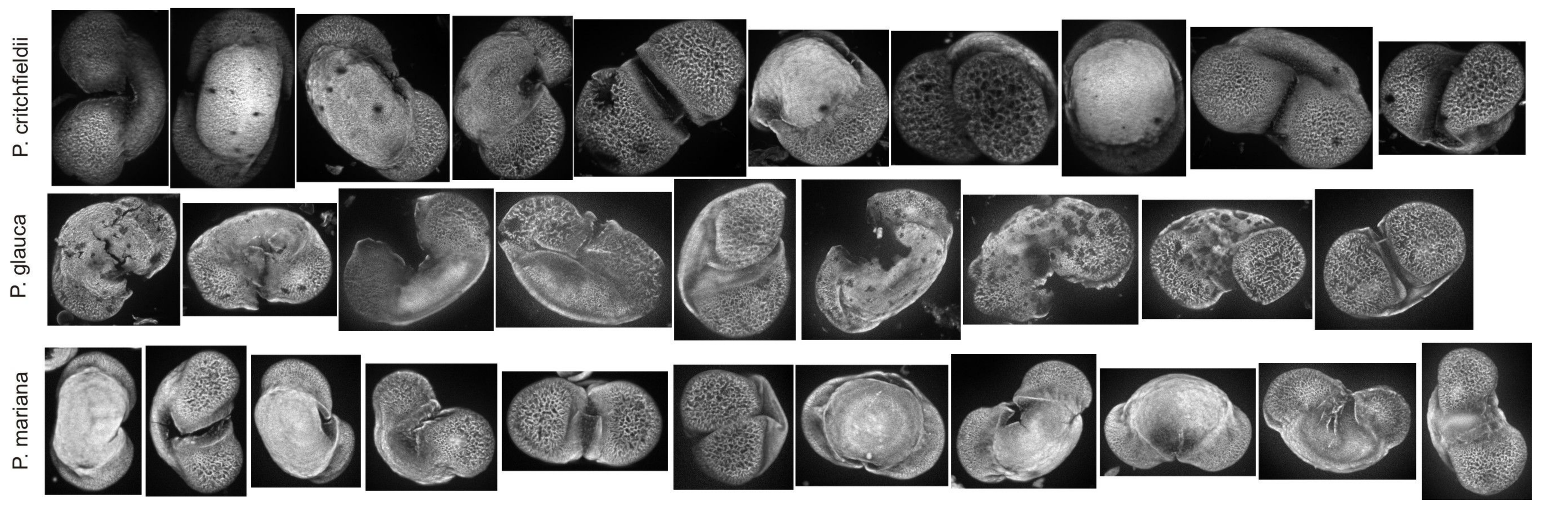
Spatially Aware Dictionary Learning and Coding for Fossil Pollen Identification

Shu Kong[1], Surangi Punyasena[2], Charless Fowlkes[1]
[1]Department of Computer Science, University of California Irvine, Irvine, California
[2]Department of Plant Biology, School of Integrative Biology, University of Illinois, Urbana, Illinois

Abstract

Pollen is one of the most ubiquitous of terrestrial fossils and has been used to test hypotheses across a broad cross-section of biological and geological sciences, from dating rock strata to understanding long-term effects of climate change on plant communities.

While high-throughput microscopic imaging allows for ready acquisition of large numbers of images of modern or fossilized pollen samples, identifying and counting by eye the number of grains of each species is painstaking work and requires substantial expertise and training.

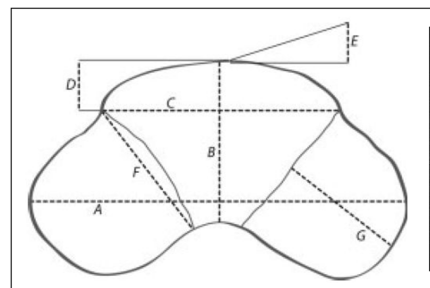


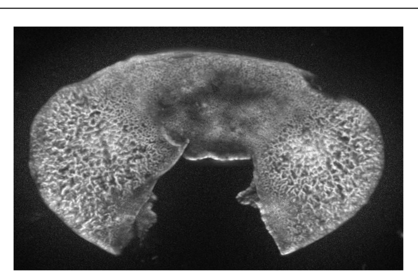
We propose a robust approach for species-level recognition of fossil pollen grains in microscopy images that exploits both global shape and local texture characteristics in a patch-based matching methodology. Specifically,

1. We introduce novel criteria for selecting informative and discriminative exemplar patches.
2. We use these selected exemplars as a dictionary basis and propose a spatially-aware sparse coding method to match testing images for identification while maintaining global shape correspondence.
3. We introduce a relaxed form that uses spatially-aware soft-thresholding during coding to accelerate the coding process for fast matching.

