Transform-Domain Image Classification

"Image Classification: Modern Approaches"

Giacomo Boracchi, Alessandro Giusti
DEIB, Politecnico di Milano

February, 14th, 2018

giacomo.boracchi@polimi.it
home.deib.polimi.it/boracchi/

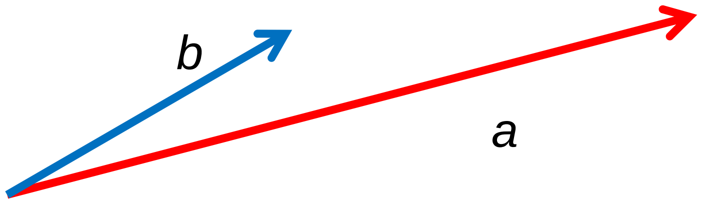
Preliminary Notation

A basis $\{e_i\}_{i=1,d}$ of a vector space V is a set of linearly independent vectors that spans (i.e. that can represent with linear combinations) the whole V.

Preliminary Notation

A basis $\{e_i\}_{i=1,d}$ of a vector space V is a set of linearly independent vectors that spans (i.e. that can represent with linear combinations) the whole V.

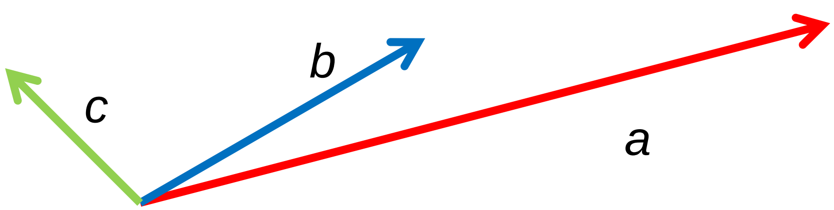
- N, the number of elements d of the basis is said the dimension of V (in principle it could not be finite)
- any other element can be represented as a linear combination of the basis atoms



Preliminary Notation

A basis $\{e_i\}_{i=1,d}$ of a vector space V is a set of linearly independent vectors that spans (i.e. that can represent with linear combinations) the whole V.

- N, the number of elements d of the basis is said the dimension of V (in principle it could not be finite)
- any other element can be represented as a linear combination of the basis atoms

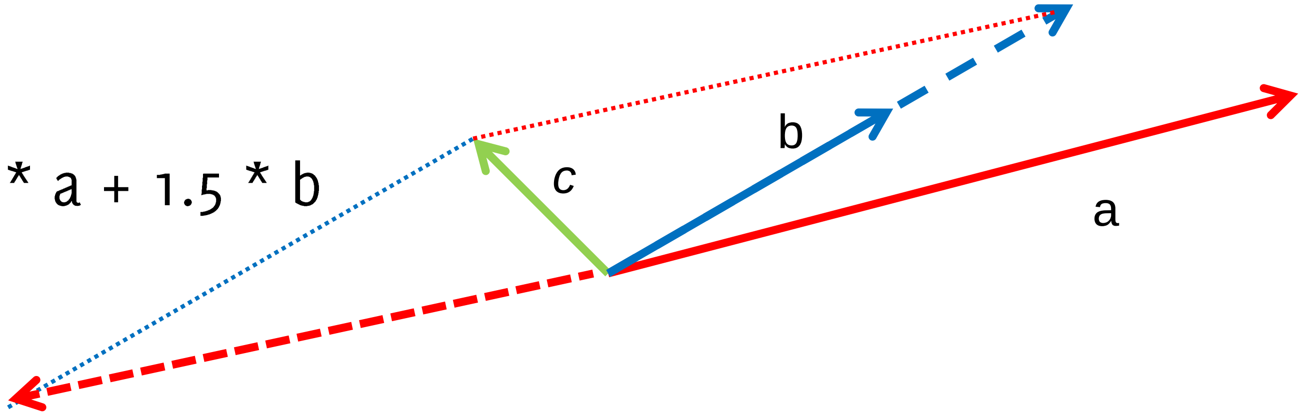


Preliminary Notation

A basis $\{e_i\}_{i=1,d}$ of a vector space V is a set of linearly independent vectors that spans (i.e. that can represent with linear combinations) the whole V.

- N, the number of elements d of the basis is said the dimension of V (in principle it could not be finite)
- any other element can be represented as a linear combination of the basis atoms

c = (-1) * a + 1.5 * b

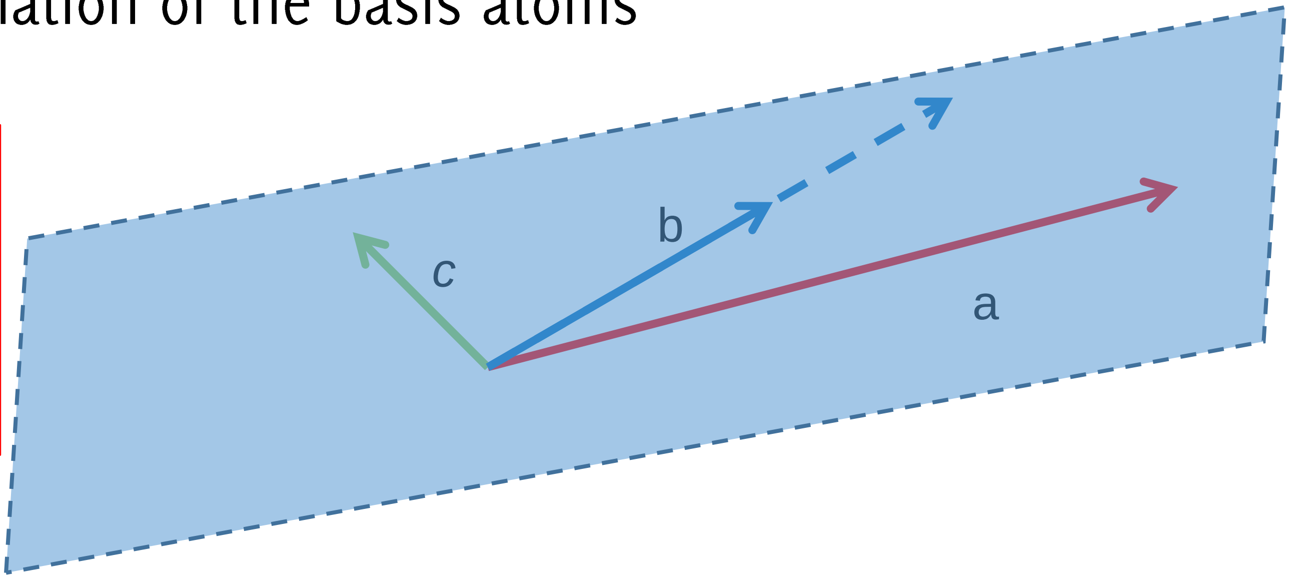


Preliminary Notation

A basis $\{e_i\}_{i=1,d}$ of a vector space V is a set of linearly independent vectors that spans (i.e. that can represent with linear combinations) the whole V.

