Text-Based Image Retrieval Using Histograms

Greg Bickerman
Harvey Mudd College
340 E Foothill Blvd
Claremont, CA 91711
gbickerman@hmc.edu

Joshua Ehrlich
Harvey Mudd College
340 E Foothill Blvd
Claremont, CA 91711
jehrlich@hmc.edu

Ryan Weaver
Claremont McKenna College
500 E 9th St
Claremont, CA 91711
rweaver11@cmc.edu

ABSTRACT

Image Retrieval is a difficult, yet important subfield of Information Retrieval. Two of the largest difficulties are relating textual queries to images, and converting images into a numerical format to allow comparison. We approach the first problem by associating a large number of terms with a small number of representative images, which can later be compared with the rest of the corpus. We approach the second problem by taking color histograms of our images, converting them into float arrays, which allow the images to be compared. The resulting image search system is not able to accurately return images matching queries exactly. However, it succeeds in finding images which resemble query terms chromatically.

1. INTRODUCTION

Image retrieval differs from text retrieval in a few fundamental aspects. In text retrieval, it is natural for both the query and the documents to be text based. In image retrieval, there are two choices for the form of queries: text-based and image-based. If the query is text-based, then the problem involves how to convert the textual query into something that can be compared with images. On the other hand, if the query is image-based, there is a different problem involving the user interface. Requiring the user to enter image queries is less useful than being able to type a text query and receiving images that match the text. In this paper we chose to focus on text-based image retrieval.

The first difficulty with text-based image retrieval is converting a textual query into a form that can be compared to images. This is important because text based image retrieval systems are more user friendly than image based image retrieval systems. The next problem in image retrieval is determining the features of an image and then building a system to compare them with features of other images. In text retrieval, features can be more easily determined. Comparing word distribution using tf-idf weightings and cosine similarity, provides a measure of similarity between a query and a document.

Unfortunately, extracting features from an image is more difficult. In image retrieval the features of an image are more difficult to determine than words in a text document. Words have similar meanings in most contexts making them ideal in determining the content of a document. Pixels on the other hand do not necessarily provide an accurate assessment of the content of an image. Computing the contextual similarity of images also presents a problem. Two images with similar features based on pixels may not be related at all.

It is important to find a solution to these problems with image retrieval. Being able to use text queries to search for images provides a better and easier user experience. Correctly identifying features of images and being able to form some measure of similarity between features of different images will provide more accurate results from image retrieval systems. Images are rapidly becoming more prevalent on the internet with advances in technology, such as digital cameras, and camera phones, which makes it important to be able to index and search more accurately for images.

In this paper we will concentrate on these problems. To incorporate textual queries into our image retrieval system, we use a set of representative images. Each representative image corresponds to a common query word. This allows text queries to be easily converted into images. To determine the features and similarities of the images we use a system involving color histograms.

2. ALGORITHM DESCRIPTION

2.1 Representative Images

In order to perform text-based image retrieval we needed some way to relate terms to images. Rather than tagging every image in our corpus with relevant keywords, which would have been time consuming and impractical, we opted for a system which only required a small number of tagged images. Within this subset, each image with a given tag was listed as a representative image for that term. We determined the similarity of each image in the larger corpus to a given term. We accomplished this by computing its similarity to each representative image of that term using image-based features. Using this set of representative images has the drawback that we are limited in the number of terms for which we can query. This limit is based on the size of our training corpus; if we find a larger corpus, will be able to search for more terms. Unfortunately, large corpora of labeled images are difficult to find.

2.2 Image Features

In order to compute the similarity of two pictures, we needed a method of representing an image numerically. We used color histograms, as they were both concise and easy to compute. For each image, we divided the spectrum of color hues into a number of slots. We then counted the number of pixels which fell into each slot. Since dark pixels look very similar regardless of hue, we weighted these counts based on pixel brightness. In this way, only visibly distinct colors had an impact on the color histogram of a given image. By converting images to an array of floats in this way, we could easily compute the similarity between two images using a simple dot product.

2.3 Indexing

Using representative images we were able to match text with a small number of images; and using color histograms, we were able to compute the similarity of all other images with those representative ones. In this way, we were able to compute a postings list for all terms corresponding to at least one representative image by calculating the maximum similarity of a given picture to any of the representative images for that term. We had considered averaging all similarities. However, we felt it was more important for a corpus image to match one representative image particularly well than for it to lie somewhere between all of them.

3. IMPLEMENTATION

We constructed our search system using a corpus of images retrieved from the Amsterdam Library of Object Images (ALOI) database. This corpus contains 1000 images of objects, each tagged with a descriptive name. To construct our index, we first selected a subset of 142 arbitrary images. We then used the words in the titles of these images as our representative terms, and assigned the corresponding pictures as representative images. In this way, we were able to index over the remaining images in the corpus, and use their titles to determine image relevance. Ultimately, we plan to use the entire ALOI corpus to construct our set of representative terms, and construct an index over a much larger corpus of images for use as an actual image retrieval system.

4. EVALUATION METHODS

We evaluate our system quantitatively by querying each term in our index and comparing the search results with our expected results from the annotations. We used two metrics to evaluate our system: r-precision and recall at 10. We chose these measures because we expect to show the first 10 images in our results but we do not know that there will always be 10 images that are relevant for each query. The r-precision measures how many of our displayed images will be of interest to the user, while the recall score measure how many of the relevant images we returned in the top results.

5. RESULTS

Unfortunately, our system performs poorly under these metrics. Averaged over all queries we have a r-precision of 14.6% and a recall of 14.2%. These scores indicate that our system

Figure 1: The top 10 results for a query for shoes

Figure 2: This image is labeled "blue/white dog" and is part of our representative set of blue images.

fails at object recognition. The only relevant results that we are returning are the ones in our representative set.

Although the system performs poorly quantitatively, it exhibits some qulaitative accomplishments. Although we do not return images that are labeled as relevant, the shape and color patterns of the returned images are similar to those of the representative images. For example in a query for shoe, we see in Figure 1 that although only the first result is actually a shoe all of the images are similar colors, and many of the images have similar shapes to the shoe. From these examples, we can see that the issue with this system is the choice of features that we extract from the images.

A second issue that we observed, after looking at sample queries, is that our annotations for the images are questionably accurate. Because we use the annotations as bags of words, we often associate pictures with words that do not accurately describe them. For example, Figure 2 shows an image that is labeled as blue, despite containing only a small amount of blue. In order to correct this problem we have two options: we could find a different set of annotated files, or try to build a more sophisticated model to match images and tags. Unfortunately, neither of these options is really feasible given the scope of this project.

6. CONCLUSIONS

In this system, we extracted features from images using only color histograms. Our search system could be greatly enhanced by comparing images using additional types of features. Color histograms are not powerful enough because they only consider the color of the image. Currently, our system can determine whether two images are the same color as each other; however, by examining additional attributes, it could also compare images using shape or texture. With enough features, our system might actually be able to match query terms to corresponding images with some degree of accuracy.

Overall, while our system performs poorly quantitatively, we are impressed with its ability to return images which resemble queried terms chromatically. This feature is not particularly useful in a search engine, as users who query for a term generally want to find images of that term, not images which share its colors; however, we feel this feature could have applications with regards to image clustering or image classification.