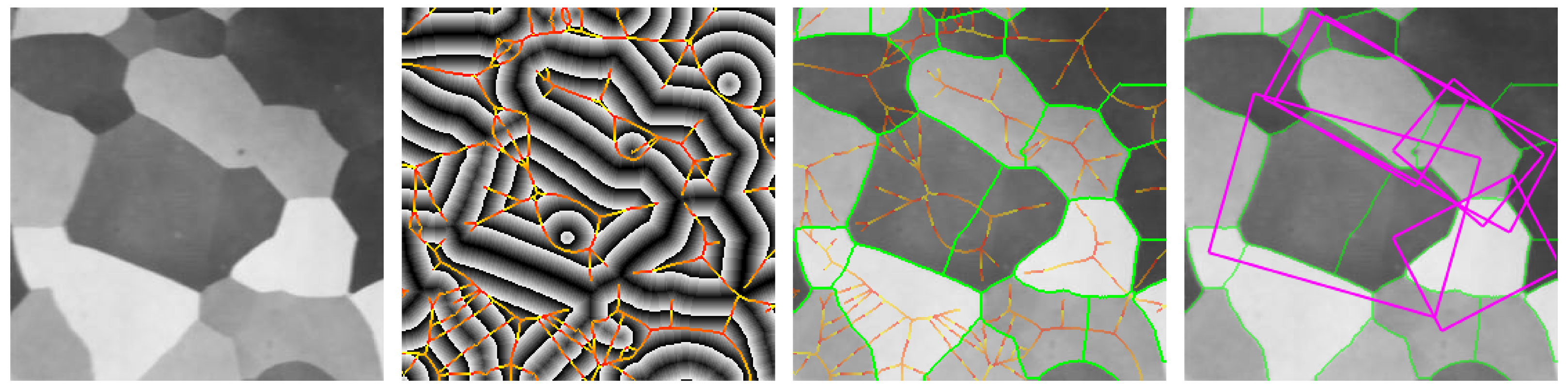
The Medial Feature Detector: Stable Regions from Image Boundaries

Yannis Avrithis and Kostas Rapantzikos
School of Electrical and Computer Engineering, National Technical University of Athens, Greece

Overview

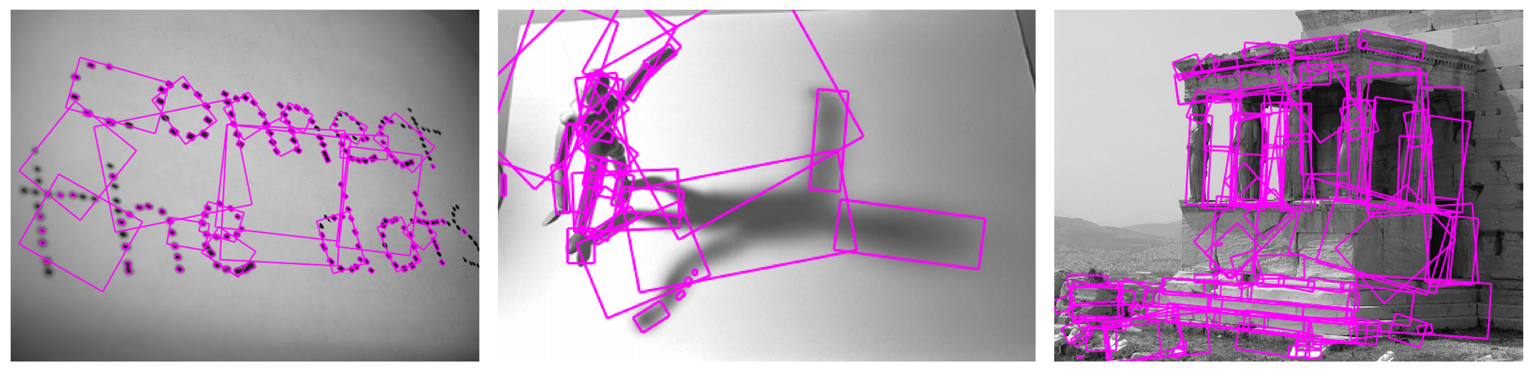
- Regions of arbitrary scale and shape, without scale space construction.
- Weighted *distance map* on image gradient, using our exact, linear-time algorithm.
- Weighted *medial axis* by generalizing chord residue, typically used in Voronoi skeletons.
- *Decompose* medial axis into a graph representing image structure in terms of peaks and saddle points.
- *Reconstruct* regions using the same propagation method, due to a duality property.