- N, the number of elements d of the basis is said the dimension of V (in principle it could not be finite)
- any other element can be represented as a linear combination of the basis atoms

We can represent all the vector in this plane

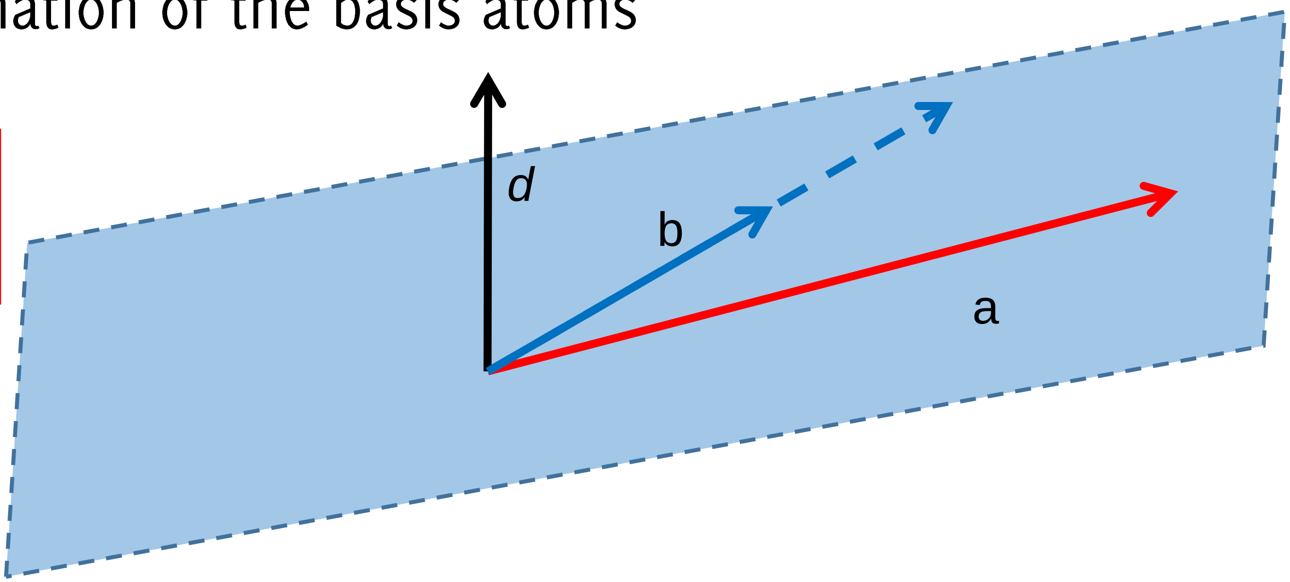


Preliminary Notation

A basis $\{e_i\}_{i=1,d}$ of a vector space V is a set of linearly independent vectors that spans (i.e. that can represent with linear combinations) the whole V.

- N, the number of elements d of the basis is said the dimension of V (in principle it could not be finite)
- any other element can be represented as a linear combination of the basis atoms

But not the orthogonal ones



Preliminary Notation

Any vector space

- admit infinite basis,
- each element of the vector space admit a unique representation w.r.t. each basis.

$$v = \sum_{i=1}^{N} a_i b_i \quad a_i \in \Re\,,\ \forall v \in V$$

In signal/image processing, basis allows us to represent signal or images by means of their coefficients

$$v \rightarrow \{a_i\}_{i=1,..,N}$$

but....

... if one needs a basis to represent a signal, how did we manage them so far?

Which is the canonical basis for digital signals/images?

116	23	33
16	3	73
5	4	30

but....

... if one needs a basis to represent a signal, how did we manage them so far?

Which is the canonical basis for digital signals/images?

116	23	33
16	3	73
5	4	30

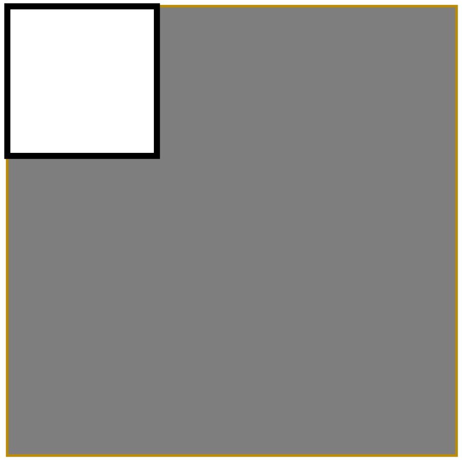
1	0	0
0	0	0
0	0	0

but....

... if one needs a basis to represent a signal, how did we manage them so far?

Which is the canonical basis for digital signals/images?

116	23	33
16	3	73
5	4	30

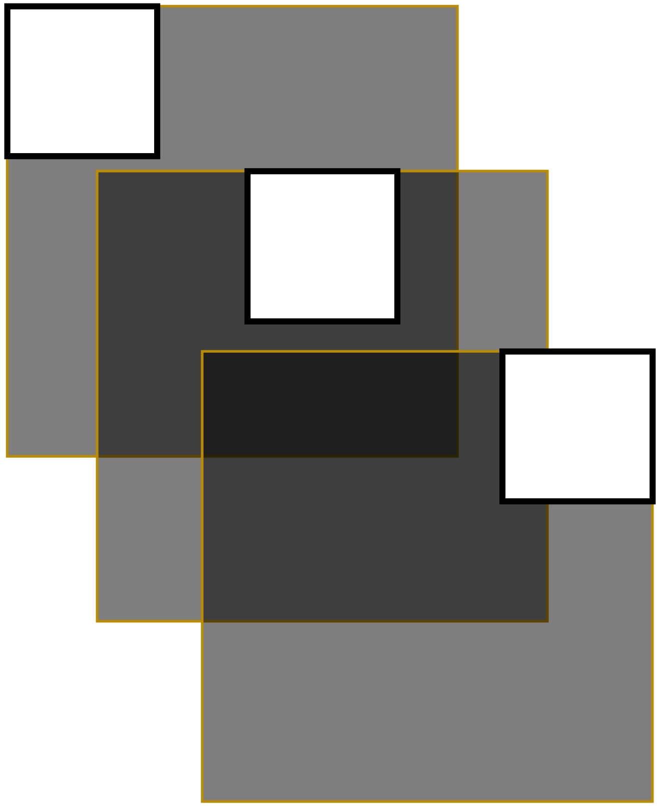


but....

... if one needs a basis to represent a signal, how did we manage them so far?

Which is the canonical basis for digital signals/images?