We demonstrate the effectiveness and efficiency of our exemplar selection and classification mechanisms, achieving 86.1% accuracy on a difficult fine-grained fossil spruce species classification task.

Alignment and Viewpoint Clustering

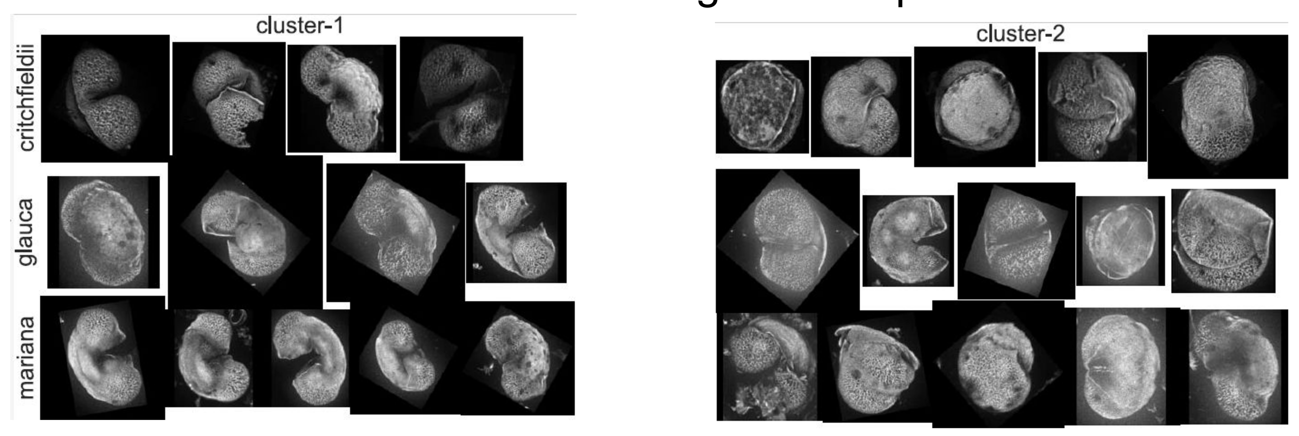




Traditional approaches to quantifying pollen shape rely on measures of overall morphology as well as textural features of the pollen grain surface [1].

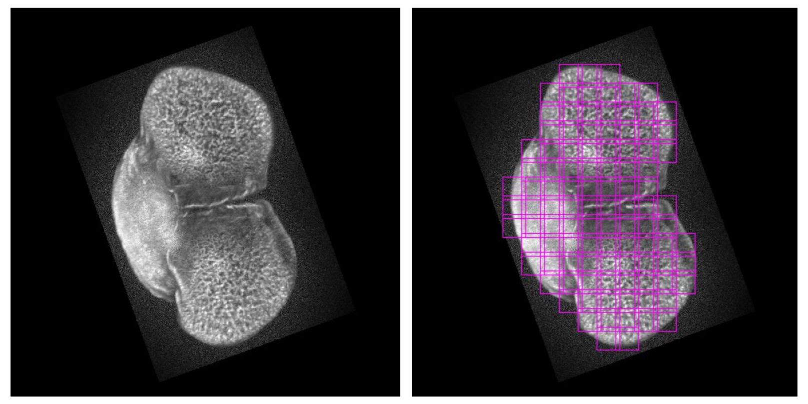
We cluster training images using k-medoids to find canonical viewpoint clusters and perform rigid 2D alignment within each cluster.

Patches in a test image are coded using patch features taken from similar spatial locations in the set of training images (after alignment) to capture both local detailed texture characteristics and global shape information.



