Using histogram representation and Earth Mover's Distance as an evaluation tool for text detection

Ana Stefania Calarasanu[1]
Jonathan Fabrizio[1], Séverine Dubuisson[2]

LRDE[1], ISIR[2]

calarasanu@lrde.epita.fr

Monday 24th August, 2015

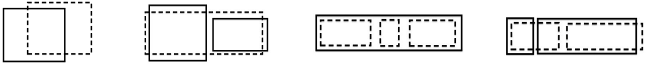
Overview

Text detection performance evaluation

- GROUND TRUTH:

Annotation levels: pixel(blue), character(red), word(green), line(magenta).

- MATCHING PROTOCOL:



Matching cases: GT (dashed) and detections (plain line).

- METRICS:

recall: *proportion of detected texts in the* GT,

precision: *proportion of accurate detections.*

Detection quantity-quality relationship

QUANTITY

how many GT objects have been detected?
how many detections have a match in the GT?

QUALITY

how much of the matched GT objects was detected?
how accurate is the detection of the objects?

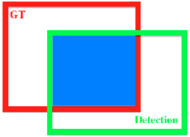
[Wolf and Jolion, 2006]

Detection quantity-quality relationship

EXAMPLE: *Coverage/Accuracy* quality measures:

$$R = \frac{\sum Cov}{nb.\ of\ GT\ objects} \qquad Cov_i = \frac{Area(G_i \bigcap D_j)}{Area(G_i)} =$$

$$P = \frac{\sum Acc}{nb.\ of\ detections} \qquad Acc_i = \frac{Area(G_i \bigcap D_j)}{Area(D_j)} =$$



DETECTOR 1

DETECTOR 2

$Recall = 0.5$

DETECTOR 3

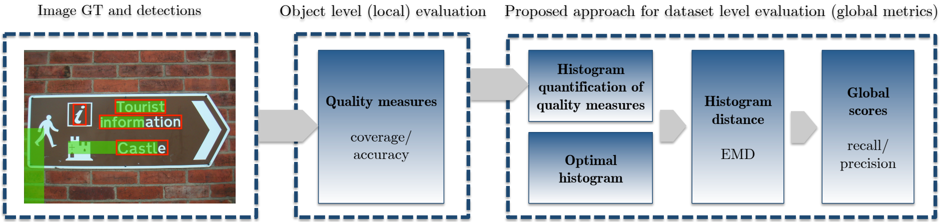
DETECTOR 4

$Precision = 0.33$

GROUND TRUTH and DETECTION text boxes.

Contributions

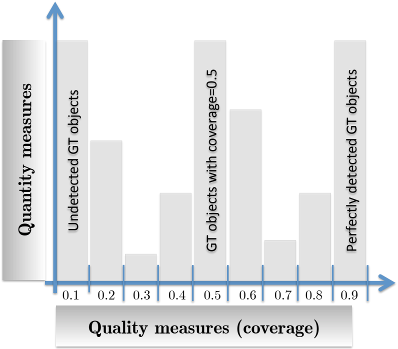
1. Capture the detection *quantity-quality* nature using histogram representation.
2. The use of histogram distances to derive global scores.

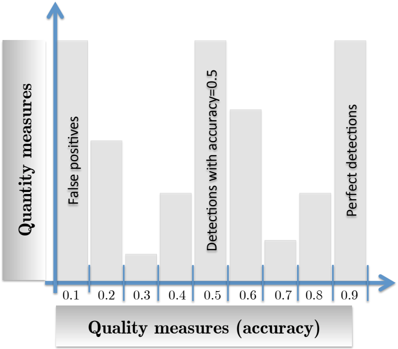


Workflow of the proposed method.

Note: the framework requires a qualitative object-level evaluation.