116	23	33
16	3	73
5	4	30

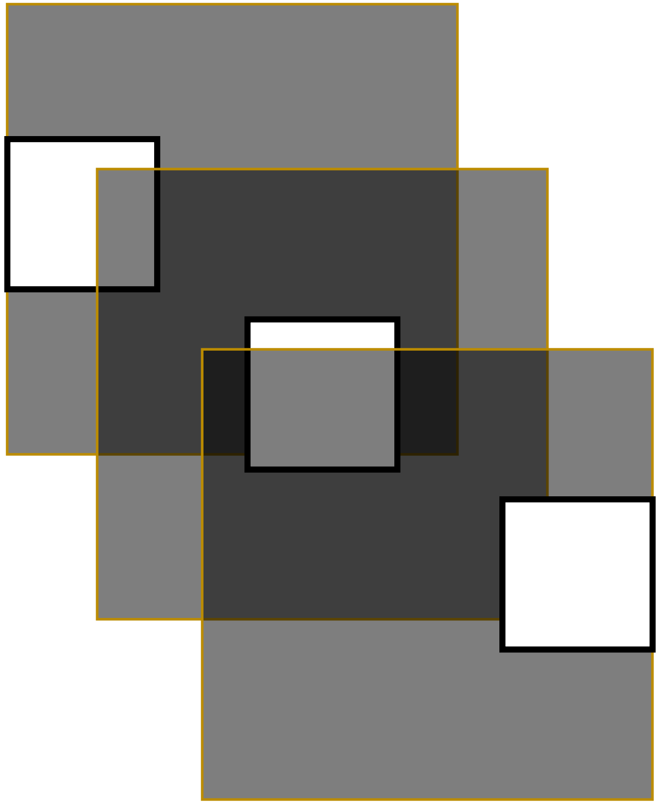


but....

... if one needs a basis to represent a signal, how did we manage them so far?

Which is the canonical basis for digital signals/images?

116	23	33
16	3	73
5	4	30

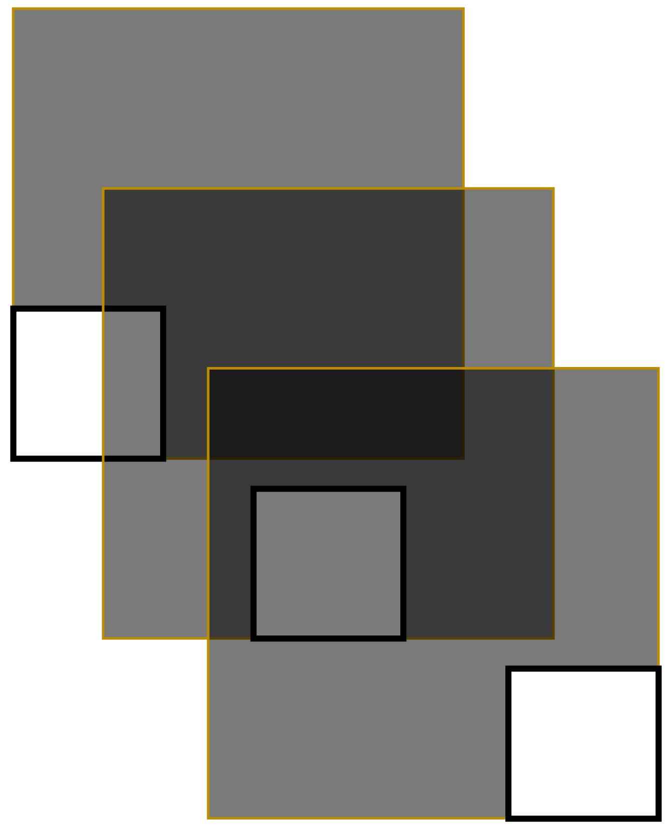


but....

... if one needs a basis to represent a signal, how did we manage them so far?

Which is the canonical basis for digital signals/images?

116	23	33
16	3	73
5	4	30



Thus the canonical basis

Uses each coefficient to represent a pixel:

- all coefficients are equally meaningful
- thus, it is not usefull at all for compression
- it corresponds to the canonical basis of $\mathbb{R}^d$

Are there basis that ease image processing tasks?

Orthonormal Basis

When V is an Hilbert space, the inner product allows us to define the orthonormal basis as a basis having orthonormal elements, i.e. $< b_i, b_j >= \delta_{i,j} \quad \forall i, j$

For orthonormal basis we have that if $\{b_i\}_{i=1,\dots N}$ is a orthonormal basis,

$$v = \sum_{i=1}^{N} a_i b_i \quad a_i \in \Re\,,\ \forall v \in V$$

Means that $a_i =< v, b_i >$

We know how to change basis, i.e. how to transform the signals/images w.r.t. a different basis.

These transforms are linear (thus invertible)

2D Fourier Transform

The (u,v)-element of the 2D Fourier basis is defined as

$$e^{-i2\pi(ux+vy)} = \cos(2\pi(ux+vy)) + i\sin(2\pi(ux+vy))$$

Each Fourier coefficient is computed with an inner product with the correspondinf function.

The Fourier basis functions are constant where $y = -\frac{ux}{v} + c$

The Fourier basis functions have unlimited support.

The Fourier Transform is invertible (it is an orthonormal transform)

Fourier domain is also called frequency domain.

2D Discrete Fourier Transform

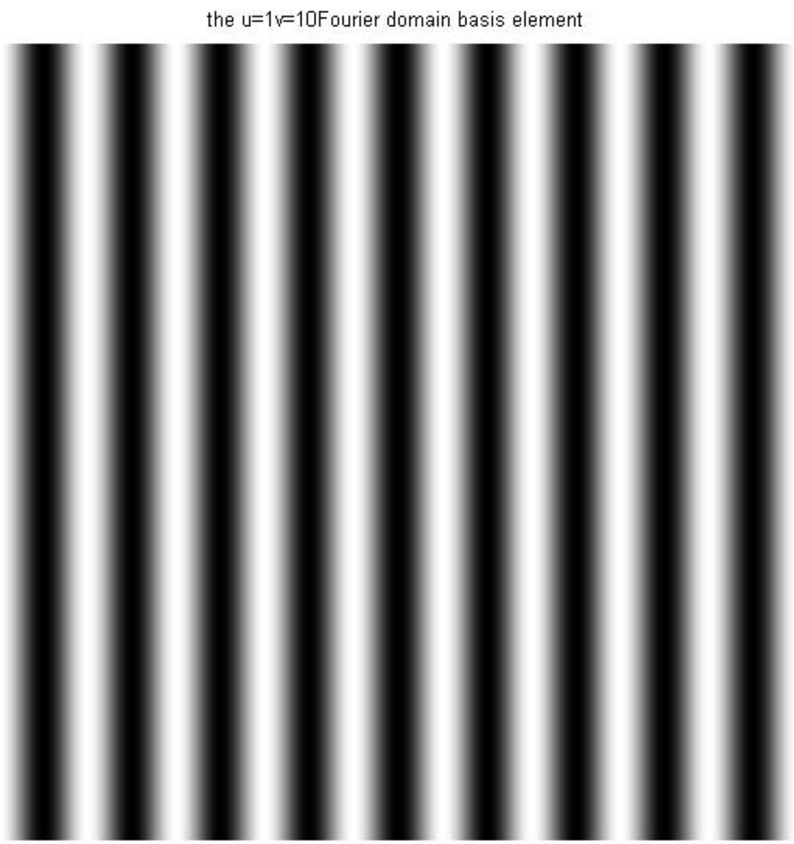
The Discrete Fourier Transform (DFT) is defined as

