Provably Fast Algorithms for Contour Tracking

Daniel Freedman and Michael S. Brandstein
Division of Engineering and Applied Sciences, Harvard University
Cambridge, MA, USA 02138
{freedman, msb}@hrl.harvard.edu

Abstract

A new tracker is presented. Two sets are identified: one which contains all possible curves as found in the image, and a second which contains all curves which characterize the object of interest. The former is constructed out of edge-points in the image, while the latter is learned prior to running. The tracked curve is taken to be the element of the first set which is nearest the second set. The formalism for the learned set of curves allows for mathematically well understood groups of transformations (e.g. affine, projective) to be treated on the same footing as less well understood deformations, which may be learned from training curves. An algorithm is proposed to solve the tracking problem, and its properties are theoretically demonstrated: it solves the global optimization problem, and does so with certain complexity bounds. Experimental results applying the proposed algorithm to the tracking of a moving finger are presented, and compared with the results of a condensation approach.

1 Introduction

This work addresses the tracking of moving contours in a video-stream. Specifically, given a sequence of images in which a known object of interest is in motion, the goal is to track the object's silhouette across time-varying images. Applications abound in medicine, surveillance, and audio-visual speech recognition for noisy environments.

A standard approach to tracking is embodied in the condensation tracker [5, 1], which relies heavily on an a priori knowledge of the dynamics of the tracked object. In this paper, a more realistic assumption is made: a priori knowledge of the space in which the tracked object lives is assumed, but not the dynamics within that space. Section 1 outlines the general approach to the problem of tracking with only knowledge of the space in which the object lives: tracking is posed a mixed continuous-combinatorial optimization problem. Section 2 states two theorems which allow this optimization problem to be attacked. The first theorem shows that it is possible to almost reach the global minimum, while the second theorem shows that it can be done efficiently. Section 3 presents proofs of these theorems. Section 4 presents complexity bounds on the speed of the algorithm. Section 5 shows results: the advantage of the proposed approach over one based on dynamics is seen in experiments.

1.1 The Problem

In tracking an object through a video-stream, the approach taken will be to focus entirely on the object's contour, or outline. Thus, the problem of tracking reduces to one of finding the "correct" curve in the image, i.e. the curve which corresponds to the object of interest. Suppose it is possible to generate two sets of curves. One set, $\tilde{E}$, represents all of the curves that can be generated from the connection of edge-points in the image; a description of how to produce such a set follows in section 1.2. The other set, $\tilde{C}$, represents all of the curves that correspond to the particular geometry of the object being tracked; that is, this set $\tilde{C}$ contains all of the information about the shape of the object's silhouette. A discussion of this set is provided at the end of section 1.2. Given these two sets, a sensible problem to solve is

$$\min_{\tilde{e}\in\tilde{E},\tilde{c}\in\tilde{C}} \|\tilde{e}-\tilde{c}\|$$

where $\|\cdot\|$ is the L_2 norm. The idea is straightforward: the minimization over the two arguments ensures that the "observed curve" $\tilde{e}$ chosen from all of the possible curves in the image best matches the model of the object being tracked, as embodied in the set $\tilde{C}$. The tracked object is taken to be $\tilde{e}^*$, the minimizing argument. An earlier attempt, on the part of the authors, to solve a related problem is contained in [4]; the approach presented in the following sections, however, is much more flexible and robust.

1.2 The Sets $\tilde{E}$ and $\tilde{C}$

The set $\tilde{E}$, of curves constructed from edge-points in the image, is generated as follows. At N equally spaced points along the detected contour of the previous frame, edge-search takes place in circular regions (in the image of the *current* frame). In each of these regions, a number of edge-points are detected; denote the set of edge-points detected in the n^{th} region by E_n. An element $\tilde{e}\in\tilde{E}$ may be constructed as follows: (1) take one edge-point

$e_n \in E_n$ from each region $n = 1, \ldots, N$; (2) smoothly interpolate the set of edge-points $e_1, \ldots, e_N$ into a curve $\tilde{e}$. (The method of interpolation will not concern us here.) Thus, the set $\tilde{E}$ is in one-to-one correspondence with the set $E \equiv E_1 \times \cdots \times E_N$. Suppose there are M edge-points detected per site, i.e. $|E_n| = M \; \forall n$ (in reality, of course, $|E_{n_1}| \neq |E_{n_2}|$); then the size of the set of observed curves is $|\tilde{E}| = |E| = M^N$.