Quality detection histograms





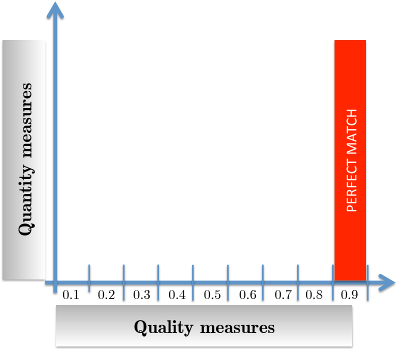
$$h_{Qual}(b) = \begin{cases} \sum_{j=1}^{m} \left\{ f_{Qual}(j) \in \left[\frac{b}{B}, \frac{b+1}{B} \right[\right\} & \text{if } b = 0, \ldots, B-2 \\ \sum_{j=1}^{m} \left\{ f_{Qual}(j) \in \left[\frac{b}{B}, \frac{b+1}{B} \right] \right\} & \text{if } b = B-1 \end{cases}$$

Optimal histogram

Optimal Histogram ($\widetilde{h_O}$) = perfect quality detection.

Global scores = $\text{dist}(h_{Qual},\widetilde{h_O})$

e.g. $Recall = \text{dist}(h_{Cov}, \widetilde{h_O})$;
$Precision = \text{dist}(h_{Acc}, \widetilde{h_O})$.



Earth Mover's Distance

Minimal cost that must be paid to transform a signature (P) into another signature (Q). [Rubner, 2000]

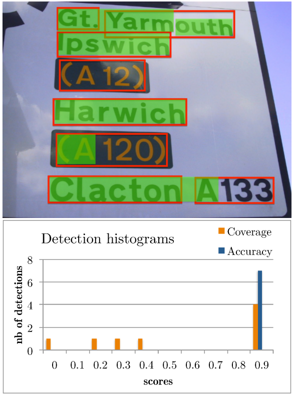
$$P = \{(p_i, w_{p_i}) \mid i \in [1, m]\} \quad Q = \{(q_j, w_{q_j}) \mid j \in [1, n]\}$$

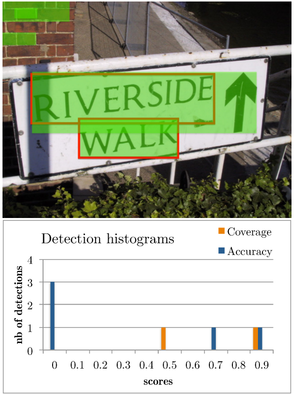
$$EMD(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}}$$

+ cross-bin distance
+ can be applied to normalized histograms
+ is a true metric [Rubner et al., 2000]

$$R = 1 - EMD(\widetilde{h_{Cov}}, \widetilde{h_O})$$
$$P = 1 - EMD(h_{Acc}, h_O)$$

Results on singular images

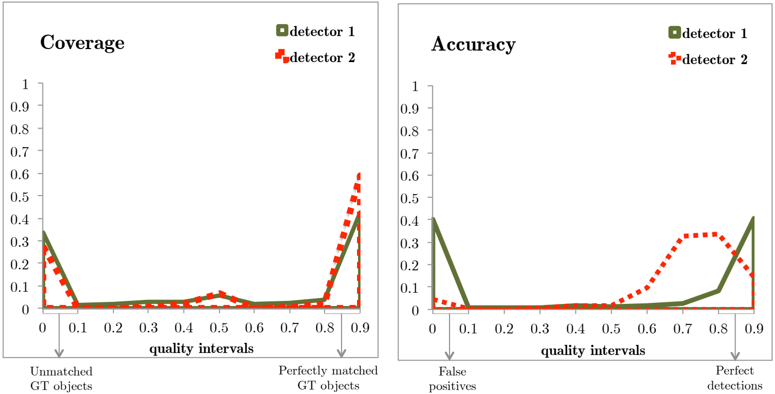




$R = 0.66,\ P = 1$ $R = 0.8,\ P = 0.42$

Two examples of GT (red rectangles) and detections (green plain rectangles) and their corresponding coverage/accuracy histograms (resp. h_{Cov} (orange) and h_{Acc} (blue)) and R/P scores.

Results on a set of images
Comparison of two detectors



Coverage and accuracy normalized histograms associated to *detector 1* ($R = 0.60$, $P = 0.58$) and *detector 2* ($R = 0.70$, $P = 0.80$).

