Information Planning and Active Data Collection

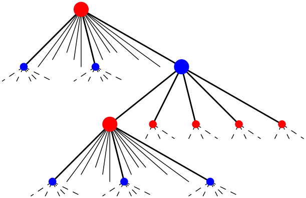
John Fisher

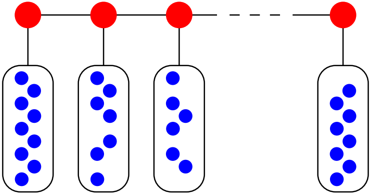
Sensing, Learning, & Inference Group
Computer Science & Artificial Intelligence Laboratory
Massachusetts Institute of Technology
http://groups.csail.mit.edu/vision/sli/

November 20, 2014

Efficient Information Planning

Tractable greedy selection achieves near-optimal performance.





Williams et al. [2007b] reduces complexity of information gathering formulated as a Markov Decision Process.

$$O([N_s 2^{N_s}]^N M^N) \rightarrow O(N N_s^3)$$

Williams et al. [2007a] the optimal information gathering rate is no greater than twice the greedy information gathering rate.

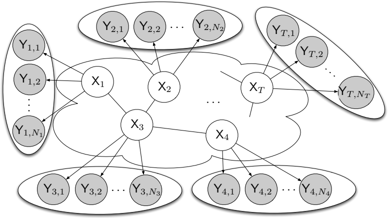
$$\frac{I\left(X;Z_N^G\right)}{I\left(X;Z_N^*\right)} \geq \frac{1}{2} \quad \forall N$$

N_s: number of sensing actions, N: planning horizon, M: measurement simulation cost.

Sequential Information Planning

- N_t measurements for each X_t, $\mathscr{V}_t = \{1, \ldots, N_t\}$, $t \in \{1, \ldots, T\}$.
- Goal

$$\mathscr{O} \in \operatorname*{arg\,max}_{|\mathscr{S}_1| \le k_1, \ldots |\mathscr{S}_T| \le k_T} f(\mathscr{S})$$



$$\mathscr{O}\left(\binom{N}{k}^T\right) \le \mathscr{O}(N^{kT}) \text{ when } N_t = N, k_t = k \forall t$$

On the Oracle Assumption

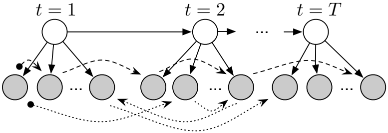
- Most of the guarantees on greedy selection assume an oracle model, *i.e.*,the complexity of reward evaluation has constant time [Nemhauser et al., 1978], [Guestrin et al., 2005], [Krause et al., 2005], [Kempe et al., 2003], [Calinescu et al., 2007], [Streeter et al., 2009].
- This generally does not hold, particularly for sequential problems (*aka* almost all problems of interest).
- Here we show
  - One can exploit sparsity in the latent variable structure and selection order reduce complexity.
  - The same reasoning leads to an efficient incremental approach inference in trees and poly-trees (with extensions to loopy graphs utilizing feedback vertex set graph decompositions).

Gaussian HMMs

Consider the Gaussian HMM governed by the dynamics:

$$X_t = A_{t-1}X_{t-1} + V_{t-1}$$
$$Y_t = C_t X_t + W_t,$$

where A_t, C_t highly sparse.



d: dimension of X_t, T: number of hidden variables, N_t: number of observations per set, m: dimension of each observation $Y_{t,u}$, k_t: number of constraints per set $(d, N_t \gg m, k_t)$

	Operation	Complexity
Propagation	$\Sigma_{t\|t-1} = A_{t-1}\Sigma_{t-1\|t-1}A_{t-1}^T + Q_{t-1}$	$\mathscr{O}(d^3)$
Update	$\Sigma_{t\|t} = \Sigma_{t\|t-1} - G_t C_t \Sigma_{t\|t-1}$	$\mathscr{O}(md^2)$
	$G_t = \Sigma_{t\|t-1} C_t^T (C_t \Sigma_{t\|t-1} C_t^T + R_t)^{-1}$	

Exploiting Sparsity

- **Update** - sparsity combined with the information form yields efficient updates (with some bookkeeping).

