Cellphone as a Perceptual Platform for Micro UAVs

Nikhil Naikal

Action Webs
Sept 15, 2010.

Perceptual Capabilities of Cell Phones

- Multiple tightly integrated sensors onboard.
  - Multiple cameras.
  - MEMS gyroscopes, accelerometers and digital compass.
  - Wireless and RF antennas.
  - Proximity and luminous intensity sensors.
  - Touch screen.
- Reasonable fast processing speeds and good memory.
- GPUs for parallel processing.
- Can potentially be used as main processing platform for small UAVs.

UAV Missions and Control

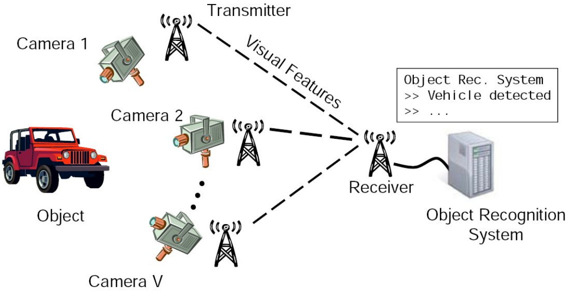
- US Army Description of most common missions performed by UAVs[1]:
  - Reconnaissance - Near real-time information about terrain, Search and rescue of friendly units, and disposition of possible enemy elements.
  - Surveillance - Area surveillance in friendly or enemy territory.
  - Situational Awareness - Provide commanders with situational awareness and mission planning information.
  - Security - Reaction time and maneuver space for the main body and area security.
  - Targeting - Target acquisition, target detection and recognition, target designation and illumination.
  - Communication Support - Voice and data communications retransmission.
  - Movement support - Convoy security, mines/IED detection.
- UAV Control
  - Currently 6 human operators for 1 UAV (Predator, Global Hawk, etc.)
  - Expert pilots for remote controller.

[1]US Army UAV field manual 2009, http://www.fas.org/irp/doddir/army/fmi3-04-155.pdf

Computer Vision Algorithms for UAV Missions and Control

- Computer vision research directions:
  - Object detection/recognition.
  - Image/video segmentation.
  - 3D reconstruction/mosaicing.
  - Object tracking.
- Can computer vision algorithms be used for aiding in UAV missions?
  - Reconnaissance - Object detection/recognition, 3-D reconstruction.
  - Surveillance - Object detection/recognition.
  - etc.
- Can computer vision algorithms aid untrained personnel to control micro UAVs?
  - Control with commands such as, "follow road", "fly until objective reached".
  - Abstracting autonomous back-end from front-end human interface.

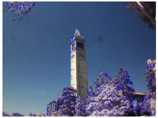
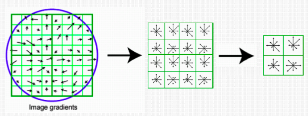
Multi-View Object Recognition



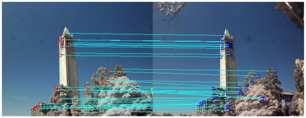
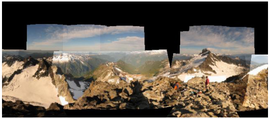
- Low cost cameras integrated with mobile platforms easily deployed.
  - Inter camera calibration usually not possible.
- Need to leverage multiple observations of objects from different vantage points.
- Problem Statement: I focus on recognition of common object over band limited communication channel.

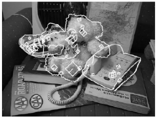
Object Recognition - Overview

- Affine invariant features such as SIFT [Lowe 2002], SURF [Bay 2006], CHoG [Chandrasekhar 2009]

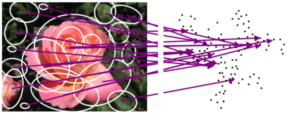
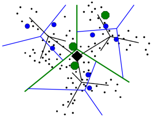




- Feature matching robust in harsh environments; popular for variety of applications.