The set $\tilde{C}$ is generated from training curves before the algorithm is run. It is assumed that this set of learned curves is a finite dimensional manifold (as it will be in all cases that will be practically encountered), has dimension σ, and may be specified parametrically as

$$\tilde{C} = \{\tilde{c}(u) : u \in U\}$$

where U is some known, σ-dimensional, real, compact, convex set (for example, $U = [0,1]^\sigma$), and $\tilde{c}(\cdot)$ is a function which maps the points in U to points in curve space. Some of the parameters may represent familiar transformations; for example, $u_1, \ldots, u_6$ could represent a subset of the affine transformations. A particular method for learning $\tilde{C}$ will be outlined in section 5; for the moment, it will be taken as given.

1.3 Recasting the Problem

Using the parametric form for $\tilde{C}$ allows the problem to be rewritten $\min_{\tilde{e} \in \tilde{E}, u \in U} \|\tilde{e} - \tilde{c}(u)\|$. However, approximating the square of the L_2 norm by its Riemann sum gives

$$\|\tilde{e} - \tilde{c}\|^2 = \int_0^L \|\tilde{e}(s) - \tilde{c}(s)\|^2 ds$$
$$\approx \frac{L}{N} \sum_{n=1}^{N} \|e_n - c_n\|^2$$

where $e_n = \tilde{e}(s_n), c_n = \tilde{c}(s_n)$, and $s_n = \frac{L(n-1)}{N-1}$. Note that $e_1, \ldots, e_N$ is simply the set of edge-points, culled from the sets $E_1, \ldots, E_N$, which were interpolated to give $\tilde{e}$; sampling $\tilde{e}$ gives back the original points. Denoting $e = (e_1, \ldots, e_N) \in \Re^{2N}$ and similarly for c, then the minimization problem may be approximated well by

$$\min_{e \in E, u \in U} \|e - c(u)\|$$

if N is sufficiently large. Note that the norm in the above is the now the normal Euclidean norm in $\Re^{2N}$, $E = E_1 \times \cdots \times E_N$ as before, and $c(\cdot) : \Re^\sigma \to \Re^{2N}$.

The recast problem is still not obviously amenable to solution, as E is still discrete and very large, while U is continuous. Below, an algorithm is proposed for solving for the global optimum. In particular, if d^* is the value of the global minimum, then the algorithm will be shown to give a value of at most $d^* + \Delta d$, for a specified Δd. Further, complexity bounds on the algorithm, in terms of M, N, and Δd will be established. The essence of the algorithm is contained in the following theorems.

2 Two Theorems on Global Optimization

Before stating the theorems, it will be necessary to make a number of definitions.

Definition: V is said to be an **ε-cover** of the compact set U if $\forall u \in U$, $\exists v \in V$ such that $\|v - u\| \leq \varepsilon$, and ε is the smallest such value. Alternatively, $\varepsilon = \max_{u \in U} [\min_{v \in V} \|v - u\|]$. (Note that the maximum is well-defined since U is compact.)

Definition: Given a compact set U, a set V satisfying $|V| < \infty$ and $V \subset U$, and a point $v \in V$, let

$$S(v, V, U) = \{u \in U : \|v - u\| \leq \|v' - u\| \; \forall v' \in V\}$$

Definition: An I-depth **tree minimization structure (TMS)** is the triple

$$(U, \{V_i\}_{i=1}^I, \{\phi_i(\cdot)\}_{i=2}^I)$$

satisfying:

1. U is a compact set
2. (a) $|V_i| < \infty$ (b) $V_i \subset U$ (c) $V_i \subset V_{i+1}$
3. For any $i \geq 2$, $\{\phi_i(v_{i-1})\}_{v_{i-1} \in V_{i-1}}$ is a partition of V_i such that (a) $v_{i-1} \in \phi_i(v_{i-1})$ (b) $S(w, V_i, U) \subset S(v_{i-1}, V_{i-1}, U) \; \forall w \in \phi_i(v_{i-1})$.

Definition: A TMS $(U, \{V_i\}_{i=1}^I, \{\phi_i(\cdot)\}_{i=2}^I)$ is said to be **convex** if $S(v, V_i, U)$ is convex for all $v \in V_i$; $i = 1, \ldots, I$.

