A Training-free Classification Framework for Textures, Writers, and Materials

Radu Timofte[1] and Luc Van Gool[1,2]
http://homes.esat.kuleuven.be/~rtimofte

[1]ESAT-VISICS /IBBT, Catholic University of Leuven, Belgium
[2]D-ITET, ETH Zurich, Switzerland

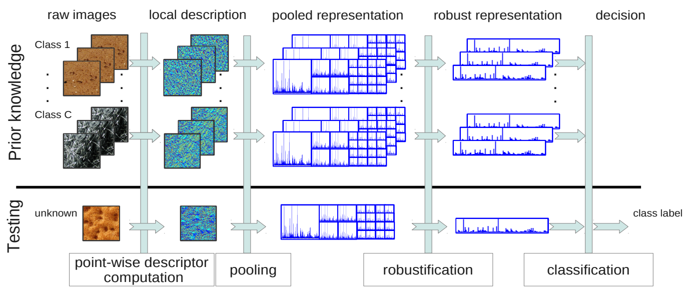

Figure 1: The scheme of our texture classification framework.

We propose a training-free texture classification scheme, outperforming methods that use training. This we demonstrate not only for traditional texture benchmarks, but also for the identification of materials and writers of musical scores. State-of-the-art methods operate using local descriptors, their intermediate representation over *trained* dictionaries, and classifiers. For the first two steps, we work with pooled local Gaussian derivative filters and a small dictionary *not* obtained through training, resp. Moreover, we build a multi-level representation similar to a spatial pyramid which captures region-level information. An extra step robustifies the final representation by means of comparative reasoning. As to the classification step, we achieve robust results using nearest neighbor classification, and s-o-a results with a collaborative strategy. Also these classifiers need no training.

Standard texture classification systems aim at i) constructing a rich representation of the image and ii) providing a classification strategy. The representation typically entails local (texture) descriptors, similarity measures, aggregating strategies, and intermediate and global (image level) descriptors. The classification strategy usually adapts its metric to the representation and aims at fixing its flaws. Class models are then built using state-of-the-art classifiers. A literature review is in the paper.

We **propose** a training-free multi-level texture classification framework (Fig. 1). It combines the robustness and simplicity of local descriptors such as BIFs [1], spatial information embedding into the global image representation in a layered fashion similar to SPM or through regions as in [3], the power of comparative reasoning [8], and s-o-a training-free classifiers [6]. Fortes of the framework are:

1) ***No need for training, and thus data independence.*** There is no need for learning a dictionary for the local descriptors (such as BIFs [1]). The system performs robustly with a fixed set of parameters on different texture, material and handwritten score datasets.

2) ***Robustness to intra-class variations.*** Robustness is provided by the local descriptors, the layered robustified representation, and the classifiers.

3) ***Layered representation embedding spatial information.*** Spatial information proved critical for object classification, and so it is for our tasks.

4) ***Robustified representations by means of comparative reasoning.*** The power of comparative reasoning (WTA-hash [8]) enhances and robustifies the representations by adding resilience to numeric perturbations.

5) ***Fast sparse and/or collaborative classification.*** Lately, sparse and collaborative representation based classifiers performed best at various tasks such as face recognition or traffic sign recognition [6].

Local Texture Descriptor (BIF) Basic Image Features (BIF) [1, 5] are defined by a partition of the filter-response space (jet space) of a set of 6 Gaussian derivative 2D filters up to 2nd order at some scale σ. The Jet space is further partitioned into 7 regions, or BIFs, corresponding to distinct types of local image symmetry. BIFs are rotation invariant. However, we can discretize orientations for BIF codes as in [5], thus obtaining Oriented Basic Image Features (oBIF). To create a more discriminative descriptor, [1] combines the descriptors at different scales on a pixelwise basis and ignores flat regions. BIF with p scales will generate 6^p distinct dictionary entries (for $p = 4$, 1296), while oBIF will generate 22^p distinct dictionary entries (for $p = 2$, 484).

Multi-Level Pooled Representation (SPM, BoR) The spatial pyramid matching (SPM) scheme pools regions at 3 or 4 pyramid levels. We continue as long as the cell/region size allows for meaningful histograms.

Table 1: Summary of texture, material, and score datasets.

Dataset	Dataset Notation	Dataset Type	Image Rotation	Controlled Illumination	Scale Variation	Significant Viewpoint	Number Classes	Sample Size	Samples per Class	Samples in Total
CUReT	D^C	texture	✓	✓			61	200×200	92	5612
UIUC	D^{UIUC}	texture	✓		✓	✓	25	640×480	40	1000
UMD	D^{UMD}	texture	✓		✓	✓	25	640×480	40	1000
Brodatz	D^B	texture					111	213×213	9	999
KTHTIPS	D^{KT}	texture		✓	✓		10	200×200	81	810
KTHTIPS2b	D^{KT2b}	texture		✓	✓	✓	11	200×200	4(×9×12)	4752
FMD	D^{FMD}	material	✓		✓	✓	10	512×384	100	1000
CVCMUSCIMA	D^{CM}	handwritten scores					50	~2000×2000	20	1000

Table 2: Comparison of our results [%] with those achieved by state-of-the-art methods. In the brackets is the number of training samples.

