CS 534: Computer Vision
Texture

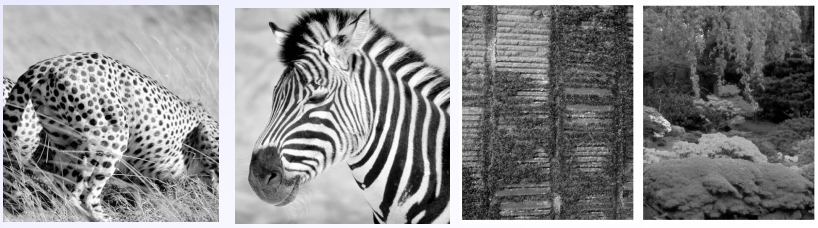
Ahmed Elgammal
Dept of Computer Science
Rutgers University

Outlines

- What is Texture
- Co-occurrence matrices for texture
- Spatial Filtering approach
- Multiresolution processing, Gaussian Pyramids and Laplacian Pyramids
- Gabor filters and oriented pyramids
- Local Binary Patterns
- CNN-based Texture models
- Texture Synthesis

Texture

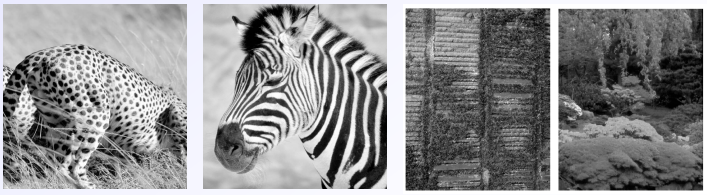
- What is texture ? Easy to recognize hard to define
  - Views of large number of small objects: grass, foliage, brush, pebbles, hair
  - Surfaces with patterns: spots, stripes, wood, skin
- Texture consists of organized patterns of quite regular sub-elements.
- Whether an effect is referred to as texture or not depends on the scale at which it is viewed.



Texture

Problems related to Texture:

- Texture analysis: how to represent and model texture
- Texture segmentation: segmenting the image into components within which the texture is constant
- Texture synthesis: construct large regions of texture from small example images
- Shape from texture: recovering surface orientation or surface shape from image texture.



Shape from Texture

- Texture looks different depending on the viewing angle.
- Texture is a good cue for shape and orientation
- Humans are very good at that.

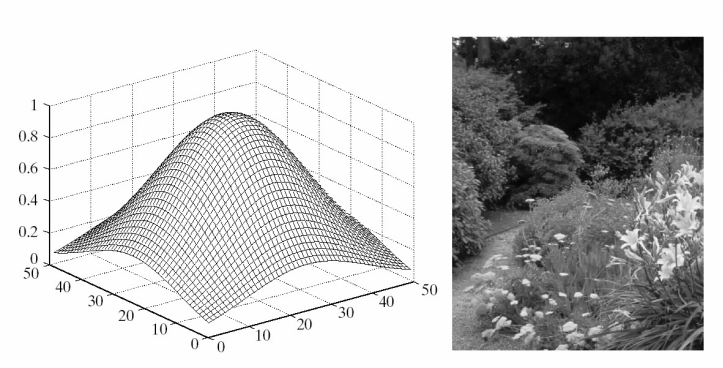


FIGURE 6.1: Although texture is difficult to define, it has some important and valuable properties. In this image, there are many repeated elements (some leaves form repeated "spots"; others, and branches, form "bars" at various scales; and so on). Our perception of the material is quite intimately related to the texture (what would the surface feel like if you ran your fingers over it? what is soggy? what is prickly? what is smooth?). Notice how much information you are getting about the type of plants, their shape, the shape of free space, and so on, from the textures. *Geoff Brightling © Dorling Kindersley, used with permission.*

Notice how the change in pattern elements and repetitions is the main difference between different textured surfaces (the plants, the ground, etc.)

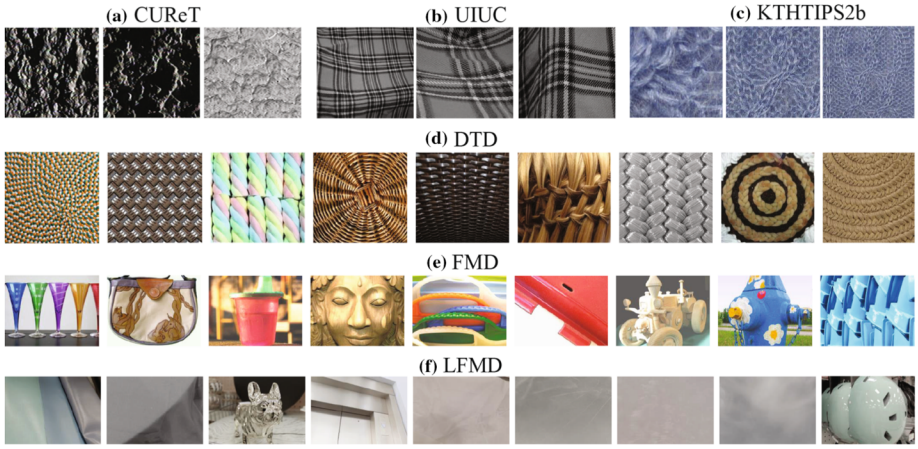
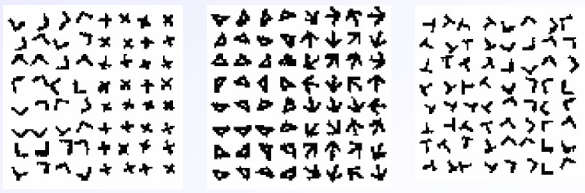


Fig. 3 Illustrations of challenges in texture recognition. Dramatic intraclass variations: **a** illumination variations, **b** view point and local nonrigid deformation, **c** scale variations, and **d** different instances from the same category. Small interclass variations make the problem harder still: **e** images from the FMD database, and **f** images from the LFMD database (photographed with a light-field camera). The reader is invited to identify the material category of the foreground surfaces in each image in (**e**, **f**). The correct answers are (from left to right): **e** glass, leather, plastic, wood, plastic, metal, wood, metal and plastic; **f** leather, fabric, metal, metal, paper, leather, water, sky and plastic. Sect. 6 gives details regarding texture databases

Figure from [1]

Representing Texture

- What we should look for ?
- Texture consists of organized patterns of quite regular subelements. "Textons"
- Find the subelements, and represent their statistics
- Reason about their spatial layout.
- Problem: There is no known canonical set of textons.



Texture Analysis

Different approaches:

- Co-occurrence matrices (classical)
- Spatial Filtering
- Random Field Models
- Local Binary Patterns
- CNN-based methods

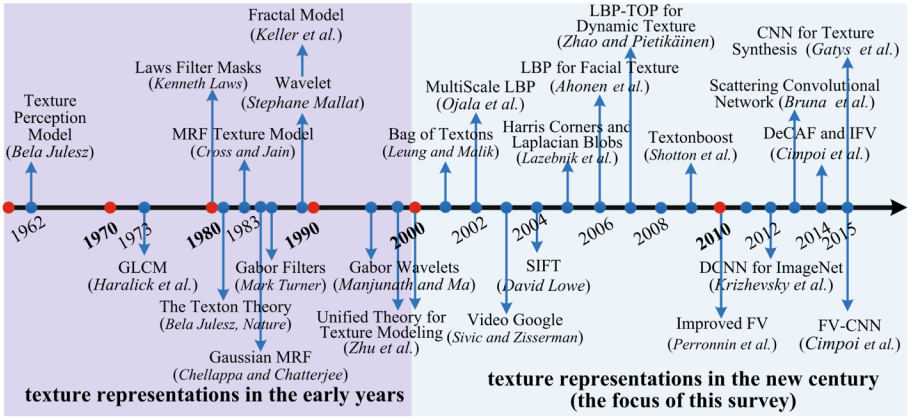


Fig. 2 The evolution of texture representation over the past decades (see discussion in Sect. 2.2)

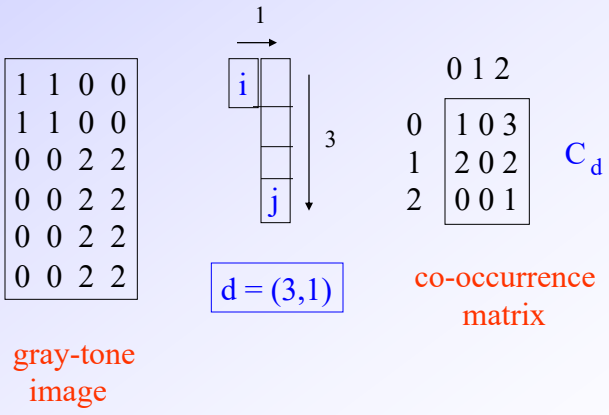
Figure from [1]

Co-occurrence Matrix Features

Objective: Capture spatial relations

A co-occurrence matrix is a 2D array C in which

- Both the rows and columns represent a set of possible image values
- $C_d(i,j)$ indicates how many times value i co-occurs with value j in a particular spatial relationship d.
- The spatial relationship is specified by a vector d = (dr,dc).