Definition: Let $H(u) = \frac{\partial c}{\partial u}$, so that $H(u) \in \Re^{2N \times \sigma}$. Let $\lambda_1(u)$ be the largest eigenvalue of the $\sigma \times \sigma$ matrix $H^T(u)H(u)$. Then for any $Y \subset U$, define $A(Y) = [\max_{u \in Y} \lambda_1(u)]^{1/2}$.

We are now ready to state the theorems which will allow us to attack the tracking problem, $\min_{e \in E, u \in U} \|e - c(u)\|$. The import of the theorems will be discussed after their formal statements. Note that theorem 1 was initially presented in [3].

Theorem 1: Let V be any ε-cover of U. Further, let $d^* = \min_{e \in E, u \in U} \|e - c(u)\|$ and let $d^\dagger = \min_{u \in U} \|e^\dagger - c(u)\|$, where $e^\dagger = \arg\min_{e \in E} (\min_{v \in V} \|e - c(v)\|)$. If $\Delta d = d^\dagger - d^*$, then

$$0 \leq \Delta d \leq \frac{3A^2(U)\varepsilon^2}{d^*} + 2A(U)\varepsilon.$$

Theorem 2: Given a convex TMS $(U, \{V_i\}_{i=1}^I, \{\phi_i(\cdot)\}_{i=2}^I)$, define:

1. $\varepsilon_i = \max_{u \in U} \min_{v \in V_i} \|v - u\|$
2. $X_1 = V_1$
3. $d_i(x_i) = \min_{e \in E} \|e - c(x_i)\|$, $x_i \in X_i$; $d_i^* = \min_{x_i \in X_i} d_i(x_i)$
4. $\bar{D}_i(x_i) = (d_i^2(x_i) - 3A^2(S(x_i, V_i, U))\varepsilon_i^2 - 2d_i(x_i)A(S(x_i, V_i, U))\varepsilon_i)^{1/2}$, $x_i \in X_i$
5. $X_{i+1} = \bigcup_{x_i \in X_i : \bar{D}_i(x_i) < d_i^*} \phi_{i+1}(x_i)$

(Note: $\{X_i\}_{i=1}^I$ can be generated by recursively applying 3-5, after starting at 2.) Then

$$\min_{e \in E, x \in X_I} \|e - c(x)\| = \min_{e \in E, v \in V_I} \|e - c(v)\|$$

and $|X_I| \leq |V_I|$.

The first theorem presents a problem whose solution is amenable, and compares the objective function value it generates with the optimal value, d^*. In particular, the problem $\min_{e \in E, v \in V} \|e - c(v)\|$ can be solved, albeit inefficiently, by exhaustive search through the two discrete sets E and V. If the e-minimizing argument is labelled $e^\dagger$, then the quantity $d^\dagger = \min_{u \in U} \|e^\dagger - c(u)\|$ is of interest; the fact that the minimizing u is never solved for does not matter, since our contour estimate is based on $e^\dagger$ rather than $c(u^\dagger)$ (see section 1.1). The theorem gives an upper bound on how far away $d^\dagger$ can be from d^*; this bound depends critically on ε, a parameter which indicates how finely V samples U.

The second theorem presents a more efficient way of solving the problem $\min_{e \in E, v \in V} \|e - c(v)\|$, as long as V can be expressed as V_I for an I-depth TMS. In this case, it is certain that the method presented in theorem 2 is at least as fast as a "naive method" (the meaning of which is made clear in section 4) would suggest. While no theoretical complexity bounds have been derived on how much faster this method is (i.e., the amount by which $|V_I|$ exceeds $|X_I|$), positive experimental results are reported in section 5.

3 Proof of the Theorems

Lemma 1: $X_i \subset V_i$.
Proof: Proceed by induction. Since $X_1 = V_1$, the lemma is satisfied trivially for $i = 1$. Suppose it is true for $i = k$: $X_k \subset V_k$. Then

$$X_{k+1} = \bigcup_{x_k \in X_k : \bar{D}_k(x_k) < d_k^*} \phi_{k+1}(x_k) \subset \bigcup_{x_k \in X_k} \phi_{k+1}(x_k) \subset \bigcup_{x_k \in V_k} \phi_{k+1}(x_k) = V_{k+1}$$

where the latter two steps follow from the induction hypothesis and property 3 of the TMS definition, respectively. Note that this also establishes the second part of theorem 2: $|X_I| \leq |V_I|$. ■

Definition: $v_{i_1} \in V_{i_1}$ is the i_1^{th} **stage ancestor** of $v_{i_2} \in V_{i_2}$ if (a) $i_1 < i_2$ and (b) $\exists$ a sequence $\{v_i\}_{i=i_1+1}^{i_2-1}$ with $v_i \in V_i$, $i = i_1 + 1, \ldots, i_2 - 1$ and $v_{i+1} \in \phi_{i+1}(v_i)$, $i = i_1, \ldots, i_2 - 1$.