input (270 × 270) dt+medial (45ms) partition (15ms) features (3ms)

- Join regions and select those that are *well enclosed by boundaries*, in agreement with the Gestalt principle of closure; *incomplete boundaries* allowed.
- Select *whole regions* and their *parts* independently.
- *Binary code* and documentation at `http://image.ntua.gr/iva/tools/mfd/`

Examples

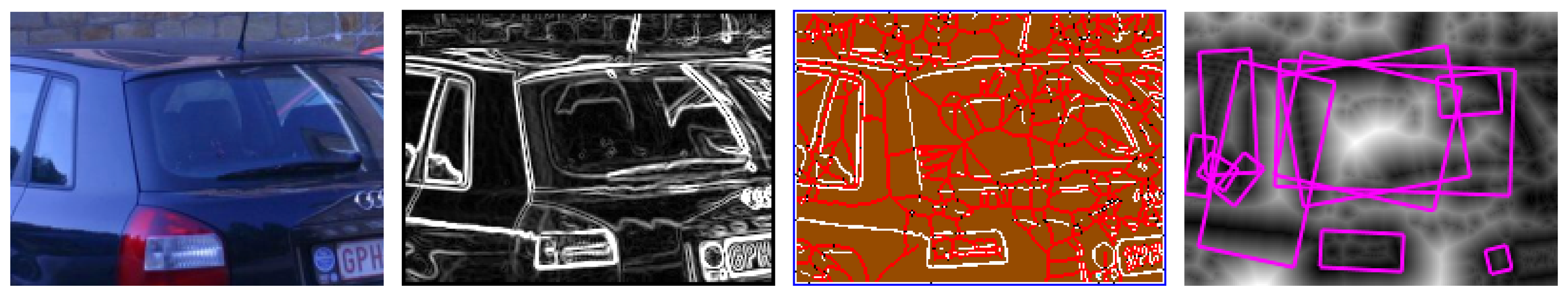


Weighted distance transform

- Given input f in domain X, define its *weighted distance map* $h = \mathcal{D}(f)$ as

$$h(x) = \bigwedge_{y \in X} d(x,y) + f(y), \quad x \in X,$$

where $\wedge$ denotes infimum and d Euclidean distance.
- Given image f_0, compute gradient magnitude $g = |\nabla f_0|$ and define $f(x) = \sigma / g(x)$ for $x \in X$, where σ is a *scale* parameter—generalizing the $0/\infty$ *indicator* function.
- *Exact group marching* (EGM) algorithm: move *all* points on the propagating front as a group, in arbitrary order; *exact* result, *linear* in the number of points.

