BİL 722
Ahmet BUĞDAY
**INTRODUCTION**

- $y = x + v$
  - $y$ noisy image (known)
  - $x$ deoised image (unknown)
  - $v$ noise (unknown)
- Use of sparse and redundant representations over trained dictionaries
- Highly 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 $D \alpha = x$

*$D \in \mathbb{R}^{n \times m}$ and $\alpha \in \mathbb{R}^{m}$

* $m$ columns atomic images

* $D$ is dictionary

  * interpret $D$ as the periodic table of the fundamental elements in the chemistry that describes our images

*$\alpha$ is sparse vector

*$\alpha$ 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
Zero-mean white and homogeneous Gaussian additive noise should be removed
Sparse and redundant representation 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
  \[ \sqrt{n} \times \sqrt{n} \text{ pixels sizes image patches} \]
* D is a \( k \times n \) size matrix with \( k > n \) implying redundant
* Assume \( D \) is known and fixed
* Every image patch represented sparsely over this dictionary
  \[ \alpha = \text{argmin} \| \alpha \|_0 \quad \text{subject to} \quad D\alpha \approx x \]
  \[ \| \alpha \|_0 \text{ count of the non-zero entries in } \alpha \ (\ell_0 \text{ norm}) \]
• Change rough constraint $D\alpha \approx x$ with clear requirement
  \[ \|D\alpha - x\|_2 \leq \varepsilon \]

* Define sparsity depth
  \[ \|\alpha\|_0 \leq L \ll n \]
  * Sparse represantation uses at most $L$ atoms from $D$ for patches

* $x$ belongs to $(\varepsilon,L,D)$ Sparseland signals

* Consider $y$ which is $x + ZMHGA$ noise with std $\sigma$

* MAP (Maximum-a-posteriori) estimator for denoising patch

\[
\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_0 \text{ subject to } \|D\alpha - y\|_2^2 \leq T
\]
Local to Global Bayesian Reconstruction

- Optimization task changed to

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

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

\[
\{\alpha \downarrow ij, X\} = \arg \min_{\alpha \downarrow ij, x} \| X - y \|_2 + \sum_{ij} \mu_{ij} \|\alpha \downarrow ij\|_0 + \sum_{ij} \|D\alpha \downarrow ij - R_{ij} X\|_2
\]

- First term; \textbf{global force} demands the proximity between \(Y\) and \(X\)
  - \(\lambda, \sigma\) has direct relationships

- Second and third term part of the image prior
  - Guarantees every patch in every location has sparse representation with bounded error
Example-Based Sparsity and Redundancy

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

$$\left\{ \hat{D}, \hat{\alpha}_{ij}, \hat{X} \right\} = \arg \min_{\hat{D}, \hat{\alpha}_{ij}, \hat{X}} \lambda \|X - Y\|_2^2 + \sum_{ij} \mu_{ij} \|\alpha_{ij}\|_0 + \sum_{ij} \|D\alpha_{ij} - R_{ij}X\|_2^2$$
Example-Based Sparsity and Redundancy

* Three unknown
  \( x \propto i_j \); sparse representation per each location
* \( D \); dictionary
* \( X \); denoised image
* Initialization for finding \( x \propto i_j \)
* \( X=Y \)
* Pre-chosen and fixed dictionary using \( Y \)
* Then
  \[
  \hat{\alpha}_{ij} = \arg \min_{\alpha} \mu_{ij} \|\alpha\|_0 + \|D\alpha - x_{ij}\|_2^2.
  \]
* 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 $D^{(0)} \in \mathbb{R}^{n \times K}$. Set $J = 1$.

Repeat until convergence,

**Sparse Coding Stage:** Use any pursuit algorithm to compute $x_i$ for $i = 1, 2, \ldots, N$

$$\min_{x} \left\{ \|y_i - Dx\|_2^2 \right\} \quad \text{subject to} \quad \|x\|_0 \leq T_0.$$ 

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

- Define the group of examples that use $d_k$, $\omega_k = \{i \mid 1 \leq i \leq N, x_i(k) \neq 0\}$.
- Compute
  $$E_k = Y - \sum_{j \neq k} d_j x^j,$$

- Restrict $E_k$ by choosing only the columns corresponding to those elements that initially used $d_k$ in their representation, and obtain $E_k^R$.
- Apply SVD decomposition $E_k^R = U \Delta V^T$. Update: $d_k = u_1$, $x_k^R = \Delta (1, 1) \cdot v_1$

Set $J = J + 1$.