$$J_{w_j|u\cup\mathscr{G}_{j-1}} = J_{w_j|\mathscr{G}_{j-1}} + C_{w_j,u}^T R_{w_j,u}^{-1} C_{w_j,u}$$
$$= \begin{bmatrix} J_{w_j|\mathscr{G}_{j-1}}(I_u,I_u) & J_{w_j|\mathscr{G}_{j-1}}(I_u,\neg I_u) \\ J_{w_j|\mathscr{G}_{j-1}}(\neg I_u,I_u) & J_{w_j|\mathscr{G}_{j-1}}(\neg I_u,\neg I_u) \end{bmatrix} + \begin{bmatrix} C_{w_j,u}(I_u)^T R_{w_j,u}^{-1} C_{w_j,u}(I_u) & \mathbf{0}_{q\times(d-q)} \\ \mathbf{0}_{(d-q)\times q} & \mathbf{0}_{(d-q)\times(d-q)} \end{bmatrix},$$

  where I_u, $|I_u| = q \ll d$.

- **Exploration.** Choose g_j as

$$J_{w_j|u\cup\mathscr{G}_{j-1}} = J_{w_j|\mathscr{G}_{j-1}} + C_{w_j,u}^T R_{w_j,u}^{-1} C_{w_j,u} = J_{w_j|\mathscr{G}_{j-1}} + \begin{bmatrix} \hat{C}_{w_j,u}^T \\ \mathbf{0}_{(d-q)\times m} \end{bmatrix} \begin{bmatrix} \hat{C}_{w_j,u} & \mathbf{0}_{m\times(d-q)} \end{bmatrix},$$

where $\hat{C}_{w_j,u} = R_{w_j,u}^{-1/2} C_{w_j,u}(I_u)$.

Operation	Speedup/measurement
Update	$\mathscr{O}\left(\frac{d^2}{\max\{m,q\}^2}\right)$
Exploration	$\mathscr{O}\left(\frac{d^3}{m\max\{m,q\}^2}\right)$

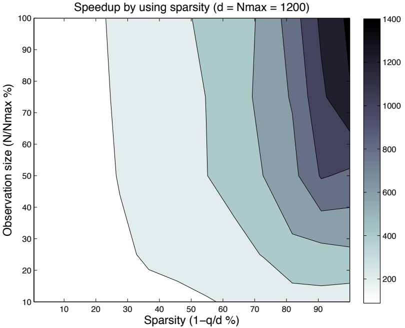
Experiment

Two hundred moving objects with different degrees of correlation.

$$X_t = A_{t-1}X_{t-1} + V_{t-1}, \forall t \in \{1,\ldots,20\}$$
$$Y_t = C_t X_t + W_t \ ,$$

where $X_t = [\ p^{1:200}_{t,(x,y,z)} \quad v^{1:200}_{t,(x,y,z)}\ ]^T$, $V_{t-1} \sim \mathcal{N}(0, Q_{t-1})$ driving and $W_t \sim \mathcal{N}(0, R_t)$ measurement noise. We of the hidden dimension and different degrees of sparsity in the measurement model.

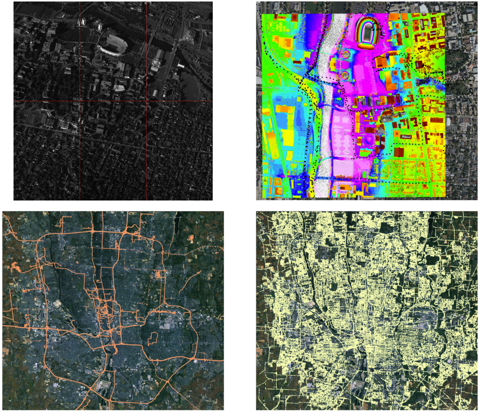
Speedup grows with observation size and sparsity.



A measurement is available for each latent variable (position, velocity). We have $N^{\max} = 1200$ measurements in total, while we consider different observation sizes, $\{10\%, 25\%, 50\%, 75\%, 100\%\}$ of $N^{\max}$. We select $k_t = 6$ measurements from each set.

Multi-Modal Fusion

3D reconstruction using multiple data sources



Multiple Data Sources

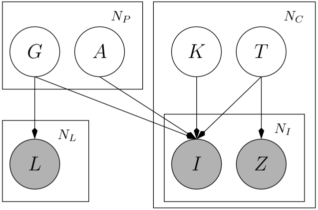
- Full motion video (FMV)
- FMV Platform GPS/INS
- LIDAR
- Open Street Map GPS
- Open Street Map Waypoints

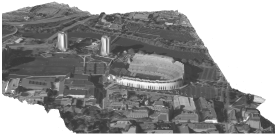
- All of these provide complementary information about the scene.
- Q. How do we combine them in a coherent way?