Lemma 2: $v^* \in V_I - X_I \Rightarrow \exists i < I$ and $v_i^* \in V_i$ such that (a) v_i^* is the i^{th} stage ancestor of v^* and (b) $\bar{D}_i(v_i^*) \geq d_i^*$.
Proof: First, note that the condition $v^* \in V_I - X_I$ only makes sense because of lemma 1, which establishes that $X_I \subset V_I$. Now, proceed by contradiction. Then if $v_1^*, \ldots, v_{I-1}^*$ are the first through $(I-1)^{th}$ stage ancestors of v^* (note the fact that there is only one ancestor at each stage due to the fact that the $\{\phi_i(\cdot)\}$ are bijections) $\Rightarrow \bar{D}_i(v_i^*) < d_i^*$ $\forall i = 1, \ldots, I-1$. Now proceed by induction. $v_1^* \in X_1$ since $X_1 = V_1$. Suppose $v_i^* \in X_i$ for $i > 1$. Then $X_{i+1} = \bigcup_{x_i \in X_i : \bar{D}_i(x_i) < d_i^*} \phi_{i+1}(x_i)$, so that $\phi_{i+1}(v_i^*) \subset X_{i+1}$. But $v_{i+1}^* \in \phi_{i+1}(v_i^*)$ by the ancestry definition, so $v_{i+1}^* \in X_{i+1}$. Thus by induction, $v^* \in X_I$. This is a contradiction, since it was assumed that $v^* \in V_I - X_I$. ■

Lemma 3: If $v_i \in V_i$ is the i^{th} stage ancestor of $v^* \in V_I$, then $v^* \in S(v_i, V_i, U)$.
Proof: If $v_i \in V_i$ is the i^{th} stage ancestor of $v^* \in V_I$, then $\exists$ a sequence $v_{i+1}, \ldots, v_{I-1}$ such that $v_{i+1} \in \phi_{i+1}(v_i)$. By property (3b) of the TMS definition $v_{k+1} \in \phi_{k+1}(v_k) \Rightarrow S(v_{k+1}, V_{k+1}, U) \subset S(v_k, V_k, U)$; repeated application for $k = i, \ldots, I-1$ gives $S(v^*, V_I, U) \subset S(v_i, V_i, U)$. But $v^* \in S(v^*, V_I, U)$ by definition, so $v^* \in S(v_i, V_i, U)$. ■

Lemma 4: Let $Q \subset U$ be a convex, compact set and let R be an ε-covering of Q. Let $e_1, e_2 \in E$, $d_i = \min_{u \in Q} \|e_i - c(u)\|$, and $\tilde{d}_i = \min_{v \in R} \|e_i - c(v)\|$. Then $d_2^2 - d_1^2 \leq \tilde{d}_2^2 - \tilde{d}_1^2 + 3A^2(Q)\varepsilon^2 + 2d_2 A(Q)\varepsilon$.
Proof: Make the following definitions: for $i = 1, 2$, let

- $u_i = \arg\min_{u \in Q} \|e_i - c(u)\|$, $\quad d_i = \|e_i - c(u_i)\|$
- $\tilde{u}_i = \arg\min_{v \in R} \|e_i - c(v)\|$, $\quad \tilde{d}_i = \|e_i - c(\tilde{u}_i)\|$
- $\hat{u}_i = \arg\min_{v \in R} \|u_i - v\|$, $\quad \hat{d}_i = \|e_i - c(\hat{u}_i)\|$

Then:

$$\begin{aligned} d_2^2 - d_1^2 &= \|e_2 - c(u_2)\|^2 - \|e_1 - c(u_1)\|^2 \\ &\leq \|e_2 - c(\hat{u}_2)\|^2 + \|c(\hat{u}_2) - c(u_2)\|^2 \\ &\quad - \|e_1 - c(\hat{u}_1)\|^2 + \|c(\hat{u}_1) - c(u_1)\|^2 \\ &= \hat{d}_2^2 - \hat{d}_1^2 + \|c(\hat{u}_1) - c(u_1)\|^2 + \|c(\hat{u}_2) - c(u_2)\|^2 \end{aligned}$$

where the inequality in the second line is a double application of the triangle inequality. Now:

