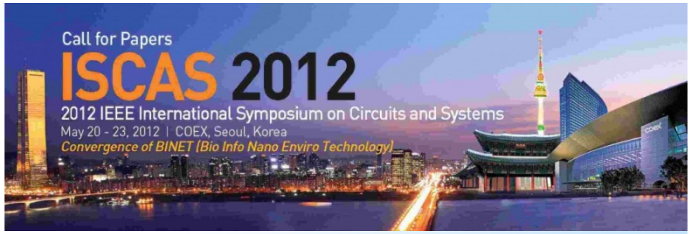


Sparsity-based Online Missing Sensor Data Recovery

Di Guo

Xiamen University, China

Xiaobo Qu, Lianfen Huang, Yan Yao, Zicheng Liu, Ming-Ting Sun

May 22, 2012

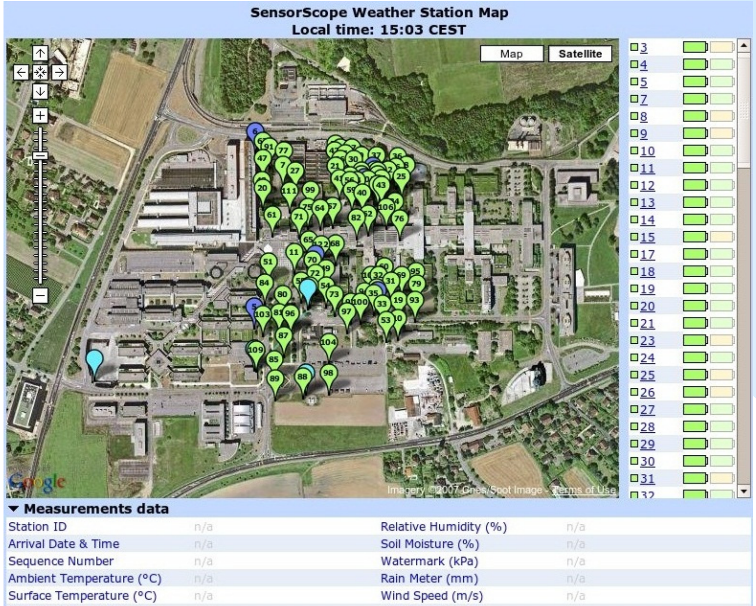
Contents

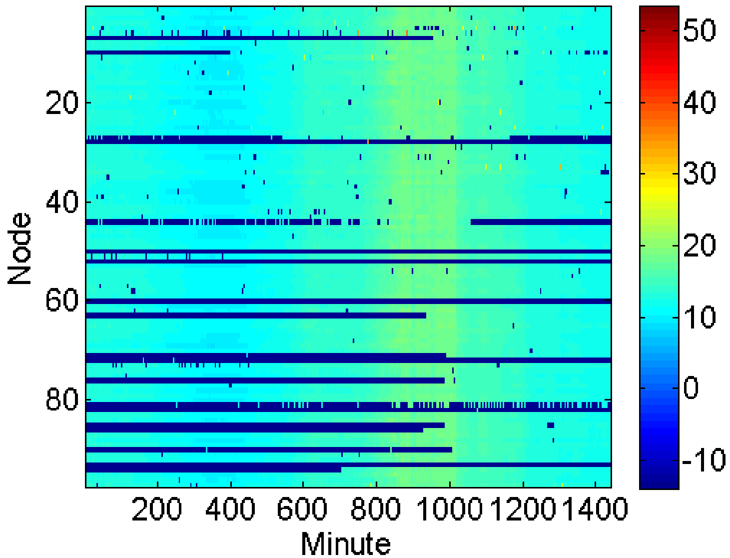
Wireless Sensor Networks

Applications: environment sensing, building, agricultural surveillance, medical care, military

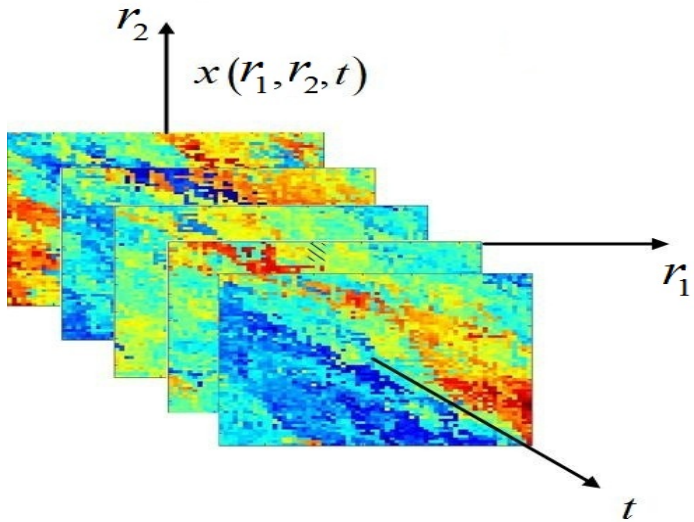
Data is missing

- Node power outage
- Hardware dysfunction
- Channel fading
- Bad environment





dark blue represent missing data



spatial-temporal sampling model

Missing data recovery

- **Retransmission**: not suitable to delay sensitive applications
- **Interpolation methods: typical ones**

