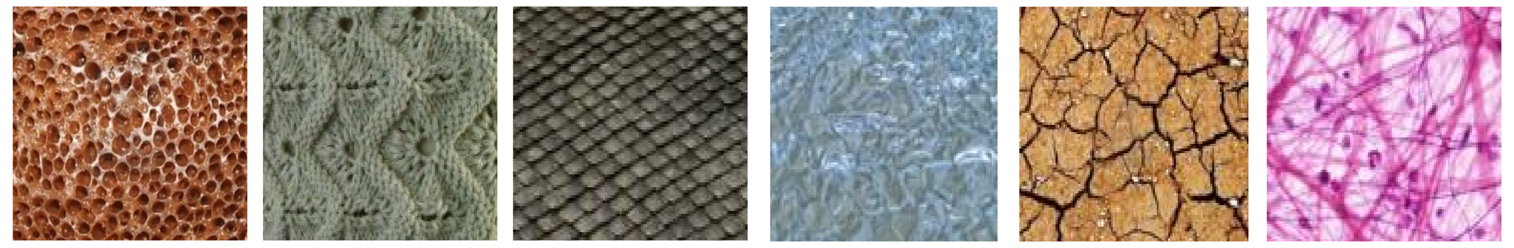
Describing Textures in the Wild

Mircea Cimpoi[1], Subhransu Maji[2], Iasonas Kokkinos[3], Sammy Mohamed[4], Andrea Vedaldi[1]

[1]Visual Geometry Group, University of Oxford {mircea, vedaldi}@robots.ox.ac.uk [2]Toyota Technological Institute smaji@ttic.edu
[3]Ecole Centrale Paris/INRIA-Saclay iasonas.kokkinos@ecp.fr [4]Stony Brook sammy.mohamed@stonybrook.edu

Describing Textures

http://goo.gl/w0E8P5



porous, dotted, freckled, honeycombed | braided, interlaced, knitted, woven, zigzagged | scaly, crosshatched, flecked studded, waffled | wrinkled, crystalline, flecked, smeared | cracked, pitted, studded | fibrous, freckled, interlaced, smeared, swirly

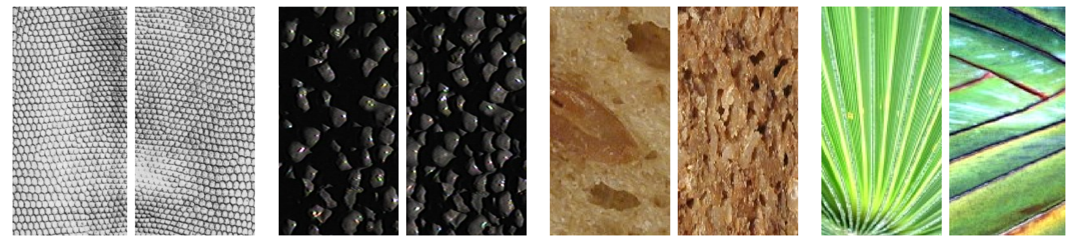
- **Goal:** automatically ***describe*** textures by using English words (e.g. interlaced, lace-like, fibrous, …)
- **Challenges:** defining, learning, and detecting multiple subjective attributes per texture
- **Applications:** human-centric texture description

Contributions

- **Describable Textures Dataset** (DTD)
- Low dimensionality texture representation,
- Above 10% accuracy improvement over existing state-of-the-art on FMD and KTH-TIPS2-b
- Coarse-to-fine strategy to cheaply label joint attributes
- Evaluation of texture representations methods on DTD

Describable Textures Dataset (DTD)

- 5640 images, 47 attributes, 120 images per attribute
- Collected in the wild (Internet: Google, Flickr)



This is **not the same** as recognizing material / instance!

