BİL 722

Ahmet BUĞDAY

Paper

* **Image denoising via learned dictionaries and sparse representation, M. Elad and M. Aharan, CVPR 2006.**

INTRODUCTION

* $y = x + v$
  * y noisy image (known)
  * x deoised image (unknown)
  * v noise (unknown)
* Use of sparse and redundant represantations over trained dictionaries
  * Higly effective and promising

INTRODUCTION

* Sparse and Redundant Representation Model
  * an ability to describe signals as linear combinations of a few atoms from a pre-specified dictionary
  * Linear system $D\propto = x$
  * $D \in \mathbb{R}\uparrow nxm$ and $\propto \in \mathbb{R}\uparrow m$
  * m columns atomic images
  * D is dictionary
    * interpret *D* as the periodic table of the fundamental elements in the chemistry that describes our images
  * $\propto$ is sparse vector
    * $\propto$ describes which atoms and what "portions" thereof were used for its construction

INTRODUCTION

* Wavelet coefficients used for sparsity and redundancy
  * 1D wavelets inappropriate for images
* Curvelet, Contourlet, Wedgelet, Bandlet, The Steerable Wavelet, ...
  * Multi scale and redundant transforms
* Sparse represantation solution NP-Hard
  * Pursuit algorithm gives approximate result

INTRODUCTION

* Inverse Problems
  * Lean on a guess to find image prior
    * Spatial smoothness
    * Low-entropy
    * Sparsity in some transform domain
* Example Based Approach
  * Learn image prior somehow
* Learning prior + sparsity and redundancy = Dictionary
* Learn dictionary using image patches

This Paper

* Zero-mean white and homogeneous Gaussian additive noise should be removed
* Sparse and redundant represantation over a trained dictionary
* Denoising and training in the same time
* K-SVD is used for training
* Global image prior forces sparsity over small patches in every location
* State-of-art performance

Local to Global Bayesian Reconstruction

* Sparseland Model
  - $\sqrt{n} \ x \sqrt{n}$ pixels sizes image patches
* D n x k size matrix k>n implying redundant
* Assume D is known and fixed
* Every image patch represented sparsely over this dictionary
  - $\propto = \arg min \| \propto \| \downarrow 0$ subject to $D \propto \approx x$
    - $\| \propto \| \downarrow 0$ count of the non-zero entries in $\propto$ ($l \downarrow 0$ norm)

Local to Global Bayesian Reconstruction

- Change rough constraint $D\propto\approx x$ with clear requirement
  - $||D\propto - x||\downarrow 2 \leq \varepsilon$
- Define sparsity depth
  - $||\propto||\downarrow 0 \leq L \ll n$
  - Sparse represantation uses at most L atoms from D for patches
- x belongs to (ε, L, D) Sparseland signals
- Consider y which is x+ZMHGA noise with std σ
- MAP (Maximum-a-posteriori) estimator for denoising patch

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0 \text{ subject to } \|\mathbf{D}\alpha - \mathbf{y}\|_2^2 \leq T$$

Local to Global Bayesian Reconstruction

* Optimization task changed to

$$\hat{\alpha} = \arg\min_{\alpha} \|\mathbf{D}\alpha - \mathbf{y}\|_2^2 + \mu\|\alpha\|_0$$

* X large image every patch of it belongs to (ε,L,D)-Sparseland model, than natural generalization of the MAP estimator

$$\{\alpha_{ij}, X\} = \arg\min_{\alpha_{ij}, x} \lambda\|X-Y\|_2^2 + \sum_{ij} \mu_{ij} \|\alpha_{ij}\|_0 + \sum_{ij} \|D\alpha_{ij} - R_{ij} X\|_2^2$$

  * First term; global force demands the proximity between Y and X
    * λ, σ has direct relationships
  * Second and third term part of the image prior
    * Guarantees every patch in every location has sparse represantation with bounded error

Example-Based Sparsity and Redundancy

* Formulation is created with the assumption of D is known
* Incorporate dictionary learning with denoising task

$$\left\{\hat{\mathbf{D}}, \hat{\alpha}_{ij}, \hat{\mathbf{X}}\right\} = \arg\min_{\hat{\mathbf{D}}, \alpha_{ij}, \mathbf{X}} \lambda \|\mathbf{X} - \mathbf{Y}\|_2^2 + \sum_{ij} \mu_{ij} \|\alpha_{ij}\|_0 + \sum_{ij} \|\mathbf{D}\alpha_{ij} - \mathbf{R}_{ij}\mathbf{X}\|_2^2$$

Example-Based Sparsity and Redundancy

* Three unknown
  - $\propto \downarrow ij$; sparse representaion per each location
  - D; dictionary
  - X; denoised image
* Initialization for finding $\propto \downarrow ij$
  - X=Y
  - Pre-chosen and fixed dictionary using Y
* Then

$$\hat{\alpha}_{ij} = \arg\min_{\alpha} \mu_{ij}\|\alpha\|_0 + \|\mathbf{D}\alpha - \mathbf{x}_{ij}\|_2^2$$

  - Orthonormal matching pursuit
    - Gathering one atom at a time
  - Stop error below T

Example-Based Sparsity and Redundancy

* Update D using K-SVD
  * Generalized K-Means
  * Works with any matching pursuit algorithm
  * Update one column at a time
  * Perform SVD
  * Penalty will drop in each update

Initialization : Set the random normalized dictionary matrix $\mathbf{D}^{(0)} \in \mathbb{R}^{n \times K}$. Set $J = 1$.