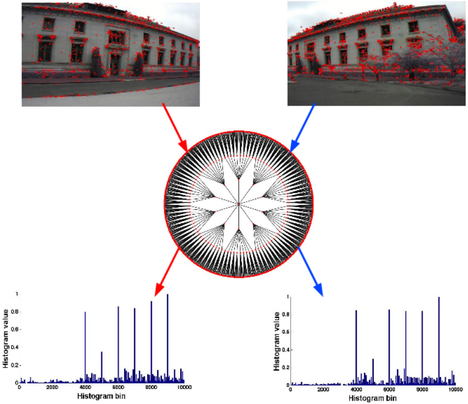


(a) Autostitch


(b) Recognition

- Scalable recognition with vocabulary tree [Nister 2006]

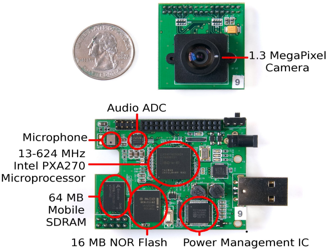
Visual Histograms



- Vocabulary tree constructed offline.
- All histograms are **nonnegative** and **sparse**.
- Multiple-view histograms share **joint sparse patterns**.
- Classification is based on a similarity measure.

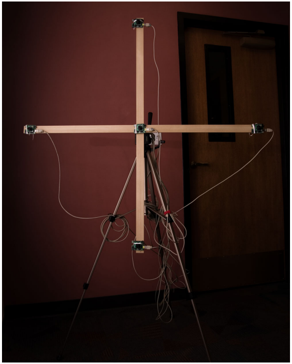
CITRIC: Wireless Smart Camera Platform

- CITRIC platform [Chen 2008]



- Available library functions
  1. Full support **Intel IPP Library** and **OpenCV**.
  2. **JPEG compression**: 10 fps.
  3. **Edge detector**: 3 fps.
  4. **Background Subtraction**: 5 fps.
  5. **SURF detector**: 10 fps.

Berkeley Multiple-view Wireless Database

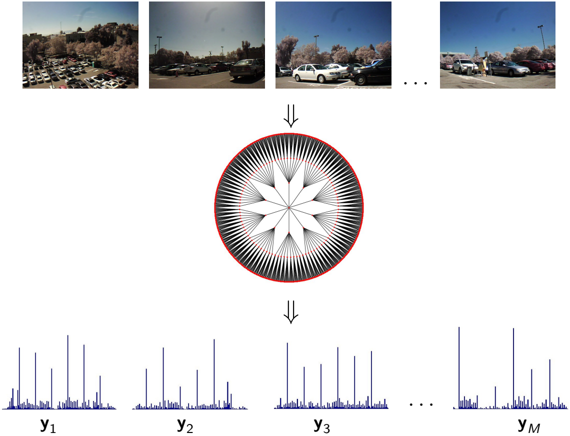


(a) Campanile: Small Baseline

(b) Campanile: Large Baseline

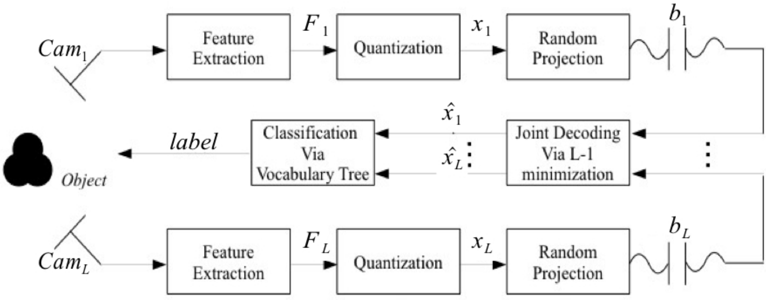
- 20 landmarks at UC Berkeley.
- 16 different vantage points (large baseline); five images at one location (small baseline).
- Low-quality camera images: resolution, focal length, dusty lenses.

Training Phase



- For each object category, $i = 1...C$, multiple histograms generated for all $j = 1...M$ training images, $Y_i = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_M\}$.
- All C subsets form training set, $Y = \{Y_1, Y_2, ..., Y_C\}$.

