Learning Dictionaries of Discriminative Image Patches

Anders Lindbjerg Dahl
abd@imm.dtu.dk, http://www.imm.dtu.dk/~abd

Rasmus Larsen
rl@imm.dtu.dk, http://www.imm.dtu.dk/~rl

DTU Informatics
Technical University of Denmark
Lyngby, Denmark

Remarkable results have been obtained using image models based on image patches, for example sparse generative models for image inpainting, noise reduction and super-resolution, sparse texture segmentation or texton models. In this paper we propose a powerful and yet simple approach for segmentation using dictionaries of image patches with associated label data. The approach is based on ideas from sparse generative image models and texton based texture modeling.

The proposed method is based on modeling an image using small image patches, which has similarities to sparse image coding [2, 6] and textons [3, 4, 7]. Sparse methods are generative and model an image by linearly combining a set of dictionary atoms. Some of the attractive properties of the sparse methods are their robustness to noise, closeness to data and simplicity in both implementation and interpretation. We aim at adopting these properties in our method, which generally target discriminative problems.

Our method is based on a dictionary of image patches denoted dictionary atoms. To each atom we associate a label atom that we use for building the segmentation image. Basically the segmentation is performed by finding the nearest dictionary atom to an image patch. The associated label atom is used for inferring the label probability to the image region covered by the image patch. Patches are overlapping so several label atoms are added to the same pixel.

In our approach we store the image intensity information and the label information in separate dictionaries. We denote the intensity dictionary $\mathbf{D} \in \mathbb{R}^{hn \times m}$, which is a matrix with m atoms – one vectorized image patch in each column. The size of the image patch is $\sqrt{n} \times \sqrt{n}$ and h is the color depth. $\mathbf{D}$ is used for modeling an image patch $\mathbf{x}$ by choosing the nearest atom in $\mathbf{d}_j \in \mathbf{D}$ such that $\mathbf{d}^* = \min_j \left|\left|\mathbf{d}_j - \mathbf{x}\right|\right|$, where $j = \{1,...,m\}$. We want to learn $\mathbf{D}$ from data to model the image textures well. This can be found by choosing $\mathbf{D}$ such that it minimizes the residual error $\mathbf{D} = \arg\min_{\mathbf{D}} \sum_{i=1}^{o} \left|\left|\mathbf{d}_i^* - \mathbf{x}_i\right|\right|_2$, where o is the number of training samples. This is a clustering problem, where a good iterative approximation based on vector quantization can be found with the k-means algorithm.

A dictionary with small residual errors on the image atoms does not necessarily have good discriminative properties. So, simultaneously to modeling intensity data well, we also want the dictionary atoms to be unique for a specific class and hereby have high discriminative power. We associate the intensity dictionary $\mathbf{D}$ with a label dictionary $\mathbf{L} \in [0;1]^{n \times m \times l}$, where l is the number of classes. The spatial extension of the label dictionary atoms is the same as the intensity dictionary, so a probability can be inferred on each pixel. Each atom $\mathbf{d}_j$ in $\mathbf{D}$ has an associated atom $\mathbf{l}_j$ in $\mathbf{L}$, and a pixel in a label atom contains the probability of all classes, so $\sum_{k=1}^{l} \mathbf{l}_k(g) = 1$ where $g \in \{1,...,n\}$ is the pixels in the label atom. Ideally a pixel in a label atom will have probability 1 for one class and 0 for all other classes, and we change an observed label atom to an ideal label atom $\hat{\mathbf{l}}$ by

$$\hat{\mathbf{l}}_k(g) = \begin{cases} 1 & \text{if } k = \max_k \mathbf{l}_k(g) \\ 0 & \text{otherwise} \end{cases} .$$

The discriminative power of the dictionary atoms can be found by the following minimization $\mathbf{L} = \arg\min_{\mathbf{L}} \sum_{i=1}^{o} \sum_{g=1}^{n} \left|\left|\hat{\mathbf{l}}_i - \mathbf{l}_i\right|\right|_1$, where each atom in $\mathbf{L}$ is optimized to contain label information for one class in each pixel. The coupling of the label and intensity dictionaries requires the optimization to be done for both dictionaries simultaneously, and we propose an iterative approach for approximating the solutions.

To build the label image we employ the label dictionary. The label image is a probability map for the class labels. Given an image $\mathbf{P} \in \mathbb{R}^{r \times c \times h}$ we construct a label image $\mathbf{P}_l \in [0;1]^{r \times c \times l}$ initially with all zeros. We build the label image by pixel wise coding $\mathbf{P}$ using the dictionary $\mathbf{D}$. This is done by visiting all pixels $\mathbf{P}(i,j)$ where we can extract a $\sqrt{n} \times \sqrt{n} \times h$ image patch. The nearest dictionary atom $\mathbf{d}_j$ is found and the corresponding label patch $\mathbf{l}_j$ is added to the $\sqrt{n} \times \sqrt{n} \times l$ image window centered at position (i,j). The final label image will in each pixel $\mathbf{P}_l(i,j)$ have the average probability of all labels that covered that pixel. The third dimension of $\mathbf{P}_l$ encodes probability for each label.

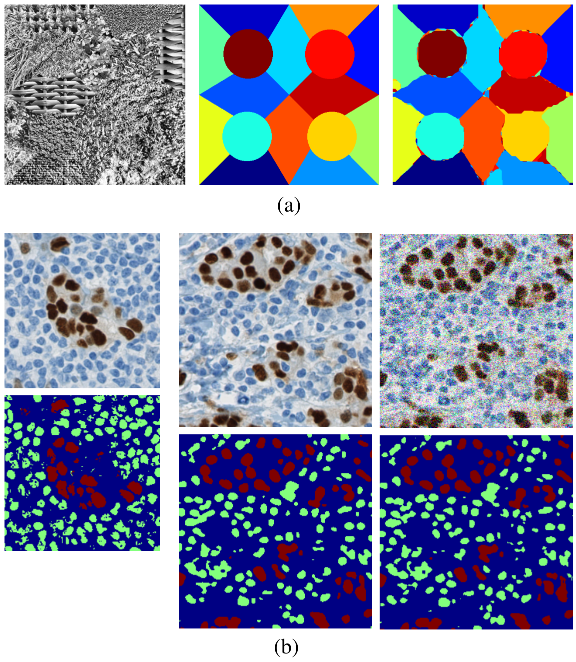

Figure 1: Example of multi class segmentation using patch based classification. (a) combination of sixteen Brodatz textures. (b) histopathological colored tissue sample of cell nuclei – *left* training sample, *center* original test image and *right* added 100 % Gaussian noise (relative to standard deviation).

We have experimentally validated our approach a number of segmentation experiments on composed and natural textures. Examples are shown in Figure 1. The experiments show the method's flexibility and robustness to noise, and we compare to similar approaches on a number of image compositions made from Brodatz textures [1], the VisTex database[1], and the MeasTex database[2]. On this dataset our procedure performs better than state of the art methods of [5, 8].

[1] P. Brodatz. *Textures: a photographic album for artists and designers*. Dover Publications, 1966.

[2] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[3] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 1981.

[4] T. Leung and J. Malik. Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons. *Int J of Comp Vis*, 43(1):29–44, 2001.

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *CVPR*. IEEE, 2008.

[6] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research*, 1997.

[7] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *CVPR*. IEEE, 2008.

[8] K. Skretting and J. H. Husøy. Texture classification using sparse frame-based representations. *EURASIP J on App Signal Processing*, 2006.

[1] http://vismod.media.mit.edu/vismod/imagery/VisionTexture/
[2] http://www.texturesynthesis.com/meastex/meastex.html