From C_d we can compute N_d , the normalized co-occurrence matrix, where each value is divided by the sum of all the values.

Co-occurrence Features

From Co-occurrence matrices extract some quantitative features:

$$Energy = \sum_i \sum_j N_d^2(i,j) \quad (7.7)$$

$$Entropy = -\sum_i \sum_j N_d(i,j) log_2 N_d(i,j) \quad (7.8)$$

$$Contrast = \sum_i \sum_j (i-j)^2 N_d(i,j) \quad (7.9)$$

$$Homogeneity = \sum_i \sum_j \frac{N_d(i,j)}{1+|i-j|} \quad (7.10)$$

$$Correlation = \frac{\sum_i \sum_j (i-\mu_i)(j-\mu_j) N_d(i,j)}{\sigma_i \sigma_j} \quad (7.11)$$

where μ_i, μ_j are the means and σ_i, σ_j are the standard deviations of the row and column

Disadvantages:

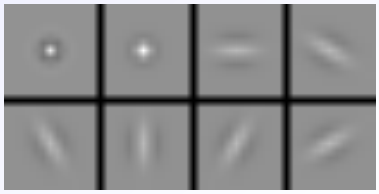
- Computationally expensive
- Sensitive to gray scale distortion (co-occurrence matrices depend on gray values)
- May be useful for fine-grain texture. Not suitable for spatially large textures.

Spatial Filtering Approaches

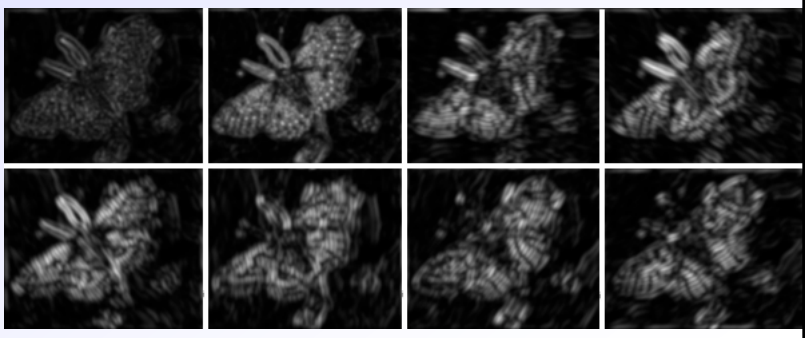
- Look for the subelements
- But what are the subelements, and how do we find them?
- Find subelements by applying filters, looking at the magnitude of the response
- Spots and bars detectors at various scales and orientations.

Typically:

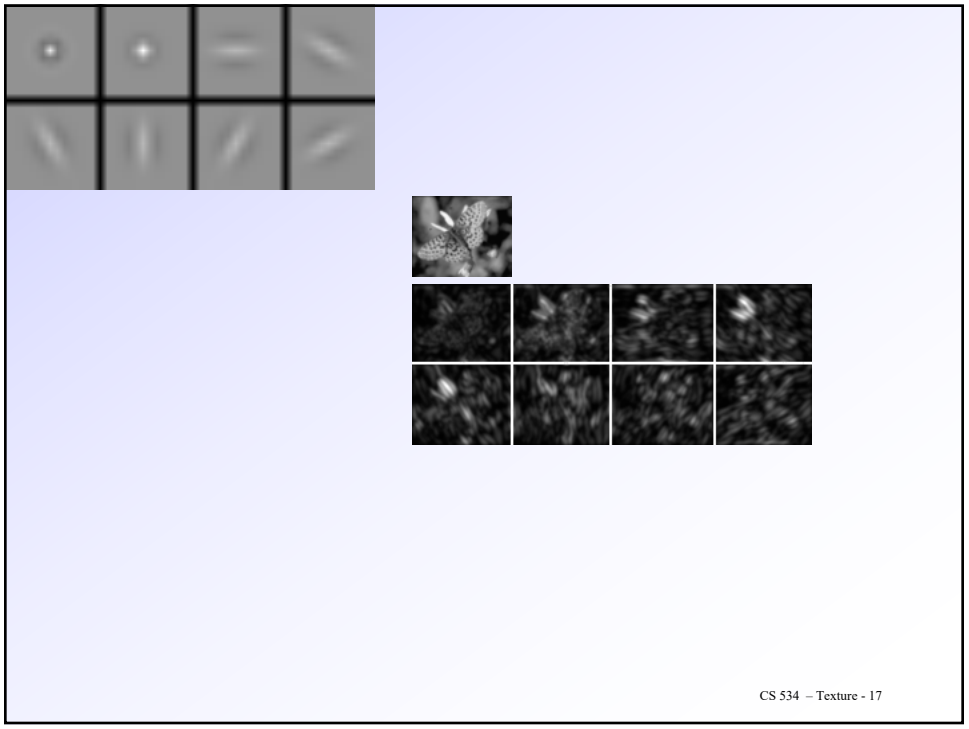
- "Spot" filters are Gaussians or weighted sums of concentric Gaussians.
- "Bar" filters are differentiating oriented Gaussians