Discriminative Patch Selection

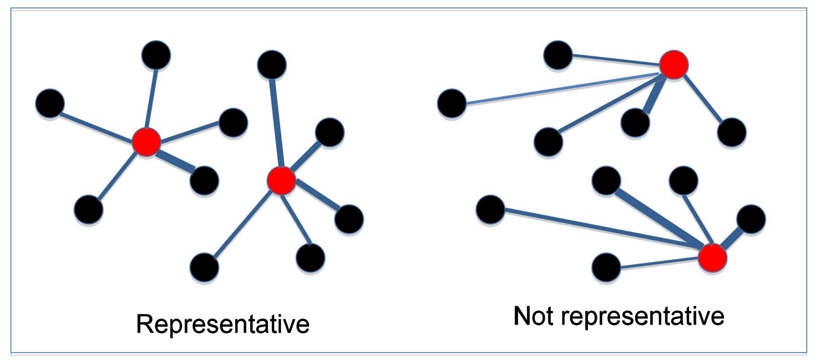
We formulate the problem of selecting a set of discriminative training patches as a submodular maximization problem based several criteria:



1.Spatially distributed in **input** space
2.Representative in **feature** space

Each training patch should be near and similar to some selected exemplar

$$\mathcal{F}_R(A) = \sum_{j \in \mathcal{V}} \max_{i \in A} \mathbf{S}_{ij}$$



3. Discriminative power

Clusters induced by exemplars should be pure with respect to class labels

$$\mathcal{F}_D(A) = \frac{1}{C} \sum_{i \in A} \max_c N_c^i - |A|$$

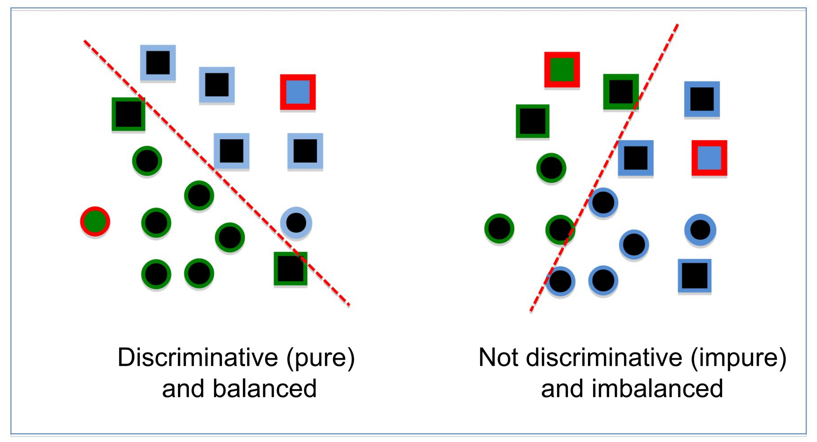
4.Class balance

Exemplars should be balanced across classes

$$\mathcal{F}_B(A) = \sum_c \log(|A_c| + 1)$$

5.Cluster compactness

Clusters should be of similar sizes

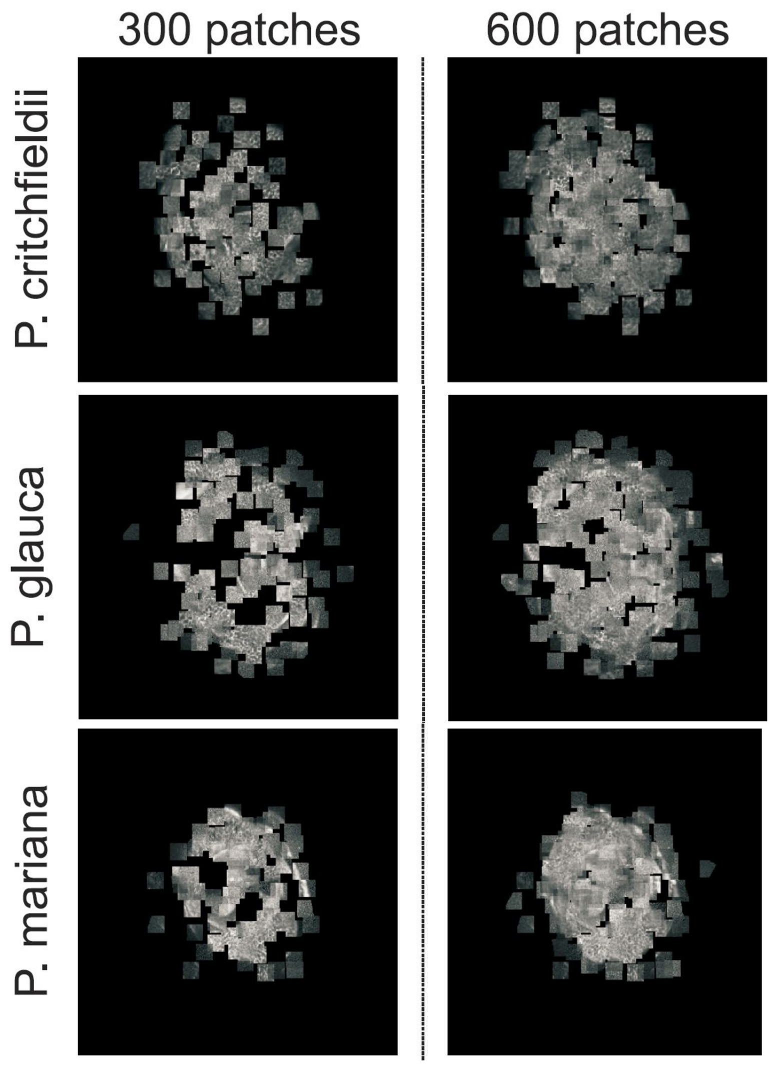


We maximize a weighted combination of these using a greedy selection algorithm with "lazy" cost updates

Input: $\mathcal{V}, \mathcal{F}, K$
Output: a subset A with $|A| \leq K$
initialize $A = \varnothing$, iteration $k = 0$
for all $i \in \mathcal{V}$, compute $\Delta(i) = \mathcal{F}(\{i\})$
while $k \leq K$ **do**
$i^* = \arg\max_{i \in \mathcal{V}/A} \Delta(i)$
compute $\Delta(i^*) = \mathcal{F}(A \cup \{i^*\}) - \mathcal{F}(A)$
if $\Delta(i^*) \geq \max_{i \in \mathcal{V}/A} \Delta(i)$ **then**
if $\Delta(i^*) < 0$ **then**
return A
else