Data Collection

- Texture vocabulary:
  - Starting point: list of 98 words in [Bhushan 97]
  - Discarded non-visual words (e.g. "jumbled" or "rhythmic")
  - Merged similar words (e.g. "corkscrewed" + "coiled" + "spiraled")
- Example images:
  - Consider each word as *key attribute*
  - Query Google (e.g. "corkscrewed textures", "coiled pattern")
  - Discard or crop images covered by less than 90% with content representing the query

Coarse-to-Fine Joint Annotation

Annotations using *Amazon MTurk*

Stage 1

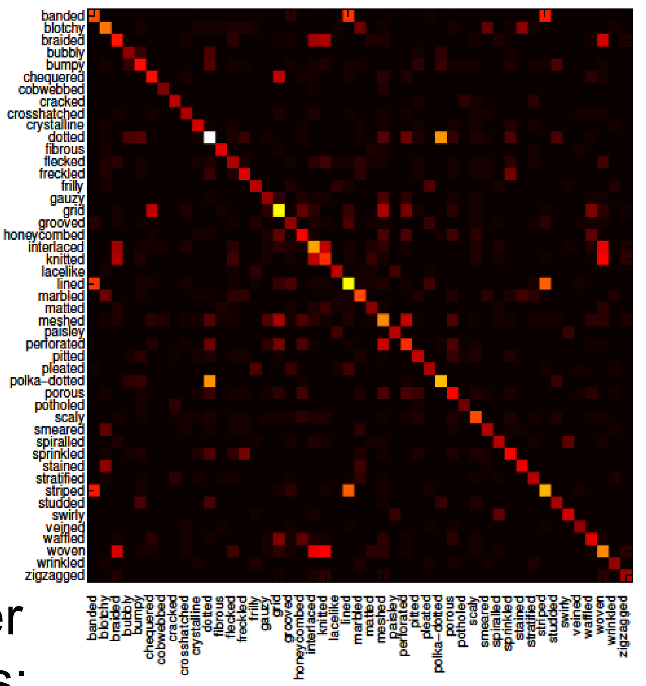
Verify key attributes.

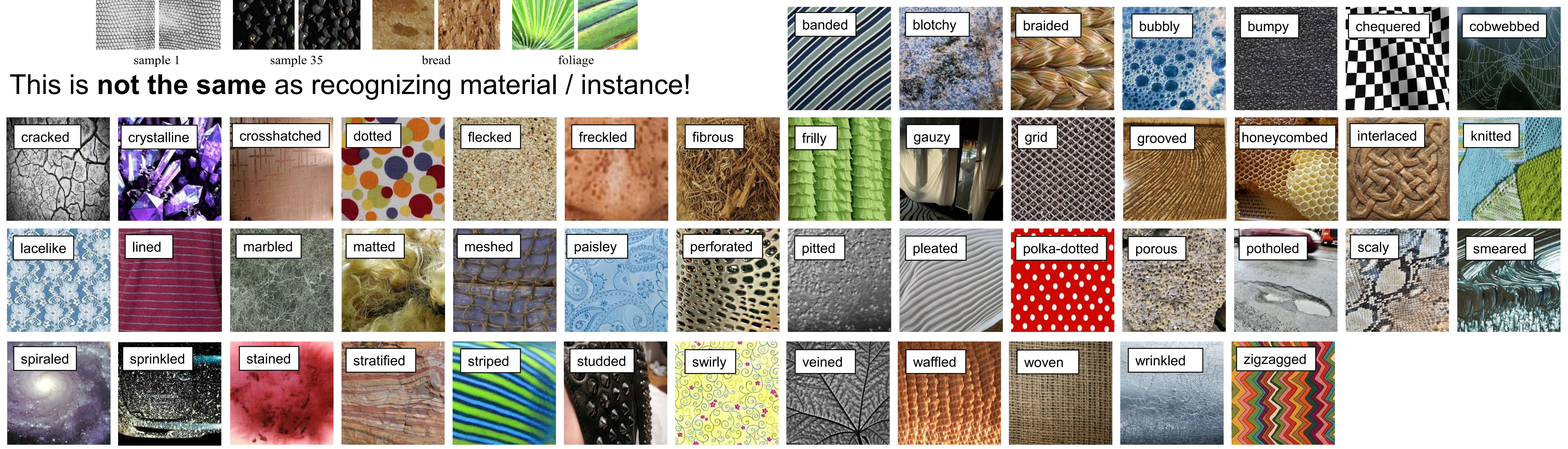
Stage 2

- Sequentially collect joint annotations based on co-occurrence probability;
- Avoid labelling low probability attributes, given key attribute;
- Using classifier scores to further

Reduce the number of annotations;

- Seek for consensus of multiple annotations (5 per image).