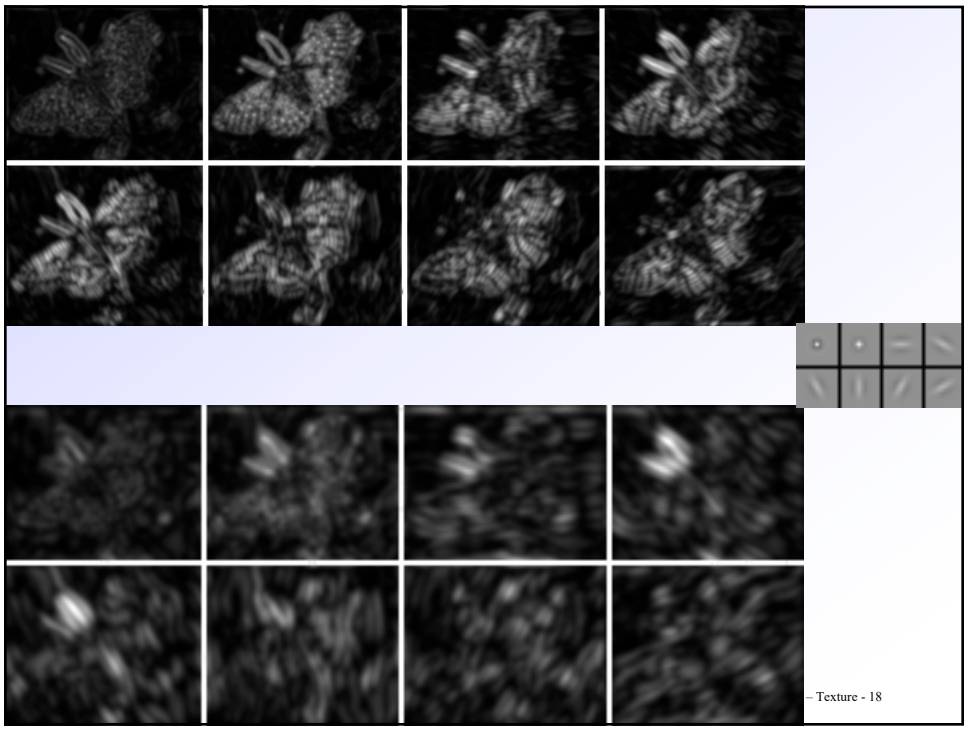


CS 534 – Texture - 15



CS 534 – Texture - 16


CS 534 – Texture - 17


– Texture - 18

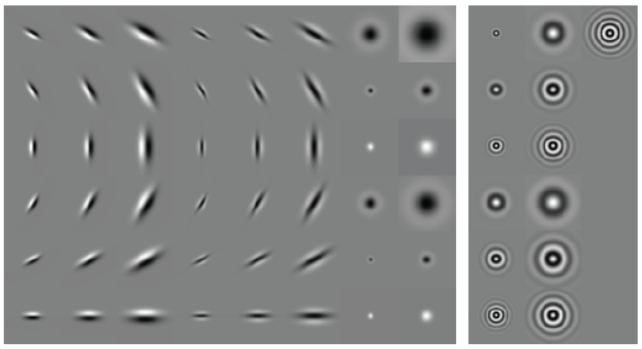


FIGURE 6.4: **Left** shows a set of 48 oriented filters used for expanding images into a series of responses for texture representation. Each filter is shown on its own scale, with zero represented by a mid-gray level, lighter values being positive, and darker values being negative. The left three columns represent edges at three scales and six orientations; the center three columns represent stripes; and the right two represent two classes of spots (with and without contrast at the boundary) at different scales. This is the set of filters used by Leung and Malik (2001). **Right** shows a set of orientation-independent filters, used by Schmid (2001), using the same representation (there are only 13 filters in this set, so there are five empty slots in the image). The orientation-independence property means that these filters look like complicated spots.

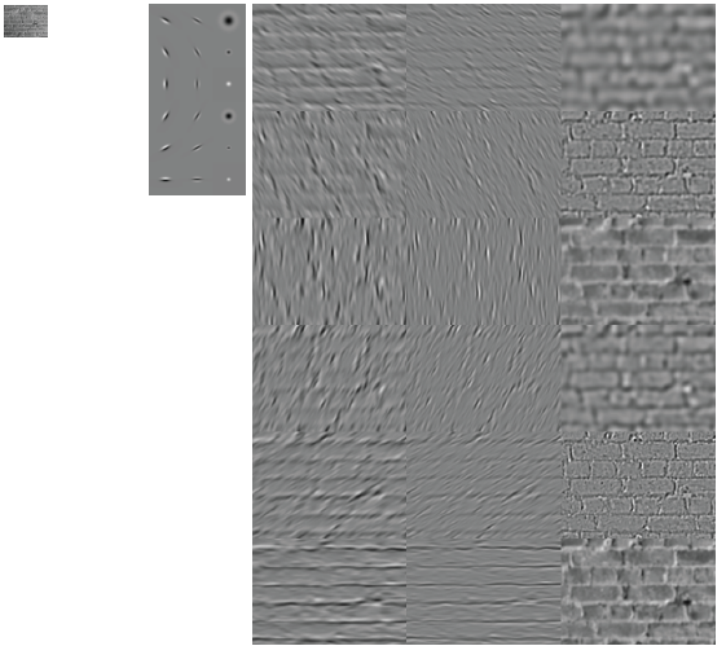


FIGURE 6.5: Filter responses for the oriented filters of Figure 6.4, applied to an image of

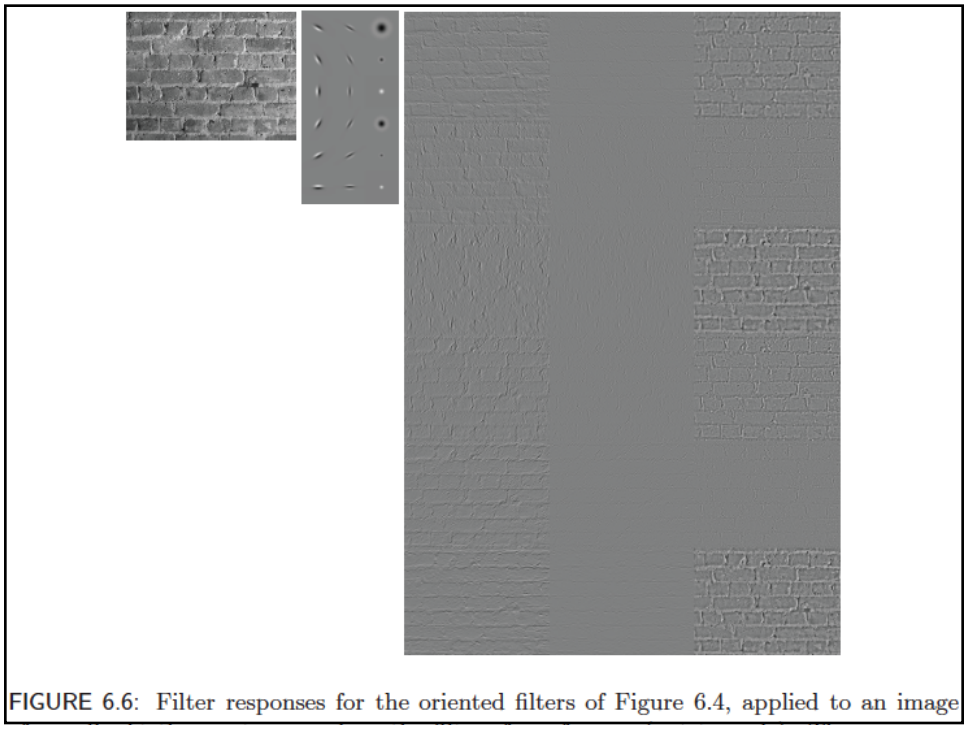
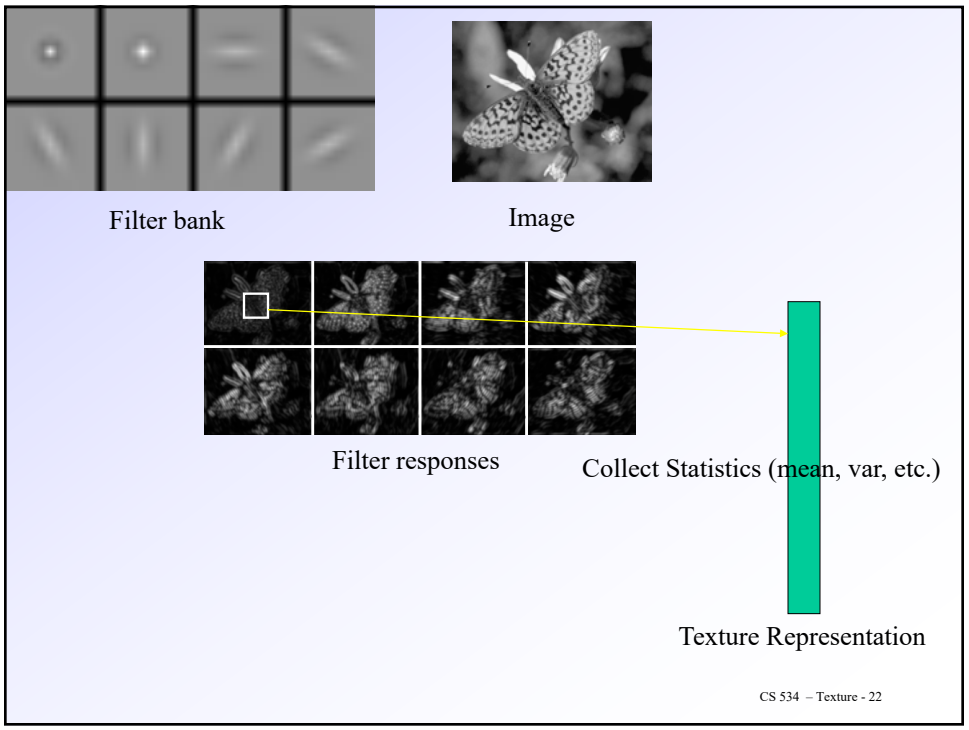