$$F[m,n] = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} f[k,l] e^{-\pi i \left(\frac{km}{M} + \frac{\ln}{N} \right)}$$

The Fourier Transform admit a fast implementation (FFT) when the signal/image sizes are powers of 2.

Frequency and Orientation of the 2D Fourier Basis Elements

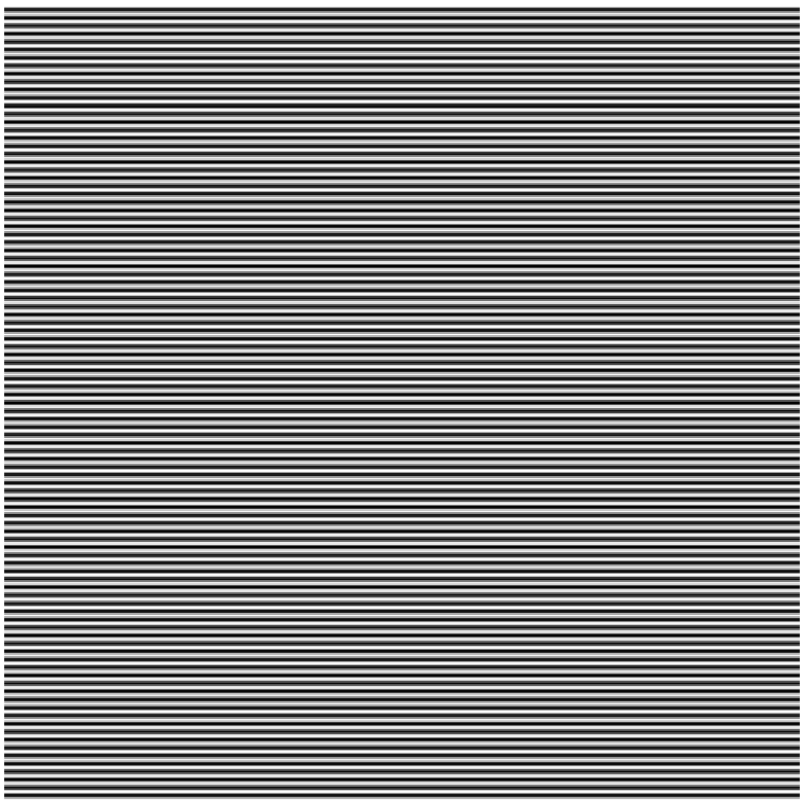
$$b_{u,v} \quad u = 1, v = 10$$



2D Fourier Basis Elements

Frequency and Orientation of the 2D Fourier Basis Elements

$$b_{u,v} \quad u = 100, v = 1$$



Frequency and Orientation of the 2D Fourier Basis Elements

$$b_{u,v} \quad u = 10, v = 10$$

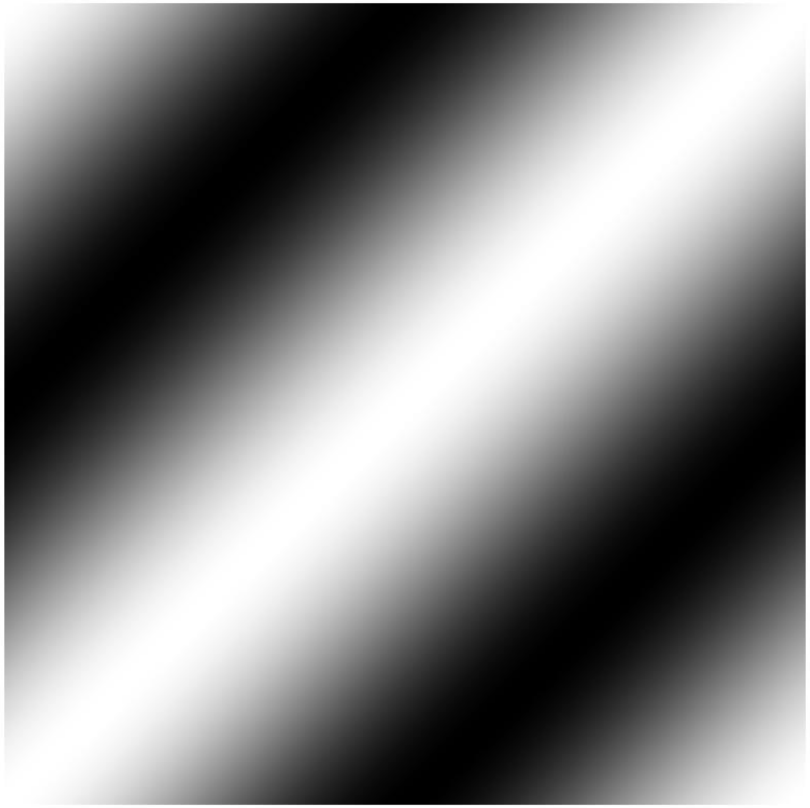
2D Fourier Basis Elements

Frequency and Orientation of the 2D Fourier Basis Elements

$$b_{u,v} \quad u = 2, v = 2$$

(note that in matlab Fourier coefficients are indexed starting from 1)

the u=2v=2Fourier domain basis element



Properties of Fourier Coefficients

The Fourier Transform is a Global Transform: every image pixel influences the value of each Fourier coefficient.

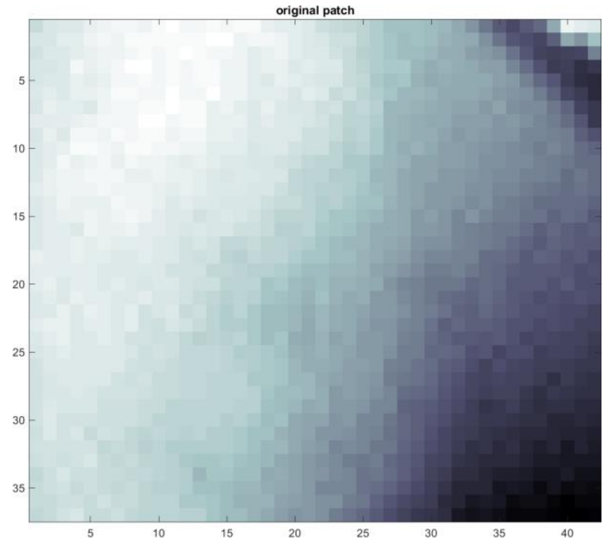
The (0,0) coefficient corresponds to the image average.

