Image Compression Using the Iteration-Tuned and Aligned Dictionary

Joaquin Zepeda, Christine Guillemot and Ewa Kijak

INRIA Rennes - Bretagne Atlantique

– ICASSP 2011 –

Tuesday May 24, 2010

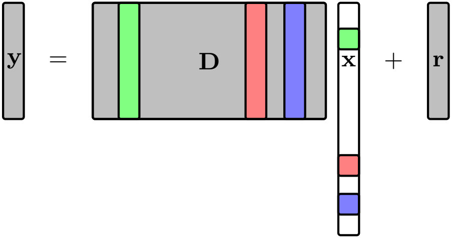
Outline

Outline

Sparse representations

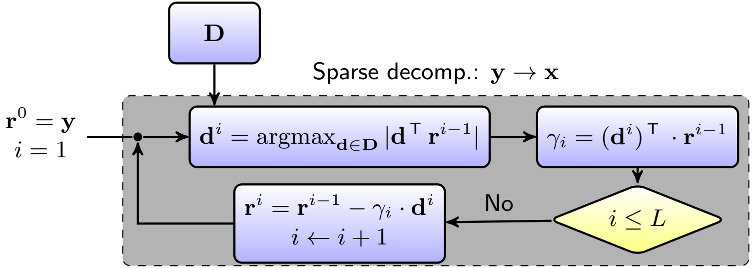
Sparse representations:

- $\mathbf{y}$: the signal vector.
- $\mathbf{D}$: the *dictionary*, OVERCOMPLETE, with columns called *atoms*.
- $\mathbf{x}$: the *sparse representation* – *fewest* atoms, *good* approximation.
- $\mathbf{r}$: approximation error or *residue*.



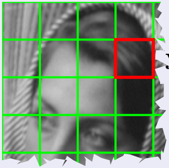
Computing sparse representations

- Very difficult to solve! *Iterative* pursuit algorithms commonly used instead, *eg*., *Matching Pursuit (MP)*:



($\mathbf{D}$ is assumed to have normalized columns.)

Application to image compression



$\mathbf{y} \approx \mathbf{Dx}$

- Representation $\mathbf{x}$ is compact version of $\mathbf{y}$.
- DESIGN ISSUES:
  1. Dictionary choice? *ITAD: New, learnt, structured dictionary.*
  2. Atom distribution across image? *New, global rate-distortion based criterion*
  3. Sparse vector encoding? DPCM encoding of block mean, uniform quantization + Huffman coding for coefficients, fixed-length code for atom indices.

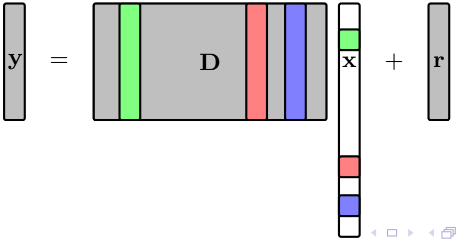
Outline

Outline

Overcompleteness

Overcompleteness

Refers to the *fat* shape of matrix $\mathbf{D}$.

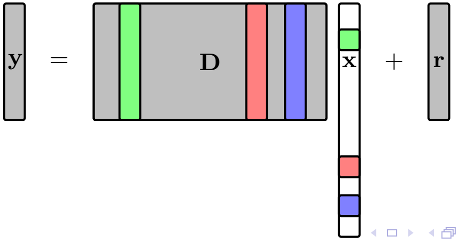


Overcompleteness

Overcompleteness

Refers to the *fat* shape of matrix $\mathbf{D}$.

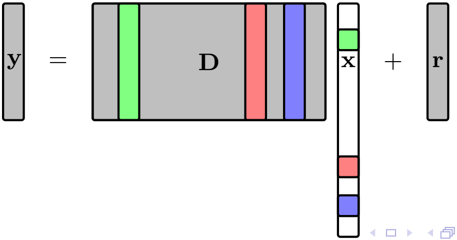
1. Overcompleteness originates signal *sparsity*,



Overcompleteness

Refers to the *fat* shape of matrix $\mathbf{D}$.