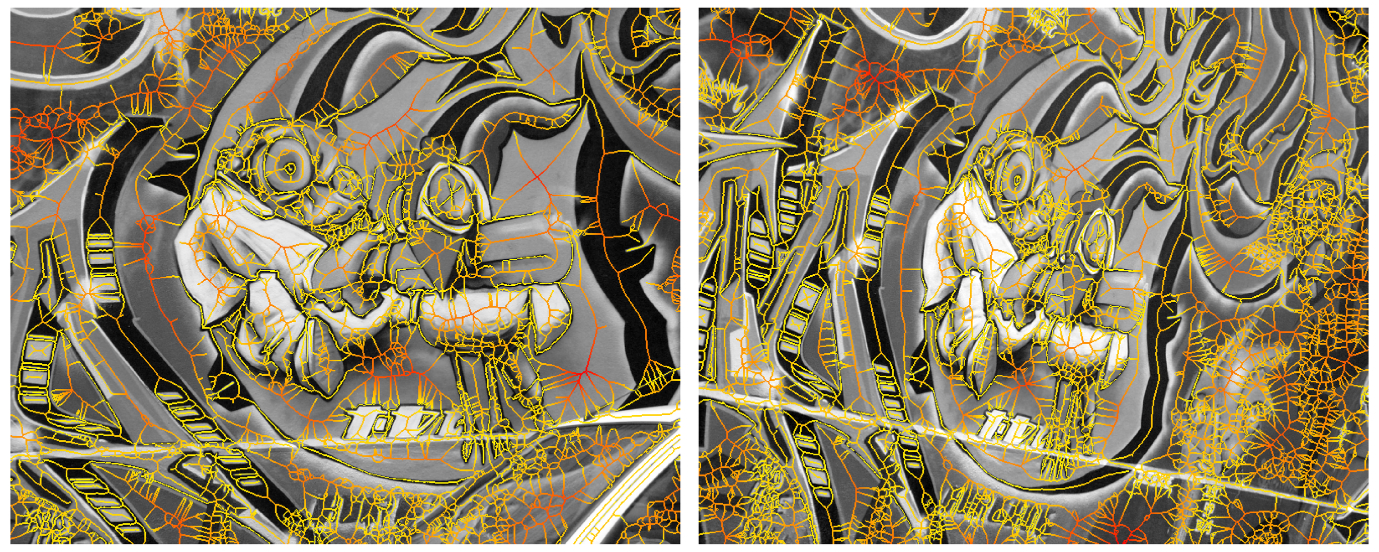


input gradient labels dt + features

- For each $x \in X$, its *source set* $S(x)$ is the set of points $y \in X$ for which $d(x,y) + f(y)$ is minimized, and is used to define the medial axis.
- The *source set* $S(f)$ of f contains all source points of f in X; $\mathcal{D}(f)$ is uniquely determined by f restricted on $S(f)$.

Weighted medial axis

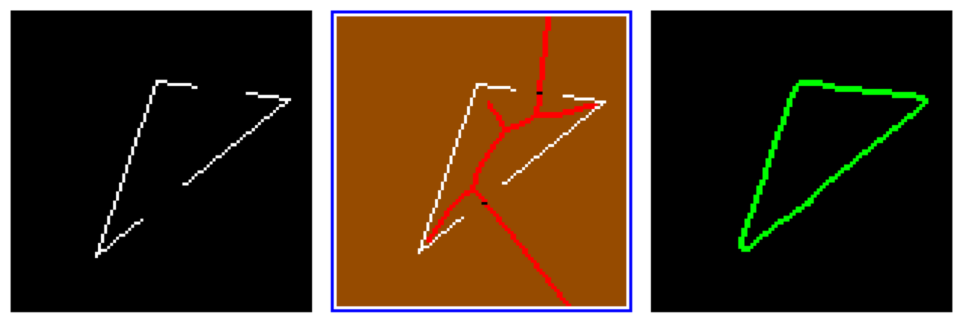
- Point $x \in X$ is a *medial point* if it has at least two distinct sources. The (weighted) *medial axis* $A(f)$ contains all such points: $A(f) = \{x \in X : |S(x)| > 1\}$.
- *Weighted medial axis* (WMA) algorithm computes $A(f)$ given $\mathcal{D}(f)$ and $S(f)$, beginning at peaks of $\mathcal{D}(f)$ and propagating based on *residue function* $r(x)$, generalizing *chord residue* by Ogniewicz and Kubler.



- Using a graph to represent the topology of the source set $S(f)$, compute $r(x)$ in constant time for each $x \in X$; hence WMA is *linear* in the number of points.
- Guarantee that $A(f)$ is connected for each component of $X \setminus S(f)$.