FIGURE 6.6: Filter responses for the oriented filters of Figure 6.4, applied to an image



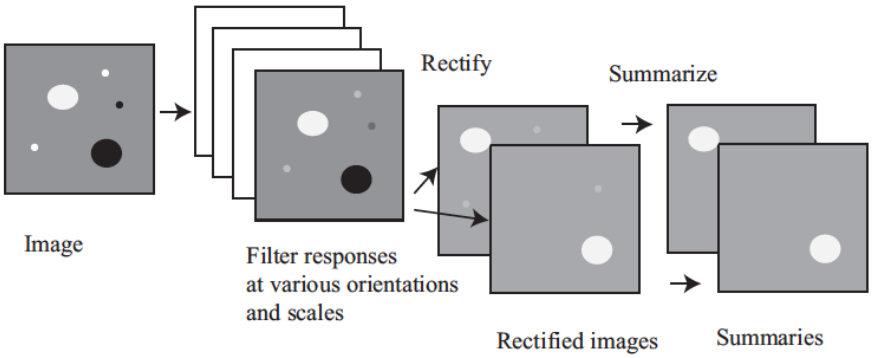
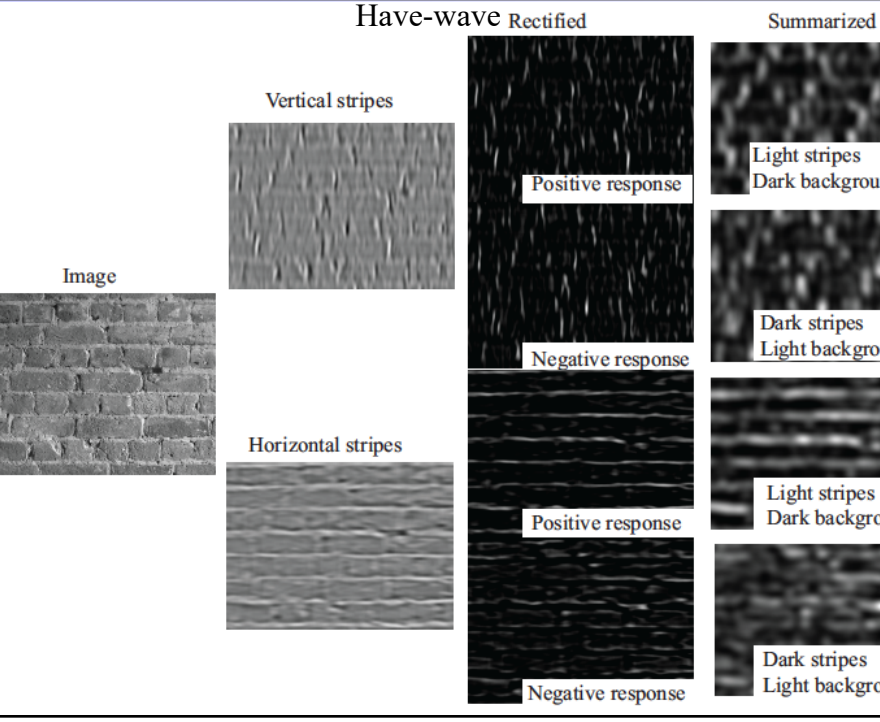
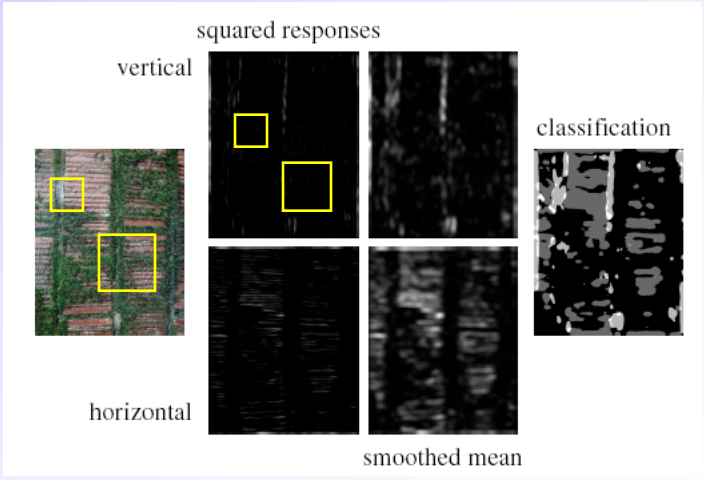


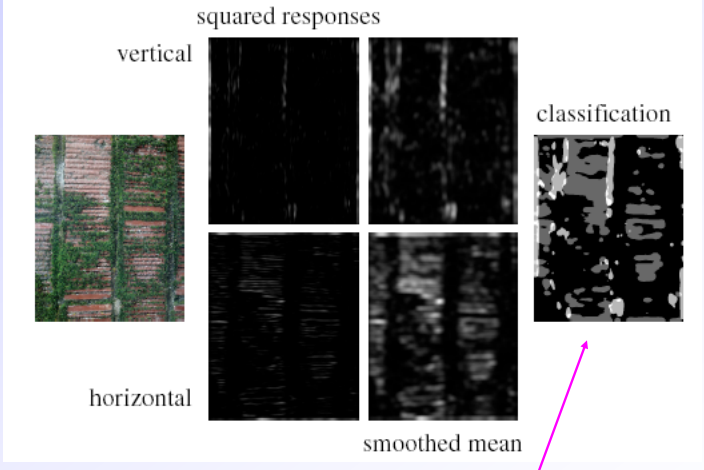
FIGURE 6.3: Local texture representations can be obtained by filtering an image with a set of filters at various scales, and then preparing a summary. Summaries ensure that, at a pixel, we have a representation of what texture appears near that pixel. The filters are typically spots and bars (see Figure 6.4). Filter outputs can be enhanced by rectifying them (so that positive and negative responses do not cancel), then computing a local summary of the rectified filter outputs. Rectifying by taking the absolute value means that we do not distinguish between light spots on a dark background and dark spots on a light background; the alternative, half-wave rectification (described in the text), preserves this distinction at the cost of a fuller representation. One can summarize either by smoothing (which will tend to suppress noise, as in the schematic example above) or by taking the maximum over a neighborhood. Compare this figure to Figure 6.7, which shows a representation for a real image.



- How many filters and at what orientations ?
- Filter responses are not unique
- Tradeoff: using more filters leads to a more detailed and more redundant representation of the image
- How to control the amount of redundant information?
- At what scale?
- There are two scales:
  - The scale of the filter
  - The scale over which we consider the distributions of the filters.
- What statistics should be collected from filters responses.



There are two scales:
- The scale of the filter
- The scale over which we consider the distributions of the filters.



Classification:
- Texture is classified into four categories: vertical, horizontal, both or neither.

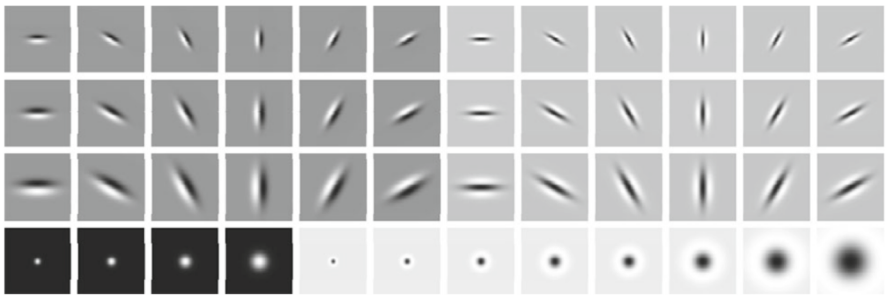


Fig. 7 The LMfilter bank has a mix of edge, bar and spot filters at multiple scales and orientations. It has a total of 48 filters: 2 Gaussian derivative filters at 6 orientations and 3 scales, 8 Laplacian of Gaussian filters and 4 Gaussian filters

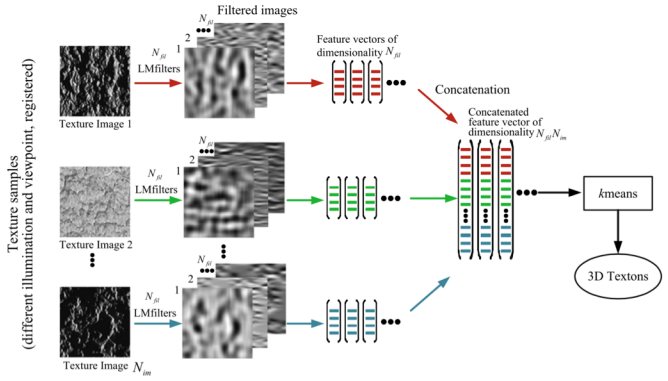


Fig. 8 Illustration of the process of 3D texton dictionary learning proposed by Leung and Malik (2001). Each image at different lighting and viewing directions is filtered using the filter bank illustrated in Fig. 7. The response vectors are concatenated together to form data vectors of length $N_{fil}N_{im}$. These data vectors are clustered using the *k*means algorithm to obtain the 3D textons

Leung and Malik 2001: Operational definition of textons: the cluster centers of the filter response vectors.