1. Overcompleteness originates signal *sparsity*,
2. yet it is computationally *complex*,

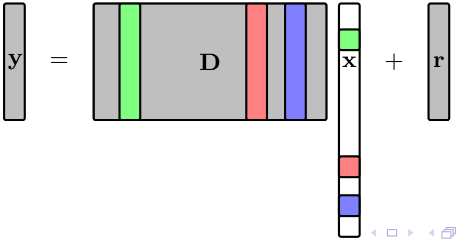


Overcompleteness

Overcompleteness

Refers to the *fat* shape of matrix $\mathbf{D}$.

1. Overcompleteness originates signal *sparsity*,
2. yet it is computationally *complex*,
3. and increases atom index coding *rate*.



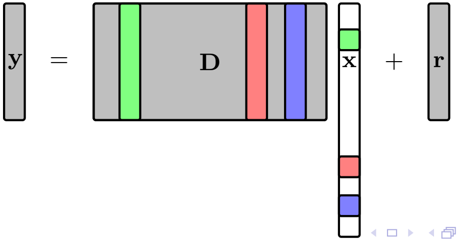
Overcompleteness

OVERCOMPLETENESS

Refers to the *fat* shape of matrix $\mathbf{D}$.

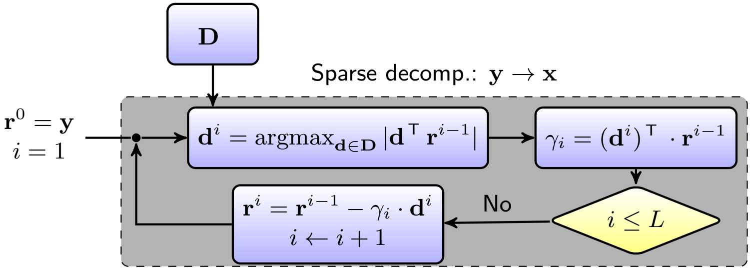
1. Overcompleteness originates signal *sparsity*,
2. yet it is computationally *complex*,
3. and increases atom index coding *rate*.

$\Rightarrow$ Structure the dictionary (constrain atom selection) to address computational and complexity issues: the *Iteration-Tuned and Aligned Dictionary.*



The Iteration-Tuned and Aligned Dictionary

Change the dictionary in each MP iteration.

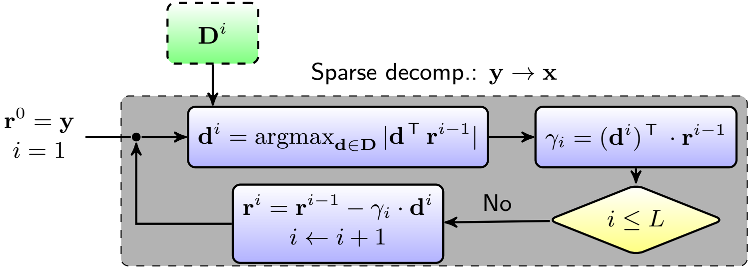


($\mathbf{D}$ is assumed to have normalized columns.)

If $\mathbf{D}$ and $\mathbf{D}^i \forall i$ all have N atoms ...

The Iteration-Tuned and Aligned Dictionary

Change the dictionary in each MP iteration.

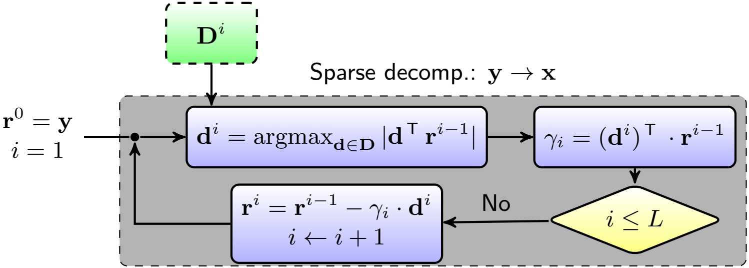


($\mathbf{D}$ is assumed to have normalized columns.)

If $\mathbf{D}$ and $\mathbf{D}^i \forall i$ all have N atoms ...

The Iteration-Tuned and Aligned Dictionary

Change the dictionary in each MP iteration.



