netherlands eScience center

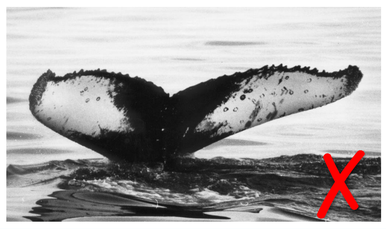
E. Ranguelova

Local Shape and Moment Invariant Descriptor for Structured Images

Objective

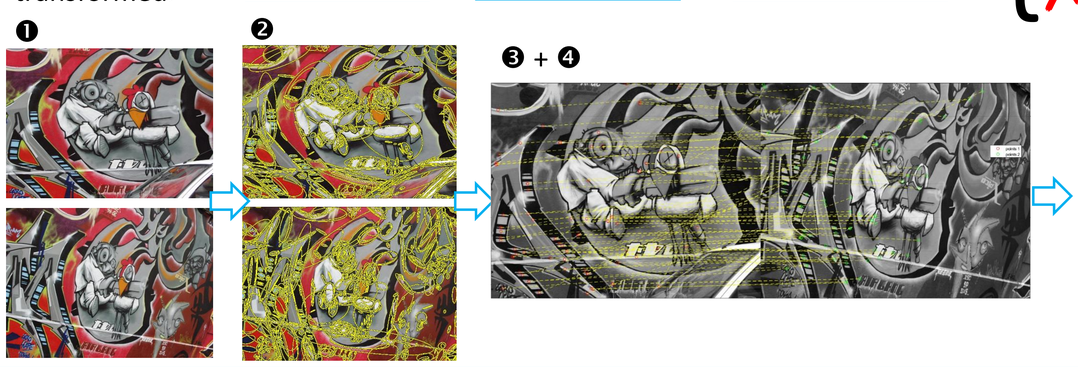
Develop a new salient region descriptor using region's shape to determine if two images, subject to geometric or photometric distortions, are depicting the same scene or object.

e. g. Photo identification application



Local features correspondence

❶ Image pair <original, transformed> → ❷ Feature Detection → ❸ Feature Description → ❹ Matching → ✓ Same scene / ✗ Not same scene

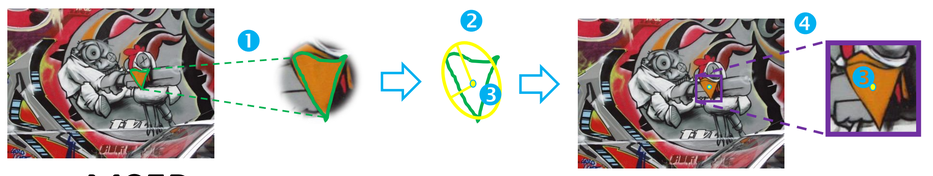


✓ Same scene

Feature descriptors

Region descriptors = Point descriptors ❸ : ❶→❷→❸→❹→❺

❶ region → ❷ equivalent ellipse → ❸ centroid → ❹ window → ❺ descriptor



e.g. MSER

e.g. SURF, SIFT

64, 128 –dimensional

The shape of the region is not used!

Shape and Moment Invariant (SMI) Descriptor

The shape of the region is used!

Shape invariants

Invariant	Definition	Description
Relative Area	$\tilde{a_i} = a_i/A$	Region's area normalized by the image area A
Ratio Axes Lengths	$r_i = v_i/\mu_i$	Ratio between ellipse minor and major axes lengths
Eccentricity	$e_i = \phi_i/\mu_i$	$e_i \in [0,1]$ (0 is a circle, 1 is a line segment)
Solidity	$s_i = a_i/a_i^c$	Proportion of the convex hull pixels, that are also in the region

Affine Moment invariants

$$M(I) = \int_{-\infty}^{+\infty} \Pi_{k,j=1}^{N} C_{kj}^{n_{kj}} \cdot \Pi_{l=1}^{N} I(x_l, y_l) dx_l dy_l,$$

$I(x,y)$ - real values image function,
n_{kj} - non-negative integers,
$C_{kj} = x_k y_j - x_j y_k$ - cross-product between points (x_k, y_k) and (x_j, y_j)

SMI descriptor: *20-dimensional* vector $\boldsymbol{SMI_i} = \{\tilde{a_i}, r_i, e_i, s_i, m_{i1}, \ldots, m_{i16}\}$