System Pipeline

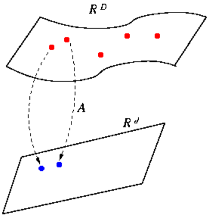


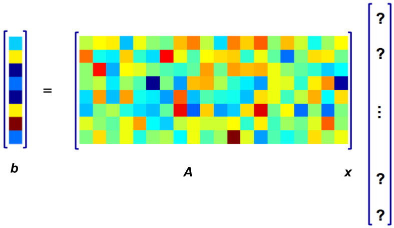
- Invariant features extracted onboard.
- Visual histogram computed for image using stored vocabulary tree, and transmitted wirelessly.
- Functions on sensor largely stabilized, thereby facilitating deployment.
- Computationally heavy functions performed by the server, and can be updated.

Random Projection to Compress Histograms

$$\mathbf{b} = A\mathbf{x}$$

Coefficients of $A \in \mathbb{R}^{d \times D}$ are drawn from zero-mean Gaussian distribution.





- Advantages of Random Projection
  1. Easy to generate and update
  2. Does not need training prior; (universal dimensionality reduction).
  3. Faster recognition speed.

Decoding via ℓ_1-Minimization

Noiseless case

Assume $\mathbf{x}$ is sufficiently k-sparse. Given triplet (D, d, k) and random A with $d > \delta(A)$ for some threshold δ, solving

$$(P_1): \quad \min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{b} = A\mathbf{x}$$

recovers the unique solution.

Noisy case

Assuming Gaussian measurement errors in $\mathbf{b}$ with bound ϵ, the solution to the convex program,

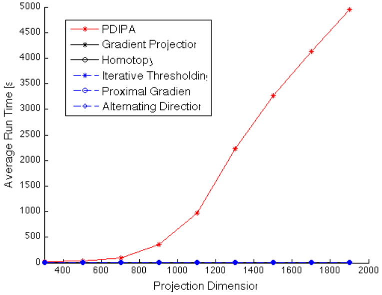
$$(P_{1,2}): \quad \min \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{e}\| = \|\mathbf{b} - A\mathbf{x}\|_2 < \epsilon$$

recovers the sparsest solution.

Compressive sensing theory shows that under broad conditions, the estimates from P_1 and $P_{1,2}$ are the sparsest solution.

Why ℓ_1-Minimization is still a difficult problem?

- General toolboxes do exist: **cvx**, **SparseLab**.
  However, interior-point methods are **very** expensive in HD space.



- Data Noise **and** Corruption

$$\mathbf{b} = A\mathbf{x} + \mathbf{e}, \quad \text{where } \|\mathbf{e}\|_2 \text{ may not be bounded!}$$

- Special structure of the data (from domain-specific knowledge)

ℓ^1-Minimization using Iterative Soft-Thresholding (IST) [Donoho 1995]

Lagrangian method

$$\begin{aligned} \mathbf{x}^* = \arg\min F(\mathbf{x}) &= \arg\min \frac{1}{2}\|\mathbf{b} - A\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \\ &\doteq \arg\min f(\mathbf{x}) + \lambda g(\mathbf{x}) \end{aligned}$$

- IST iteratively approximates the composite objective function

$$\begin{aligned} \mathbf{x}^{(k+1)} &\approx \arg\min_{\mathbf{x}}\{f(\mathbf{x}^{(k)}) + (\mathbf{x} - \mathbf{x}^{(k)})^T\nabla f(\mathbf{x}^{(k)}) + \frac{\nabla^2 f(\mathbf{x}^{(k)})}{2}\|\mathbf{x} - \mathbf{x}^{(k)}\|_2^2 + \lambda g(\mathbf{x})\} \\ &= \arg\min_{\mathbf{x}}\{(\mathbf{x} - \mathbf{x}^{(k)})^T\nabla f(\mathbf{x}^{(k)}) + \frac{\alpha^{(k)}}{2}\|\mathbf{x} - \mathbf{x}^{(k)}\|_2^2 + \lambda g(\mathbf{x})\} \end{aligned}$$