  (1) K-Nearest-Neighbor (KNN)

  (2) Kriging

Intercommunity: linear combination of available data

Different weight: KNN: distance between neighbors;

Kriging: data statistics（variogram）

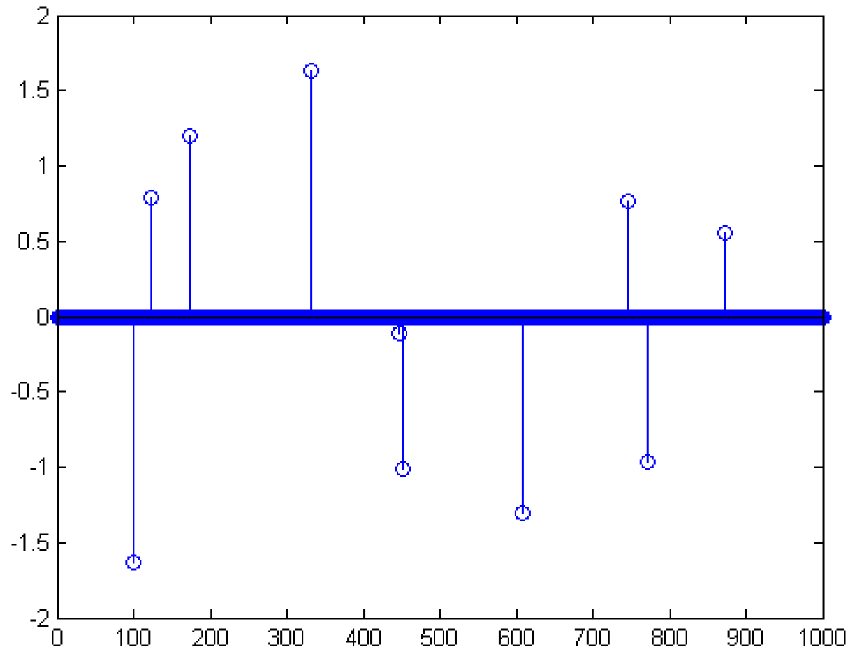
Proposed method

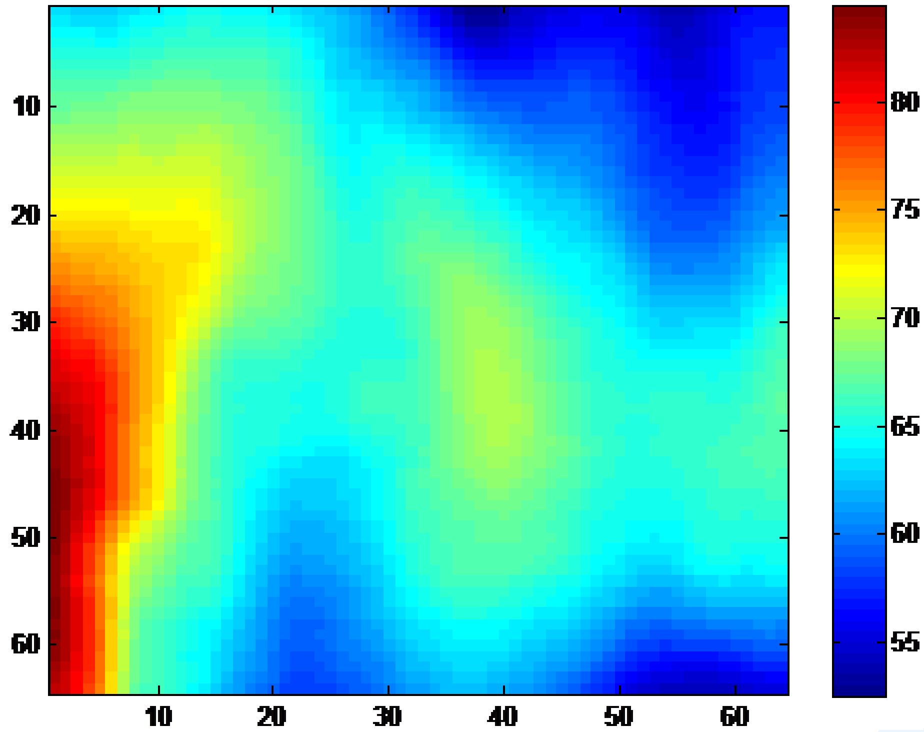
- Sparse linear combination of atoms

$$\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\alpha} = \sum_{j=1}^{N} \psi_j \, \alpha_j$$