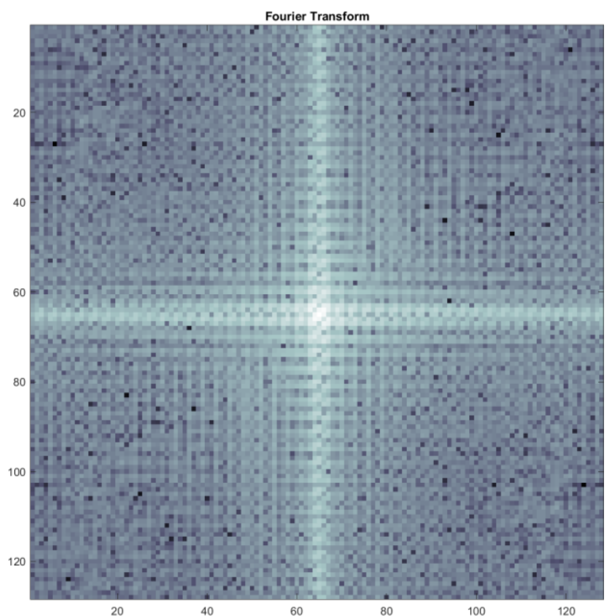
Thus not all coefficients are "equally meaningful" in the image representation.

Fourier Transform Properties

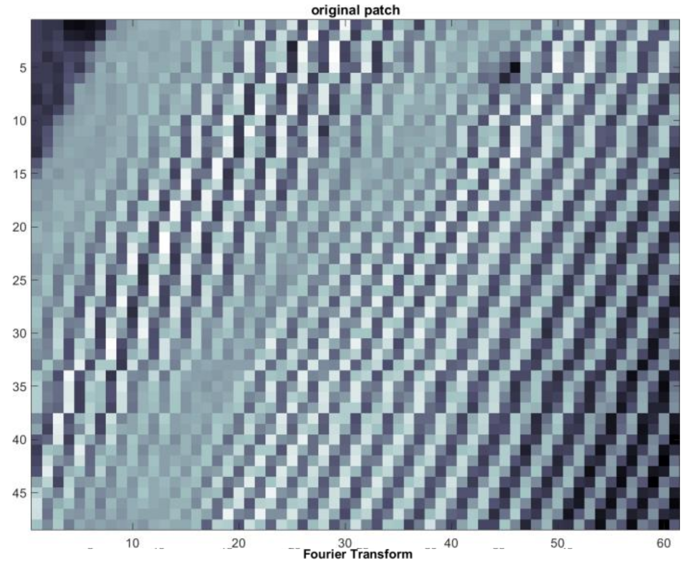
Property	Function	Fourier Transform	
	$f(t)$	$\hat{f}(\omega)$	
Inverse	$\hat{f}(t)$	$2\pi f(-\omega)$	(2.15)
Convolution	$f_1 \star f_2(t)$	$\hat{f}_1(\omega)\hat{f}_2(\omega)$	(2.16)
Multiplication	$f_1(t)f_2(t)$	$\frac{1}{2\pi}\hat{f}_1 \star \hat{f}_2(\omega)$	(2.17)
Translation	$f(t-u)$	$e^{-iu\omega}\hat{f}(\omega)$	(2.18)
Modulation	$e^{i\xi t}f(t)$	$\hat{f}(\omega-\xi)$	(2.19)
Scaling	$f(t/s)$	$\lvert s\rvert\hat{f}(s\omega)$	(2.20)
Time derivatives	$f^{(p)}(t)$	$(i\omega)^p\hat{f}(\omega)$	(2.21)
Frequency derivatives	$(-it)^p f(t)$	$\hat{f}^{(p)}(\omega)$	(2.22)
Complex conjugate	$f^*(t)$	$\hat{f}^*(-\omega)$	(2.23)
Hermitian symmetry	$f(t) \in \mathbb{R}$	$\hat{f}(-\omega) = \hat{f}^*(\omega)$	(2.24)

