Shape and Moment Invariants Local Descriptor for Structured Images

Anonymous Submission

Anonymous Affiliation

Abstract

Keywords: image matching, shape descriptors, moment invariants,

1 Introduction

A pair of images $< I1, I2 >$, related with an photometric P and affine geometric transformation G: $I2 = I1.G.P$.

2 Related Work

Salient region detectors Maximally Stable Extremal Regions (MSER), [Matas et al., 2002]. The Data-driven Morphology Salient Regions Detector (DMSR) have been demonstrated to outperform MSER in ..., [Ranguelova, 2016]. Here, we propose to use a Binary detector (BIN) using the data-driven binarization explained in [Ranguelova, 2016], with either all or only regions with large area ($A_{reg.} \geq f_A.A_{Im.}$) are used.

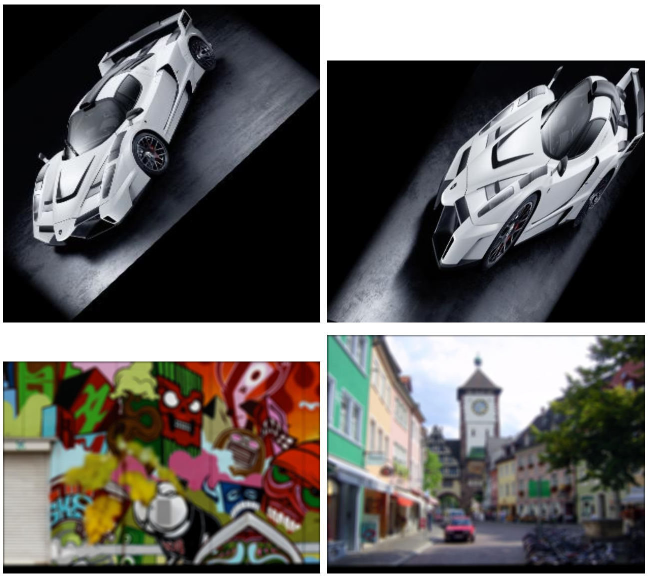

Figure 1: “Is it the same object or scene?” Matching two images under different transformation using local interest regions detected by MSER.
Top image pair (scale and viewpoint): SURF descriptor yields false negative (similarity score 0.096), while the proposed SMI descriptor - true positive (0.89).
Bottom image pair (blur): SURF gives false positive (0.27), while SMI - true negative (-0.11).

Region descriptors *State of the art in region descriptor.* Moment invariants are a group of efficient invariant object descriptors. Flusser et al. introduced a general framework for the derivation of Affine Moment Invariants (AMIs) using graph representation [Flusser and Suk, 1993; Suk and Flusser, 2004; Flusser et al., 2009].

Detector-descriptor combination

3 Image matching with Shape and Moment Invariant descriptor

We propose a set of several Shape and Moment Invariants (SMI) to encode each salient region into a feature vector (descriptor) used for the region matching. The SMI descriptor contains two parts: *simple shape invariants* and *moment invariants* $SMI_i = \{S_i, M_i\}$.

Simple shape invariants A binary shape of a region R_i can be described by a set of simple properties defined over the original shape or the equivalent up to the second order moments ellipse E_i. These properties are: the region's area a_i, the area of the region's convex hull a_i^c, the length of the major and minor axes of E_i, μ_i and ν_i and the distance between the foci of the ellipse ϕ_i. From these basic properties, a set of shape affine invariants are defined in Table 1.

Invariant	Definition	Description
Relative Area	$\tilde{a}_i = a_i/A$	region's area normalized by the image area A
Ratio Axes Lengths	$r_i = \nu_i/\mu_i$	ratio between E_i 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.

Table 1: Simple shape invariants.

The simple shape invariants part of SMI_i is then $S_i = \{\tilde{a}_i, r_i, e_i, s_i\}$.

Affine Moment Invariants If $f(x,y)$ is a real-valued image with N points, the AMI functional is defined by

$$I(f) = \int_{-\infty}^{\infty} \prod_{k,j=1}^{N} C_{kj}^{n_{kj}} . \prod_{l=1}^{N} f(x_l, y_l) dx_l dy_l, \tag{1}$$

where n_{kj} are non-negative integers, $C_{kj} = x_k y_j - x_j y_k$ is the cross-product (graph edge) of points (nodes) (x_k, y_k) and (x_j, y_j), [Suk and Flusser, 2004]. For full details of the AMI's theory the reader is referred to [Flusser et al., 2009]. We use the set of 16 irreducible AMIs of $N = 4$th order as implemented by the authors in an open source MATLAB software [Suk, 2004]. The AMI part of SMI_i is $M_i = \{m_{i1}, \ldots, m_{i16}\}$.