Local Descriptor Comparison on DTD

- Bag of Visual Words approach
- 470 dimensional vocabularies, built using K-means
- 10 visual words per texture
- Filter banks, SIFT, LBP and image patches as local descriptors
- SVM with several kernels: linear, Hellinger, χ^2 and exponential χ^2

	Kernel			
Feature	Linear	Hellinger	add-χ^2	exp-χ^2
MR8	25.8 (0.1)	28.9 (0.5)	31.9 (0.5)	35.6 (0.4)
LM	18.1 (0.9)	24.1 (0.1)	29.2 (0.5)	33.4 (0.5)
Patch 3x3	13.9 (1.0)	20.1 (0.5)	23.1 (0.3)	26.5 (0.4)
Patch 7x7	16.7 (0.8)	24.1 (0.3)	28.6 (0.5)	32.3 (0.5)
LBP^{u2}	8.9 (0.7)	9.7 (0.7)	12.4 (0.5)	19.7 (0.5)
LBP-VQ	19.6 (1.0)	22.7 (0.4)	26.5 (0.4)	31.2 (0.2)
SIFT	**31.2 (0.3)**	**38.6 (1.0)**	**41.4 (1.5)**	**44.1 (1.7)**
root-SIFT	30.6 (0.6)	38.1 (1.2)	40.6 (1.4)	43.3 (1.9)

State of the Art on Texture Datasets

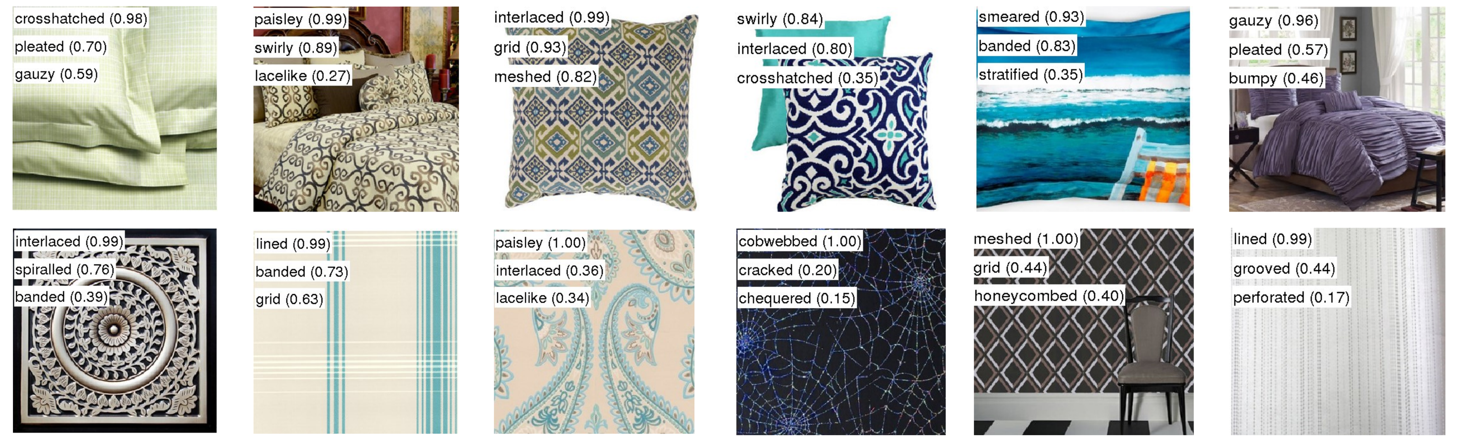
- Experiment with various encodings on top of best performing local descriptor (SIFT)
- Improved Fisher Vector (IFV) and Deep Convolutional Activation Feature (DeCAF) are tuned for object recognition, but perform very well on textures
- Combined, lead to state-of-the-art results on all datasets

