Using Statistical Topic Models to Organize and Visualize Large-Scale Architectural Image Databases

Patrick Paczkowski, Julie Dorsey, and Holly Rushmeier
Yale University

1 Introduction

Digital libraries have now become by far the preferred method for storing large image collections, making it increasingly important to improve issues of search and discovery within these databases. Content-based annotations of images, for classification purposes, have been a particularly significant area of research; however, even the best of such methods are prone to error – potentially associating incorrect text to image features and overlooking significant features. In the broader field of classification and search, Newman et al. demonstrated benefits of using statistical topic models to enrich document metadata. [Newman et al. 2007] This assumes that a document can be described with sufficient accuracy using a "bag of words" model, i.e. described by the words it contains (instances from a large vocabulary of words), and the frequency with which they occur. [Horster et al. 2007] applied this idea to large-scale image datasets, using SIFT features as data for creating a topic model.

Here, we extend this idea of effectively organizing and visualizing large-scale image collections based on their visual content. We focus on architecturally-themed images, due to the serviceability of such a tool in the field of architectural design. A key differentiator from previous work is that we do not attempt to assign semantic meaning to the images. We instead remain in the visual image space at each stage, relying on visual "words" and topics. Moreover, we attempt to use a more intuitive visual vocabulary that is naturally suited to a particular database. An overview of our method follows.

2 Technical Approach

First, we had to settle on a visual vocabulary for our dataset, which would be able to provide a sufficiently accurate description of images in the database, and consisting of words that can be accurately detected and extracted from the images. We made an important observation, namely that a natural way of visually deconstructing a building is by describing its exterior as a composition of architectural elements (e.g. windows, archways, etc). Many of these can be approximated by one or more basic geometric shapes (i.e. rectangles, triangles, ellipses); we used this fact, in conjunction with using a Canny edge detector and Hough transforms [Duda and Hart 1972], as the foundation for detecting architectural elements within the images in our dataset. We created a vector descriptor for each detected architectural element (image region) using positional and dimensional information, along with histograms of edge segments within the region. Initially, we clustered all the elements found in the dataset by noting the presence neighboring architectural elements within the original image, as the relative position of these elements is often significant. Subsequently, we applied hierarchical k-means clustering, as outlined by [Nister and Stewenius 2006], to obtain the final set of visual words for our database.

Based on these clustered features, we are able to now represent each image by the instances and frequencies of visual words, or architectural elements, appearing within it. This representation allows us to use Latent Dirichlet Allocation (LDA) [Blei et al. 2003] to describe each of our images as a mixture of topics, which in turn is used to meaningfully organize the dataset. We have tested our pipeline on 15,976 images acquired from the Visual Resources Collection of the Yale Library, and found roughly 35–40 of 100 generated topics to be meaningful. Each topic is visualized using around 16 images from the dataset for which that topic is most dominant (see Figure 1).

We administered a user-study to validate the coherence of these topics – in particular, relative to sets of random images. Details can be found in our poster draft under auxiliary materials.

References


Figure 1: Four topics obtained using our classification procedure, each represented by images for which that specific topic is most dominant.