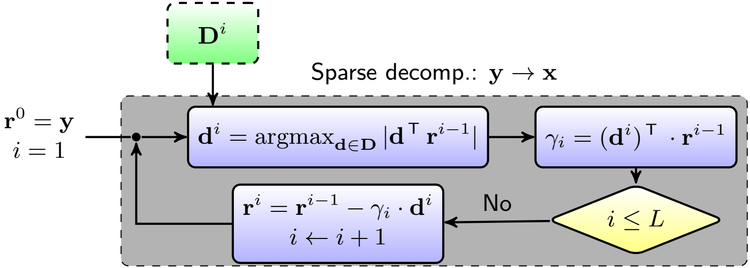
($\mathbf{D}$ is assumed to have normalized columns.)

If $\mathbf{D}$ and $\mathbf{D}^i \forall i$ all have N atoms ...

$\Rightarrow$ *ITD* $L\times$ *more overcomplete.*

The Iteration-Tuned and Aligned Dictionary

Change the dictionary in each MP iteration.



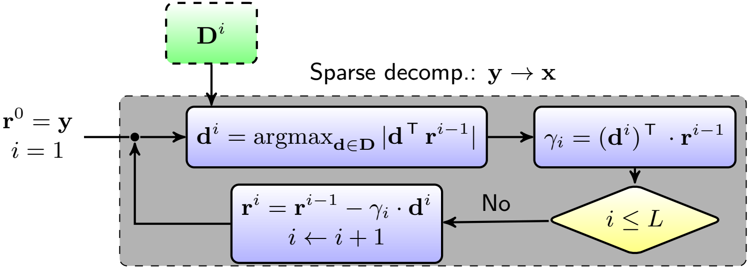
($\mathbf{D}$ is assumed to have normalized columns.)

If $\mathbf{D}$ and $\mathbf{D}^i \forall i$ all have N atoms ...

- $\Rightarrow$ *ITD* $L\times$ *more overcomplete.*
- $\Rightarrow$ *Comparable complexity under MP.*

The Iteration-Tuned and Aligned Dictionary

Change the dictionary in each MP iteration.

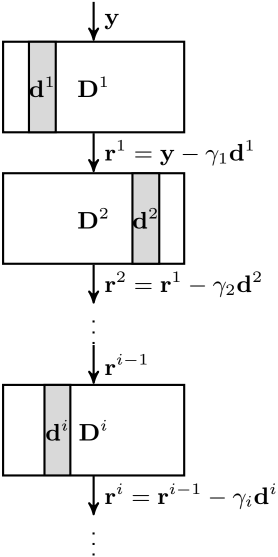


($\mathbf{D}$ is assumed to have normalized columns.)

If $\mathbf{D}$ and $\mathbf{D}^i \forall i$ all have N atoms ...

- $\Rightarrow$ *ITD $L\times$ more overcomplete.*
- $\Rightarrow$ *Comparable complexity under MP.*
- $\Rightarrow$ *Comparable atom index coding rate $log_2(N)$ (fixed-length code).*

The Iteration-Tuned and Aligned Dictionary



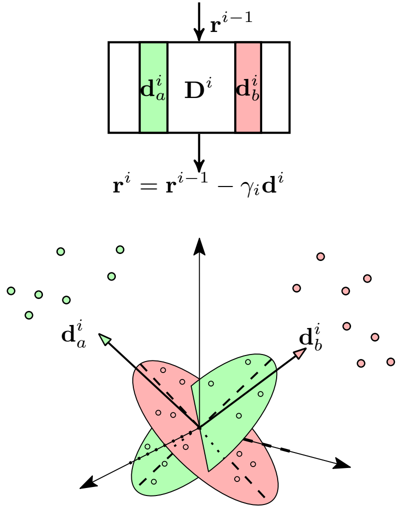
Iteration-Tuned Dictionary (ITD)

- *Layered* structure, one $\mathbf{D}^i$ per layer i.
- *Training?* Top-down approach simplifies training.