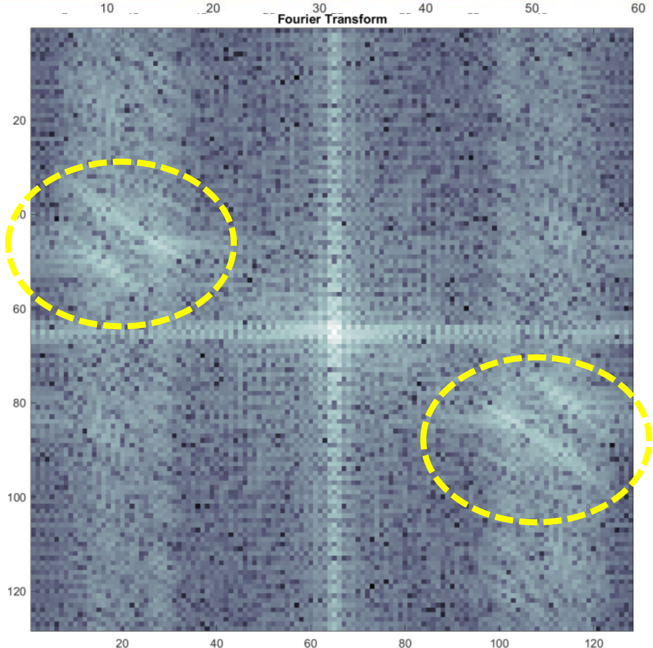
Example


original patch
5
10
15
20
25
30
35
40


Fourier Transform
20
40
60
80
100
120

Example





Example

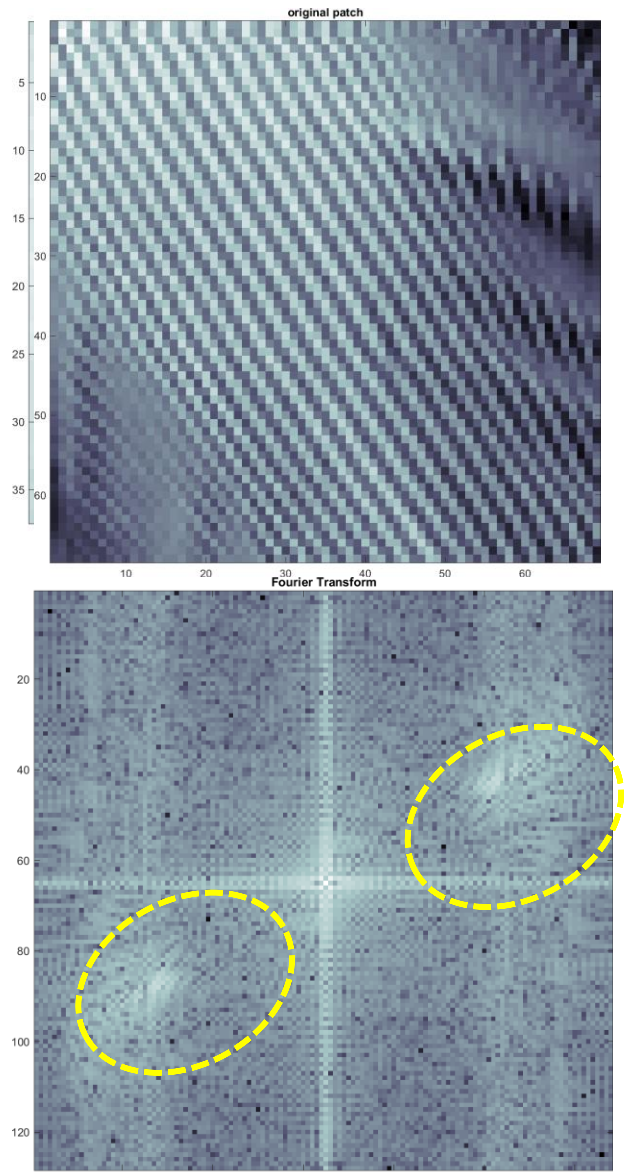


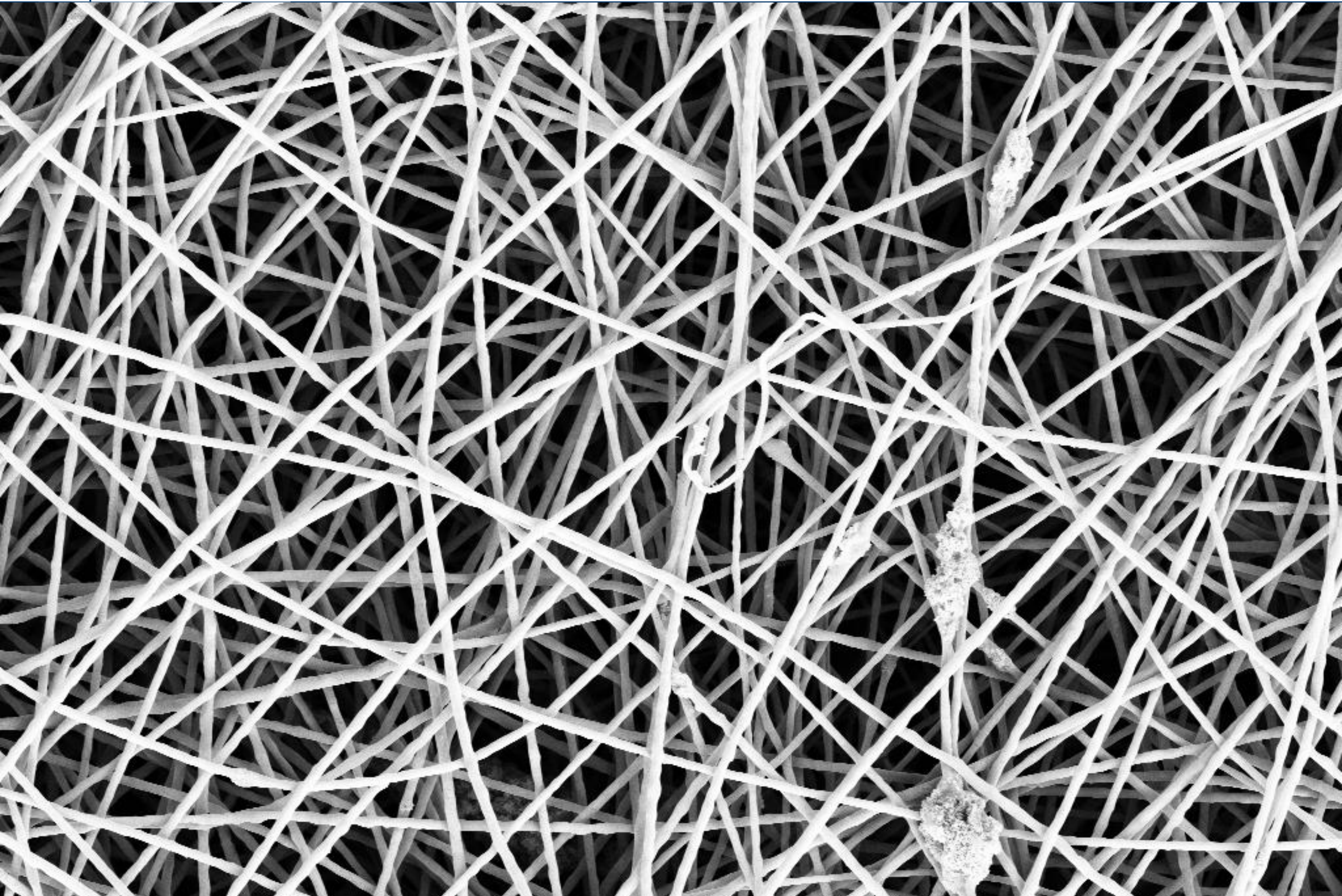
Image Classification by Transform-domain features

The overall energy in specific spectral bands can be used for classification purposes, since it is a discriminative information