- Weight relies on the available data

Sparsity





$$\|\boldsymbol{\alpha}\|_0 \ll N,\ \mathbf{x} \in \mathbb{R}^N$$

$$S/_N = {}^{240}/_{4096}\,,\quad \frac{\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_S\|_2^2}{\|\boldsymbol{\alpha}\|_2^2} < 10^{-5}$$

Model

$$\mathbf{f}_n = \mathbf{\Phi}_n \boldsymbol{\alpha}_n \Rightarrow \begin{pmatrix} \mathbf{f}_n^{\Lambda_n} \\ \mathbf{f}_n^{\bar{\Lambda}_n} \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi}_n^{\Lambda_n} \\ \mathbf{\Phi}_n^{\bar{\Lambda}_n} \end{pmatrix} \boldsymbol{\alpha}_n$$

$$\arg\min_{\boldsymbol{\alpha}_n} \|\boldsymbol{\alpha}_n\|_1 \quad \text{s.t.} \quad \mathbf{f}_n^{\Lambda_n} = \mathbf{\Phi}_n^{\Lambda_n} \boldsymbol{\alpha}_n$$

Assumption: Gaussian noise

$$\hat{\boldsymbol{\alpha}}_n = \arg\min_{\boldsymbol{\alpha}_n} \frac{1}{2} \left\| \mathbf{f}_n^{\Lambda_n} - \mathbf{\Phi}_n^{\Lambda_n} \boldsymbol{\alpha}_n \right\|_2^2 + \lambda \|\boldsymbol{\alpha}_n\|_1$$

Maximum a posteriori probability

Output: $\mathbf{A}_n = \mathbf{\Phi}_n \hat{\boldsymbol{\alpha}}_n$

Key: How to reduce recovery error?
(1) Dictionary, (2) Available data consistency

