Gradient Difference for Candidate Pixel Detection

(d) Local Gradient Resultant (LGR) images for 2D and 3D images



(e) Max cluster results given by K-means clustering on LGR images. Candidate pixels detection based on local gradient difference.

The Proposed Method

- COLD for Extracting Spatial Proximiy of Candidate Pixels

$$\theta = \tan^{-1}\left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\right)$$

$$r = abs\left(\sqrt{(y_{i+1} - y_i)^2 + (x_{i+1} - x_i)^2}\right)$$

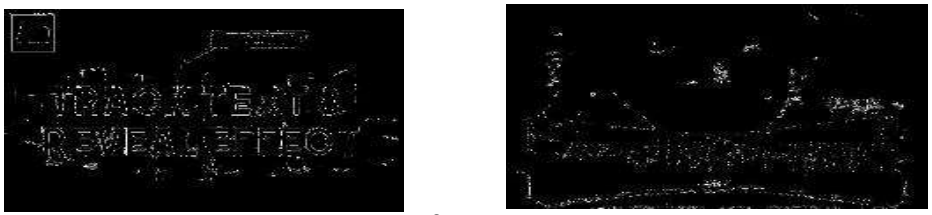
- Here $(x_i, y_i)$and (x_{i+1}, y_{i+1}) denote the coordinates of a pair pixels. When we draw points for all the pairs in polar domain (θ, r) it results in a distribution.

The Proposed Method

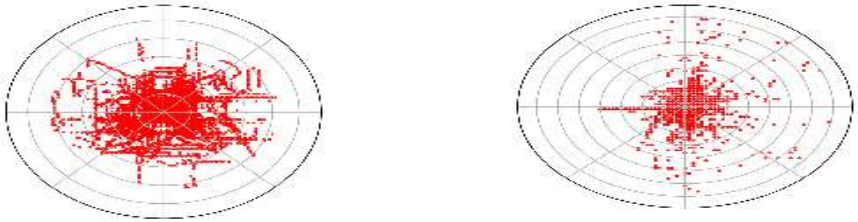
- COLD for Extracting Spatial Proximiy of Candidate Pixels

(a) Connected component labeling for candidate pixels images

(b) Traversing in 360 directions to find stroke pixels for each connected component. Yellow color denotes direction and red pixel denote stroke pixels.



(c) Stroke pixels for the 2D and 3D images



(d) Cold distribution for the stroke pixels of 2D and 3D images. Studying spatial distribution of stroke pixels through COLD.

The Proposed Method

- Mass Features Extracted from COLD for Classification
  - The radius is mean distance of stroke pixel pair
- *mass(x_a)* for a ring a, where $x_a \in \{x_1, x_2, \ldots, x_{n-1}, x_n\}$ is defined as a summation of a series of mass base weighted by *p(a)* over *n* rings.
- Here $\{x_1, x_2, \ldots, x_{n-1}, x_n\}$ is the number of pixels in rings from range$(1, n)$, where n is equal to 8.
- The mass is defined as follows:
- $mass(x_a) = \sum_{k=1}^{a-1}(n-a) \times p(a) + \sum_{k=a}^{n} a \times p(a), a \in (1, n)$
- Where $p(a)$ is the probability of number of pixels in ring a
- $p(a) = \frac{x_a}{\sum_{i=1}^{n} x_i} \quad a \in (1, n)$
- Finally, we combine all mass features for all rings a $\{a \in (1, n)\}$ to obtain the feature vector for the input image.
- This process results in the feature vector containing 8 features. The features are fed to NN for classification.

Experimental Results

Table: Details of different datasets for evaluation

Datasets	Type	2D	3D	Total
Our dataset	Image	400	400	800
Standard Natural Scene	Image	130	126	256
Our	Line	513	505	1018
IIIT5k	Line	317	305	632
COCO-Text	Line	472	530	992
ICDAR 2013	Line	123	74	197
ICDAR 2015	Line	111	90	201
Zhong et al.	Line	200	216	416
Ali et al.	Non-Text Images	250	250	500

Experimental Results

Ablation Study

Table: Confusion matrix and average classification rate for the key steps of the proposed method on our dataset at image level(in %).