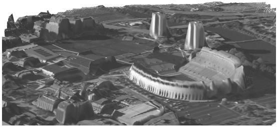
Multi-Modal Fusion

3D reconstruction using multiple data sources

- Challenge: We need a model for integration.
- Approach: Formulate as inference in a graphical model.
- Reality: Handles some aspects really well, others require new algorithmic and theoretical developments.







- Data Integration versus Queries are separated in such models.
- Uncertainty is explicitly represented.

Multi-Modal Fusion

3D reconstruction using multiple data sources

Multi-Modal Fusion

3D reconstruction using multiple data sources

- However, construction of such a model is often merely an intermediate step to more complex reasoning.

Multi-Modal Fusion

3D reconstruction using multiple data sources

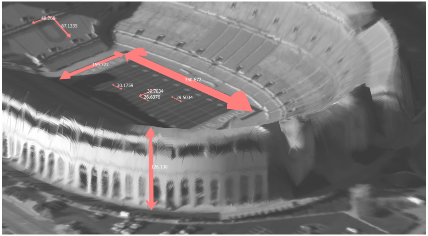
Multi-Modal Fusion

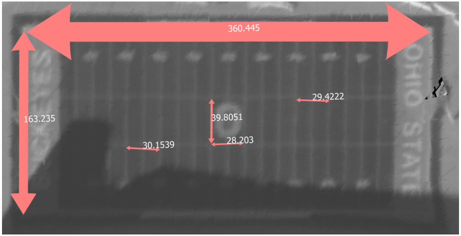
3D reconstruction using multiple data sources

- We can reason over additional content of the scene.

Multi-Modal Fusion

3D reconstruction using multiple data sources

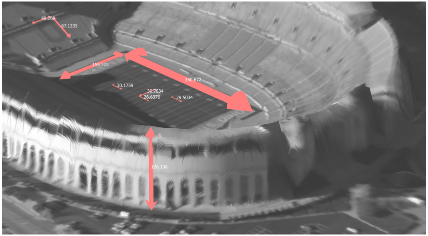


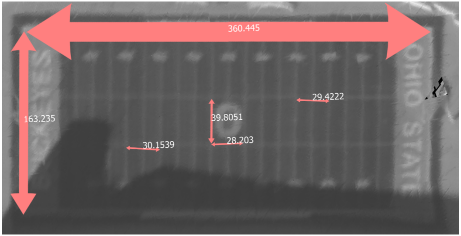


- Mensuration: we can measure physical dimensions.

Multi-Modal Fusion

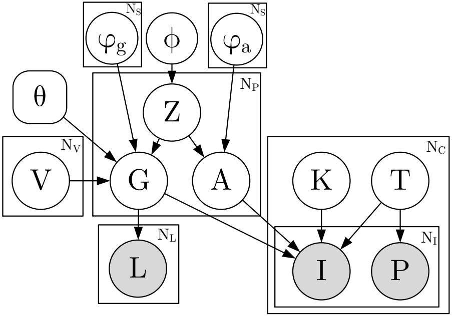
3D reconstruction using multiple data sources





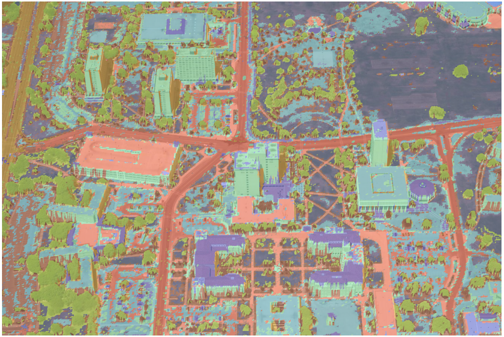
- Mensuration: we can measure physical dimensions.
- In short, this is an intermediate step to higher-level reasoning, *i.e.*,asking questions about the scene.