The Iteration-Tuned and Aligned Dictionary

Alignment of Residual Subspaces:

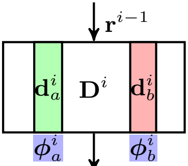
- Residual subspaces of a given atom are of reduced dimensionality.
- Union of residual subspaces spans entire space.
- Use rotation matrix to



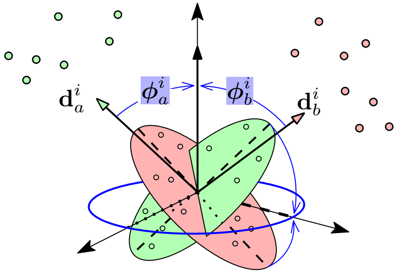
The Iteration-Tuned and Aligned Dictionary

Alignment of Residual Subspaces:

- Residual subspaces of a given atom are of reduced dimensionality.
- Union of residual subspaces spans entire space.
- Use rotation matrix to
  - $\Rightarrow$ Rotate residual subspaces to align them.
  - $\Rightarrow$ Align also their principal components.

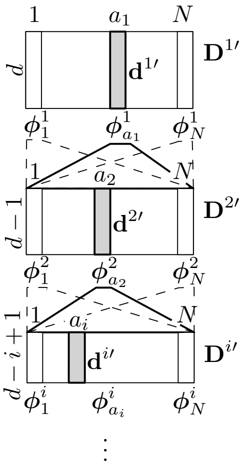


$$\mathbf{r}^i = \phi_a^{i\,\mathsf{T}} \left(\mathbf{r}^{i-1} - \gamma_i \mathbf{d}^i\right)$$



The Iteration-Tuned and Aligned Dictionary (ITAD)

- One alignment matrix per atom.
- Each dictionary exists in *reduced residual space*: $\mathbf{D}^{i\prime} \in \mathbb{R}^{d-i+1}$



Outline

Outline

1. Introduction – Sparse representations
2. Design issues
   - Dictionary choice
   - Atom distribution across image
3. Results

Block sparsity selection

- Standard approach (local):

$$\underset{L}{\operatorname{argmin}}\, L \text{ s.t. } |\mathbf{y} - \tilde{\mathbf{y}}^L|^2 \leq d \cdot \epsilon^2.$$

- Need to choose the sparsities L_n for each block $\mathbf{y}_n$, $n = 1, \ldots, B$.
- Global rate-distortion based formulation:

$$\underset{L_1,\ldots,L_B}{\operatorname{argmin}} \sum_{n=1}^{B} |\mathbf{y}_n - \tilde{\mathbf{y}}_n^{L_n}|^2 \text{ s.t. } \sum_{n=1}^{B} \mathrm{R}(\mathcal{Y}_n^{L_n}) \leq \Psi,$$

Approximate solution

1. Initialize all sparsities to zero: $\forall n, L_n = 0$.
2. Find block offering the largest reduction in distortion per bit

$$\beta = \underset{n}{\operatorname{argmax}} \frac{|\mathbf{y}_n - \tilde{\mathbf{y}}_n^{L_n}|^2 - |\mathbf{y}_n - \tilde{\mathbf{y}}_n^{L_n+1}|^2}{\mathrm{R}\left((a_{n,L_n+1}, \tilde{\gamma}_{n,L_n+1})\right)},$$

3. Incorporate new atom to the approximation $\tilde{\mathbf{y}}_\beta$ of the chosen block, $L_\beta \leftarrow L_\beta + 1$.
4. Repeat image bit budget is exhausted.

Outline

Dataset

- The data set: 545 different subjects from the FERET dataset.
  - *Train set:* 445 training images
  - *Test set:* 100 test images.

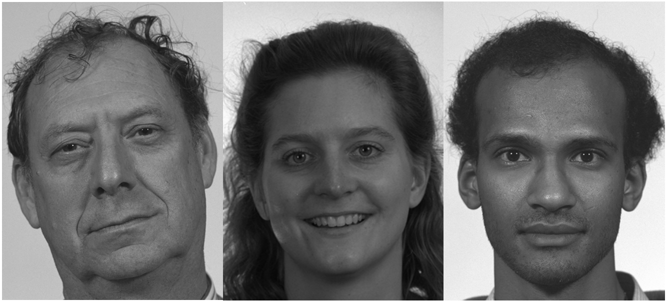
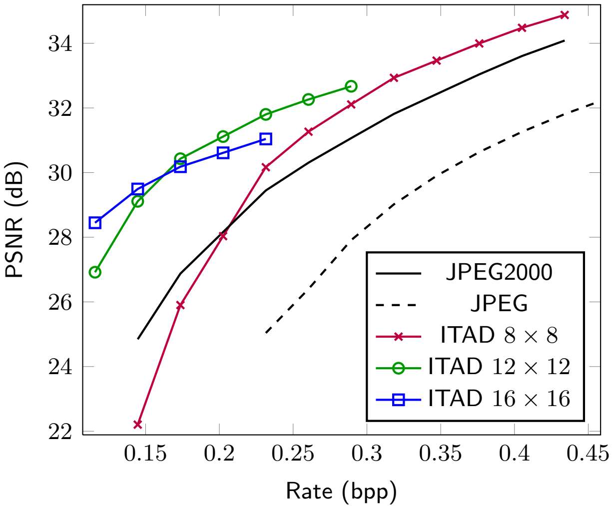