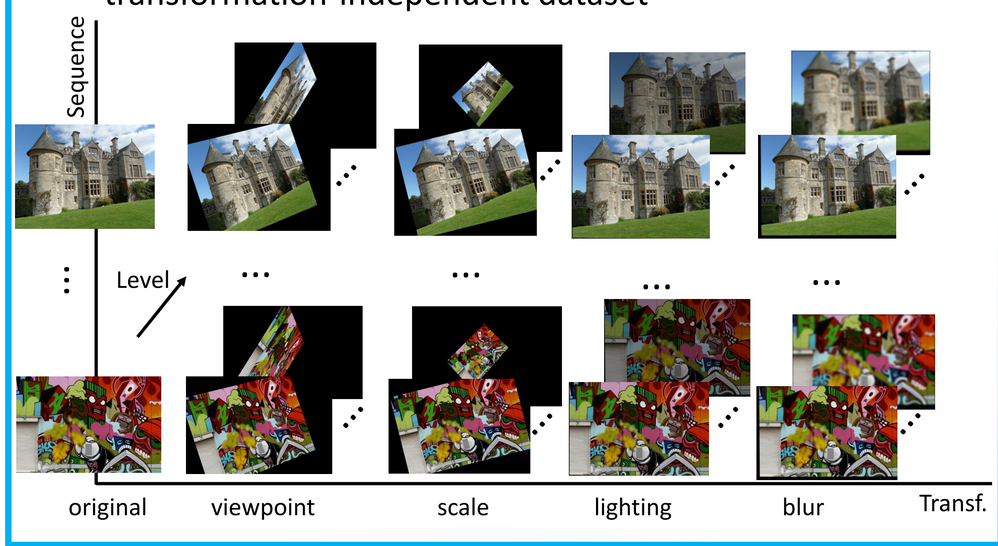
Matching: Sum of Squared Differences between two image descriptor sets → affine transformation estimate → image transformation → average correlation between original and transformed images is final similarity score → thresholding → Same/Not same

Data

Oxford (VGG) dataset: 4 sequences x 6 images = 24
Every sequence represents only one transformation

viewpoint scale lighting blur

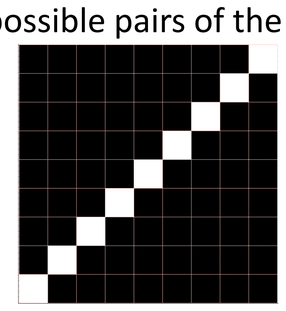
OxFrei dataset: 9 sequences x 21 images = 189
All sequences have all transformations → transformation-independent dataset



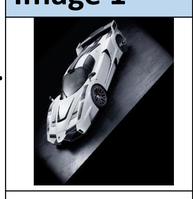
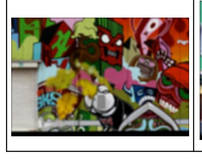
Results

Is it the same scene? **True = white | False = black**.
Ground truth -all possible pairs of the OxFrei dataset.

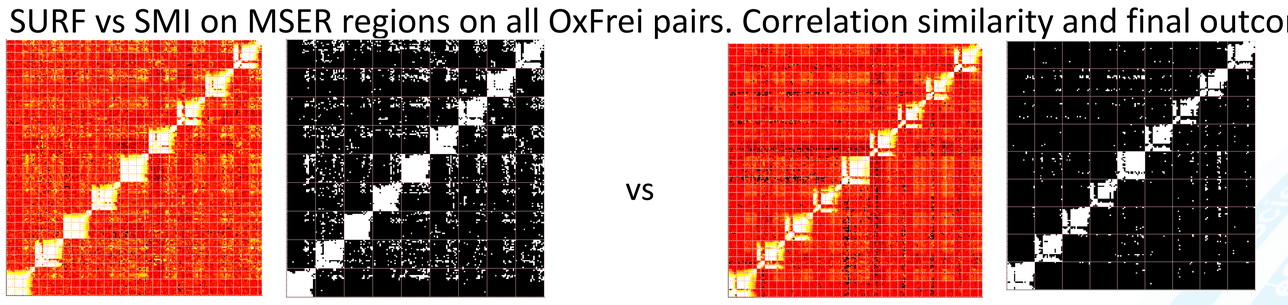
189 x 189 pixels
Pixel : image pair
Block : sequence



SMI vs SURF descriptor on MSER regions.

Image 1	Image 2	SURF	SMI
		False Negative Score 0.096	True Positive Score 0,89
		False Positive Score 0.27	True Negative Score - 0,11

SURF vs SMI on MSER regions on all OxFrei pairs. Correlation similarity and final outcome.



Performance of detector – descriptor combinations on both datasets.

Dataset	Oxford			OxFrei		
Detector – descriptor	Accuracy	Precision	Recall	Accuracy	Precision	Recall
MSER – SURF	0.97	0.97	0.89	0.90	0.53	0.83
MSER – SMI	0.96	0.98	0.85	0.95	0.83	0.74
BIN (All) – SURF	0.95	0.95	0.85	0.85	0.41	0.63
BIN (All) – SMI	0.89	1	0.58	0.91	0.73	0.32
BIN (Largest) – SMI	0.93	0.93	0.77	0.85	0.38	0.52

Conclusions

SMI descriptor is best to use when the number of False Positives should be minimal.
Detector – descriptor MSER – SMI has the best overall performance for both datasets.

www.eScienceCenter.nl

github.com/NLeSC/SalientDescriptor-matlab

E.Ranguelova@esciencecenter.nl