Dataset		SIFT		DeCAF	IFV + DeCAF	Previous best
	IFV	BOVW	VLAD			
CUReT	99.6±0.4	98.1±0.9	99.1±0.6	98.9±0.4	***99.8±0.2***	99.4
UMD	99.2±0.4	98.1±0.8	99.4±0.4	97.4±0.7	***99.5±0.3***	99.7±0.3
UIUC	97.2±0.8	94.4±1.3	97.3±0.9	95.5±0.9	***99.0±0.5***	99.4±0.4
KTH-TIPS	99.7±0.4	98.6±1.0	99.2±0.8	98.4±0.8	***99.8±0.2***	99.4±0.4
KTH-TIPS2*a*	82.5±5.3	74.8±5.4	77.6±4.3	77.7±2.0	***84.3±1.8***	73.0±4.7
KTH-TIPS2*b*	69.3±0.9	58.4±2.2	61.7±2.2	70.4±1.8	***76.0±2.9***	66.3
FMD	58.1±1.7	49.5±1.9	54.8±1.8	57.6±1.2	***65.6±1.4***	57.1
DTD	58.6±2.0	53.6±1.5	57.3±1.5	52.5±1.3	***64.7±1.6***	--
DTD(AP)	60.3±2.8	52.2±2.2	58.5±2.4	51.3±1.6	***66.7±2.3***	--
DTD-J(AP)	60.6±2.4	53.6±1.9	58.9±1.9	51.7±1.5	***66.5±1.9***	--

Describable Attributes as Representation

- Use the scores from the 47 classifiers trained on DTD as a meaningful, low dimensionality descriptor.
- Low dimensionality allows to apply an RBF kernel
- DTD descriptor learned on IFV + DeCAF, alone, exceeds previous state-of-the-art on FMD and KTH-TIPS2-b
- Combined with IFV and DeCAF results in more than 10% above previous best.

Feature	KTH-TIPS2-b	FMD
$DTD(IFV)_{LIN}$	64.07 +/- 3.07	45.70 +/- 1.33
$DTD(IFV)_{RBF}$	67.68 +/- 2.18	50.94 +/- 1.46
$DTD(FVCAF)_{LIN}$	70.31 +/- 0.91	53.72 +/- 2.16
$DTD(FVCAF)_{RBF}$	***72.45 +/- 2.30***	***57.74 +/- 1.68***
$IFV+DTD_{RBF}$	76.17 +/- 1.21	65.12 +/- 1.86
$DeCAF+DTD_{RBF}$	74.92 +/- 1.18	64.86 +/- 2.24
IFV+DeCAF	76.10 +/- 3.14	65.90 +/- 1.50
$IFV+DeCAF+DTD_{RBF}$	**77.44 +/- 2.16**	**68.28 +/- 1.48**



Conclusions

- Introduced a large texture dataset, exhaustively labelled with joint subjective attributes
- Proposed a low dimensionality, meaningful, texture descriptor based on describable texture attributes
- Set new state-of-the art on challenging material datasets

References

[Bhushan 97] N. Bhushan, A. Rao, and G. Lohse. *The texture lexicon: Understanding the categorization of visual texture terms and their relationship to texture images*. Cognitive Science, 21(2):219–246, 1997

Acknowledgements

This research is based on work at the 2012 CLSP Summer Workshop. It was partially supported by NSF Grant #1005411, ODNI via the JHU HLTCOE and Google Research. Mircea Cimpoi was supported by ERC grant VisRec no. 228180 and Iasonas Kokkinos by ANR-10-JCJC-0205.