Dictionary

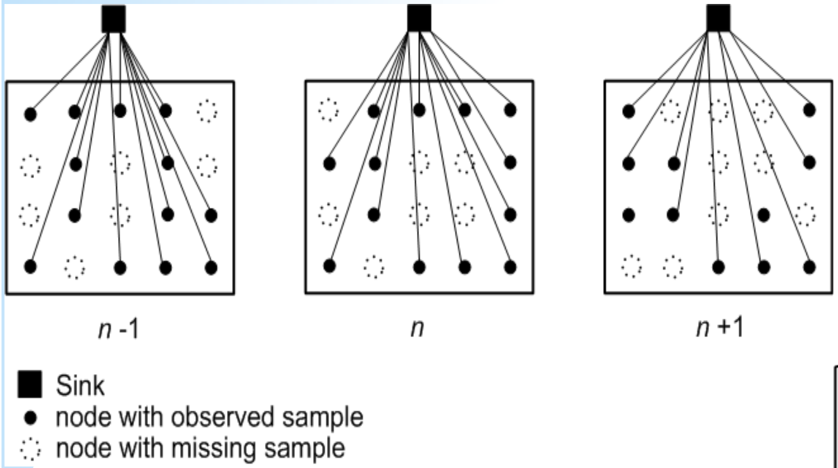
Features of WSN data

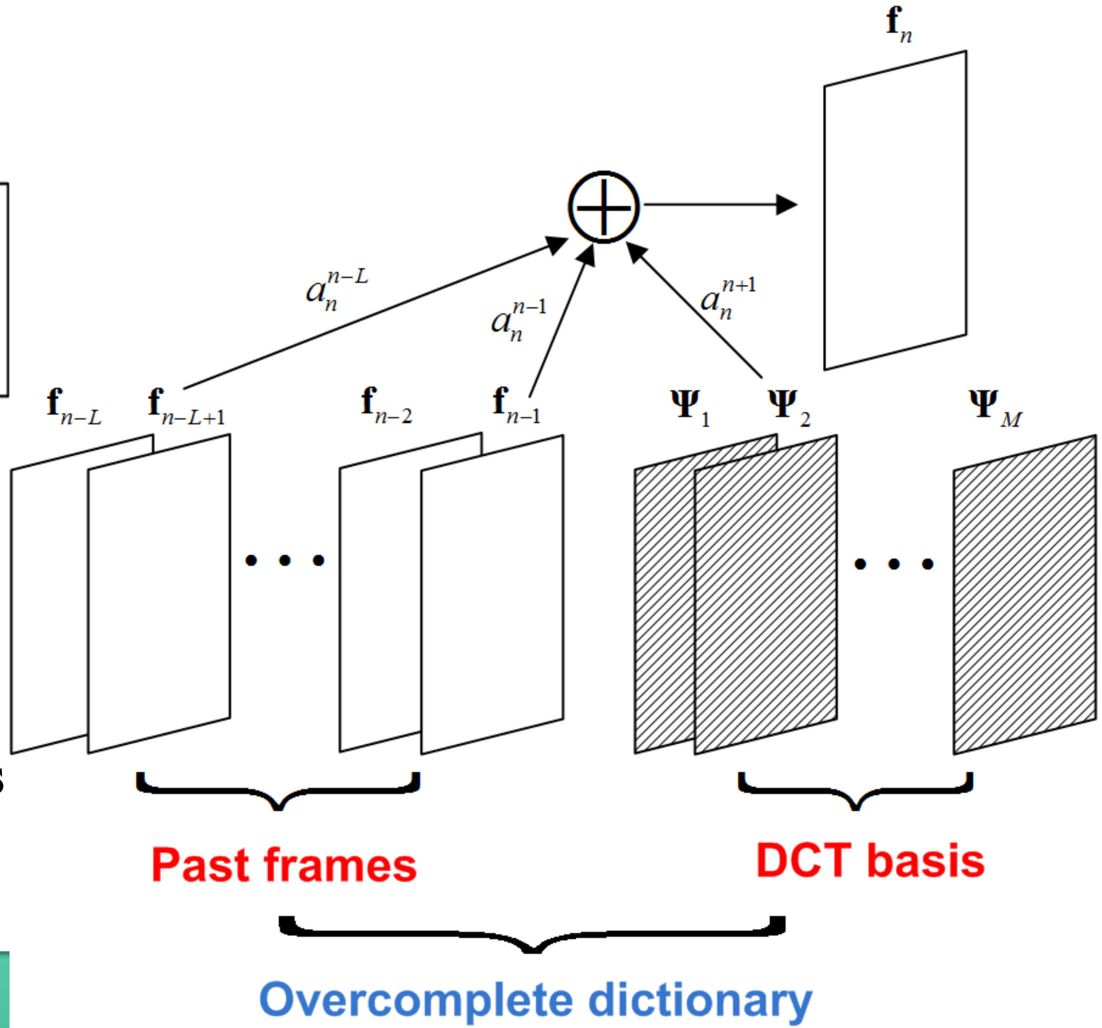
- **smooth, few boundaries**
- **weak spatial correlation**
- **strong temporal correlation**

Example: surface sunshine duration

- **Spatial domain： DCT basis**
- **Temporal spatial domain： a few past frames + DCT basis (overcomplete dictionary)**

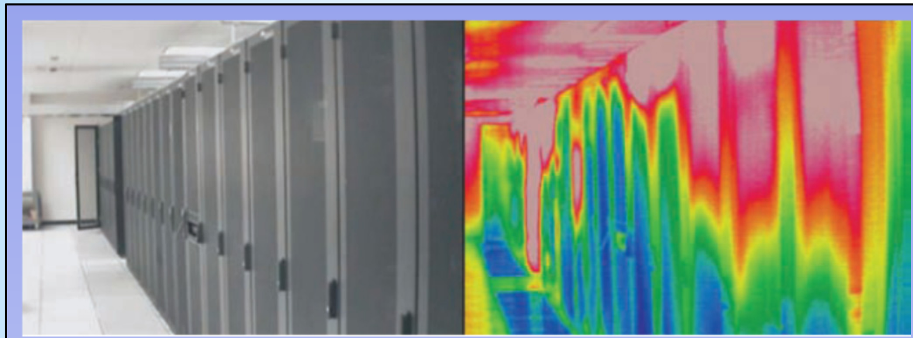
Sparsity-based online data recovery





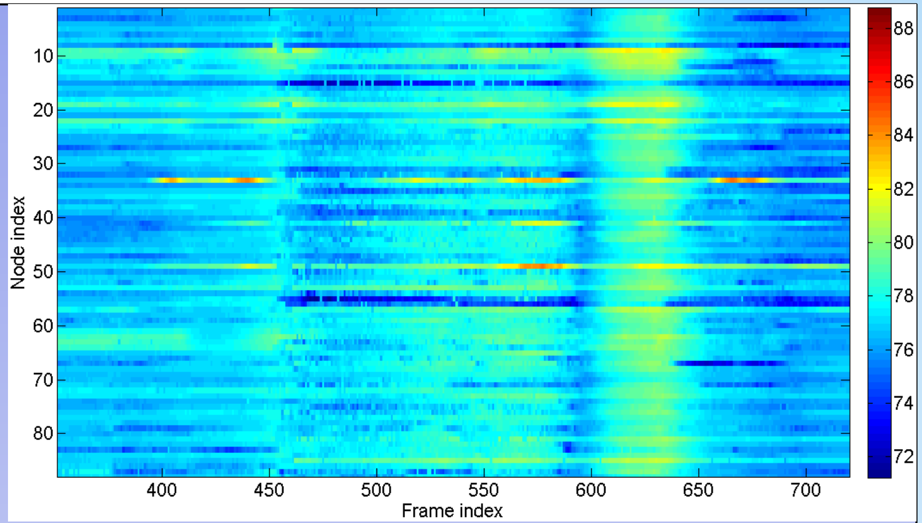
Motivation:
temporal correlation among frames