1. $\hat{d}_1 \geq \tilde{d}_1$ by definition, so $-\hat{d}_1^2 \leq -\tilde{d}_1^2$

2. Expanding $\hat{d}_2^2 = \|e_2 - c(\hat{u}_2)\|^2$ gives

$$\hat{d}_2^2 = \|e_2 - [c(u_2) + H(\breve{u}_2)(\hat{u}_2 - u_2)]\|^2$$

where $H(u) = \frac{\partial c}{\partial u}$ and $\breve{u}_2 \in Q$. This is the multi-variable mean value theorem, which is valid due to the convexity of Q [2]. Thus,

$$\hat{d}_2^2 = \|e_2 - c(u_2)\|^2 + \|H(\breve{u}_2)(\hat{u}_2 - u_2)\|^2 + 2(e_2 - c(u_2))^T H(\breve{u}_2)(\hat{u}_2 - u_2)$$

(a) $\|e_2 - c(u_2)\|^2 = d_2^2$

(b) $\|H(\breve{u}_2)(\hat{u}_2 - u_2)\|^2 = (\hat{u}_2 - u_2)^T H^T(\breve{u}_2) H(\breve{u}_2)(\hat{u}_2 - u_2)$. Since R is an ε-cover of Q, $\exists v$ such that $\|u_2 - v\| \leq \varepsilon$. But by definition $\hat{u}_2 = \arg\min_{v \in R} \|u_2 - v\|$; thus, $\|\hat{u}_2 - u_2\| \leq \varepsilon$. But then

$$(\hat{u}_2 - u_2)^T H^T(\breve{u}_2) H(\breve{u}_2)(\hat{u}_2 - u_2) \leq \lambda_1(\breve{u}_2)\varepsilon^2 \leq \left(\max_{u \in Q} \lambda_1(u)\right)\varepsilon^2 \equiv A^2(Q)\varepsilon^2$$

where $\lambda_1(\cdot)$ and $A(\cdot)$ are defined as before.

(c) Finally,

$$(e_2 - c(u_2))^T H(\breve{u}_2)(\hat{u}_2 - u_2) \leq |(e_2 - c(u_2))^T H(\breve{u}_2)(\hat{u}_2 - u_2)| \leq \|e_2 - c(u_2)\| \|H(\breve{u}_2)(\hat{u}_2 - u_2)\|$$

where the latter inequality is due to the Cauchy-Schwartz inequality. But $\|e_2 - c(u_2)\| = d_2$ and from the previous argument $\|H(\breve{u}_2)(\hat{u}_2 - u_2)\| \leq A(Q)\varepsilon$. Thus, $(e_2 - c(u_2))^T H(\breve{u}_2)(\hat{u}_2 - u_2) \leq d_2 A(Q)\varepsilon$.

3. Using the mean value theorem once again

$$\|c(\hat{u}_i) - c(u_i)\|^2 = \|c(u_i) + H(\breve{u}_i)(\hat{u}_i - u_i) - c(u_i)\|^2 = \|H(\breve{u}_i)(\hat{u}_i - u_i)\|^2 \leq A^2(Q)\varepsilon^2$$

Thus,

$$d_2^2 - d_1^2 \leq d_2^2 - \tilde{d}_1^2 + 3A^2(Q)\varepsilon^2 + 2d_2 A(Q)\varepsilon. \quad \blacksquare$$

Lemma 5: $\bar{D}_i(v_i) \geq d_i^* \Rightarrow \min_{e \in E} \|e - c(u)\| \geq d_i^* \quad \forall u \in S(v_i, V_i, U)$.

