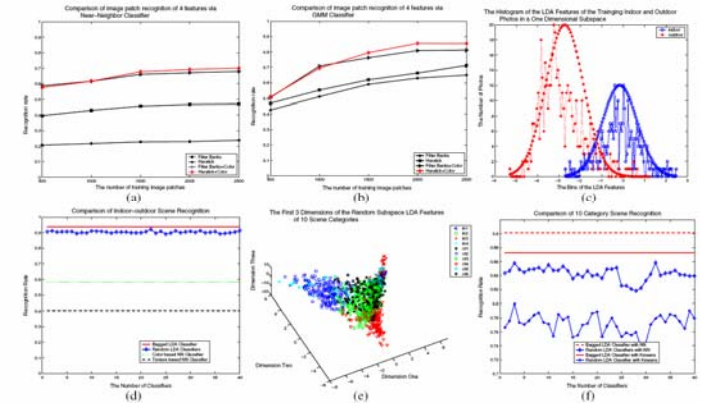
A Two Level Approach for Scene Recognition

Le Lu[1] Kentaro Toyama[2] Gregory D. Hager[1]
[1]Computer Science Department, Johns Hopkins University, Baltimore, MD 21218, USA
[2]Microsoft Research, Redmond, WA 98052, USA

Abstract

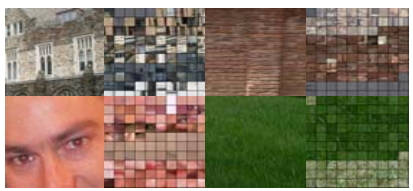
Classifying pictures into one of several semantic categories is a classical image understanding problem. In this paper, we present a stratified approach to both binary (outdoor-indoor) and multiple category of scene classification. We first learn mixture models for 20 basic classes of local image content based on color and texture information, then apply them to a test image, and produce 20 probability density response maps (PDRM) indicating the likelihood that each image region was produced by each class. We further extract some very simple features from those PDRMs, and use them to train a bagged LDA classifier for 10 scene categories. To test this classification system, we created a labeled database of 1500 photos taken under very different environment and lighting conditions, using different cameras, and from 43 persons over 5 years. The classification rate of outdoor-indoor classification is 93.8%, *and the classification rate for 10 scene categories is* 90.1%.

3, Scene Classification: Use bootstrapping and random subspace method to generate random LDA classifiers and bagging for classification

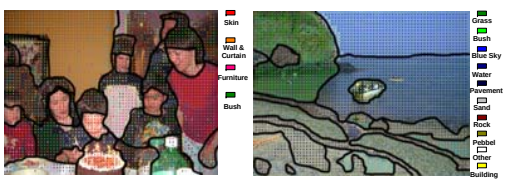


(a) Comparison of the image patch based recognition of 4 kinds of features (filter banks feature, Haralick texture feature and their joint features with color) via Nearest-Neighbor Classifier.
(b) Comparison of the image patch based recognition of 4 kinds of features via GMM Classifier.
(c) The 1D feature histogram distributions of indoor-outdoor photos after LDA projection.
(d) The comparison of indoor-outdoor recognition rates of 4 methods.
(e) The first 3D feature point distributions of 10 category photos after LDA projection.
(f) The comparison of 10 categories recognition rates of 4 methods.

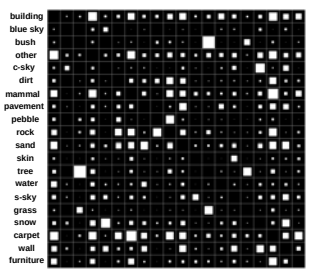
1, Local Image Processing: Learn a discriminative density model for patch level texture-color description of 20 materials



Examples of cropped subimages of building, building under direct sun, human skin, and grass respectively, and their randomly sampled image patches (25 by 25 pixels).



The local pixel's material labeling results of an indoor photo and an outdoor photo. The colored dots represent the material label and the boundaries are manually overlayed for illustration.

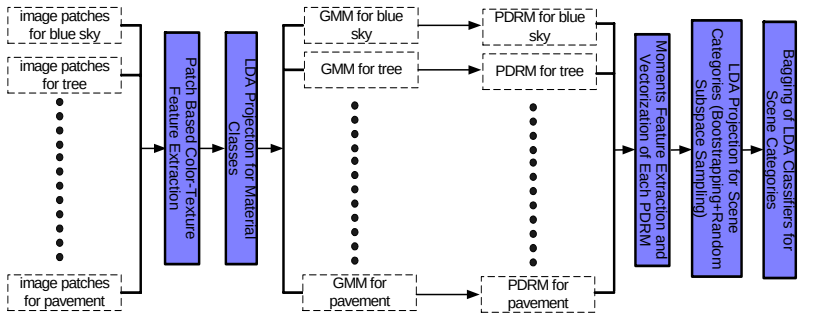


The pairwise confusion matrix of 20 material classes. The indexing order of the confusion matrix is shown on the left of the matrix. The indexing order is symmetrical.

Labeled Image Patches for each of 20 materials

Patch-based Discriminative Gaussian Mixture Density Model for each of 20 materials

Probability Density Response Maps for each of 20 materials



Patch-Level Processing

Image-Level Procesing + Classification

2, Global Image Processing: Compute 20 materials' support maps in an image level and their according moment features

(a) Photo 1459# (b) Confidence map (c) Blue sky (d) Cloud Sky (e) Water (f) Building (g) Skin

4, What can be wrong?

Example 1: man sitting in front of a curtain with artificial tree leaves

Example 2: playground incorrectly classified as carpet

Contributions:

1, *An* **efficient, yet effective, approach** *for scene recognition for* **both indoor-outdoor and multiple photo categories**, *based upon* **a challenging photo database**.
2, *A* **discriminative density model** *that achieves a good balance of discrimination and smoothness over large scale image patch data.*
3, *The use of* **moment features of PDRMs** *as* **an effective image-level representation** *for scene classification, and the* **bagging method** *to combine* **bootstrapping** *and* **random subspace** *generated LDA classifiers (useful especially for 10 category scene recognition).*

Discussion:

1, **Supervise Learning VS. Unsupervised Learning**: Although we have used supervised methods to create the local image patch classifiers, a practical system would like learn at least some of these classifiers using unsupervised methods. However we believe that the supervised material detectors provide the best scene recognition performance, and as such provide a "benchmark" against which unsupervised methods can be evaluated.
2, **Online Learning**: We also plan to investigate a hybrid approach where classified images are used as labeled data to compute an initial LDA projection, which is then subsequently refined with new, unlabeled images using iterative LDA-GMM.
3, **Non-parametric Discriminant Analysis** (proposed in [34]) will be tested as a means to generalize our LDA approach to a more comprehensive image database which may contain thousands of various kinds of photos.