Classes	Proposed without LGD		Proposed without COLD		Proposed with density		Proposed Method	
	2D	3D	2D	3D	2D	3D	2D	3D
2D	61.25	38.75	76.25	23.75	77.5	22.5	85.0	15.0
3D	47.5	52.5	41.25	58.75	27.5	72.5	21.25	78.75
Average	56.875		67.5		75.0		81.875	

Experimental Results

Evaluating the Proposed Classification

Table: Classification at image level, text line level and comparative study with the existing methods in (%)

Dataset	Our dataset-image level						Standard dataset-image level					
Methods	Zhong et al.		Xu et al.		Proposed		Zhong et al.		Xu et al.		Proposed	
Classes	2D	3D	2D	3D	2D	3D	2D	3D	2D	3D	2D	3D
2D	76.5	23.5	58.7	41.3	83.0	17.0	52.0	48.0	56.4	33.6	78.4	21.6
3D	41.5	58.5	31.9	68.1	22.0	78.0	42.4	57.6	39.2	60.8	26.5	72.5
Average	67.5		63.4		80.5		54.8		58.6		75.45	

Dataset	Our dataset –Text line level						Standard dataset-Text line level					
Methods	Zhong et al.		Xu et al.		Proposed		Zhong et al.		Xu et al.		Proposed	
Classes	2D	3D	2D	3D	2D	3D	2D	3D	2D	3D	2D	3D
2D	32.7	67.2	66.2	33.7	91.1	8.9	53.8	46.1	46.6	53.4	87.7	12.3
3D	46.3	53.6	28.3	71.6	11.2	88.8	32.8	67.2	31.2	68.8	16.5	83.5
Average	43.2		70.48		89.95		60.53		57.7		85.6	

Dataset	Zhong et al dataset						Ali et al., Non-Text Image dataset	
Methods	Zhong et al.		Xu et al.		Proposed		Proposed Method	
Classes	2D	3D	2D	3D	2D	3D	2D	3D
2D	67.0	23.0	74.8	25.2	92.9	7.1	82.0	18.0
3D	28.4	65.6	26.0	64.0	12.8	87.2	24.0	76.0
Average	66.3		69.4		90.05		78.0	

Experimental Results

Validating the Proposed Classification

Table: Text detection performance in terms of F-measure of different methods for our and standard full dataset at image level before and after classification. BC denotes before classification and AC denotes after classification.

Methods	Our Dataset-image level				Standard dataset-image level			
	BC	AC			BC	AC		
	2D + 3D	2D	3D	Avg	2D + 3D	2D	3D	Avg
PSEnet	67.9	73.3	66.9	70.1	73.3	88.6	64.0	76.3
FOTS	59.2	70.3	56.9	63.6	64.3	75.1	60.5	67.8
DB	60.5	68.2	59.1	63.6	66.6	80.8	61.4	71.1

Table: Text recognition performance in terms of character recognition rate of different methods for our and standard dataset at line levels before and after classification. BC denotes before classification and AC denotes after classification.

Methods	Our Dataset-Text line level				Standard dataset-Text line level			
	BC	AC			BC	AC		
	2D + 3D	2D	3D	Avg	2D + 3D	2D	3D	Avg
ASTER	79.0	97.0	85.7	91.3	88.5	96.1	85.4	90.7
MORAN	87.2	94.1	87.6	90.8	89.7	96.0	86.4	91.2

Conclusion and Future Work

- We have proposed a new method for the classification of 2D and 3D text in natural scene images.
- The proposed method employs a local gradient difference for detecting candidate pixels from input images.
- The COLD approach used for representing the spatial relationship between candidate pixels in 2D and 3D images.
- The proposed method estimates mass for extracting such observations from each ring over the COLD distribution.
- The extracted mass features are fed to a Neural Network classifier for the classification of 2D and 3D images.
- Experiments on classification, text detection and recognition show that the proposed classification method is effective and useful.
- However, the reported results are still low. Our next target is to investigate new features for improving the proposed method classification.

Thank you for your patience

Questions and Suggestions