	$D^C(46)$	$D^B(3)$	$D^{KT}(41)$	$D^{UIUC}(20)$	$D^{UMD}(20)$	$D^{KT2b}(1)$	$D^{CM}(10)$	$D^{FMD}(50)$
1. Our Results	**99.42**	**97.26**	**99.35**	**99.01**	**99.54**	**66.26**	**99.80**	**55.78**
3. VZ-Joint [7]	98.03	92.90(*)	92.40(*)	97.83		53.30(**)		
5. Lezebnik *et al.*	72.50(*)	88.15	91.30(*)	96.03				
7. J.Zhang *et al.*	95.30	95.90	96.10	98.70				
9. Crosier and Griffin [1]	98.60		98.50	98.80				
12. Xu *et al.*-WMFS				98.60	98.68			
14. L.Liu *et al.*-SRP [4]	99.37	97.16	99.29	98.56	99.30			48.2
15. L.Liu *et al.*-ELBP	97.29					58.10		
16. Kong and Wang [3]		96.61	99.32		99.32			
17. PRIP02 [2]							77.00	
23. Hu *et al.*								54.00

Another proposed approach [3] uses multi-levels, similar to SPM, for creating orderless region parts, allowing for overlap. The images are represented by sets of regions, called Bag-of-Regions (BoRs). BoRs cover a much larger variance in scale, translation, rotation, viewpoint, illumination by enlarging the training pool. For the test image represented as BoR, the classification score is computed for each class and region. At image level (or BoR level) the label is taken as the class with the best cumulative score over the BoR.

Robustified Representation - (WTA-hash) The power of comparative reasoning was exploited and a Winner Take All (WTA) hash technique proposed in [8]. WTA-hash transforms the input feature space into binary codes and in the resulting space the Hamming distance closely correlates with the rank similarity measures. The rank correlation measures are resilient to perturbations in numeric values and WTA-hash brings perturbation robustness to the original feature space representation.

Sparse and Collaborative Classification - (SRC, CRC, INNC) For classification we use robust classifiers in the sense of not requiring parameter tuning for different datasets: Nearest Neighbor Classifier (NNC), Sparse Representation Classifier (SRC), Collaborative Representation Classifier (CRC), and Iterative Nearest Neighbors Classifier (INNC) [6].

We have shown that training-free pipelines can outperform several s-o-a texture classification methods. We are conservative in our **experiments**, in that further fine-tuning would be possible, i.e. we only went up to the point where the methods would outperform or get on par with the s-o-a, training-based methods. We only report results using the basic image features (BIFs) and its variants as local descriptors [1, 5], SPM with one level and a simple Bag-of-Regions model [3] as intermediate representations, and classifiers such as NNC, SRC, CRC, or INNC [6].

We were somewhat surprised by the strong performance of these methods, regardless of the dataset and/or task. We believe that adding training at any level in our framework can improve the performance further.

The proposed approach is computationally simple. To a large extent, it also is training-free and data-independent. The system is validated for texture, material, and writer classification on several benchmarks. We obtain results that are at least on-par, but sometimes substantially better than state-of-the-art performance (Tables 1,2).

Details about the framework and the benchmarks are in the paper.

[1] M. Crosier and L.D. Griffin. Using basic image features for texture classification. *IJCV*, 88(3):447–460, 2010.

[2] A. Fornés, A. Dutta, A. Gordo, and J. Lladós. The ICDAR 2011 music scores competition: Staff removal and writer identification. In *ICDAR*, 2011.

[3] S. Kong and D. Wang. Multi-level feature descriptor for robust texture classification via locality-constrained collaborative strategy. *CoRR*, 2012.

[4] L. Liu, P.W. Fieguth, D.A. Clausi, and G. Kuang. Sorted random projections for robust rotation-invariant texture classification. *Pattern Recognition*, 2012.

[5] A.J. Newell and L.D. Griffin. Multiscale histogram of oriented gradient descriptors for robust character recognition. In *ICDAR*, 2011.

[6] R. Timofte and L. Van Gool. Iterative nearest neighbors for classification and dimensionality reduction. In *CVPR*, 2012.

[7] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *PAMI*, 31(11):2032–2047, 2009.

[8] J. Yagnik, D. Strelow, D.A. Ross, and R.-S. Lin. The power of comparative reasoning. In *ICCV*, 2011.