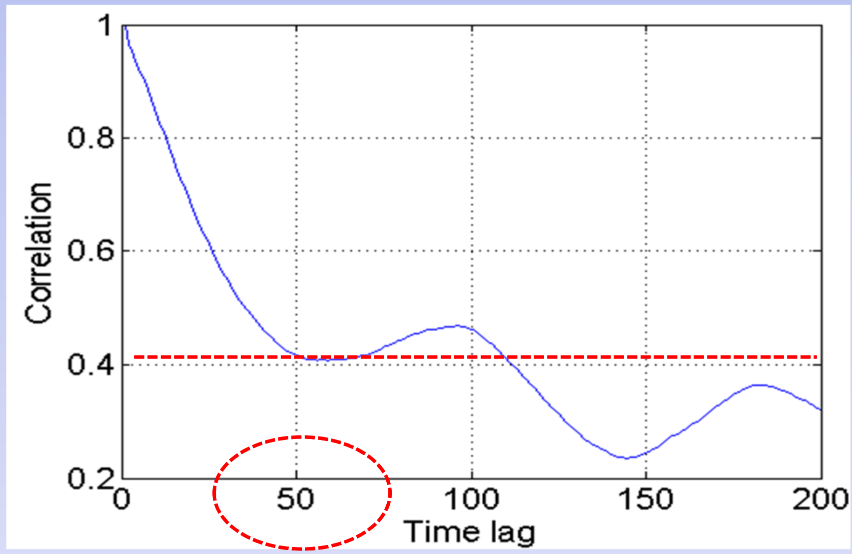
Proposed approach:
Sparse Recovery using Overcomplete Dictionary (SROD): Using a sparse linear combination of the overcomplete dictionary to represent the current frame.



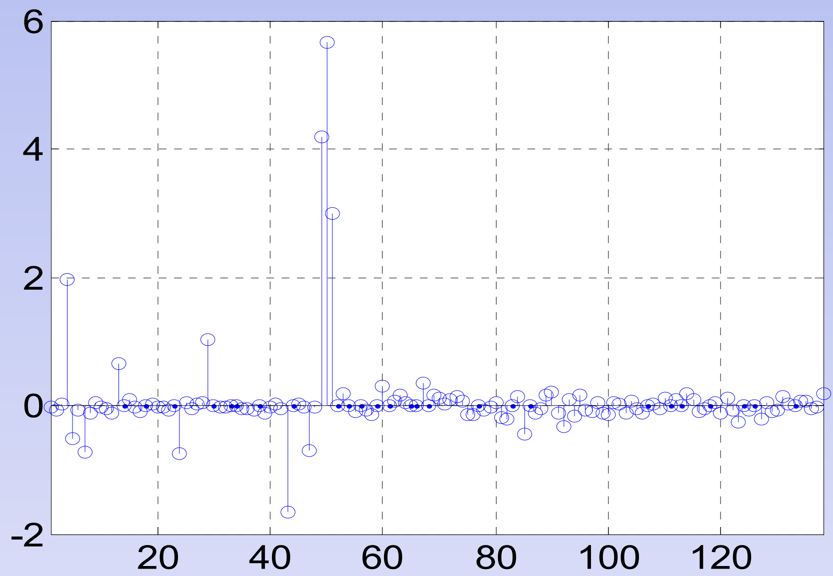
Thermal image of an data center

(data from Microsoft)