$A = A \cup \{i^*\}, \quad k = k + 1$
end if
end if
end while

The selected exemplar patches are used as a dictionary for coding-based classification. Our discriminative selection criteria yield a substantial improvement in performance over random patch selection.

dictionary size	300	512	600
Random Selection	77.66	76.49	77.23
Discriminative Selection	81.75	81.60	82.34

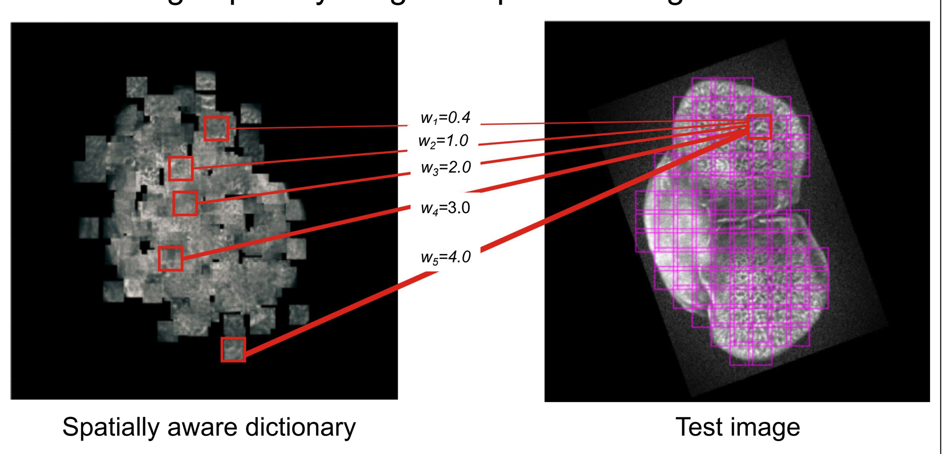


Visualizing the selected exemplars we can see that these patches not only capture local texture information, but also convey the global shape and average size of the three species.

Spatially Aware Coding for Matching

Given the dictionary of selected exemplar patches for each class, we encode patches extracted from a test image using a weighted variant of sparse coding where the weights depend on the spatial offset of patch relative to the location of the dictionary element in the training image.

Identifying a test image by matching exemplars through spatially weighted sparse coding



We investigated two variations of spatial weighting:

$$\mathbf{a}^* = \underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda_1 \|\operatorname{diag}(\mathbf{w})\mathbf{a}\|_1$$

$$\mathbf{a}^* = \underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda_2 \|\operatorname{diag}(\mathbf{w})\mathbf{a}\|_2^2 + \lambda_1 \|\mathbf{a}\|_1$$

Weighting **w** encourages use of exemplar patches that are nearby spatially.