Figure 1: Sample images from the FERET dataset.

Full Image Codec – quantitative

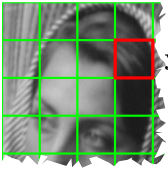


- Experimental (measured rate / distortion) rate-distortion curves.
- Atom/coefficient encoding: *very simple* (non-optimized), yet outperforms JPEG2000.
- Gain comes from ITAD transform.

Full Image Codec – qualitative

Original	JPEG2000	JPEG	ITAD

Conclusion



$$\mathbf{y} \approx \mathbf{D}\mathbf{x}$$

$$\mathbf{x} \iff \{(a_i, \gamma_i)\}_{i=1}^{L}$$

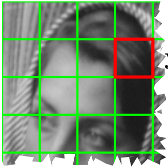
Design issues

1. What transformation to apply to $\mathbf{y}$?
2. Block sparsity selection ?
3. Atom / coefficient quantization?

Results

Large improvement over JPEG2000 (class of facial images).

Conclusion



$\mathbf{y} \approx \mathbf{D}\mathbf{x}$

$$\mathbf{x} \iff \{(a_i, \gamma_i)\}_{i=1}^{L}$$

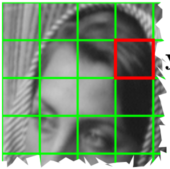
Design issues

1. What transformation to apply to $\mathbf{y}$? *ITAD.*
2. Block sparsity selection ?
3. Atom / coefficient quantization?

Results

Large improvement over JPEG2000 (class of facial images).

Conclusion



$\mathbf{y} \approx \mathbf{D}\mathbf{x}$

$\mathbf{x} \iff \{(a_i, \gamma_i)\}_{i=1}^{L}$

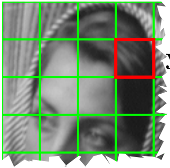
Design issues

1. What transformation to apply to $\mathbf{y}$? *ITAD.*
2. Block sparsity selection ? *New global, rate-distortion-based approach*.
3. Atom / coefficient quantization?

Results

Large improvement over JPEG2000 (class of facial images).

Conclusion



$\mathbf{y} \approx \mathbf{D}\mathbf{x}$

$\mathbf{x} \iff \{(a_i, \gamma_i)\}_{i=1}^{L}$

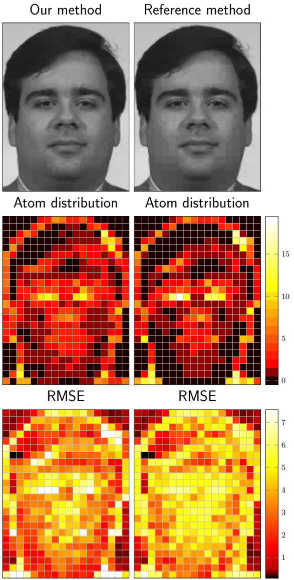
Design issues

1. What transformation to apply to $\mathbf{y}$? *ITAD.*
2. Block sparsity selection ? *New global, rate-distortion-based approach*.
3. Atom / coefficient quantization? Standard approach.

Results

Large improvement over JPEG2000 (class of facial images).

