Hierarchical Representation of Videos with Spatio-Temporal Fibers

Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat

INRIA

March 25, 2014

Outline

1. Video Segmentation
2. Fibers
3. Segmentation Results

Video Volume : Stack of Successive Frames

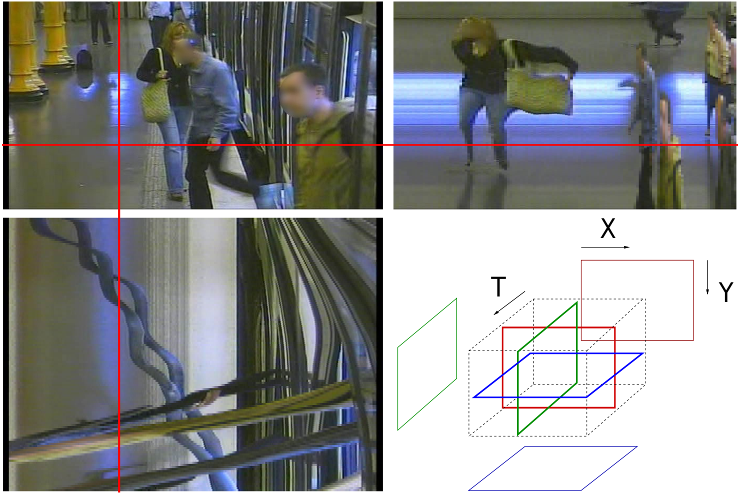


Figure: XT and YT slices from a Video.

Video Segmentation

Temporal coherency

Aim : Incorporate long term temporal coherency, along with spatial coherency.

A reliable video representation algorithm should have following traits :

- Region homogeneity.
- Pixelic precision on edges.
- No oversegmentation.
- Temporal coherency.
- Robustness and reliability.
- Efficient Computation.

Fibers

Fibers

Lines of homogeneous intensity across a Video Volume.

- Straight Fibers - Points of static background.
- Curved Fibers - Points on moving objects.

Aim

Segment video in terms of fibers to obtain dense segmentation with correspondences for all pixels.

SOA

SOA approaches and information provided by them.

Method	Tracks	Density	Reliability
[Brox 2010]	✓ (sparse)	×	×
[Grundmann 2010]	×	✓	×
[Tsai 2011]	×	✓	×
Ours	✓	✓	✓

Table: State-of-the-art approaches and the information provided by them.

Exact neighborhood

As opposed to computing rough bounding boxes, our approach provides exact neighborhood system using mesh.

Formalization

A **fiber F** $= \left(\{T_i\}_{i \in [1,m]} \,,\ \mathcal{M} \right)$ is a set of m trajectories T_i, spatially connected with a triangular mesh $\mathcal{M}$.

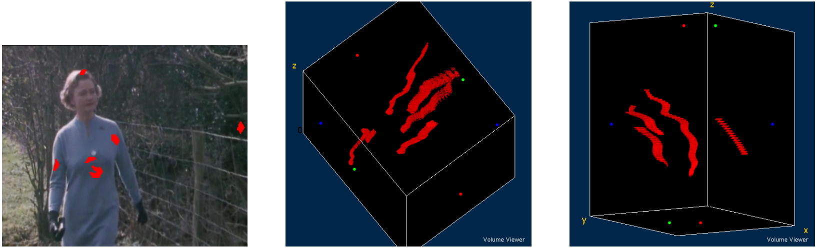
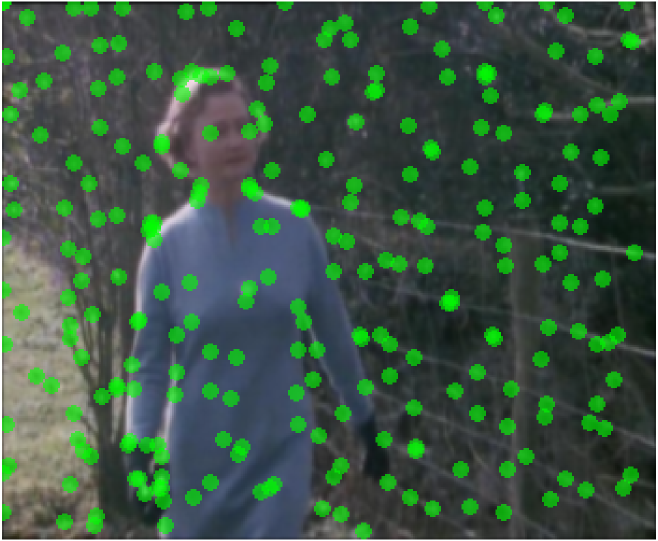


Figure: Curved Fibers in a Video Volume.

Curved Fiber Search

Compute Points of Interest in Images : Color Corners.



Curved Fiber Search

Lay a sub-pixellic mesh and segment this mesh into **multiple** regions.