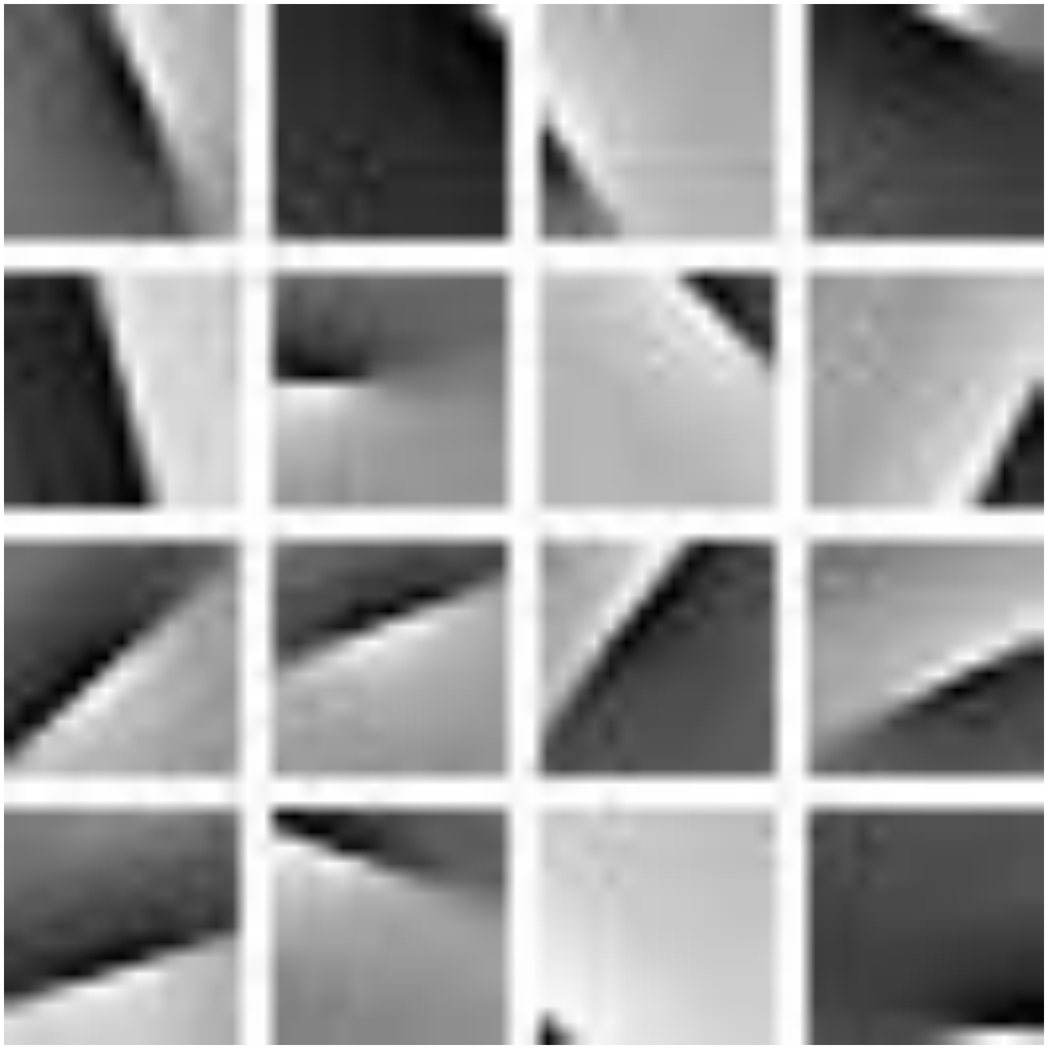
Which of the coefficients is useful have to be defined manually, as hand-crafted features

It is possible to learn class-specific or discriminative representations. However, these can not be a basis (which are not flexible enough) and we have to resort to redundant set of generators: dictionaries yielding sparse representations

Data-Driven (Sparse) Representations (test image example)



Example of
Learned atoms



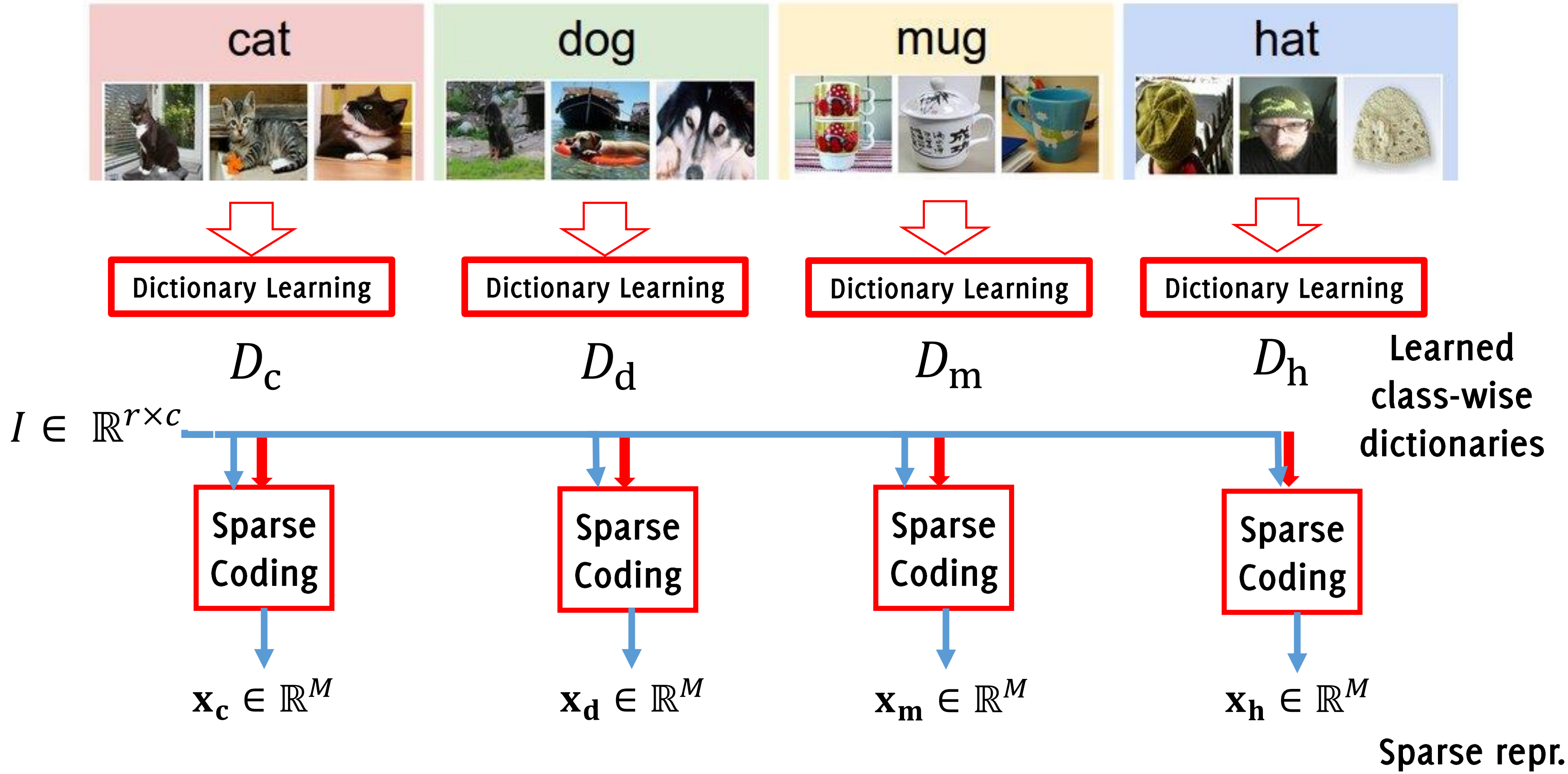
Example of
training patches

Carrera et al. Defect Detection in SEM Images of Nanofibrous Materials, ***IEEE TII 2016***

