Title
Visual Localization and Mapping with Multiple View Features

Authors
Jason Meltzer
Rakesh Gupta
Ming-Hsuan Yang
et al.

Publication Date
2004

Center for Embedded Networked Sensing

Visual Localization and Mapping with Multiple View Features

Jason Meltzer[1], Rakesh Gupta[2], Ming-Hsuan Yang[2], Stefano Soatto[1]
[1]UCLA Vision Lab [2]Honda Research Institute

Introduction: How do we exploit motion for visual recognition and navigation?

Multiple view feature descriptors

- **Establish correspondence between different views of same scene**
  Useful for 3D reconstruction, stereo, recognition, and tracking.
- **Describe appearance of locations in space from multiple viewpoints**
  More robust to changes in ***illumination and viewpoint*** than descriptors derived from single images.
- **Kernel principal component analysis**
  Models changes of ***appearance*** of locations in space with changes in viewpoint and illumination.

Application to robot navigation

- **Structure from motion generates 3D map and egomotion**
  - Image ***feature tracker*** follows points on images.
  - ***Extended Kalman filter*** develops 3D structure and egomotion.
  - Multiple view descriptors generated at the same time.
- **Novel viewpoints of same scene recognized via feature descriptors and geometry.**

Proposed Solution: Incorporate multiple views!

• Local image variability is modeled by a group deformation of the domain and a non-linear combination of basis elements (kernel principal components).

$$f \circ I(x) = \sum_{i=1}^{k} a_i y_i(Ax + b)$$

• A ***kernel function*** computes dot products in the high dimensional space $\boldsymbol{F}$.

$$x \in \Omega \qquad f : \mathbf{R}^n \to \mathbf{R}^N \qquad y : \mathbf{R}^m \to \boldsymbol{F}$$

• ***Kernel principal component analysis*** is used to develop the descriptor.

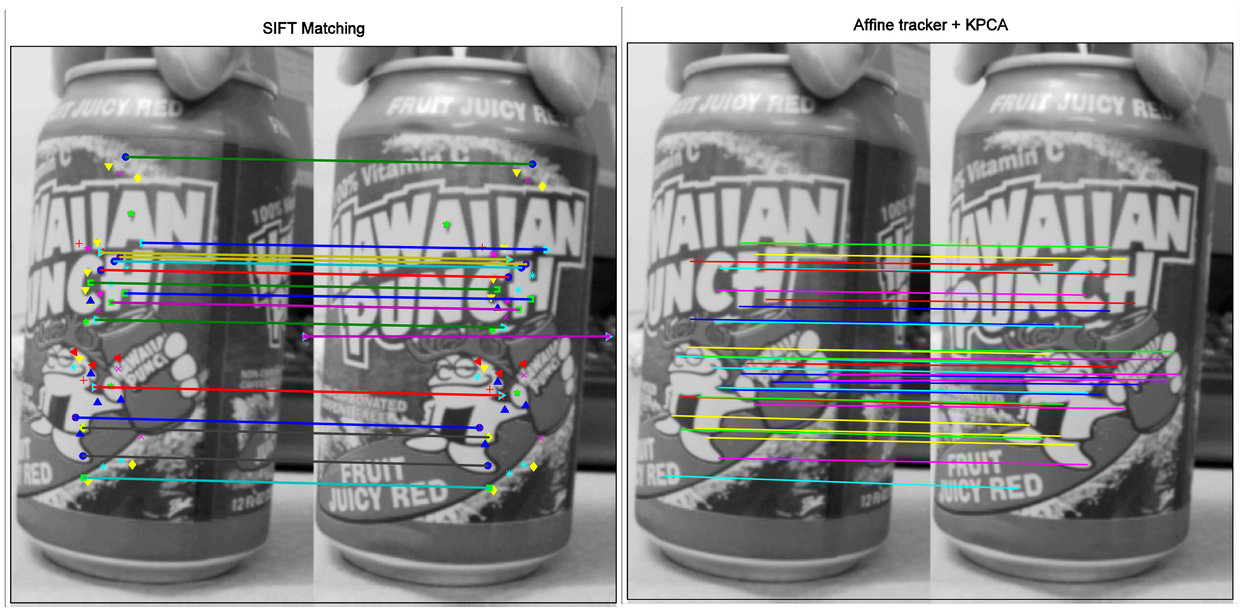
$$k(\mathbf{w}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{w} - \mathbf{y}\|^2}{2s^2} \right)$$

• Match quality is determined by the residual of the projection onto the principal components.

$$\left\| y(\mathbf{y}) - \sum_{i=1}^{k} \frac{\langle y(\mathbf{y}), j_i \rangle}{\langle j_i, j_i \rangle} \right\|^2$$

• Structure of a scene and motion of the camera are determined by an ***extended Kalman filter***. The descriptors are developed at the same time.

• When matching scenes is required, the descriptors are used to establish ***correspondence*** and the geometric transformation is recovered.



Existing approaches based on single-view descriptors fail to match under wide variations in shape (e.g. strongly non-planar scenes), pose, or illumination. Comparison with SIFT feature descriptor [Lowe et al., 2000] shown above.