Proof: For any $e_1, e_2 \in E$, lemma 4 states that $d_2^2 - d_1^2 \leq d_2^2 - \tilde{d}_1^2 + 3A^2(Q)\varepsilon^2 + 2d_2 A(Q)\varepsilon \Rightarrow -d_1^2 \leq -\tilde{d}_1^{\,2} + 3A^2(Q)\varepsilon^2 + 2d_2 A(Q)\varepsilon$. But $\tilde{d}_2 \geq d_2 \Rightarrow d_1^2 \geq \tilde{d}_1^2 - 3A^2(Q)\varepsilon^2 - 2\tilde{d}_2 A(Q)\varepsilon$. Now, choose

$$e_1 = \arg\min_{e \in E} \left(\min_{u \in Q} \|e - c(u)\|\right)$$
$$e_2 = \arg\min_{e \in E} \left(\min_{v \in R} \|e - c(v)\|\right)$$

Then $\tilde{d}_2 \leq \tilde{d}_1$ (see the definitions of $\tilde{d}_i$). Thus

$$d_1 \geq (\tilde{d}_2^2 - 3A^2(Q)\varepsilon^2)^{1/2} - 2\tilde{d}_2 A(Q)\varepsilon$$

Let $Q = S(v_i, V_i, U)$ and $R = \{v_i\}$ (i.e., a single element set). Then $\tilde{d}_2 = d_i(v_i)$ and $d_1 = \min_{e \in E, u \in S(v_i, V_i, U)} \|e - c(u)\|$, so that the above inequality becomes

$$\min_{e \in E, u \in S(v_i, V_i, U)} \|e - c(u)\| \geq \bar{D}_i(v_i)$$

Finally

$$\bar{D}_i(v_i) \geq d_i^* \Rightarrow \min_{e \in E, u \in S(v_i, V_i, U)} \|e - c(u)\| \geq d_i^*$$
$$\Rightarrow \min_{e \in E} \|e - c(u)\| \geq d_i^* \; \forall u \in S(v_i, V_i, U). \quad \blacksquare$$

Proof of Theorem 1: For any $e_1, e_2 \in E$, lemma 4 states that $d_2^2 - d_1^2 \leq d_2^2 - \tilde{d}_1^2 + 3A^2(Q)\varepsilon^2 + 2d_2 A(Q)\varepsilon$. However, $d_2 \leq \tilde{d}_2$ so that $d_2^2 - d_1^2 \leq \tilde{d}_2^2 - \tilde{d}_1^2 + 3A^2(Q)\varepsilon^2 + 2d_2 A(Q)\varepsilon$. Now, let $Q = U$ and $R = V$; further, let $e_1 = e^* = \arg\min_{e \in E}[\min_{u \in U} \|e - c(u)\|]$ and $e_2 = e^\dagger = \arg\min_{e \in E}[\min_{v \in V} \|e - c(v)\|]$. Then by definition $\tilde{d}_2 \leq \tilde{d}_1$, so that $\tilde{d}_2^2 - \tilde{d}_1^2 \leq 0$, and thus

$$d^{\dagger 2} - d^{*2} \leq 3A^2(U)\varepsilon^2 + 2d^\dagger A(U)\varepsilon$$

$$d^\dagger - d^* \leq \frac{3A^2(U)\varepsilon^2 + 2d^\dagger A(U)\varepsilon}{d^* + d^\dagger} = \frac{\frac{3A^2(U)\varepsilon^2}{d^\dagger} + 2A(U)\varepsilon}{1 + \frac{d^*}{d^\dagger}}$$

However, $1/(1 + \frac{d^*}{d^\dagger}) \leq 1$ since $d^* \geq 0$, so that

$$\Delta d \leq \frac{3A^2(U)\varepsilon^2}{d^\dagger} + 2A(U)\varepsilon \leq \frac{3A^2(U)\varepsilon^2}{d^*} + 2A(U)\varepsilon. \quad \blacksquare$$

Proof of Theorem 2: Proceed by contradiction. Suppose $\min_{e \in E, x \in X_I} \|e - c(x)\| \neq \min_{e \in E, v \in V_I} \|e - c(v)\|$; in particular, since by Lemma 1 $X_I \subset V_I$, suppose that $d_I^* \equiv \min_{e \in E, x \in X_I} \|e - c(x)\| > \min_{e \in E, v \in V_I} \|e - c(v)\|$. Let $v^* = \arg\min_{v \in V_I}(\min_{e \in E} \|e - c(v)\|)$; then the previous supposition implies that $v^* \in V_I - X_I$. By Lemma 2, there exists an i^{th} stage ancestor $v_i^* \in V_i$ of v^*, for some $i < I$, such that $\bar{D}_i(v_i^*) \geq d_i^*$. But by Lemma 5, it follows that $\min_{e \in E} \|e - c(u)\| \geq d_i^* \quad \forall u \in S(v_i^*, V_i, U)$. Lemma 3 asserts that $v^* \in S(v_i^*, V_i, U)$; thus, in particular $\min_{e \in E} \|e - c(v^*)\| \geq d_i^*$. Finally, note that $d_i^* \geq d_I^*$; thus, $\min_{e \in E} \|e - c(v^*)\| \geq d_I^*$. This is a contradiction. $\blacksquare$

