Table 1: Databases of images from the sixteen-category dataset

Database	Image Categories
I	(Side-viewed) Elephant, Horse, Leopard, Rhino
II	Cat face, Dog face, Leopard face, Cow head
III	(Side-viewed) Bike, Motorbike, Cannon, Car

and beard. Our model can also capture different face orientations of the same person. For example, the first two images share the top two attributes but the first image does not have the third attribute of the second image, due to the difference in face orientations.

In general attributes are not spatially localized as we might have hoped, mainly because there are some limitations in our model. For instance, there is no prior bias that they should be spatially localized. Moreover, images are not precisely aligned and there are no latent variables in the model that are responsible for attribute positions, so attributes need to model global translations. Lastly, the basic IBP cannot model correlations in feature usage.

6 Evaluation

We evaluated our model on the following vision tasks and compared them to simple baselines.

6.1 Image Retrieval

We begin by evaluating our model on an image retrieval task: given a single image, a retrieval system must provide a ranking of images in the database. Since we do not have relevance information from the users, we consider two images as relevant if and only if they are in the same category. We built three databases of images from the sixteen-category dataset. Each database consists of 100 images from 4 similar categories, shown in Table 1. We seperated 20 images from each category and used them as queries to our retrieval system later on.

The score of an image in the database is the dot product of that image's feature vector and the test image's feature vector (i.e., the linear kernel). In our baseline method, we used a 1089-dimensional vector of HOG features to represent an image. In our attributed-based method, we represented each image with a vector of attribute weights. First, we ran the infinite sparse factor analysis model on images from each category so as to find their attributes. Given all those attributes, we used Gibbs sampling to derive an attribute weight vector for each image in the database. Figure 9 shows attributed-based representations of images in a database, where images from different categories seem to have different attribute assignments. To infer the attribute weight vector of an image query, we ran the Gibbs sampler again, given attribute definition matrix, HOG features of images in the database along with their attribute weight vectors.

We submitted 80 image queries (20 per category) for each database. We compared the two rankings returned by our baseline and attributed-based models. Table 2 shows the mean average precision of the rankings for each image category. We can see that our attributed-based features can significantly improve retrieval effectiveness in all cases. Overall, we see a 30.11% improvement in the ranking performance.

6.2 Transfer Learning

Transfer learning can be viewed as generalization of knowledge; the knowledge learned in one situation might be useful in another. In the domain of object detection, transfer learning is especially important because the object space is extremely large, and it is often the case that we are in a situation where we see things we have not seen before. In this subsection, we test the ability of our attributed-based models to perform transfer learning. In particular, given a large number of images in one category as well as background images, we would like to improve the detection rate of objects of similar type. In the transfer learning paradigm, we only have a few of examples of the target class.