Abstract

This thesis is devoted to the Earth Mover's Distance and its use within content-based image retrieval (CBIR). The major CBIR problem discussed is the *pattern problem*: Given an image and a query pattern, determine if the image contains a region which is visually similar to the pattern; if so, find at least one such image region. The four main themes of this work are: (i) partial matching, (ii) matching under transformation sets, (iii) combining (i) and (ii), and (iv) effective pruning of unnecessary, expensive distance/matching computations.

The first pattern problem we consider is the *polyline shape search problem* (PSSP): Given *text* and *pattern* planar polylines, find all approximate occurrences of the pattern within the text, where such occurrences may be scaled and rotated versions of the pattern. For a text and a pattern with n and m edges, respectively, we present an $O(m^2n^2)$ time, $O(mn)$ space PSSP algorithm. A major strength of our algorithm is its generality, as it can be applied for any shape pattern represented as a polyline.

The main distance measure studied in this thesis is the *Earth Mover's Distance* (EMD), which is an edit distance between distributions that allows for partial matching, and which has many applications in CBIR. A discrete distribution is just a set of (point,weight) pairs. In the CBIR context, the weight associated with a particular point in a feature space is the amount of that feature present in the image. The EMD between two distributions is proportional to the minimum amount of work needed to change one distribution into the other, where one unit of work is the amount necessary to move one unit of weight by one unit of *ground distance*. We give a couple of modifications which make the EMD more amenable to partial matching: (i) the *partial* EMD in which only a given fraction of the weight in one distribution is forced to match weight in the other, and (ii) the τ-EMD which measures the amount of weight that cannot be matched when weight moves are limited to at most τ ground distance units.

An important issue addressed in this thesis is the use of efficient, effective lower bounds on the EMD to speed up retrieval times. If a system can quickly prove that the EMD is larger than some threshold, then it may be able to avoid an EMD computation and decrease

its query time. We contribute lower bounds that are applicable in the partial matching case in which distributions do not have the same total weight. The efficiency and effectiveness of our lower bounds are demonstrated in a CBIR system which measures global color similarity between images.

Another important problem in CBIR is the EMD *under transformation* ($\text{EMD}_{\mathcal{G}}$) *problem*: find a transformation of one distribution which minimizes its EMD to another, where the set of allowable transformations $\mathcal{G}$ is given. The problem of estimating the size/scale at which a pattern occurs in an image is phrased and efficiently solved as an $\text{EMD}_{\mathcal{G}}$ problem in which transformations scale the weights of a distribution by a constant factor.

For $\text{EMD}_{\mathcal{G}}$ problems with transformations that modify the points of a distribution but not its weights, we present a monotonically convergent iteration called the *FT iteration*. This iteration may, however, converge to only a locally optimal EMD value and transformation. The FT iteration is very general, as it can be applied for many different (ground distance, transformation set) pairs, and it can be modified to work with the partial EMD, as well as in some cases in which transformations change both distribution points and weights. We apply the FT iteration to the problems of (i) illumination-invariant object recognition, and (ii) point feature matching in stereo image pairs. We also present algorithms that are guaranteed to find a globally optimal transformation when matching equal-weight distributions under translation (i) on the real line with the absolute value as the ground distance, and (ii) in any finite-dimensional space with the Euclidean distance squared as the ground distance.

Our pattern problem solution is the SEDL (Scale Estimation for Directed Location) content-based image retrieval system. Three important contributions of this system are (1) a general framework for finding both color and shape patterns, (2) the previously mentioned novel scale estimation algorithm using the EMD, and (3) a directed (as opposed to exhaustive) search strategy. We show that SEDL achieves excellent results for the color pattern problem on a database of product advertisements, and the shape pattern problem on a database of Chinese characters. A few promising pattern locations are efficiently computed at query time without having to examine image areas that obviously do not contain the pattern. SEDL uses the τ-EMD to help eliminate false positives resulting from difficulties in trading off, for example, color and position distances to measure visual similarity.