HOW IT WORKS:

1. EXTRACT VISUAL WORDS

SCALE INVARIANT FEATURE TRANSFORM

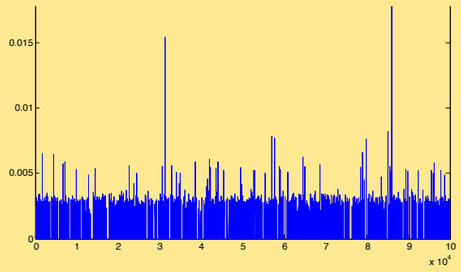
$$\begin{pmatrix} x_1 & \cdots & x_N \\ y_1 & \cdots & y_N \\ s_1 & \cdots & s_N \\ \theta_1 & \cdots & \theta_N \end{pmatrix}_{4\times N} \qquad (v_1|\ldots\ldots|v_N)_{128\times N}$$

The features of an image are detected and described by a 128-dimensional vector. The feature space is then quantized, creating “visual words”.

FEATURES

WORDS

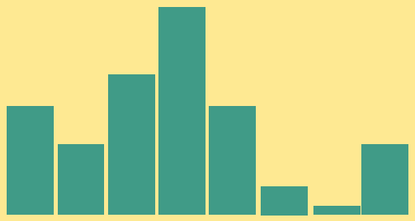
HISTOGRAM



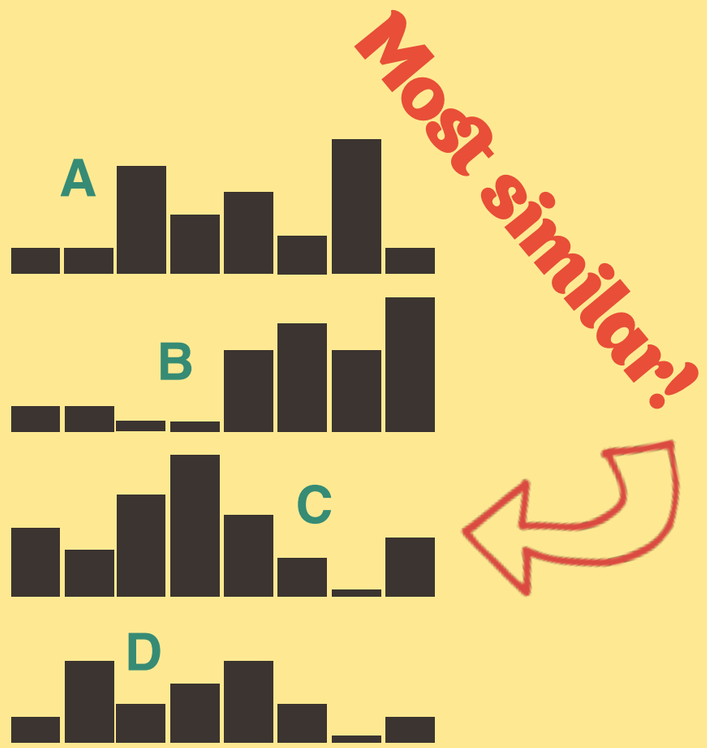
2. GOOGLE STYLE SEARCH

```
matches = search(Database for Words);
```

The tf-idf weighted histograms are then used to search the database for similar histograms. This is the same method as Google uses for text search, replacing words with “visual words”.

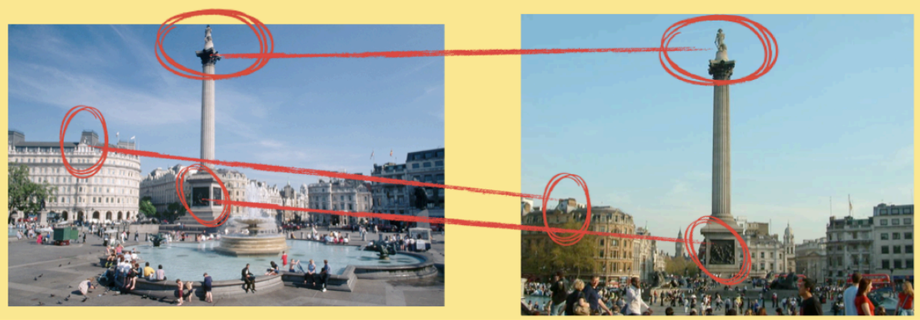


QUERY



DATABASE

3. SPATIALLY VERIFY



$$X_{db} = \begin{bmatrix} A & \vec{t} \\ 0 & 1 \end{bmatrix} X_{query}$$

The match from the database and query image must depict the same object, so there should be a spatial correspondence between the visual words of the two images.