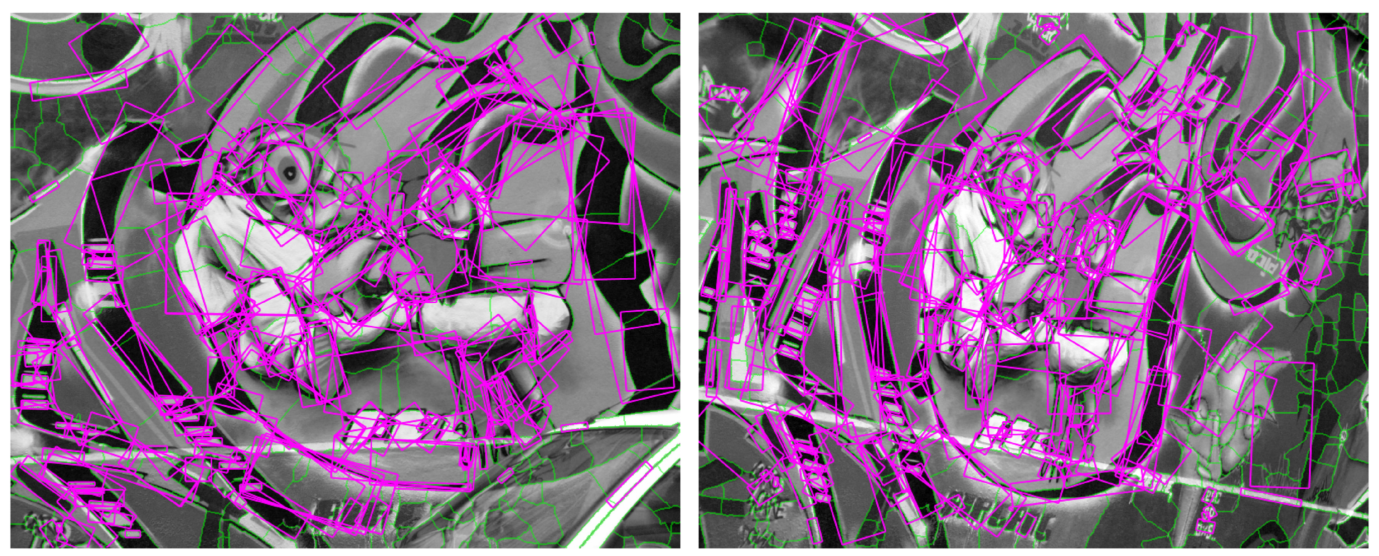
Feature detection

- Decompose $A(f)$ and construct weighted graph $\mathcal{G}(f)$ such that: (a) vertices correspond to peaks of $\mathcal{D}(f)$; (b) edges correspond to local minima along $A(f)$ (*saddle points* of $\mathcal{D}(f)$); (c) edge weights defined as the height at those points.
- *Medial axis decomposition* (MAD) algorithm constructs $\mathcal{G}(f)$ given $h = \mathcal{D}(f)$ and $A(f)$; equivalent to *watershed* of $-h$ on $A(f)$, with peaks as markers.



- *Partition* image: invoke EGM with input $g(x) = -h(x)$ if $x \in A(f)$, and $+\infty$ otherwise, *backpropagating* from medial to boundaries; equivalent to *watershed* of $-h$ on X.
- Each edge e of $\mathcal{G}(f)$ is generated at saddle point $x(e)$ with boundary gap $w(x(e))$; then component κ with area $a(\kappa)$ and incident edge set $E(\kappa)$ has *shape fragmentation factor*

