**TEXT LOCALIZATION IN REAL-WORLD IMAGES USING AN EFFICIENTLY PRUNED SEARCH**

Lukáš Neumann and Jiří Matas
Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University in Prague

neumal@cmp.felk.cvut.cz  
matas@cmp.felk.cvut.cz

http://cmp.felk.cvut.cz/~neumal/  
http://cmp.felk.cvut.cz/~matas/

**INTRODUCTION**

- An efficient method for text localization and recognition in real-world images is proposed.
- The search and recognition of text sequences is real-time (200ms on an 640x480 image).
- The key role of the character grouping stage is demonstrated.
- A novel selector of Maximally Stable Extremal Regions (MSERs) [1] is introduced; experimental validation shows that 95.7% characters are detected.
- State-of-the-art results in text localization on the standard ICDAR 2003 dataset.
- The only method that has published both text localization and recognition results on the ICDAR dataset.

### EXTENDED MAXIMALLY STABLE EXTREMAL REGIONS

- Individual characters detected as MSER++ (MSERs with low margin in multiple projections, inclusion relation is taken into account).
- Computational complexity linear in number of pixels, robustness to geometric and illumination conditions and variations in foreground and background.

### VERIFICATION FUNCTIONS

- Verification function \( v(w) \rightarrow \{0,1\} \) estimates whether sequence of image regions \( w \) is text.
- Text localization methods aim to find an optimal verification function \( v(w) \) to maximize precision \( p \) / recall \( r \) ("T" denotes words in ground truth, \( S \) set of all sequences in an image).
- Verification function \( v(w) \) for sequences of arbitrary length is decomposed into \( N \) individual verification functions \( v_1(w), v_2(w), \ldots, v_N(w) \) for sequences of specific length 1, 2, ..., \( N \).
- Unary verification function \( v(w) \) is an SVM classifier/character classifier.
- The verification function \( v(w) \) is implemented by several pairwise rules (height ratio, mutual distance, ...) and a rule preventing region inclusion (taking MSER lattice into account).
- The verification function \( v(w) \) creates text line approximation and checks consistency of the approximation with regions in the triplet.

### EFFICIENTLY PRUNED EXHAUSTIVE SEARCH

- The method iteratively builds sets of text sequences of length 1, 2, 3, ...
- Search is efficiently pruned by the verification functions.
- Conflicting sequences (each region can be in one sequence only) removed, longer sequences preferred.

### EXPERIMENTS

- The method was evaluated on the ICDAR 2003 dataset [2] (249 images).
- Hypotheses-verification framework [3] was used to perform text recognition.

### CONCLUSIONS

State-of-the-art results:
- Text localization: precision 0.65, recall 0.64.
- Text recognition: precision 0.42, recall 0.41.
- 0.83s on ICDAR dataset, on average.
- 0.2s for 640x480 resolution.

Limitations of the proposed method:
- Reflection of a flash is too strong on letter "n", unsupported text line shape, multiple letters joint in one MSER.

### REFERENCES