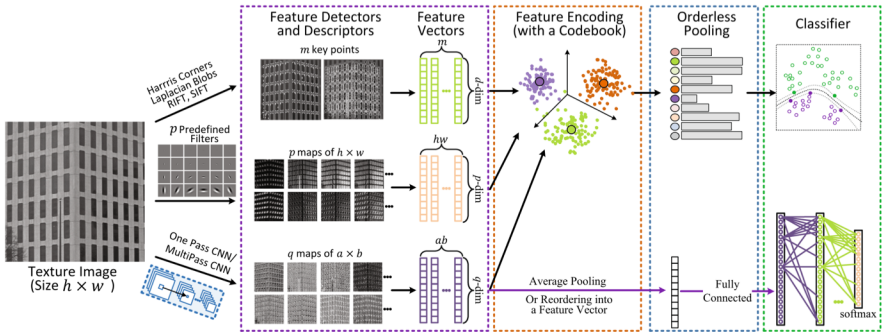
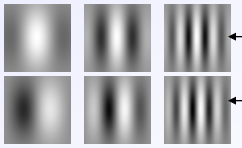


Fig. 5 General pipeline of the BoW model. See Table 1, and also refer to Sect. 3 for detail discussion. Features are computed from handcrafted detectors for descriptors like SIFT and RIFT, and densely applied local texture descriptors like handcrafted filters or CNNs. The CNN features can also be computed in an end-to-end manner using finetuned CNN models. These local features are quantized to visual words in a codebook

Gabor Filters

- Fourier coefficients depend on the entire image (Global): We lose spatial information.
- Objective: Local Spatial Frequency Analysis
- Gabor kernels: look like Fourier basis multiplied by a Gaussian
  - The product of a symmetric Gaussian with an oriented sinusoid
  - Gabor filters come in pairs: symmetric and antisymmetric
  - Each pair recover symmetric and antisymmetric components in a particular direction.
  - (k_x,k_y) :the spatial frequency to which the filter responds strongly
  - σ: the scale of the filter. When σ = infinity, similar to FT
- We need to apply a number of Gabor filters at different scales, orientations, and spatial frequencies.



$$G_{symmetric}(x,y) = \cos(k_x x + k_y y)\exp-\left\{\frac{x^2+y^2}{2\sigma^2}\right\}$$

$$G_{antisymmetric}(x,y) = \sin(k_x x + k_y y)\exp-\left\{\frac{x^2+y^2}{2\sigma^2}\right\}$$

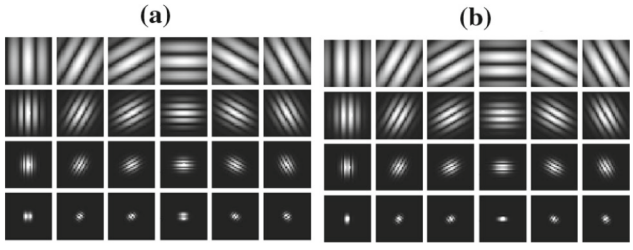
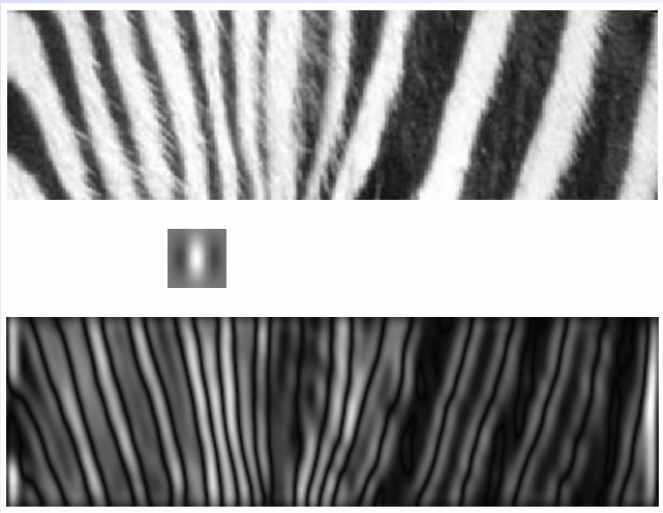


Fig. 6 Illustration of the Gabor wavelets used in Manjunath and Ma (1996). **a** Real part, **b** Imaginary part



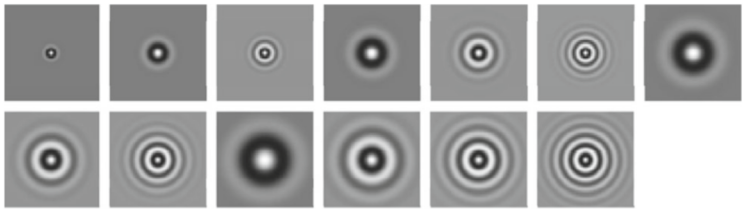


Fig. 9 Illustration of the rotationally invariant Gabor-like Schmid filters used in Schmid (2001). The parameter (σ, β) pair takes values (2,1), (4,1), (4,2), (6,1), (6,2), (6,3), (8,1), (8,2), (8,3), (10,1), (10,2), (10,3) and (10,4)

Scaled representations: Multiresolution

Use a multiresolution representation (Image Pyramid)

- Search over scale
- Spatial Search
- Feature Tracking

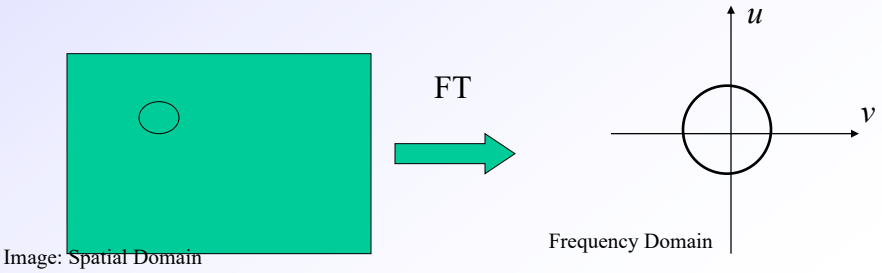
Examples:

- Search for correspondence
  - look at coarse scales, then refine with finer scales
- Edge tracking
  - a "good" edge at a fine scale has parents at a coarser scale
- Control of detail and computational cost in matching
  - e.g. finding stripes
  - terribly important in texture representation

Gaussian Filter and Smoothing

Gaussian Filter is Low-Pass Filter:

- Recall: Convolution in the image domain is equivalent to multiplication in the Frequency domain.
- Recall: FT of a Gaussian with sd=σ is a Gaussian with sd=1/σ
- Therefore, convolving an image with a Gaussian with sd=σ is equivalent to multiplying it's FT with a Gaussian with sd=1/σ
- Therefore we will get rid of high frequencies.
- Smoothing with a Gaussian with a very small σ ⇒ get rid of highest spatial frequencies
- Smoothing with a Gaussian with a very large σ ⇒ averaging