4 Complexity

The complexity of the optimization procedure is as follows. With no modification, the problem $\min_{e \in E, v \in V} \|e - c(v)\|$ can be solved with brute force substitution with complexity $O(M^N|V|)$ since $|E| = M^N$. (Note: it can be shown that for any compact set U, there exists a *finite* ε-cover V of that set; thus, the notation $|V|$ makes sense.) However, if the problem is solved as $\min_{v \in V}[\min_{e \in E} \|e - c(v)\|]$ and it is noted that

$$\min_{e \in E} \|e - c(v)\|^2 = \min_{e_1 \in E_1, \ldots, e_N \in E_N} \sum_{n=1}^{N} \|e_n - c_n(v)\|^2$$

$$= \sum_{n=1}^{N} \min_{e_n \in E_n} \|e_n - c_n(v)\|^2$$

then the complexity is reduced to $O(MN|V|)$ (since the latter step has a complexity of $O(MN)$). Further, using a result from computational geometry, it can be shown that each minimization of the form $\min_{e_n \in E_n} \|e_n - c_n(v)\|$ can be performed with $O(\log M)$ complexity, leading to an overall complexity of $O(N|V|\log M)$. (Note: in order to gain this $\log$ factor, it is necessary to incur $O(M \log M)$ in overhead to calculate the relevant Voronoi diagram; however, this is negligible in the scheme of things.) It is useful to convert the complexity $O(N|V|\log M)$ into an expression which depends on M, N, and Δd. Use a dimensional argument. Let V be an ε-covering of U; then using something akin to sphere-packing, it is clear that $vol(U) \approx |V|\varepsilon^\sigma$, where $\sigma = dim(U) = dim(C)$. That is, $|V| \propto \varepsilon^{-\sigma}$. Now, assuming that Δd is fairly small, then it can be shown that ε is fairly small, so that the upper bound on Δd from theorem 1 is proportional to ε (that is, the term in ε^2 drops out). In this case, the algorithm has complexity $O(N\Delta d^{-\sigma} \log M)$.

Thus far, only the results of theorem 1 have been employed. Incorporating theorem 2 allows for $|V_I|$ to be replaced by $|X_I|$, that is, for the complexity to be written $O(N|X_I|\log M)$. It is assumed that the latter will be *much* smaller than the former, although as yet no formal result to illustrate this has been achieved; at the very least, $|X_I| \leq |V_I|$ is known. In terms of the more relevant parameter Δd, it is hoped that a result may be proven to show that the complexity using the algorithm described in theorem 2 is of the form $O(N\Delta d^{-\beta} \log M)$, where $\beta < \sigma$; the difference between β and σ will depend heavily on the behaviour of the manifold $\tilde{C}$.

5. Results and Conclusions

An experiment is performed in which a moving finger is tracked. Clutter is in the form of both the background writing (much of which is small, and therefore leads to many extraneous edges) as well as the self-clutter of the doubled over finger. The motion of the finger illustrates two different kinds of tracking: flexing, which is a highly nonrigid type of motion, and translation. Furthermore, the motion is relatively fast (flexing takes just over half a second). The tracker sucessfully follows the finger for 202 frames, or 6.7 seconds. Results are shown in figure 1; a full video sequence can be viewed at website http://himmel.hrl.harvard.edu/daniel/research.html. In terms of speed, typical values for $|V_I|$ and $|X_I|$ were about 10^6 and 10^3 respectively; theorem 2 does indeed provide considerable speed-up in the complexity.

$\tilde{C}$ was learned in the following manner. (1) Each training curve, represented as the $2D$ coefficients in a pair of D^{th} order Legendre polynomial expansions (one each for x and y), was transformed into its euclidean-similarity invariant, represented in the same basis. (2) A one dimensional manifold was learned by smoothly interpolating through all of the invariants. This degree of freedom is captured in the variable u_1. (3) 4 extra dimensions were then added, corresponding to the group of euclidean similarity transformations: translation in both x- and y-directions, rotations, and scaling. These degrees of freedom are represented by u_2, u_3, u_4, u_5. Thus, both $\tilde{C}$ and C are five-dimensional manifolds. U is chosen to be $[0, 1]^5$ for convenience.