Detection quantity-quality relationship

DETECTOR 1

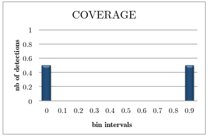
DETECTOR 2

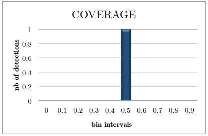
$Recall = 0.5$

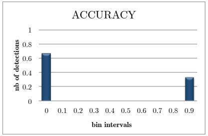
DETECTOR 3

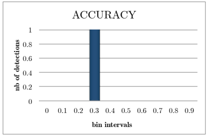
DETECTOR 4

$Precision = 0.33$









GROUND TRUTH and DETECTION text boxes.

Conclusions

- intuitive visual representation of detection results
- better delimitation of the quantity from the quality aspects
- easy comparison between detectors
- powerful similarity measure (EMD) to depict global scores

Future works

- available tool online

References

Rubner, Y., Tomasi, C., and Guibas, L. (2000).
The earth mover's distance as a metric for image retrieval.
IJCV, 40(2):99–121.

Wolf, C. and Jolion, J.-M. (2006).
Object count/area graphs for the evaluation of object detection and segmentation algorithms.
IJDAR, 8(4):280–296.

Using histogram representation and Earth Mover's Distance as an evaluation tool for text detection

Ana Stefania Calarasanu[1]
Jonathan Fabrizio[1], Séverine Dubuisson[2]

LRDE[1], ISIR[2]

calarasanu@lrde.epita.fr

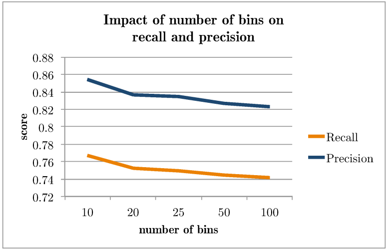
Monday 24th August, 2015

Results: ICDAR2013 Set

Impact of tuning the number of bins

Method	Recall	Precision
EMD_{10bins}	0.7667	0.8799
EMD_{20bins}	0.7526	0.8713
EMD_{25bins}	0.7495	0.8693
EMD_{50bins}	0.7441	0.8659
$EMD_{100bins}$	0.7413	0.8642

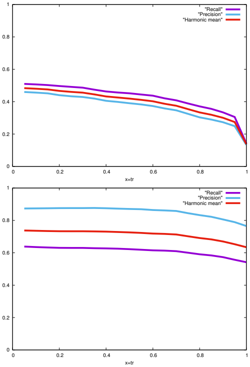
Bin size impact on recall and precision scores.



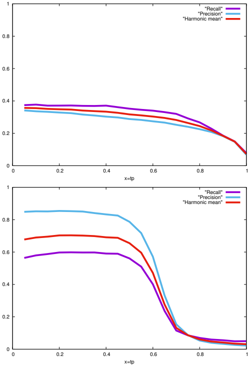
Variation of R_G and P_G scores depending on the number of bins B

Observation: stabilization of these two global scores when number of bins sufficiently large.

Comparison to AUC plots



(c) varying constraint t_r



(d) varying constraint t_p

Performance plots generated with *DetEval* tool [Wolf and Jolion, 2005] (recall in purple, precision in blue); top: *detector 1* ($R_{OV} = 0.37$, $P_{OV} = 0.32$); bottom: *detector 2* ($R_{OV} = 0.49$, $P_{OV} = 0.69$).

Earth Mover's Distance detailed

Let $P = \{(p_i, w_{p_i})\}_{i=1}^{m}$ and $Q = \{(q_j, w_{q_j})\}_{j=1}^{n}$ be two signatures where p_i and q_j are the position of ith, respectively jth element and w_{p_i} and w_{q_j} their weights. The EMD searches for a flow $F = [f_{ij}]$ between p_i and q_j, that minimizes the cost to transform P into Q:

$$COST(P, Q, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}, \tag{1}$$

where d_{ij} is the ground distance between clusters p_i and q_j; the cost minimization is done under the following constraints:

$$f_{ij} \geq 0, \quad \sum_{j=1}^{n} f_{ij} \leq w_{p_i}, \quad \sum_{i=1}^{m} f_{ij} \leq w_{q_j}, \;\; i \in [1, m], \; j \in [1, n]$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = min(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}), \;\; i \in [1, m], \;\; j \in [1, n]$$

The EMD distance is then defined as:

$$EMD(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \tag{2}$$