  where the hessian $\nabla^2 f(\mathbf{x})$ is approximated by a diagonal matrix αI.
- Denote $\mathbf{u}^{(k)} = \mathbf{x}^{(k)} - \frac{1}{\alpha^{(k)}}\nabla f(\mathbf{x}^{(k)})$, then

$$\mathbf{x}^{(k+1)} \approx \arg\min_{\mathbf{x}}\{\frac{1}{2}\|\mathbf{x} - \mathbf{u}^{(k)}\|_2^2 + \frac{\lambda}{\alpha^{(k)}}g(\mathbf{x})\}.$$

- When $g(\mathbf{x}) = \|\mathbf{x}\|_1$, a closed-form solution exists *element-wise*

$$x_i^{(k+1)} = \arg\min_{\mathbf{x}_i}\{\frac{(x_i - u_i^{(k)})^2}{2} + \frac{\lambda|x_i|}{\alpha^{(k)}}\} = \text{soft}(u_i^{(k)}, \frac{\lambda}{\alpha^{(k)}})$$

More References

1. **Primal-Dual Interior-Point** Methods
   - Log-Barrier [Frisch 1955, Karmarkar 1984, Megiddo 1989, Monteiro-Adler 1989, Kojima-Megiddo-Mizuno 1993]
2. **Homotopy** Methods:
   - Homotopy [Osborne-Presnell-Turlach 2000, Malioutov-Cetin-Willsky 2005, Donoho-Tsaig 2006]
   - Polytope Faces Pursuit (PFP) [Plumbley 2006]
   - Least Angle Regression (LARS) [Efron-Hastie-Johnstone-Tibshirani 2004]
3. **Gradient Projection** Methods
   - Gradient Projection Sparse Representation (GPSR) [Figueiredo-Nowak-Wright 2007]
   - Truncated Newton Interior-Point Method (TNIPM) [Kim-Koh-Lustig-Boyd-Gorinevsky 2007]
4. **Iterative Thresholding** Methods
   - Soft Thresholding [Donoho 1995]
   - Sparse Reconstruction by Separable Approximation (SpaRSA) [Wright-Nowak-Figueiredo 2008]
5. **Proximal Gradient** Methods [Nesterov 1983, Nesterov 2007]
   - FISTA [Beck-Teboulle 2009]
   - Nesterov's Method (NESTA) [Becker-Bobin-Candés 2009]
6. **Alternating Direction** Methods [Yang-Zhang 2009, Figueiredo-Bioucas-Dias 2010]
   - YALL1 [Yang-Zhang 2009]

References:
Yang, *et al.*, *Fast ℓ_1-minimization algorithms and an application in robust face recognition*. Preprint, 2010.
http://www.eecs.berkeley.edu/~yang/software/l1benchmark/

Joint Decoding

- Multi-view scenario gives rise to Sparse Innovation Model (SIM):

$$\begin{aligned} \mathbf{x}_1 &= \tilde{\mathbf{x}} + \mathbf{z}_1, \\ &\vdots \\ \mathbf{x}_L &= \tilde{\mathbf{x}} + \mathbf{z}_L. \end{aligned}$$

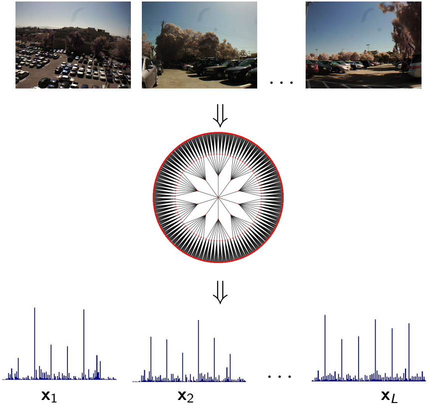
  $\tilde{\mathbf{x}}$ is called the **joint sparse** component, and $\mathbf{z}_i$ is called an **innovation**.

