FINDING COLOR AND SHAPE PATTERNS IN IMAGES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Scott Cohen
May 1999

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Leonidas Guibas
(Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Carlo Tomasi

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jeffrey Ullman

Approved for the University Committee on Graduate Studies:

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* ($\mathrm{EMD}_{\mathcal{G}}$) *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 $\mathrm{EMD}_{\mathcal{G}}$ problem in which transformations scale the weights of a distribution by a constant factor.

For $\mathrm{EMD}_{\mathcal{G}}$ 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.

Acknowledgements

I am deeply indebted to my adviser Professor Leonidas Guibas for guiding me through the Ph.D. program. Amazingly to me, he always seemed to know just the right reference for every discussion. Whenever I was stuck in some detail, he would help me see the bigger picture and offer useful comments and suggestions (although it would usually take me a while to truly understand and appreciate his words). I am also very grateful to Professor Carlo Tomasi, whom I consider my secondary adviser, for all his time and help throughout the years. He made me feel like a part of his group, and he taught me to appeal to the basics and intuition before mechanical manipulations. I also thank Professor Jeffrey Ullman and Professor Leon Simon for serving on my orals committee. I learned how to be a better teacher from these superb educators. Some of my fondest undergraduate class memories are from Professor Simon's real analysis and differential equations classes. I thank Professor Gene Golub for playing the roles of my sponsor and adviser in an undergraduate research program many years ago, for recommending me to the Stanford computer science Ph.D. program, and for serving on my orals committee.

I have always been very fortunate to have excellent teachers who instilled a love of learning within me. I recognize the contributions of all my teachers in helping me reach my current position. There are too many to give all by name, but I especially thank my high school mentor Lawrence Zimmerman for his problem solving lessons that have stayed with me all these years.

I have also been very fortunate to have the support of many close friends during my time here at Stanford. Again, there are too many to give all by name, but I would like to mention Krista Lentine, Scott Putterman, Mark Ruzon, João Comba, David Hoffman, Alok Shah, John Stoffel, Jeff Chen, Ann Chan, Joseph Silvagnoli, and Nikhyl Singhal. I also thank Tong Zheng, a very special person in my life who was always there for me. Her unconditional love and support helped get me through the most difficult times. Finally, I would be remiss if I did not say how much I enjoyed being a part of Professor Tomasi's computer vision group, including all the vislunches, vispics, and friendships with its members.

I owe special thanks to a couple of vision group members for very direct contributions to my work. I thank Yossi Rubner for his Earth Mover's Distance code, and much more importantly, for a framework upon which to build, explore, and extend. I also thank Stan Birchfield for his feature extraction code. Other members of the academic community outside of Stanford helped as well. I thank David Slater for his database of objects imaged under varying illumination, Madirakshi Das for her database of color advertisements, Oliver Lorenz for his database of Chinese characters, and Robert Ogniewicz for his medial axis code. For my first three years in the Ph.D. program, I was supported by an NSF Fellowship (grant CCR-9215219); for my last three years, I was supported by the DARPA grant for Stanford's image retrieval project (contract DAAH04-94-G-0284).

Finally, I thank my uncle Lenny, from whom I undoubtedly get my academic inclination, for his constant encouragement, support, and good advice from the very beginning of my education.

Dedicated in loving memory to my grandmother Ann

Contents

List of Figures