The K-SVD Algorithm
Example-Based Sparsity and Redundancy

* Given all $\alpha_{ij}$ and updated $D$

$$\hat{X} = \arg\min_x \lambda \|X - Y\|_2^2 + \sum_{ij} \|D\hat{\alpha}_{ij} - R_{ij}X\|_2^2$$

* Solution is

$$\hat{X} = \left(\lambda I + \sum_{ij} R_{ij}^T R_{ij}\right)^{-1} \left(\lambda Y + \sum_{ij} R_{ij}^T D\hat{\alpha}_{ij}\right)$$
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 on noise level
    * \( \sigma = 10 \) \( L \) is 2.96
    * \( \sigma = 20 \) \( L \) is 1.12
RESULTS

* Compared with

* Redundant DCT 64x256 D
* 8x8 image patch
  * n=64, k=256
  * 0.24 db advantage
<table>
<thead>
<tr>
<th>$\sigma/PSNR$</th>
<th>Lena</th>
<th>Barb</th>
<th>Boats</th>
<th>Fgrpt</th>
<th>House</th>
<th>Peppers</th>
<th>$\sigma_{PSNR}$</th>
<th>Lena</th>
<th>Barb</th>
<th>Boats</th>
<th>Fgrpt</th>
<th>House</th>
<th>Peppers</th>
<th>$\sigma_{PSNR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/42.11</td>
<td>43.23</td>
<td>43.58</td>
<td>43.29</td>
<td>43.67</td>
<td>42.99</td>
<td>43.14</td>
<td>43.05</td>
<td>42.99</td>
<td>44.07</td>
<td>44.47</td>
<td>43.33</td>
<td>43.00</td>
<td>43.33</td>
<td>0.012</td>
</tr>
<tr>
<td>5/34.15</td>
<td>38.49</td>
<td>38.60</td>
<td>37.79</td>
<td>38.08</td>
<td>36.97</td>
<td>38.08</td>
<td>36.68</td>
<td>36.65</td>
<td>38.65</td>
<td>39.37</td>
<td>37.31</td>
<td>37.78</td>
<td>0.014</td>
<td>0.017</td>
</tr>
<tr>
<td>10/28.13</td>
<td>35.61</td>
<td>35.47</td>
<td>34.03</td>
<td>34.42</td>
<td>33.58</td>
<td>33.64</td>
<td>32.45</td>
<td>32.39</td>
<td>35.35</td>
<td>35.98</td>
<td>33.77</td>
<td>34.28</td>
<td>0.017</td>
<td>0.027</td>
</tr>
<tr>
<td>15/24.61</td>
<td>33.90</td>
<td>33.70</td>
<td>31.86</td>
<td>32.37</td>
<td>31.70</td>
<td>31.73</td>
<td>30.14</td>
<td>30.06</td>
<td>33.64</td>
<td>34.32</td>
<td>31.74</td>
<td>32.22</td>
<td>0.024</td>
<td>0.035</td>
</tr>
<tr>
<td>20/22.11</td>
<td>32.66</td>
<td>32.38</td>
<td>30.32</td>
<td>30.83</td>
<td>30.38</td>
<td>30.36</td>
<td>28.60</td>
<td>28.47</td>
<td>32.39</td>
<td>33.20</td>
<td>30.31</td>
<td>30.82</td>
<td>0.031</td>
<td>0.027</td>
</tr>
<tr>
<td>25/20.17</td>
<td>31.69</td>
<td>31.32</td>
<td>29.13</td>
<td>29.60</td>
<td>29.37</td>
<td>29.28</td>
<td>27.45</td>
<td>27.26</td>
<td>31.40</td>
<td>32.15</td>
<td>29.21</td>
<td>29.73</td>
<td>0.037</td>
<td>0.036</td>
</tr>
<tr>
<td>50/14.15</td>
<td>28.61</td>
<td>27.79</td>
<td>25.48</td>
<td>25.47</td>
<td>26.38</td>
<td>25.95</td>
<td>24.16</td>
<td>23.24</td>
<td>28.26</td>
<td>27.95</td>
<td>25.90</td>
<td>26.13</td>
<td>0.049</td>
<td>0.058</td>
</tr>
<tr>
<td>75/10.63</td>
<td>26.84</td>
<td>25.80</td>
<td>23.65</td>
<td>23.01</td>
<td>24.79</td>
<td>23.98</td>
<td>22.40</td>
<td>19.97</td>
<td>26.41</td>
<td>25.22</td>
<td>24.00</td>
<td>23.69</td>
<td>0.061</td>
<td>0.060</td>
</tr>
<tr>
<td>100/8.13</td>
<td>25.64</td>
<td>24.46</td>
<td>22.61</td>
<td>21.89</td>
<td>23.75</td>
<td>22.81</td>
<td>21.22</td>
<td>18.30</td>
<td>25.11</td>
<td>23.71</td>
<td>22.66</td>
<td>21.75</td>
<td>0.070</td>
<td>0.046</td>
</tr>
</tbody>
</table>
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

* 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