For purposes of comparison, a condensation tracker [1] was also run on the flexing part of the sequence. Training used the same sequence as the subset tracker, and a reduced dimensional space derived from applying the Karhunen Loeve transform (in this case, ten dimensions) was used as the space to learn the dynamical model for flexing. A comparison of the condensation algorithm with the subset tracker shown in figure 2. Although both trackers succeed in tracking for the entire 24 frame (0.8 second) sequence, the contours from the subset tracker are demonstrably clearer and crisper than those from the condensation tracker. Further insight into the different performance of the two trackers is obtained by examining a 20 frame (0.7 second) sequence in which the finger is almost completely still. As can be seen in figure 3, the subset tracker is successful in finding the static finger; the condensation tracker, by contrast, is much less successful. In particular, while the condensation tracker never entirely loses lock, it gives results which do not correspond very closely to the finger's true silhouette. This is due to the fact that the dynamical model used in condensation is learned for a flexing motion, and no "pause" motion is included in the training sequence. While it may be argued that a dynamical model could include both types of motion (indeed, possibly such a switching model could even be learned from a larger training sequence), this misses the point. The advantage of an algorithm which makes no use of dynamical models is *precisely* that the exact type of motion that will be encountered in a given application often cannot be anticipated. If all modes of motion could be entirely anticipated, then the tracking problem would be a much simpler one. In this case, one very simple non-learned

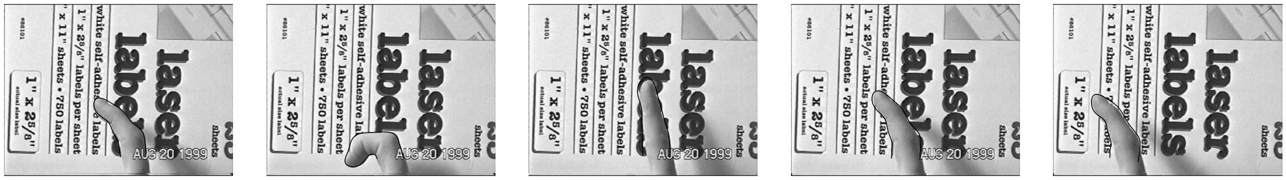

Figure 1. Tracking a flexing and translating finger.

Figure 2. Condensation tracker (above) vs. subset tracker (below).

mode, namely no motion at all, was tracked; the condensation tracker is unable to deal with this, while the subset tracker, which does not rely on dynamics, is successful.

Figure 3. Condensation tracker (above) vs. subset tracker (below).

There are two principal directions for future research. First, the development of an algorithm for learning a multidimensional manifold would be of benefit. In the current experiments, a one-dimensional manifold was learned in invariant space; however, it is quite likely that the true manifold was of higher dimension. (In this scenario, the one-dimensional manifold is simply a subset of the higher dimensional manifold.) Success in this area would also allow for more efficient implementation of the algorithm. Second, the algorithm may be extended to the task of object localization, in which an object is to be located within a single still image, rather than a video-sequence . Edge-search can no longer be initiated at the previous frame's contour estimate, as there is no such estimate; thus, in principle any edge-point in the image may be potentially part of the relevant curve. In other words, for each n, the set E_n of edge-points at the n^{th} site must include all edge-points detected in the image. Thus, $|E_n| = M$ is very much larger than in the case of tracking. The ability to search through the resulting huge space of observed curves $\tilde{E}$ relies on the $\log M$ term in the complexity, as discussed in section 4.

References

[1] A. Blake and M. Isard. Condensation - conditional density propagation for visual tracking. *Int. J. Comp. Vis.*, 29(1):5–28, 1998.

[2] W. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, San Diego, 2nd edition, 1986.

[3] D. Freedman and M. Brandstein. Methods of global optimization in the tracking of contours. In *Proceedings of the 33rd Asilomar Conference on Signals, Systems, and Computers*, 1999.

[4] D. Freedman and M. Brandstein. A subset approach to contour tracking in clutter. In *Proceedings ICCV*, pages 242–247, 1999.

[5] M. Isard and A. Blake. Contour tracking by stochastic propagation of conditional density. In *Proceedings ECCV*, pages 343–356, 1996.