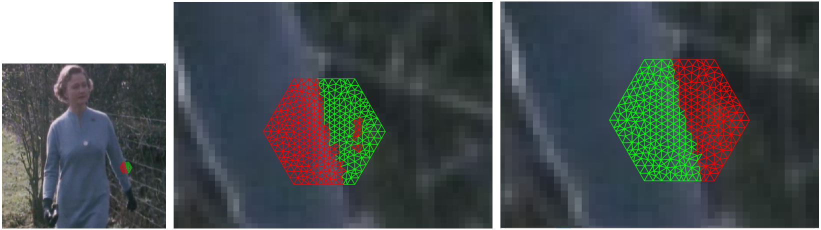


Figure: Segmentation of mesh into two regions

- Segmentation : K-Means + Graph Cuts.
- Correspondences of a **submesh vertices** across video volume determines a fiber.

Curved Fiber Traits

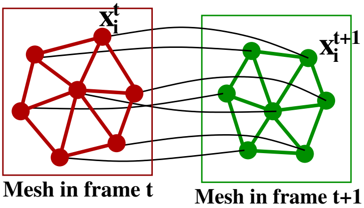


Figure: Temporal Correspondences of a mesh; Obtained from optical flow.

Inspect following traits

Color Match Error

$$\sum_{i \in \mathcal{M}} w_i \left\| I_t(\mathbf{x}_i^t) - I_{t+1}(\mathbf{x}_i^{t+1}) \right\|^2$$

Motion Variation

$$\frac{1}{|\mathcal{M}_t|} \int_{\mathcal{M}_t} \left\| \nabla_{\mathbf{x}} \mathbf{m}^t(\mathbf{x}) \right\|^2 d\mathbf{x} \simeq \sum_{i \sim j} w_{ij} \left\| \mathbf{m}_i^t - \mathbf{m}_j^t \right\|^2$$

Optional Flow Regularization

$$\inf_A \int_{\mathcal{M}} \left\| A\,\mathbf{x}(t) - \mathbf{x}(t+1) \right\|^2 d\mathbf{x}$$

Curved Fiber Traits (cont.)

- Split (both spatially and temporally) on the evidence of **incoherency** in optical flow (opf) : High $variance(opf)$ or $variation(opf)$.

Reliability/Rank of A Fiber

$$reliability \propto cornerness + temporal\ length \quad (1)$$

$$cornerness = exp(-\gamma / ||str(s)||_2) \quad (2)$$

$$str(s) = ||\lambda_1 - \lambda_2||_2 \quad (3)$$

λ_1 and λ_2 are eigenvalues of the 2D structure tensor.

Hierarchial Fiber Merging

- Two fibers are similar if they have similar average **speed**.
- Compute distance matrix for all fibers.

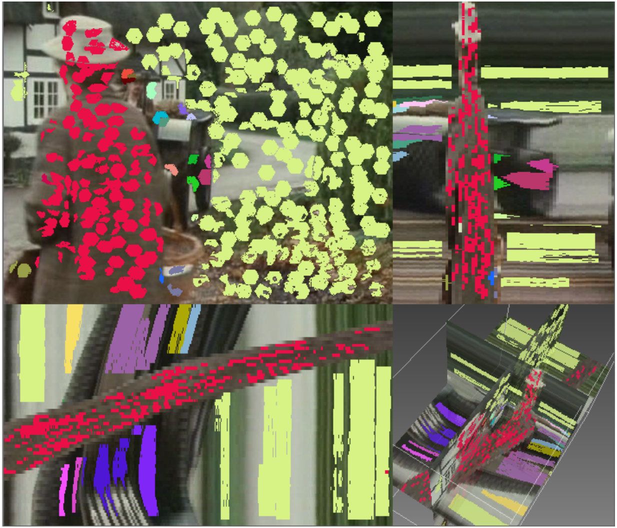
Distance between two fibers

$$d(F_i, F_j) = (1/O[F_i, F_j]) * \sum_{O[F_i,F_j]} ((x^i_f - x^i_{f+1}) - (x^j_f - x^j_{f+1}))^2 \quad (4)$$

$O[F_i, F_j]$: Overlap time span for two fibers F_i and F_j.

- Fuse fibers having distance lesser than τ.
- Increase τ with hierarchy.

Example : Fibers merged at a hierarchy



Full video coverage using fibers

Extend Fibers to homogeneous regions :

- Associate *w.r.t.* geodesics between fibers & rest of the video (Dijkstra).

Rationale

- Labels should be same if color and motion at *p* and *q* are similar.
- Motion is reliable only at corners/edges.