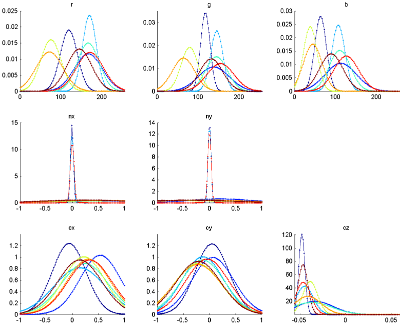
Adding Contextual Variables

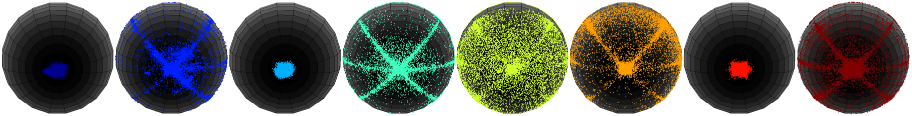


- Context is a loaded term.
- In this setting, it represents learned local and global priors on appearance and geometry.
- Can be shown to reduce the VoI of measurements (this is expected).
- Questions:
  - Is there context to exploit?
  - Can we learn it from data?
  - Can integrate VoI analysis that trades off measurement VoI versus contextual VoI?

Categories of Surface Normals

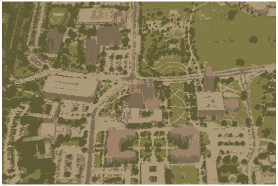




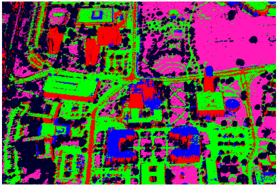


Categories of Appearance

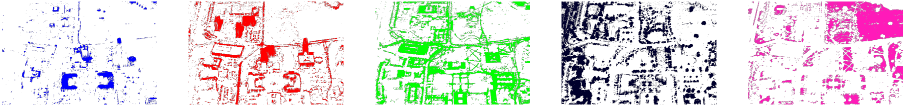
Original



Cluster Mean



Labels (K=5)



References I

P. L. Bartlett, M. I. Jordan, and J. D. Mcauliffe. Convexity, classification, and risk bounds. Journal of the American Statistical Association, 2003.

M. Basseville. Distance measures for signal processing and pattern recognition. Signal Processing, 18(4):349–369, Dec 1989.

J. M. Bernardo. Expected information as expected utility. The Annals of Statistics, 7 (3):686–690, may 1979. ISSN 0090-5364. URL http://links.jstor.org/sici?sici=0090-5364(197905)7:3<686:EIAEU>2.0.CO;2-#.

A. W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. Computer VisionECCV98, 1:311–326, 1998. doi: 10.1007/BFb0055655. URL http://www.springerlink.com/index/D334346K230720X5.pdf.

K. Ikeuchi and B. K. P. Horn. Shape from shading. chapter Numerical shape from shading and occluding boundaries, pages 245–299. MIT Press, Cambridge, MA, USA, 1989. ISBN 0-262-08183-0. URL http://dl.acm.org/citation.cfm?id=93871.93880.

References II

A. Krause and C. Guestrin. Near-optimal nonmyopic value of information in graphical models. In Uncertainty in Artificial Intelligence, July 2005.

D. V. Lindley. On a measure of the information provided by an experiment. The Annals of Mathematical Statistics, 27(4):986–1005, dec 1956. ISSN 0003-4851. URL http://links.jstor.org/sici?sici=0003-4851%28195612%2927%3A4%3C986%3AOAMOTI%3E2.0.CO%3B2-9.

X. Nguyen, M. J. Wainwright, and M. I. Jordan. On surrogate loss functions and f-divergences. Annals of Statistics, 2009.

G. Papachristoudis and J. W. Fisher III. Theoretical guarantees on penalized information gathering. In Proc. IEEE Workshop on Statistical Signal Processing, August 2012. URL publications/papers/Papachristoudis12SSPWorkshop.pdf. **.

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. International Journal of Computer Vision, 9(2):137–154, 1992. ISSN 09205691. doi: 10.1007/BF00129684. URL http://www.springerlink.com/index/10.1007/BF00129684.

References III

J. L. Williams, J. W. Fisher III, and A. S. Willsky. Performance guarantees for information theoretic active inference. In M. Meila and X. Shen, editors, Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics, pages 616–623, March 2007a. URL `publications/papers/WilFis07AISTATS.pdf`.

J. L. Williams, J. W. Fisher III, and A. S. Willsky. Approximate dynamic programming for communication-constrained sensor network management. IEEE Transactions on Signal Processing, 55(8):3995–4003, August 2007b. URL `publications/papers/williams07a.pdf`.

F. Zhao, J. Shin, and J. Reich. Information-driven dynamic sensor collaboration. Signal Processing Magazine, IEEE, 19(2):61 –72, mar 2002. ISSN 1053-5888. doi: 10.1109/79.985685.

**Outgrowth of Supervised Student Research