Temporal correlation of frames



Coefficients

Simulation

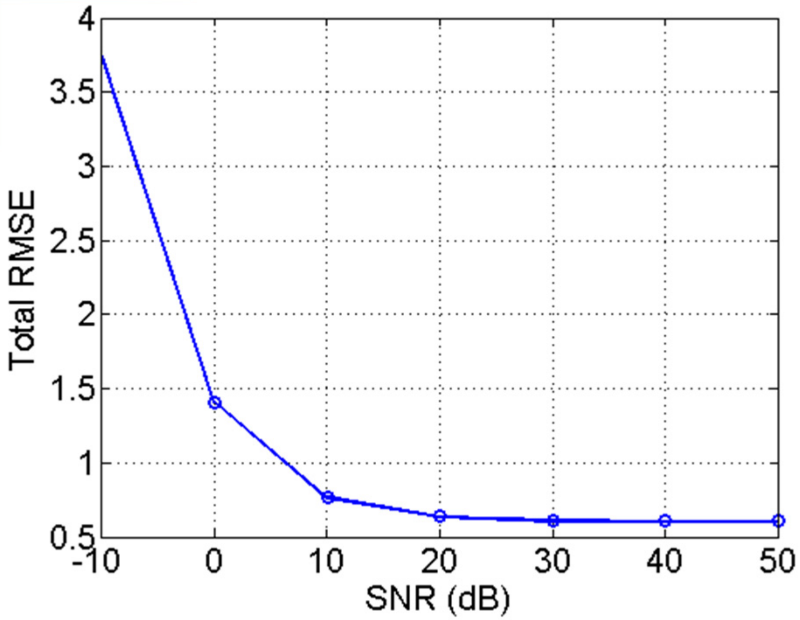
Methods / Mean	*KNN*				*SROD*			
	10%		20%		10%		20%	
	5	10	5	10	5	10	5	10
MAE_frame	1.31	1.40	1.54	1.88	0.88 (32.8%)	1.11 (20.7%)	1.19 (22.7%)	1.49 (20.7%)
MAE_node	1.48	1.48	1.75	1.80	1.06 (28.4%)	1.21 (18.2%)	1.39 (20.6%)	1.50 (16.7%)
RMSE_frame	0.66	0.69	0.66	0.78	0.43 (34.8%)	0.53 (19.7%)	0.47 (28.8%)	0.58 (25.6%)
RMSE_node	0.68	0.67	0.69	0.77	0.43 (36.8%)	0.52 (23.5%)	0.47 (31.9%)	0.56 (27.3%)
Total RMSE	**0.76**	**0.75**	**0.73**	**0.84**	**0.47 (38.2%)**	**0.57 (25.0%)**	**0.51 (30.1%)**	**0.63 (25.0%)**

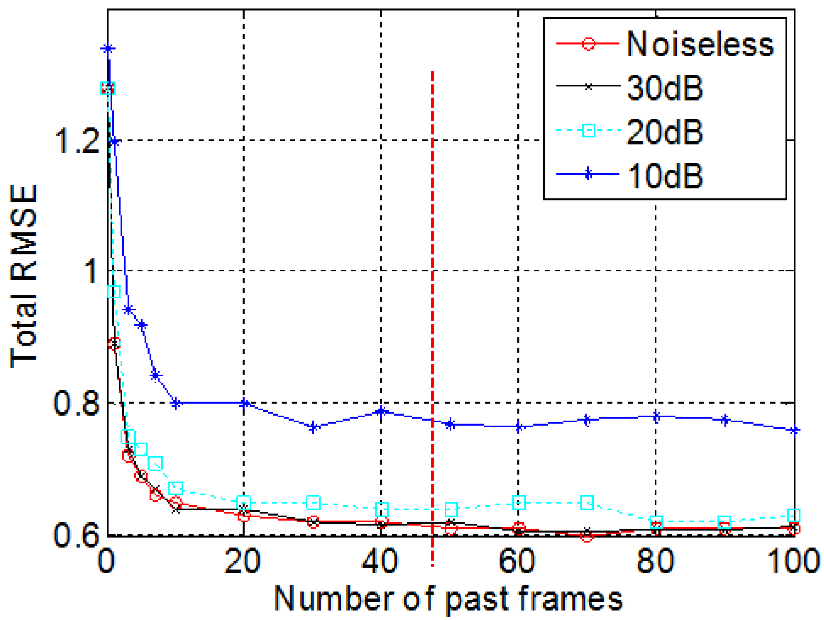
3D-KNN: anisotropic temporal spatial correlation

Data missing rate: 10%, 20%

Burst missing length: 5, 10

Robustness to Noise

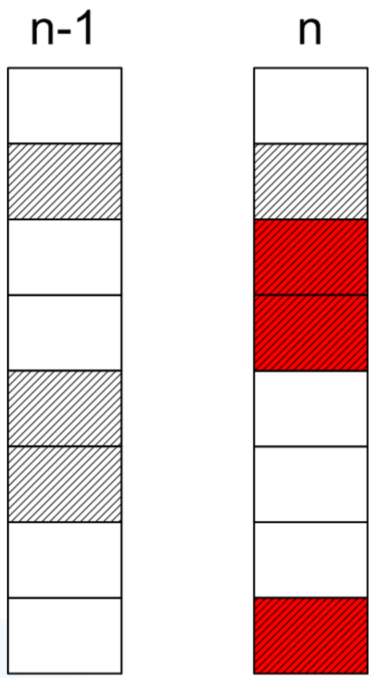




Error propagation

- **Problem: the recovery error of last frame may propagate**
- **Possible solution:**

  Leverage the available data of current frame to correct the recovery error in the last frame in some degree.