Hence, the final descriptor for the i-th region is a 20 element feature vector $SMI_i = \{\tilde{a}_i, r_i, e_i, s_i, m_{i1}, \ldots, m_{i16}\}$.

Matching Lets $\{SMI_i^1\}, i = 1, \ldots, n_1$ and $\{SMI_j^2\}, j = 1, \ldots, n_2$ be the $n_1 \times 20$ and $n_2 \times 20$ matrices with rows the SMI descriptors for the n_1 and n_2 regions detected via MSER or BIN (all/largest) detector in the pair of images to compare. We compare exhaustively every pair of local SMI descriptors SMI_i^1 and SMI_j^2 with Sum of square differences metric. The matching threshold for selection of the strongest matches is mt, the max ratio threshold for rejecting ambiguous matches is mr, the confidence of a match is mc and only unique matches are considered. After matching of all descriptor pairs, we select the top quality matches above a matching cost threshold ct. From those, we estimate in ititerations the affine transformation $\tilde{T}$ between the two sets of points being the centroids of the two matching regions sets as average of nr runs with allowed max point distance md. The two images are then transformed $J2 = I1.\tilde{T}$, $J1 = I2.\tilde{T}^{-1}$ and a correlation ($cor[X,Y] = cov[X,Y]/\sqrt{var[X]var[Y]}$) between the original and transformed images is used for confirmation of a true match. If the average correlation similarity between both images and their transformed versions $(cor[I1,J1] + cor[I2,J2])/2$ is above a similarity threshold st, we declare the original image pair $< I1, I2 >$ to be depicting (partially) the same scene.

4 Performance Evaluation

VGG dataset, [Mikolajczyk et al., 2005]. OxFrei dataset, [Ranguelova, 2016]. Used parameters: $mt = mr = 1$, $f_A = 2e-3$ (for BIN largest), $it = 1000$, $nr = 10$, $mc = 95$, $md = 8px$, $ct = 0.025$, $st = 0.25$.

4.1 VGG dataset

The performance results on the VGG dataset are summarized in Table 2.

4.2 OxFrei dataset

The performance results on the OxFrei dataset are summarized in Table 3.

Det. + descr.	TP	TN	FP	FN	Acc.	Prec.	Recall
MSER + SURF	128	428	4	16	0.965	0.969	0.889
MSER + **SMI**	122	430	2	22	0.958	0.98	0.847
BIN + SURF	122	426	6	22	0.951	0.953	0.847
BIN (All) + **SMI**	84	432	0	60	0.89	1	0.58
BIN (Largest) + **SMI**	112	424	8	32	0.93	0.93	0.77

Table 2: Performance of salient region detectors and descriptors on the VGG dataset.

Det. + descr.	TP	TN	FP	FN	Acc.	Prec.	Recall
MSER + SURF	3309	28848	2904	660	0.90	0.53	0.83
MSER + **SMI**	2957	31162	590	1012	0.95	0.83	0.74
BIN + SURF	2513	28198	3554	1456	0.85	0.41	0.63
BIN (All) + **SMI**	1275	31298	454	2694	0.91	0.73	0.32
BIN (Largest) + **SMI**	2079	28474	3278	1890	0.85	0.38	0.52

Table 3: Performance of salient region detectors and descriptors on the OxFrei dataset.

5 Conclusion

References

[Flusser and Suk, 1993] Flusser, J. and Suk, T. (1993). Pattern recognition by affine moment invariants. *Pattern Recognition*, 26(1):167 – 174.

[Flusser et al., 2009] Flusser, J., Suk, T., and Zitova, B. (2009). *Moments and Moment Invariants in Pattern Recognition*. Wiley.

[Matas et al., 2002] Matas, J., Chum, O., Urban, M., and Pajdla, T. (2002). Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In *Proceedings BMVC*, pages 36.1–36.10.

[Mikolajczyk et al., 2005] Mikolajczyk, K. et al. (2005). A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72.

[Ranguelova, 2016] Ranguelova, E. (2016). A Salient Region Detector for Structured Images. In *Proceedings of IEEE/ACS 13th Int. Conf. of Computer Systems and Applicaitons (AICCSA)*, pages 1–8.

[Suk, 2004] Suk (2004). Matlab codes: Affine moment invariants. http://zoi_zmije.utia.cas.cz/mi/codes.

[Suk and Flusser, 2004] Suk, T. and Flusser, J. (2004). Graph method for generating affine moment invariants. In *17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, August 23-26, 2004.*, pages 192–195.