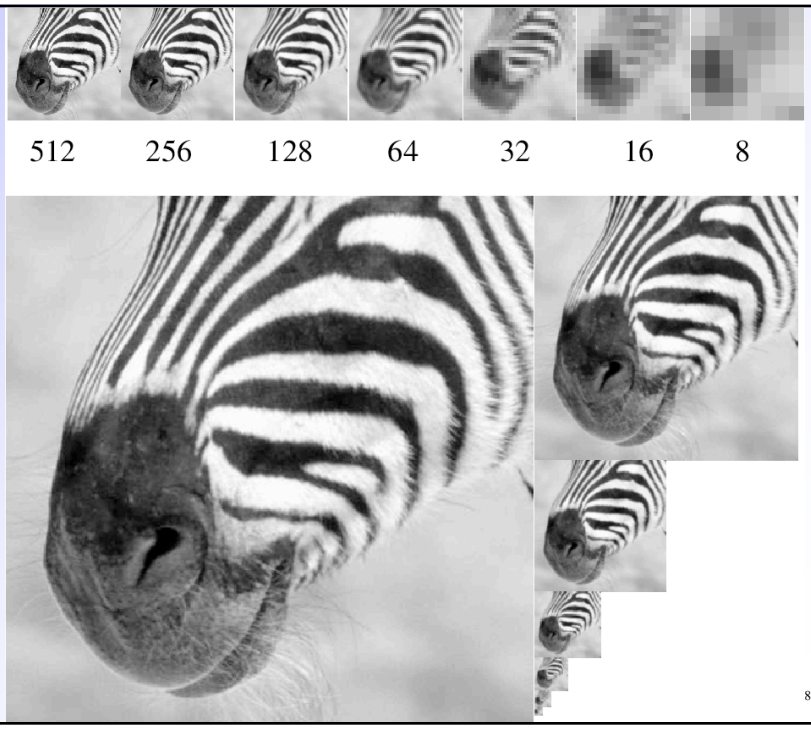


The Gaussian pyramid

- Smooth with Gaussians, because
  - a Gaussian*Gaussian=another Gaussian
- Forming a Gaussian Pyramid:
  - Set the finest scale layer to the image
  - For each layer going up (coarser)
    - Obtain this layer by smoothing the previous layer with a Gaussian and subsampling it

$$P_{\text{Gaussian}}(I)_{n+1} = S^{\downarrow}(G_\sigma * P_{\text{Gaussian}}(I)_n)$$
$$P_{\text{Gaussian}}(I)_1 = I$$

CS 534 – Texture - 37



The Laplacian Pyramid

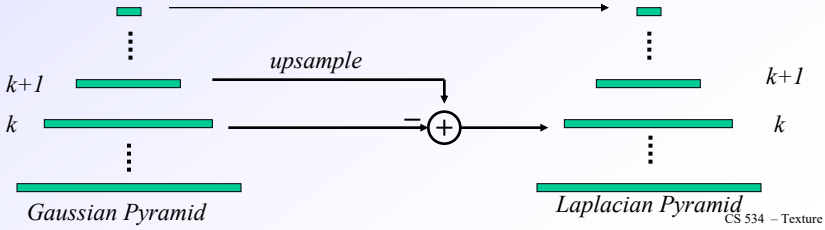
- Gaussians are low pass filters, so response is redundant
- A coarse level layer of the Gaussian pyramid predicts the appearance of the next finer layer
- Laplacian Pyramid
  - preserve differences between upsampled Gaussian pyramid level and Gaussian pyramid level
  - band pass filter - each level represents spatial frequencies (largely) unrepresented at other levels

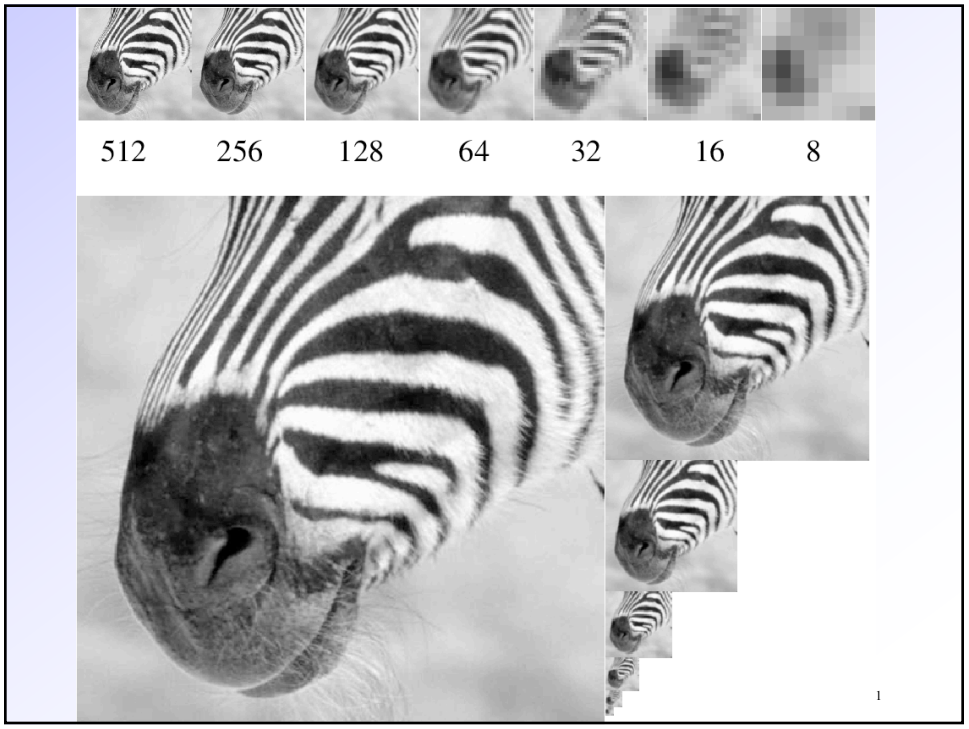
$$P_{\text{Laplacian}}(I)_m = P_{\text{Gaussian}}(I)_m$$
$$P_{\text{Laplacian}}(I)_k = P_{\text{Gaussian}}(I)_k - S^{\uparrow}(P_{\text{Gaussian}}(I)_{k+1})$$

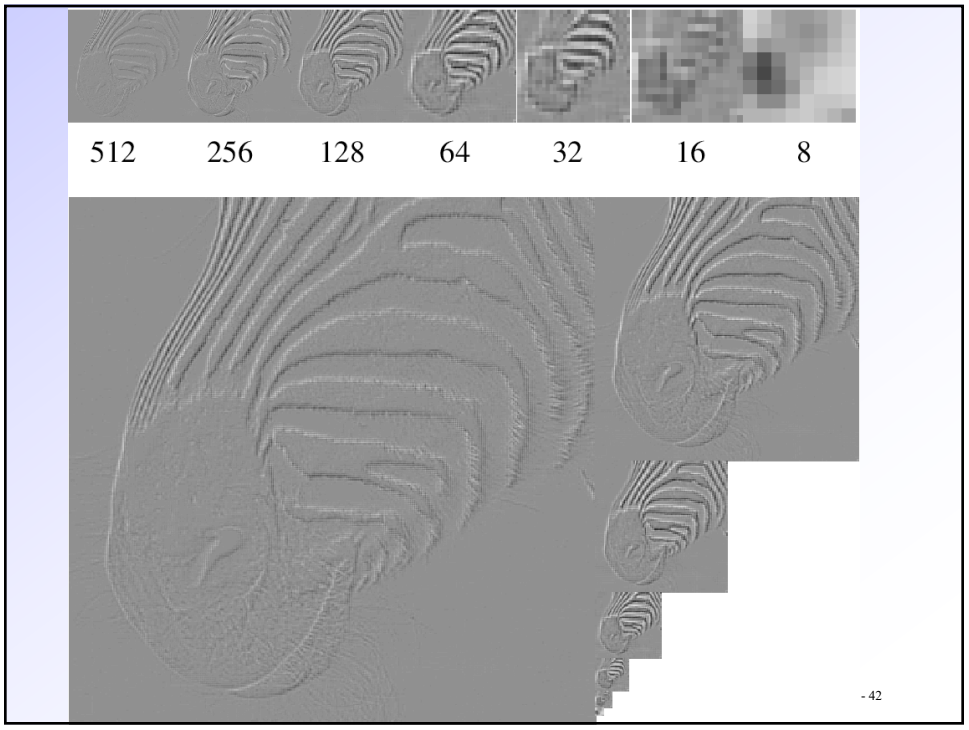
Laplacian Pyramid

- Building a Laplacian Pyramid:
  - Form a Gaussian pyramid
  - Set the coarsest layer of the Laplacian pyramid to be the coarsest level of the Laplacian pyramid
  - For each layer going from next to coarsest to finest (top to bottom):
    - Obtain this layer by upsampling the coarser layer and subtracting it from this layer of the Gaussian pyramid.