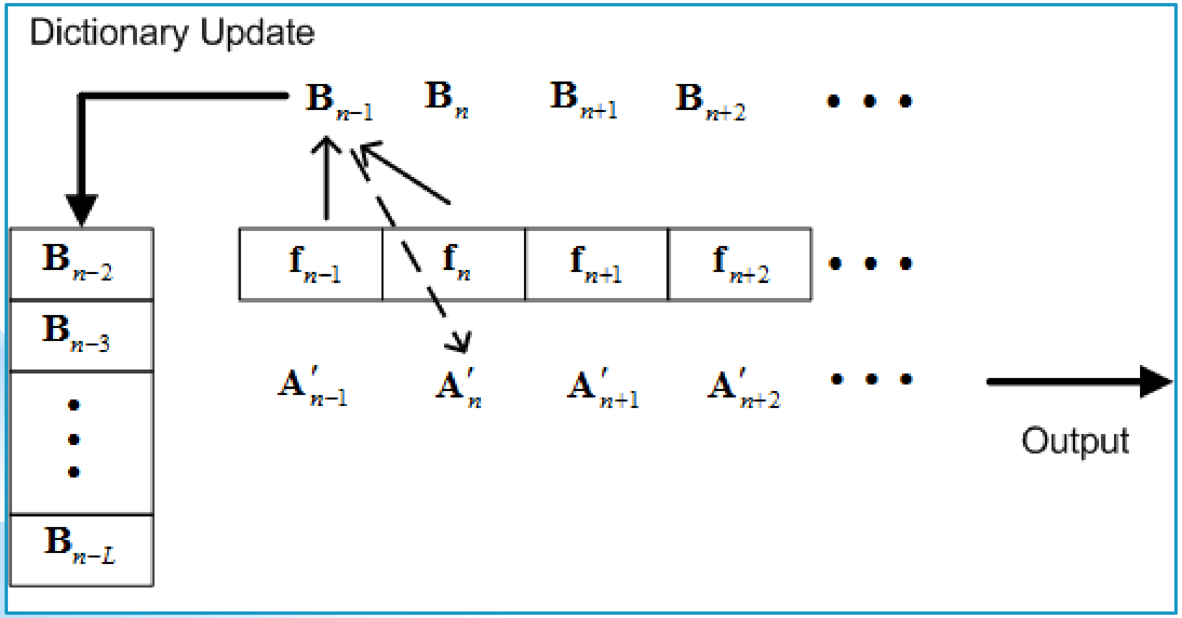
Last frame missing,
but current frame available

Recovery with Corrected Dictionary (RCD)

Neighboring data consistency

$$\hat{\boldsymbol{\alpha}}_{n-1} = \arg\min_{\boldsymbol{\alpha}_{n-1}} \frac{1}{2}\left\|\mathbf{f}_{n-1}^{\Lambda_{n-1}} - \boldsymbol{\Phi}_{n-1}^{\Lambda_{n-1}}\boldsymbol{\alpha}_{n-1}\right\|_2^2 + \lambda\left\|\boldsymbol{\alpha}_{n-1}\right\|_1 + \boxed{\frac{\mu^2}{2\sigma_n^2}\left\|\mathbf{f}_n^{\bar{\Lambda}_{n-1}\cap\Lambda_n} - \boldsymbol{\Phi}_{n-1}^{\bar{\Lambda}_{n-1}\cap\Lambda_n}\boldsymbol{\alpha}_{n-1}\right\|_2^2}$$

Update one atom of the dictionary $\boxed{\mathbf{B}_{n-1} = \boldsymbol{\Phi}_{n-1}\hat{\boldsymbol{\alpha}}_{n-1}}$



