3D Rotational Video Stabilization using Manifold Optimization

Chao Jia and Brian L. Evans

THE UNIVERSITY OF TEXAS AT AUSTIN

WNCG Wireless Networking & Communications Group

1. Problem

Video Stabilization

- Remove unwanted jitter in videos shot by handheld cameras

General Steps

Camera motion estimation → Motion smoothing → Frame synthesizing

Camera Motion Model: 3D rotational model

2. Special Orthogonal Group SO(3)

Special Orthogonal Group SO(3)

- All 3D rotation matrices lie on SO(3)
- An embedded submanifold in $\mathbb{R}^9$ (dimension = 3)

Geodesic distance

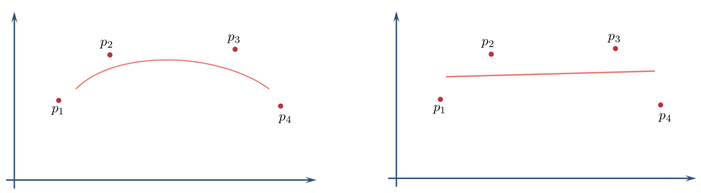
- The length of the shortest curve between two elements
- A natural extension from Euclidean distance

$$d_g(\mathbf{R}_m, \mathbf{R}_n) = (1/\sqrt{2})||\text{logm}(\mathbf{R}'_m\mathbf{R}_n)||_F$$

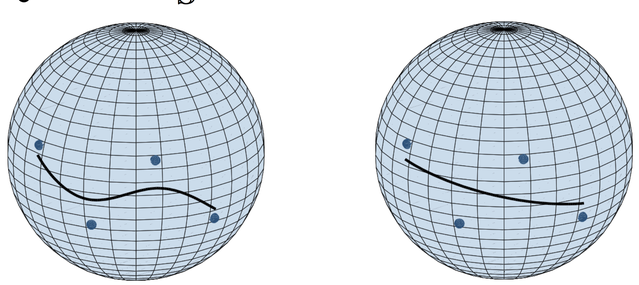
3. Motion Smoothing as Regression

A balance between fitting and smoothness

- Regression on $\mathbb{R}^2$



- Regression on $\mathbb{S}^2$



For video stabilization we need to apply regression on SO(3).

4. Problem Formulation

$$\min_{\{\mathbf{R}_n^{new}\}} \sum_{n=1}^{N} d_g^2(\mathbf{R}_n^{old}, \mathbf{R}_n^{new}) + \alpha \sum_{n=1}^{N-1} d_g^2(\mathbf{R}_n^{new}, \mathbf{R}_{n+1}^{new})$$

Penalty on misfit — *Penalty on 1st-order difference*

Optimization on the sequence of rotation matrices

- Solution set is still an embedded submanifold (dimension = 3N)

How to solve it?

- Constrained optimization on the embedded Euclidean space (*slow, non-linear, non-convex*)
- Unconstrained optimization on the manifold (*better numerical properties*)

5. Manifold Optimization

Gradient Related Algorithms

- Steepest gradient descent
- Newton's method

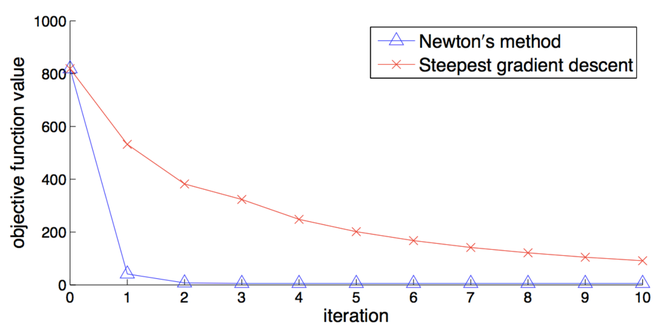
Similar convergence properties as in Euclidean space

Difficulties

- Computation of gradient and Hessian (omitted here)

Convergence example

- 478 frames; 2.93 sec/iteration for Newton's method
- Converge in only 2 iterations



6. Experimental Results

Test Videos

❶ Video shot by a walking forward person (no intentional motion)

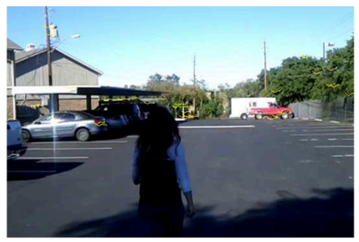
❷ Video shot while panning the camera (intentional motion)

Compared Methods

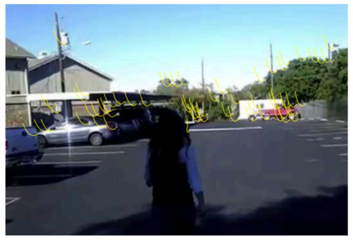
a) L_1 regression on 2D affine motion model (Euclidean space)

b) Local Low-pass filtering on SO(3)

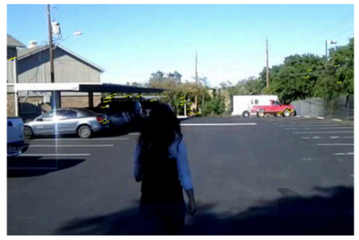
Original Video



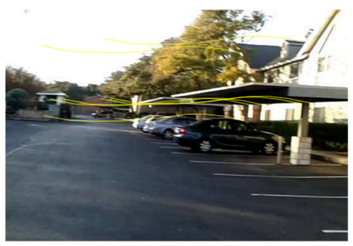
Result of (b)



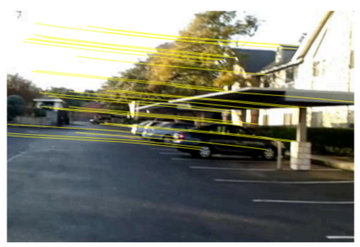
Result of (a)



Our result

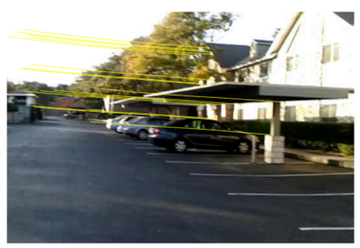


Original Video



Result of (b)

Result of (a)



Our result

- Research is supported in part by a gift from Texas Instruments in Dallas, Texas.
- Web page for the project is http://users.ece.utexas.edu/~bevans/projects/dsc/index.html