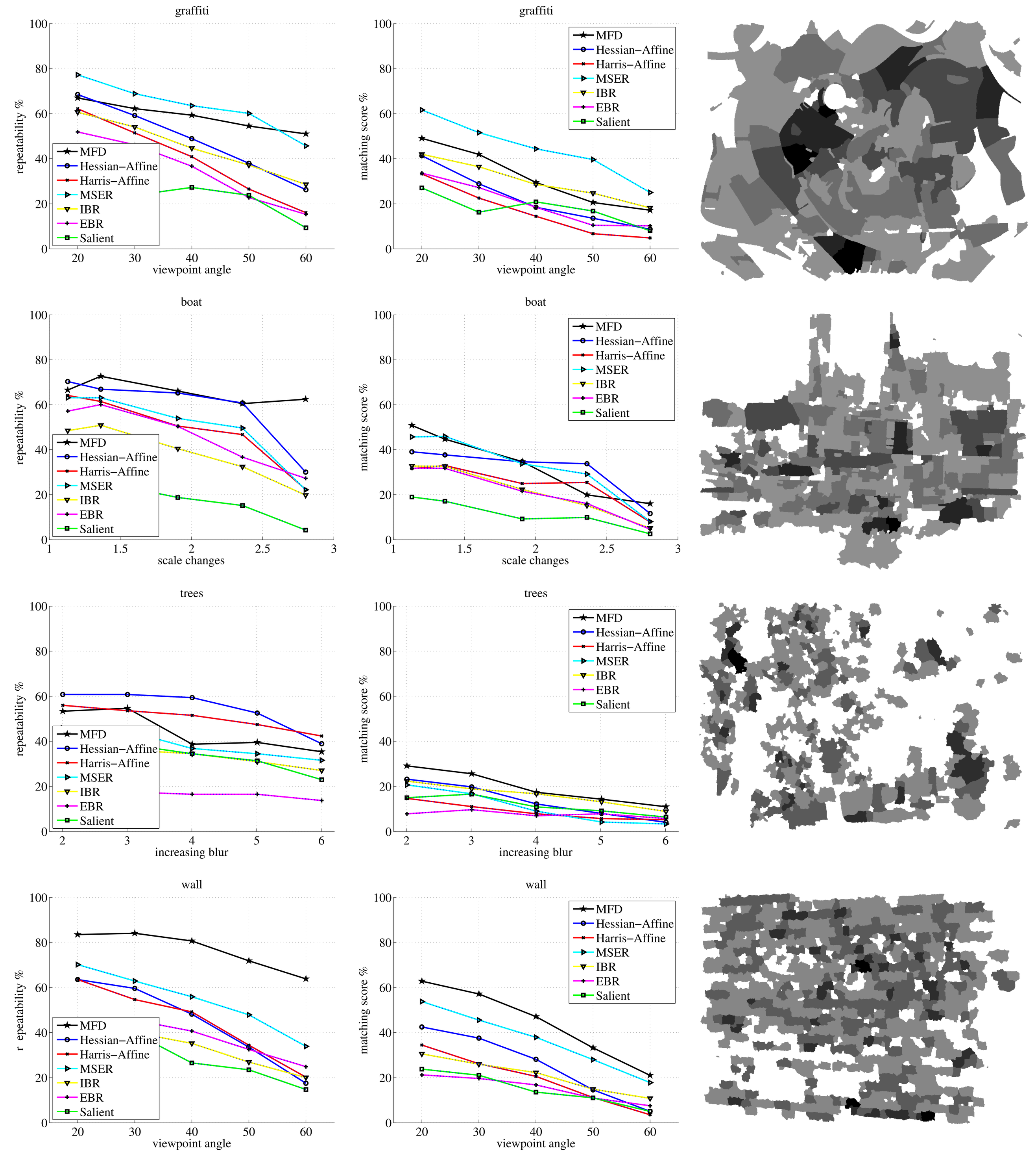
$$\phi(\kappa) = \frac{1}{a(\kappa)} \sum_{e \in E(\kappa)} w^2(x(e)).$$



- Join components in non-increasing order of edge weights and select component κ as a *feature* if $\phi(\kappa) < \tau$ where threshold $\tau > 0$ controls *selectivity*.

Matching experiment

- *Matching* across viewpoint, zoom, rotation, light, and blur; measure performance in terms of *repeatability* and *matching score*, using SIFT descriptors for all detectors.



graffiti
800 × 640
530 features
0.55 sec

boat
850 × 680
665 features
0.72 sec

trees
1000 × 700
594 features
1.05 sec

wall
1000 × 700
876 features
1.08 sec

Retrieval experiment

- Larger scale *retrieval* using *bag of words* (BoW); ranking by TF-IDF and *fast spatial matching* (FastSM); measure performance in terms of *mean Average Precision* (mAP).
- *Oxford 5K* dataset, comprising 5,062 images with 55 queries. 200K vocabulary from the same dataset, constructed separately for each detector.

Detector	Features ($\times 10^6$)	Space (MB)	Query (ms) Index	Query (ms) Rerank	mAP Index	mAP Rerank
MFD	9.32	68	3.96	8.2	**0.580**	**0.617**
Hessian-affine	29.03	126	6.72	44.9	0.573	0.614
MSER	13.33	76	4.30	13.2	0.544	0.589
SURF	**4.24**	**34**	**3.00**	**1.7**	0.531	0.536
SIFT	11.13	82	4.61	8.4	0.457	0.495

- Indexing space and query time depend on the average number of features per image; the objective is highest mAP with reasonable space/time requirements.

Current and future work

- Exploit exact region shape; other types of features; GPU implementation.
- Generic image segmentation; edge detection and grouping.
- Shape-based object detection.

Contact: {iavr,rap}@image.ntua.gr