B_{n-1}: updated last frame using RCD

A'_n: recovered current frame using SROD

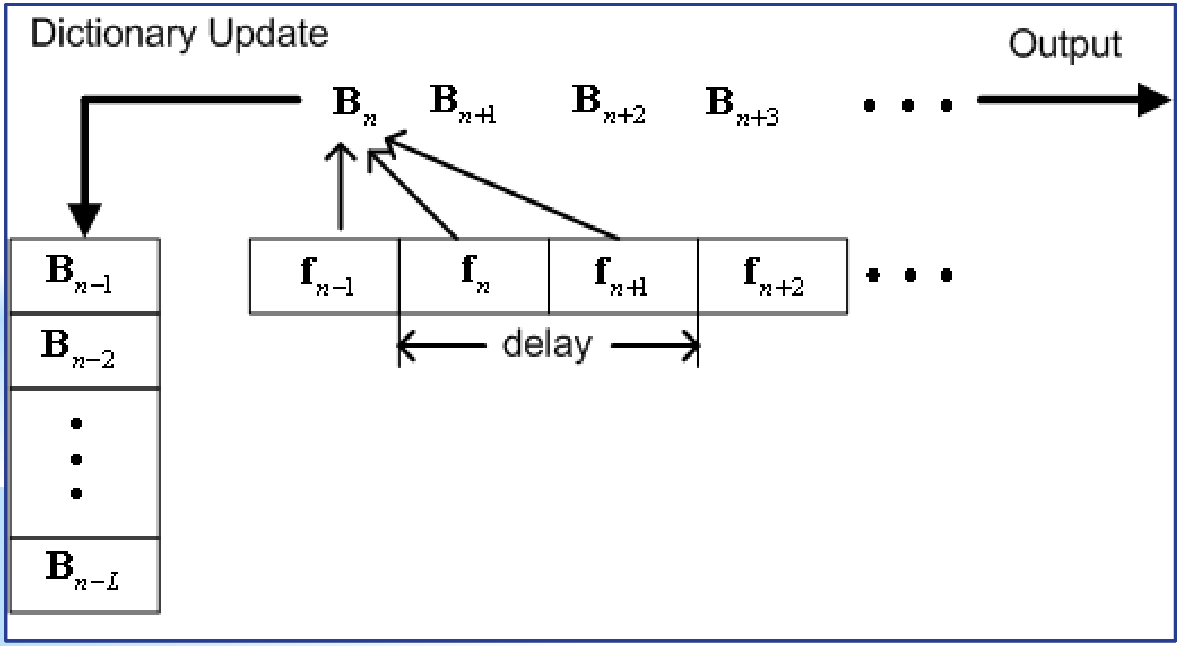
Recovery with future frame compensation (RFFC)

- **If delay is not a major concern:**

Neighboring data consistency

$$\hat{\boldsymbol{\alpha}}_n = \arg\min_{\boldsymbol{\alpha}_n} \frac{1}{2}\left\|\mathbf{f}_n^{\Lambda_n} - \boldsymbol{\Phi}_n^{\Lambda_n}\boldsymbol{\alpha}_n\right\|_2^2 + \lambda\left\|\boldsymbol{\alpha}_n\right\|_1 + \boxed{\frac{\mu}{2\sigma_{n+1}^2}\left\|\mathbf{f}_{n+1}^{\bar{\Lambda}_n\cap\Lambda_{n+1}} - \boldsymbol{\Phi}_n^{\bar{\Lambda}_n\cap\Lambda_{n+1}}\boldsymbol{\alpha}_n\right\|_2^2}$$

Current frame $\mathbf{B}_n = \boldsymbol{\Phi}_n\hat{\boldsymbol{\alpha}}_n$



B_n: recovered current frame

Simulation

Three proposed sparsity-based recovery method compare with corresponding 3-D KNN

Missing rate: 20%, burst missing length: 1

Methods / Mean	*KNN*			*Proposed*		
	KNN	*KNN-CD*	*KNN-FFC-1*	*SROD*	*RCD*	*RFFC-1*
MAE_frame	1.55	1.54 (0.6%)	1.46 (5.8%)	0.97 (37.4%)	0.95 (38.7%)	0.79 (49.0%)
MAE_node	1.80	1.79 (0.6%)	1.73 (3.9%)	1.25 (30.6%)	1.24 (31.1%)	1.08 (40.0%)
RMSE_frame	0.66	0.66 (-)	0.62 (6.1%)	0.38 (42.4%)	0.37 (43.9%)	0.30 (54.5%)
RMSE_node	0.69	0.69 (-)	0.65 (5.8%)	0.39 (43.5%)	0.38 (44.9%)	0.32 (53.6%)
Total RMSE	**0.72**	**0.72 (-)**	**0.69 (4.2%)**	**0.43 (40.3%)**	**0.42 (41.7%)**	**0.36 (50.0%)**

- ❖ **Error reduce by 40%**
- ❖ **RFFC reduce error by 10% over SROD**

Conclusion

- **Propose sparsity-based online data recovery method**
- **Construct an overcomplete dictionary: past frames + DCT basis**
- **Recovery performance significantly outperforms KNN**
- **Robust to certain noise**
- **RCD may reduce error propagation**
- **RFFC can further improve recovery performance**

Future work

- Test missing pattern from the perspective of wireless communication
- Extract data feature using data mining
- Design dictionary and optimization algorithms

Acknowledgement

Data from

Dr. Jie Liu in Microsoft

Funds from

- Tsinghua-Qualcomm Joint Research Program
- National Natural Science Foundation of China (No. 61001142)
- China Scholarship Council

Thank you

Any questions?