Global vs. local block sparsity selection



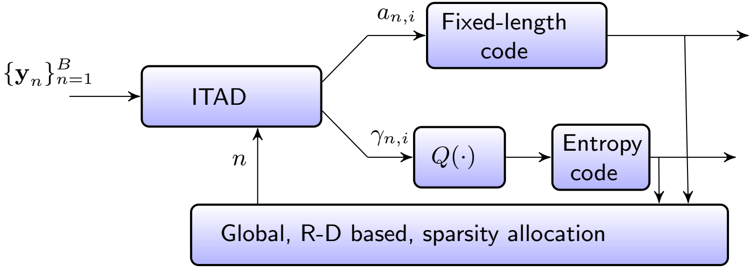
Proposed, global rate-distortion sparsity selection vs. reference RMSE-threshold method.

- More uniform atom distribution.
- Less uniform error distribution.

	Our method	Reference method
Rate	0.5 bpp	0.5 bpp
PSNR (dB)	36.40	35.77
Mean sparsity	2.07	1.92

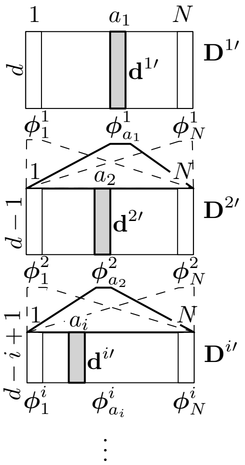
Full Image Codec – proposed

- *Block mean:* encoded with DPCM + entropy encoder.
- *Mean-removed blocks* $\mathbf{y}$ are encoded with ITAD transform.
- *Very simple codec!* Gain due to ITAD transform.



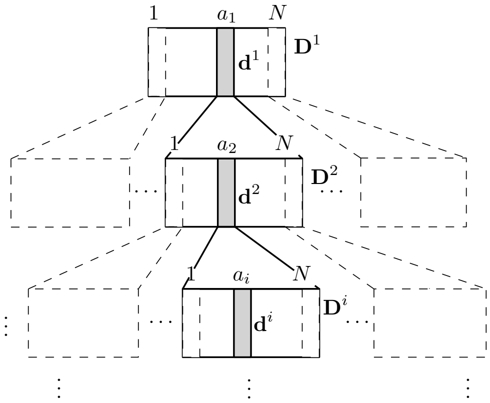
The Iteration-Tuned and Aligned Dictionary

- One alignment matrix per atom.
- Each dictionary exists in *reduced residual space*: $\mathbf{D}^{i\prime} \in \mathbb{R}^{d-i+1}$
- *Signal space* dictionaries are $\mathbf{D}^{i} = (\phi^{1}_{a_1} \cdot \ldots \cdot \phi^{i-1}_{a_{i-1}})\mathbf{D}^{i\prime} \in \mathbb{R}^{d}$
- ... and define a *tree structure*.

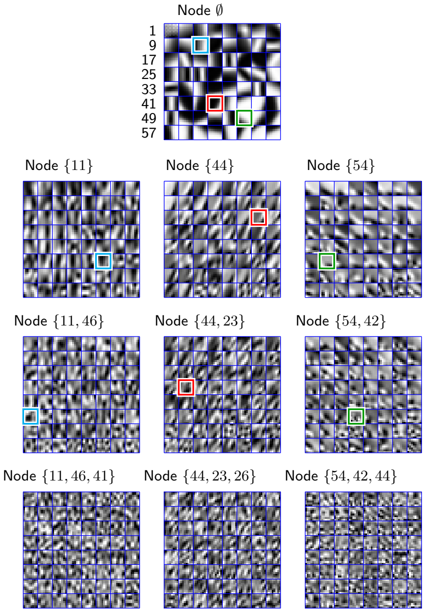


The Iteration-Tuned and Aligned Dictionary

- One alignment matrix per atom.
- Each dictionary exists in *reduced residual space*: $\mathbf{D}^{i\prime} \in \mathbb{R}^{d-i+1}$
- *Signal space* dictionaries are $\mathbf{D}^{i} = (\phi^{1}_{a_1} \cdot \ldots \cdot \phi^{i-1}_{a_{i-1}})\mathbf{D}^{i\prime} \in \mathbb{R}^{d}$
- ... and define a *tree structure*.



ITAD Example



- Three paths through the ITAD (signal-space) tree, layers $1, \ldots, 4$.
- 64 atoms per component dictionary / 8×8 blocks.
- Dictionaries display frequential hierarchy / parent atom dependence.