Repeat until convergence,

Sparse Coding Stage: Use any pursuit algorithm to compute $\mathbf{x}_i$ for $i = 1, 2, \ldots, N$

$$\min_{\mathbf{x}} \{ \|\mathbf{y}_i - \mathbf{D}\mathbf{x}\|_2^2 \} \quad \text{subject to} \quad \|\mathbf{x}\|_0 \leq T_0.$$

Codebook Update Stage: For $k = 1, 2, \ldots, K$

- Define the group of examples that use $\mathbf{d}_k$, $\omega_k = \{i | \, 1 \leq i \leq N, \; \mathbf{x}_i(k) \neq 0\}$.
- Compute

$$\mathbf{E}_k = \mathbf{Y} - \sum_{j \neq k} \mathbf{d}_j \mathbf{x}^j,$$

- Restrict $\mathbf{E}_k$ by choosing only the columns corresponding to those elements that initially used $\mathbf{d}_k$ in their representation, and obtain $\mathbf{E}_k^R$.
- Apply SVD decomposition $\mathbf{E}_k^R = \mathbf{U}\mathbf{\Delta}\mathbf{V}^T$. Update: $\mathbf{d}_k = \mathbf{u}_1, \mathbf{x}_R^k = \mathbf{\Delta}(1,1) \cdot \mathbf{v}_1$

Set $J = J + 1$.

The K-SVD Algorithm

Example-Based Sparsity and Redundancy

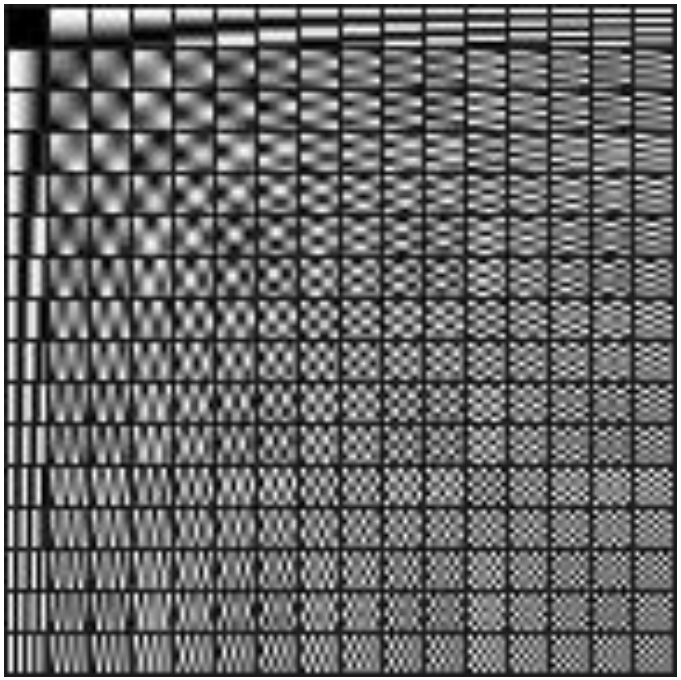
* Given all $\propto \downarrow ij$ and updated D

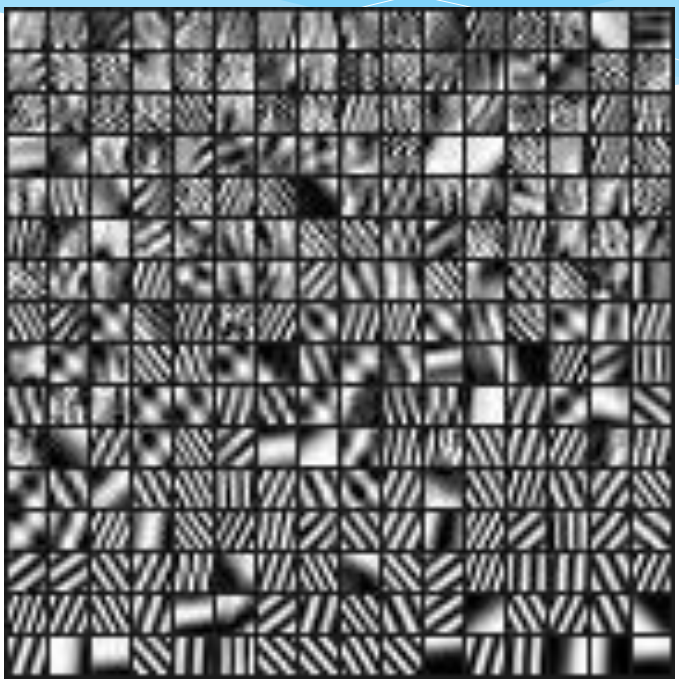
$$\hat{\mathbf{X}} = \arg\min_{\mathbf{x}} \ \lambda\|\mathbf{X} - \mathbf{Y}\|_2^2 + \sum_{ij} \|\mathbf{D}\hat{\alpha}_{ij} - \mathbf{R}_{ij}\mathbf{X}\|_2^2$$