To expedite coding, we propose a fast approximation that upper-bounds the exact reconstruction cost for undercomplete dictionaries and can be computed in one step using soft thresholding.

$$\|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 \Rightarrow \|\mathbf{\Omega}\mathbf{x} - \mathbf{a}\|_2^2$$

SACO-I

$$\mathbf{\Omega} \equiv (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T$$
$$\mathbf{u} = \mathbf{\Omega}\mathbf{x}$$
$$a_i^* = \operatorname{sgn}(u_i) \cdot \max(0, |u_i| - \lambda_1 w_i)$$
$$\mathbf{a}^* = [a_1^*, \dots, a_i^*, \dots, a_m^*]^T$$

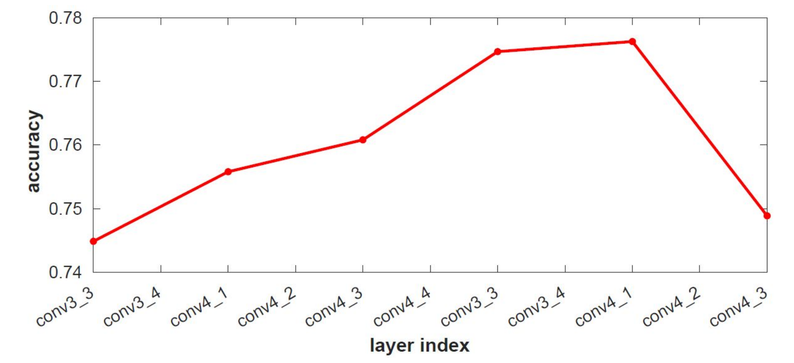
SACO-II

$$\mathbf{\Omega} \equiv (\mathbf{D}^T\mathbf{D} + \lambda_2 \operatorname{diag}(\mathbf{w})^2)^{-1}\mathbf{D}^T$$
$$\mathbf{u} = \mathbf{\Omega}\mathbf{x}$$
$$a_i^* = \operatorname{sgn}(u_i) \cdot \max(0, |u_i| - \lambda_1)$$
$$\mathbf{a}^* = [a_1^*, \dots, a_i^*, \dots, a_m^*]^T.$$

Experiments

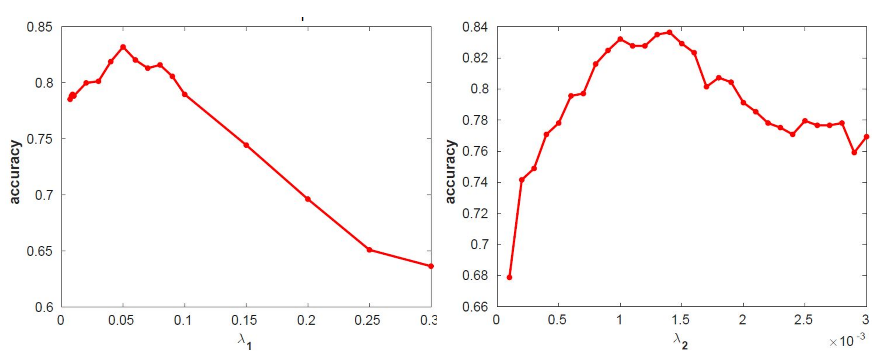
	#train	#test	#total
P. critchfieldii	65	43	108
P. glauca	65	355	420
P. mariana	65	287	352
Summary	195	685	880

Our dataset consists of pollen from three species of spruce imaged using a Zeiss Apotome fluorescence microscope to produce high-resolution, three-dimensional image stacks. We use maximum intensity projection over the top half of the grain image stack to produce a single in focus 2D image.



We use the VGG19 CNN model [2] to extract features representing each image patch. Performance varies depending on which layer of the feature hierarchy is used. Hand-designed feature (e.g., SIFT) perform significantly worse.

The spatially aware coding penalty improves classification performance and outperforms strong baseline methods including sparse coding (SRC) and standard CNN feature-based image classification.



SRC	VGG19+SVM	FV+SVM	SACO-I	SACO-II
62.04	65.11	61.46	83.21	86.13

References:
[1] SW Punyasena, DK Tcheng, C Wesseln, PG Mueller, "Classifying black and white spruce pollen using layered machine learning", New Phytologist, 2012.
[2] K Simonyan and A Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", ICLR, 2015.
[3] S. Kong, S. Punyasena, C. Fowlkes, "Spatially aware dictionary learning and coding for fossil pollen recognition", http://arxiv.org/abs/1605.00775

Acknowledgements: This work was supported by NSF grants DBI-1262547 and IIS-1253538.