

# Visual Websearching Using Iconic Queries

Michael S. Lew Nicu Sebe  
Leiden Institute of Advanced Computer Science  
Leiden University  
{mlew, nicu}@liacs.nl

## Abstract

*Imagine a professional Web site designer who constantly has to come up with innovative looks for the client's homepages? How does he find the new content? The major search engines such as Hotbot (<http://www.hotbot.com>) allow us to find text on the Web, but typically have few or no capabilities for finding visual media. In this article we discuss methods for finding visual media on the WWW. The emphasis is on iconic queries which are essentially drag-and-drop visual concepts or simple semantics. We describe our method for finding static feature sets and then describe a novel method called active feature sets, which chooses a feature set based on the context in the image.*

## 1 Introduction

An emerging trend is to try to have the computer understand how people see the world [Picard 1996]. However, understanding the world is a fundamental computer vision problem which has withstood decades of research. Petkovic [1997] has called this finding "simple semantics." From recent literature, this generally means finding computable image features which are correlated with visual concepts. The key distinction is that we are not trying to fully understand how human intelligence works. This would imply creating a general model for understanding all visual concepts. Instead, we are satisfied to find features which describe some small, but useful domains of visual concepts.

## 2 ImageScape System

The ImageScape system [Lew, et al. 1997] consists of the following modules:

- collection of text, images, audio, and video from the WWW

- compression of the image database
- semantic object detection in images
- computer sketching of images
- matching between the icons/sketches with the database images

URL: <http://ind134a.wi.leidenuniv.nl:2001>

In this paper our discussion is limited to the visual matching techniques which include the iconic and sketch queries. There are other interesting WWW image search engines which have been introduced in the research literature and are not described here due to space limitations. An example and starting point would be the WebSeek [Smith and Chang 1997] system from Columbia University, which finds similar images and performs automatic text based category classification.

### 2.1 Detecting Visual Concepts

For each visual concept we wish to detect, a large set of positive and negative examples is collected. We measure a variety of texture, color, and shape features, and find the subset of the features which maximizes the discriminatory power. This subset is called the *normal set* and is derived from the examples as follows:

1. Assume that there are  $M$  scalar features, each of which has been normalized to  $0..255$
2. Measure the distribution of the positive examples  $\mathbf{F}[x, y], x=1..M, y=0..255$
3. Measure the distribution of the negative examples  $\mathbf{G}[z, y], z=1..M, y=0..255$
4. Calculate the Kullback relative information,  $\mathbf{K}[x]$  from  $\mathbf{F}$  and  $\mathbf{G}$ .
5. Calculate the correlation between features,  $\mathbf{C}[x,z]$  from  $\mathbf{F}$  and  $\mathbf{G}$ .
6. Define the  $N$  most informative features as the  $N$  features which maximize the Kullback relative information and minimize the correlation between

features. This set of features is denoted the normal set,  $D_m[u], u=1...N$

## 2.2 Active Feature Sets

The previous algorithm had the limitation that it used a static feature set. Static feature sets may be inappropriate for specific classes of imagery. For example, using color to find skin is appropriate if the image is color, but inappropriate if the image is grayscale. This observation led to the use of active feature sets (AFS). Once we have the first feature set (Normal Set), we re-classify the ground truth test set to determine in which images the misclassifications occur. In order to identify these images and classify them, we design two more feature sets. The second feature set (Decision Set) is trained to discriminate between the correctly classified images and the misclassified images. Using this feature set, we attempt to determine the situations when the normal set was inappropriate. The third feature set (Outlier Set) is trained to discriminate between the misclassified images and the non-concept images. Therefore, if the normal set was found to be inappropriate by the decision set, then we apply the outlier set as shown in Figure 2.1.

This method resulted in lower overall misdetection rates of 0.04 for sky, 0.11 for water, 0.13 for human faces/skin, 0.23 for trees/grass, and 0.25 for stone/sand. An example query is shown in Figure 2.5. We define it as an *active feature set* because it does not use a *single static feature set*. It determines the context of the input image and then chooses an appropriate feature set for the situation.

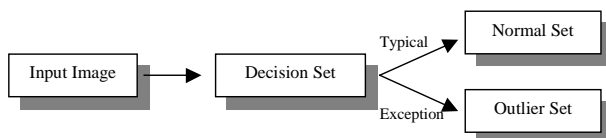


Figure 2.1. Using an active feature set

An example is shown in Figure 2.2 of an iconic query in ImageScape and the results.

## 2.3 Matching Hand Drawn Sketches

Matching hand-drawn sketches to images can be reduced to a shape matching problem, that is, matching the shape of the sketch to the edges (using the Sobel operator) of the candidate image from the database. In our shape matching framework we have split the shape matching process into two parts: global shape matching is addressed by moment invariants and local shape matching is measured by the deformation energy from active contours (snakes). The first 7 moment invariants give us a fast filter for reducing the candidates while

preserving global shape similarity. Keeping only the top N image candidates from the moment invariants shape filter, we sort the remaining image candidates by the deformation energy between the user sketch and each database image.

## 3 Summary

In this article we described the primary visual concept detection methods within the ImageScape system. First, an initial feature set is trained from a set of positive and negative examples of the visual concept. Second, we introduce a new method for visual concept detection which learns which feature set to use depending on the input image content.

## References

- Lew, M., K. Lempinen, and N. Huijismans, "Webcrawling Using Sketches," VISUAL'97, San Diego, December, 1997, pp. 77-84.
- Petkovic, D., "Challenges and Opportunities for Pattern Recognition and Computer Vision Research in Year 2000 and Beyond," Proc. of the Int. Conf. on Image Analysis and Processing, September, Florence, vol. 2, pp. 1-5, 1997.
- Picard, R.. "A Society of Models for Video and Image Libraries." IBM Systems Journal. 1996.
- Smith, J. R. and S.F. Chang, "Visually Searching the Web for Content," IEEE Multimedia, 1997, pg. 12-20.



Figure 2.2. A query and results for a water below a person