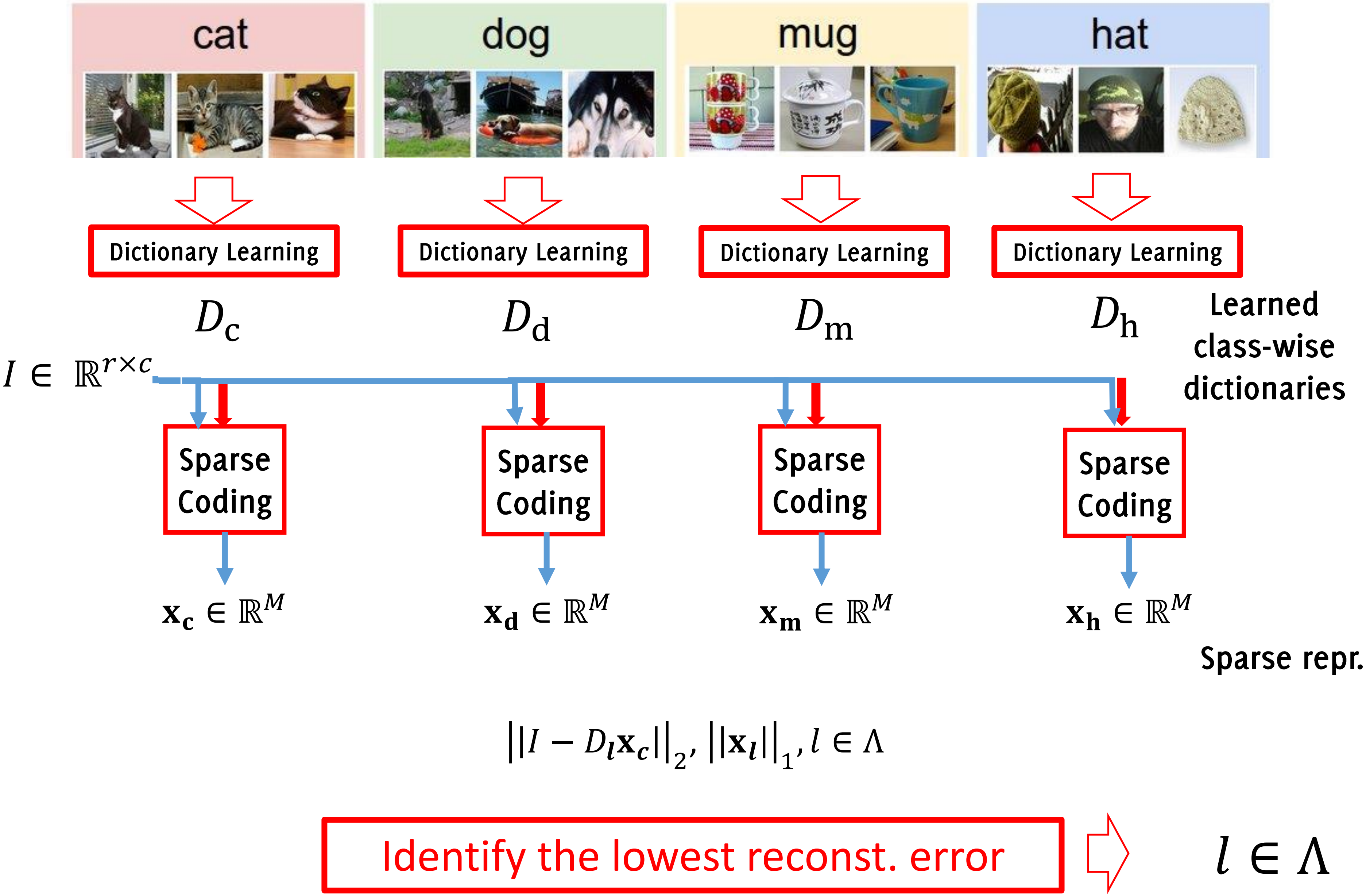
IC by Learned Sparse Representations

Dictionary Learning	Dictionary Learning	Dictionary Learning	Dictionary Learning	
D_c	D_d	D_m	D_h	**Learned class-wise dictionaries**

IC by Learned Sparse Representations



IC by Learned Sparse Representations



IC by Learned Sparse Representations (2)

Dictionary Learning | Dictionary Learning | Dictionary Learning | Dictionary Learning

D_{c} D_{d} D_{m} D_{h}

Learned class-wise dictionaries

$$D = [D_{\mathrm{c}}, D_{\mathrm{d}}, D_{\mathrm{m}}, D_{\mathrm{h}}]$$

Dictionary of all classes

IC by Learned Sparse Representations (2)

Dictionary Learning | Dictionary Learning | Dictionary Learning | Dictionary Learning

D_{c} D_{d} D_{m} D_{h}

Learned class-wise dictionaries

$$D = [D_{\mathrm{c}}, D_{\mathrm{d}}, D_{\mathrm{m}}, D_{\mathrm{h}}]$$

Dictionary of all classes

$I \in \mathbb{R}^{r \times c}$ → Sparse Coding

$$\mathbf{x} \in \mathbb{R}^{M} \; s.t. \; I \approx D\mathbf{x} \; and \; ||\mathbf{x}||_{\mathbf{0}} < \tau$$

Sparse repr.

Identify the lowest reconst. error → $l \in \Lambda$

IC by Learned Sparse Representations (3)

Dictionary Learning → D_c

Dictionary Learning → D_d

Dictionary Learning → D_m

Dictionary Learning → D_h

Learned class-wise dictionaries

$$D = [D_\mathrm{c}, D_\mathrm{d}, D_\mathrm{m}, D_\mathrm{h}]$$

Dictionary of all classes

$I \in \mathbb{R}^{r \times c}$ → **Sparse Coding**

$$\mathbf{x} \in \mathbb{R}^M \; s.t. \; I \approx D\mathbf{x} \; and \; ||\mathbf{x}||_0 < \tau$$

Sparse repr.

Classifier → $l \in \Lambda$

IC by Learned Sparse Representations

Advantages:

- Entirely data-driven classification
- Interpretability (it's a synthesis model, reconstruct data)
- Can be used for one-class classification (anomaly detection)

Cons:

- Dictionary learning and sparse coding is meant for relatively low-dimensional data ($d \approx 10^2 - 10^3$).
- Not effective in many visual recognition tasks (e.g. on natural images) which are easily performed by humans
- Synthesis models are typically more computationally demanding than analysis ones at test time
- Representations learned in an unsupervised manner (although task-driven dictionary learning algorithms exist)