* Solution is
*

$$\hat{\mathbf{X}} = \left(\lambda\mathbf{I} + \sum_{ij} \mathbf{R}_{ij}^T\mathbf{R}_{ij}\right)^{-1} \left(\lambda\mathbf{Y} + \sum_{ij} \mathbf{R}_{ij}^T\mathbf{D}\hat{\alpha}_{ij}\right)$$

Dictionary





* **The overcomplete DCT dictionary (left). The trained dictionary for 'Barbara' with $\sigma = 15$, after 10 iterations (right).**

Computational Complexity

* O(nLS) operations per pixel
  * n block dimension (64)
  * S iteration number (10)
  * L depends noise level
    * $\sigma = 10$ L is 2.96
    * $\sigma = 20$ L is 1.12

RESULTS

* Compared with
  * Portilla J., Strela, V., Wainwright, M.J. and Simoncelli, E.P. (2003) Image denoising using scale mixtures of gaussians in the wavelet domain IEEE Transactions on Image Processing, Vol. 12, No.11, pp. 1338-1351, November
* Redundant DCT 64x256 D
* 8x8 image patch
  * n=64, k=256
* 0.24 db advantage

RESULTS

$\sigma/PSNR$	Lena		Barb		Boats		Fgrpt		House		Pepers		σ_{PSNR}	
2/42.11	43.23	**43.58**	43.29	**43.67**	42.99	**43.14**	**43.05**	42.99	44.07	**44.47**	43.00	**43.33**	0.012	0.017
5/34.15	38.49	**38.60**	37.79	**38.08**	36.97	**38.08**	**36.68**	36.65	38.65	**39.37**	37.31	**37.78**	0.014	0.017
10/28.13	**35.61**	35.47	34.03	**34.42**	33.58	**33.64**	**32.45**	32.39	35.35	**35.98**	33.77	**34.28**	0.017	0.027
15/24.61	**33.90**	33.70	31.86	**32.37**	31.70	**31.73**	**30.14**	30.06	33.64	**34.32**	31.74	**32.22**	0.024	0.035
20/22.11	**32.66**	32.38	30.32	**30.83**	**30.38**	30.36	**28.60**	28.47	32.39	**33.20**	30.31	**30.82**	0.031	0.027
25/20.17	**31.69**	31.32	29.13	**29.60**	**29.37**	29.28	**27.45**	27.26	31.40	**32.15**	29.21	**29.73**	0.037	0.036
50/14.15	**28.61**	27.79	**25.48**	25.47	**26.38**	25.95	**24.16**	23.24	**28.26**	27.95	25.90	**26.13**	0.049	0.058
75/10.63	**26.84**	25.80	**23.65**	23.01	**24.79**	23.98	**22.40**	19.97	**26.41**	25.22	**24.00**	23.69	0.061	0.060
100/8.13	**25.64**	24.46	**22.61**	21.89	**23.75**	22.81	**21.22**	18.30	**25.11**	23.71	**22.66**	21.75	0.070	0.046

RESULTS

* Left orginal image
* Right Noisy image (24.6 dB,σ=15)
* Middle denoised image

Cons

* Higher noise power
  * Weaker result
  * Deteriorates faster
* Time comparison ?
* Textures ?
* Less result image

Pros

* Work with all matching pursuit algorithm
* Global Bayesian
* A new approach K-SVD used for D training
* Good result in less noise (noise < 50db)

REFERENCES

* R. Rubinstein, T. Peleg, and M. Elad, K-SVD Dictionary Learning for the Analysis Co-Sparse Model , ICASSP, Kyoto, Japan, March 25-30, 2012.
* M. Aharon, M. Elad, and A.M. Bruckstein, The K-SVD Algorithm, Proceedings of SPARSE05, Rennes, France, November 2005.
* M. Aharon, M. Elad, and A.M. Bruckstein, " The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation", the *IEEE Trans. On Signal Processing*, Vol. 54, no. 11, pp. 4311-4322, November 2006.
* Michael Elad, Mario A.T. Figueiredo, On the Role of Sparse and Redundant Representations in Image Processing, *IEEE*
* Ron Rubinstein, Alfred M. Bruckstein and Michael Elad, Dictionaries for Sparse Representation Modeling, IEEE PROCEEDINGS