- Joint recovery of SIM

$$\begin{aligned} \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_L \end{bmatrix} &= \begin{bmatrix} A_1 & A_1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ A_L & 0 & \cdots & 0 & A_L \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}} \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_L \end{bmatrix} \\ \Leftrightarrow \quad \mathbf{b}' &= A'\mathbf{x}' \in \mathbb{R}^{dL}. \end{aligned}$$

- Joint sparsity $\tilde{\mathbf{x}}$ is automatically determined by ℓ^1-Minimization.

Multi-View Classification



- Multi-view relevance score assigned to query category as,

$$m(X, Y_i) = \text{median}_{\mathbf{x}_k \in X}\, s(\mathbf{x}_k, Y_i),$$

where,

$$s(\mathbf{x}_k, Y_i) = \min_{\mathbf{y}_j \in Y_i} \left\| \frac{\mathbf{x}_k}{\|\mathbf{x}_k\|_1} - \frac{\mathbf{y}_{ij}}{\|\mathbf{y}_{ij}\|_1} \right\|_1.$$

Small Baseline Experiments

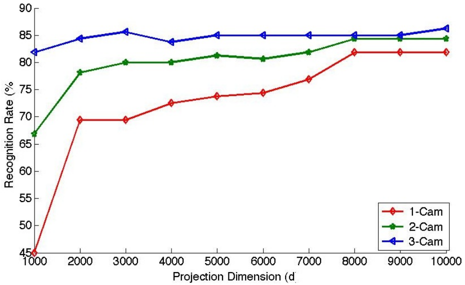


Table: Small-baseline recognition rates without histogram compression. The best rates are marked in bold face.

Expt.	# Train Images	# Test Images	SIFT Rate(%)	SURF Rate(%)	CHoG Rate(%)
1 Cam	160	160	71.25	80.62	**81.88**
2 Cam	160	320	72.5	81.25	**84.38**
3 Cam	160	480	73.75	81.88	**86.25**

Large Baseline Experiments

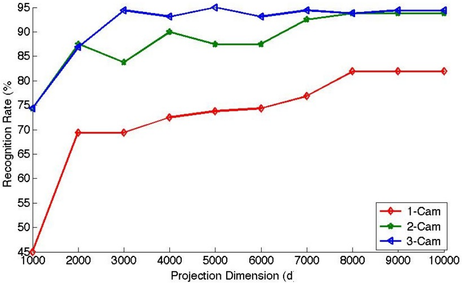


Table: Large-baseline recognition rates without histogram compression. The best rates are marked in bold face.

Expt.	# Train Images	# Test Images	SIFT Rate(%)	SURF Rate(%)	CHoG Rate(%)
1 Cam	160	160	71.25	80.62	**81.88**
2 Cam	160	320	76.88	88.13	**93.75**
3 Cam	160	480	83.13	90.00	**94.88**

Distributed Object Recognition in Band-Limited Smart Camera Networks

1. To harness the smart camera capacity, the system is separated in two components: **distributed feature extraction** and **centralized recognition**.

2. Multiple view information boosts recognition rates.

3. Drawn from Compressive Sensing theory to formulate distributed codec scheme.

4. Wireless cameras need not be calibrated. Further, system flexible to addition/omission of cameras and mobile platforms.

Future work (near future)

- Extending to video sequences.
  - Scenario: Car broken down in mountains, needs to be found. UAV on "detection" mode, to find car.
  - Multiple images obtained from video stream.
  - Signal has slowly varying sparse support.
  - Developing mathematical methods to speed up ℓ^1-minimization for time varying sparse signal.
- Multiple camera images to recover 2.5-D or 3-D maps.
  - Sparse support represents common features between multiple images.
  - Using structure from motion methods to recover 3-D representations.
- Identifying "good" features in the training process using geometric relationships between training images.
  - Developing methods to identify strong visual features in the training process.
  - This will potentially make visual histograms more sparse.