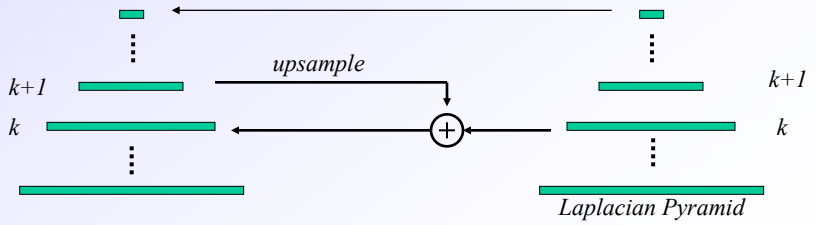
$$P_{\text{Laplacian}}(I)_m = P_{\text{Gaussian}}(I)_m$$
$$P_{\text{Laplacian}}(I)_k = P_{\text{Gaussian}}(I)_k - S^{\uparrow}(P_{\text{Gaussian}}(I)_{k+1})$$




512 256 128 64 32 16 8
1

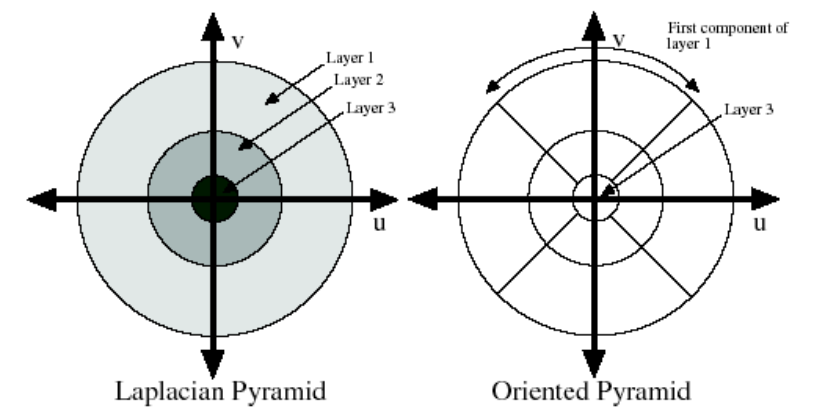

512 256 128 64 32 16 8
- 42

- Synthesis: Obtaining an Image from a Laplacian Pyramid:
  - Start at the coarsest layer
  - For each layer from next to coarsest to finest
    - Upsample the current image and add the current layer to the result



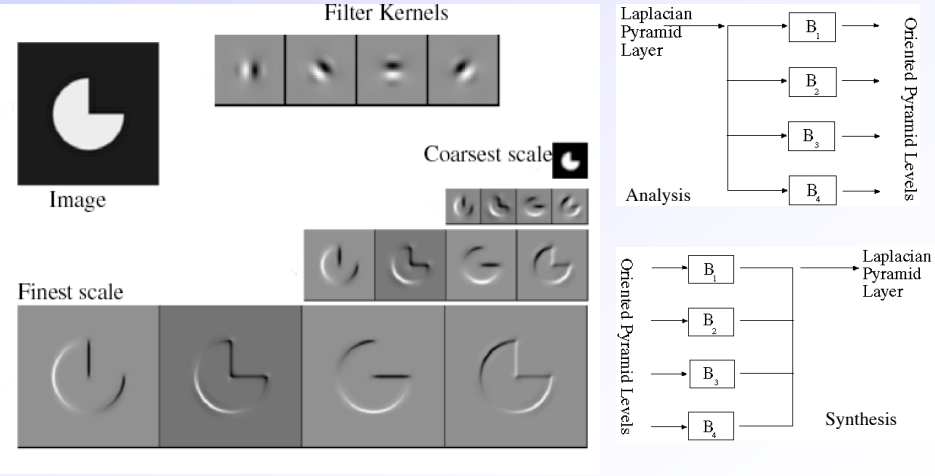
CS 534 – Texture - 43

- Laplacian pyramid layers are band-pass filters.
- Laplacian pyramid is orientation independent
- Apply an oriented filter to determine orientations at each layer
- Look into spatial frequency domain:



CS 534 – Texture - 44

Oriented Pyramids



Reprinted from "Shiftable MultiScale Transforms," by Simoncelli et al., IEEE Transactions on Information Theory, 1992, copyright 1992, IEEE

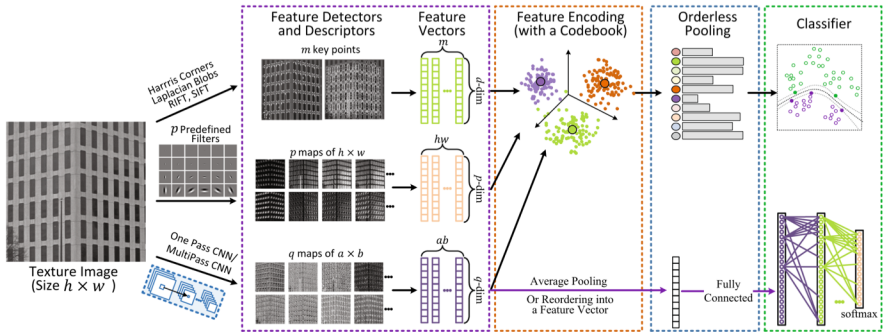
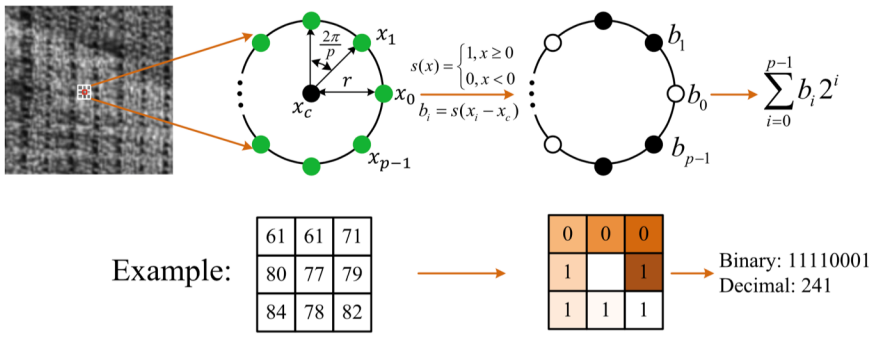


Fig. 5 General pipeline of the BoW model. See Table 1, and also refer to Sect. 3 for detail discussion. Features are computed from handcrafted detectors for descriptors like SIFT and RIFT, and densely applied local texture descriptors like handcrafted filters or CNNs. The CNN features can also be computed in an end-to-end manner using finetuned CNN models. These local features are quantized to visual words in a codebook

LBP: Local Binary Patterns - Ojala et al. 1996-2002



Example:

61	61	71
80	77	79
84	78	82

→

0	0	0
1		1
1	1	1

→ Binary: 11110001
Decimal: 241

Fig. 12 A circular neighborhood used to derive an LBP code: a central pixel x_c and its p circularly and evenly spaced neighbors on a circle of radius r