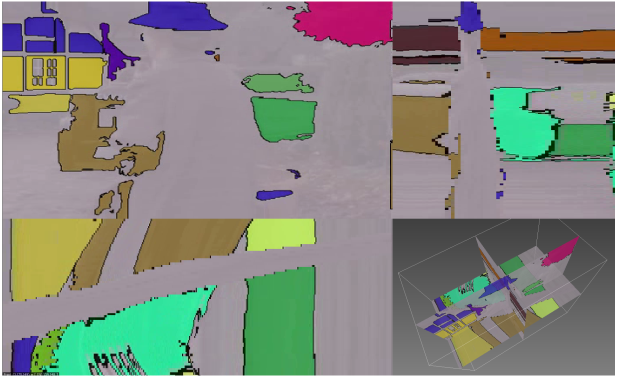
- Cost between pixel *p* and *q* is (λ : cornerness) :

$$\exp\left(\frac{-\alpha}{\|I(p)-I(q)\|_2}\right)+\lambda_{p,q}\exp\left(\frac{-\beta}{\|\mathbf{m}(p)-\mathbf{m}(q)\|_2}\right)$$

- *Influence Zone* of a fiber : Pixels associated to this fiber.

Leak Example

Local assessment of color and motion can cause **leaks**.



Fiber Association Assessment : Incorporating long term motion coherency

Fiber Association

To be a part of a fiber, trajectory of the pixel should be similar to the fiber.

- Trajectory of a pixel T_p : Trajectory of the closest fiber vertex.
- H_p (Color histogram of p) should be homogeneous.

Trajectory Coherency

Earth Movers Distance to quantify color homogeneity of a trajectory.

$$coh := \min_i \; EMD(H_p, H_{\delta_i}) \tag{5}$$

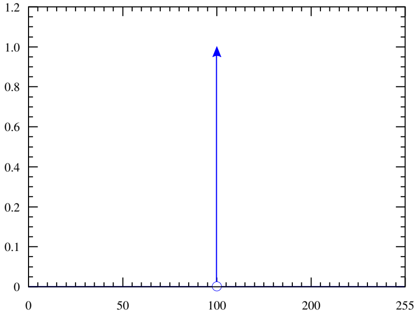


Figure: Histogram of a color homogeneous trajectory should be close to a Dirac Peak Histogram.

Fiber Footprints

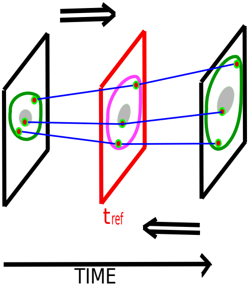


Figure: Reference Frame and Fiber Footprint (magenta).

For each fiber, project all pixels currently belonging to it onto a reference frame (using Trajectories associated with a pixel) : *fiber footprint*

Fiber Association Assessment

Figure: Fiber Footprint and Distance map (after temporal assessment).

- Use Dijkstra on fiber footprint with cost function from (6).

$$Cost(p) = \exp\left(\frac{-\alpha}{||coh||_2}\right) + \exp\left(\frac{-\beta}{||C(f) - C(p)||_2}\right) \quad (6)$$

Sample fiber footprints

Figure: Distance Map : Fiber association assessment based on (6).

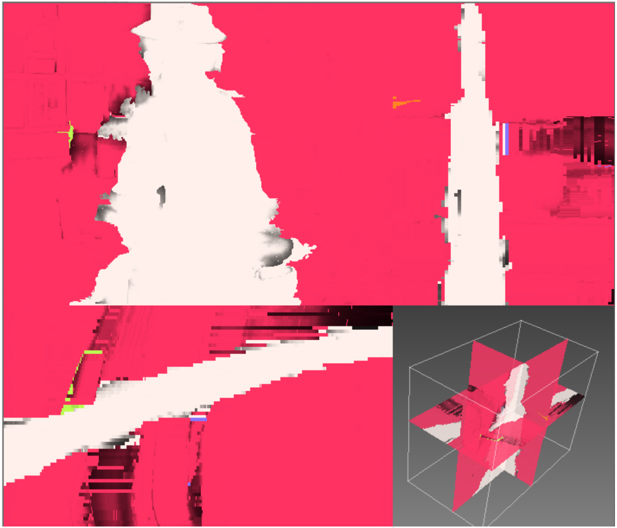
Computational Speed

Complexity

Information Propagation : $O(|V| log|\omega|)$

- Optical Flow for Curved Fibers : Werlberger'09 BMVC (1 to 3 seconds for 1024x768)

Output : Highest hierarchy



In-painting Demo

Only 7 mouse clicks (green dots in left image) needed for the full video to remove disturbing girl.

Figure: **Left**: Original Video. **Right** : Our Result **(no artifact!)**. Hole is filled by extending the straight fibers in time.

In-painting Videos

X-T

Future Work

- Implementation for streaming video segmentation.
- Application to activity recognition and video compression.
- Tackling low level issues such as shadows at fiber level.

Thanks

Poster Board # 10

See you at the Poster Board.

Thomas Brox and Jitendra Malik.
Object segmentation by long term analysis of point trajectories.
In ECCV. Springer, 2010.

Matthias Grundmann, V. Kwatra, Mei Han and Irfan Essa.
Efficient hierarchical graph-based video segmentation.
In CVPR, 2010.

David Tsai, Matthew Flagg, Atsushi Nakazawa and James M. Rehg.
Motion Coherent Tracking Using Multi-label MRF Optimization.
International Journal of Computer Vision, December 2011.