Rotation Invariant in LBP

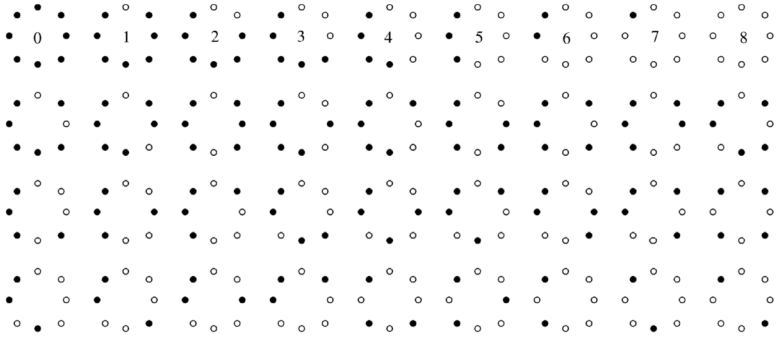
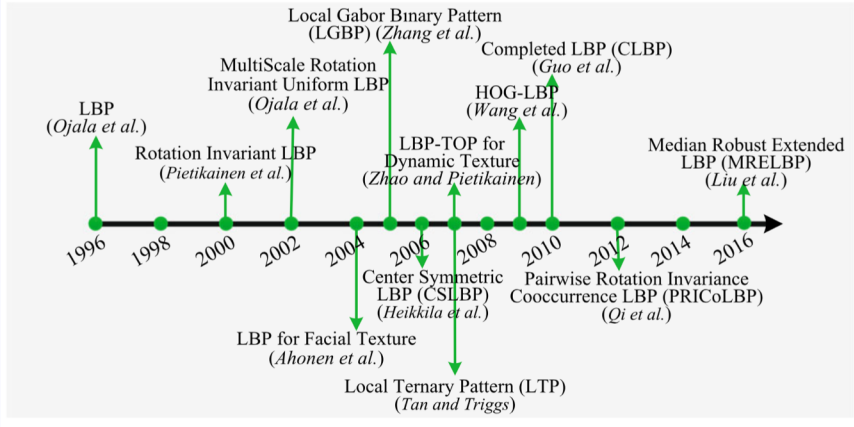


Fig. 2. The 36 unique rotation invariant binary patterns that can occur in the circularly symmetric neighbor set of $LBP^{ri}_{8,R}$. Black and white circles correspond to bit values of 0 and 1 in the 8-bit output of the operator. The first row contains the nine "uniform" patterns and the numbers inside them correspond to their unique $LBP^{riu2}_{8,R}$ codes.

All patterns that can be rotated (shifted) to each other are grouped as one pattern

$$LBP^{i}_{P,R} = min\{ROR(LBP_{P,R}, i) \quad | \quad i = 0, 1, \ldots, \mathrm{P}-1\}, \tag{8}$$

Figure from Ojala et al. 2002



Texture Classification using CNNs

- Using pretrained generic CNN models
- Using finetuned CNN models
- Using handcrafted deep convolutional networks

Approach	Highlights
Using Pretrained Generic CNN Models (Cimpoi et al. 2016) (Sect. 4.1)	Traditional feature encoding and pooling; New pooling such as bilinear pooling (Lin and Maji 2016; Lin et al. 2018) and LFV (Song et al. 2017)
AlexNet (Krizhevsky et al. 2012)	Achieved breakthrough image classification result on ImageNet; The historical turning point of feature representation from handcrafted to CNN
VGGM (Chatfield et al. 2014; Cimpoi et al. 2016)	Similar complexity as AlexNet, but better texture classification performance
VGGVD (Simonyan and Zisserman 2015)	Much deeper than AlexNet; Much Larger model size than AlexNet and VGGM; Much better texture recognition performance than AlexNet and VGGM
GoogleNet (Szegedy et al. 2015)	Much deeper than AlexNet; Small pretrained model size; Not often used in texture classification
ResNet (He et al. 2016)	Significantly deeper than VGGVD; Smaller model size (ResNet 101) than AlexNet
Using Finetuned CNN Models (Sect. 4.2)	End-to-end learning
TCNN (Andrearczyk and Whelan 2016)	Using global average pooling; Combining outputs from multiple CONV layers
BCNN (Lin et al. 2015; Lin and Maji 2016)	Introducing a novel and orderless bilinear feature pooling method; Generalizing Fisher Vector and VLAD; Good representation ability; Very high feature dimensionality
Compact BCNN (Gao et al. 2016)	Adopting Random Maclaurin Projection or Tensor Sketch Projection to reduce the dimensionality of bilinear features (e.g. from 262144 (512^2) to 8192); Maintain similar performance to BCNN;
FASON (Dai et al. 2017)	Combining the ideas of TCNN (Andrearczyk and Whelan 2016) and Compact BCNN (Gao et al. 2016)
NetVLAD (Arandjelovic et al. 2016)	Plugging a VLAD like layer in a CNN network at the last CONV layer
DeepTEN (Zhang et al. 2017)	Similar to NetVLAD (Arandjelovic et al. 2016), integrating an encoding layer on top of CONV layers; Generalizing orderless pooling methods such as VLAD and FV in a CNN trained end to end
Texture Specific Deep Convolutional Models (Sect. 4.3)	
ScatNet (Bruna and Mallat 2013)	Use Gabor wavelets for comvolution; Mathematical interpretation of CNNs; Features being stable to deformations and preserving high frequency information;
PCANet (Chan et al. 2015)	Inspired by ScatNet (Bruna and Mallat 2013), using PCA filters to replace Gabor wavelets;Using LBP and histogramming as feature pooling; No local invariance

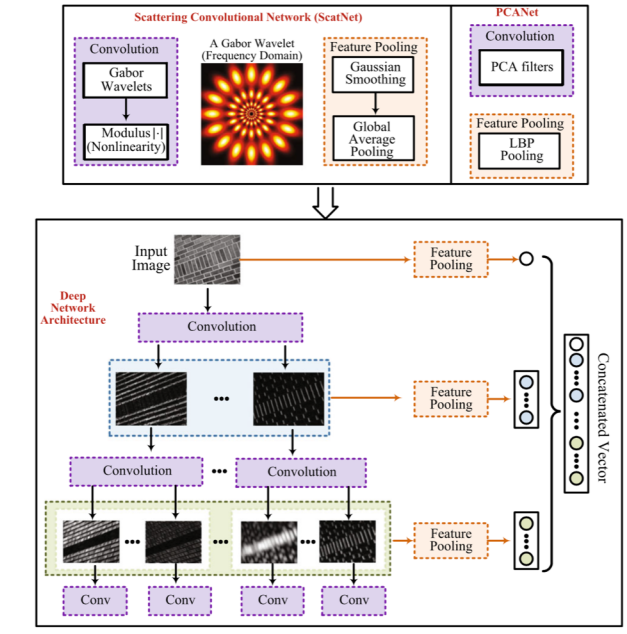
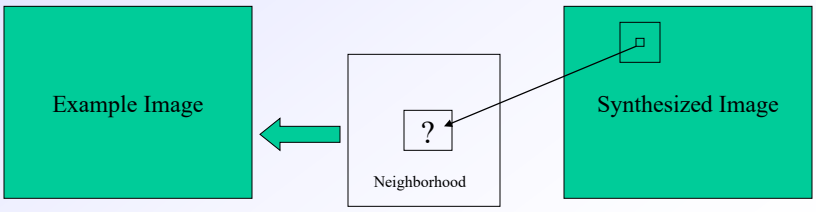


Fig. 21 Illustration of two similar handcrafted deep convolutional networks: ScatNet (Bruna and Mallat 2013) and PCANet (Chan et al. 2015)

Texture synthesis

- Variety of approaches.
- Example: Synthesis by Sampling Local Models: Efros and Leung 1999 (Nonparametric texture matching)
  - Use image as a source of probability model
  - Choose pixel values by matching neighborhood, then filling in



Find Matching Image Neighborhood and chose value uniformly randomly from these matches


```
Choose a small square of pixels at random from the example image
Insert this square of values into the image to be synthesized
Until each location in the image to be synthesized has a value
  For each unsynthesized location on
    the boundary of the block of synthesized values
    Match the neighborhood of this location to the
      example image, ignoring unsynthesized
      locations in computing the matching score
    Choose a value for this location uniformly and at random
      from the set of values of the corresponding locations in the
      matching neighborhoods
  end
end
```

Algorithm 6.4: Non-parametric Texture Synthesis.

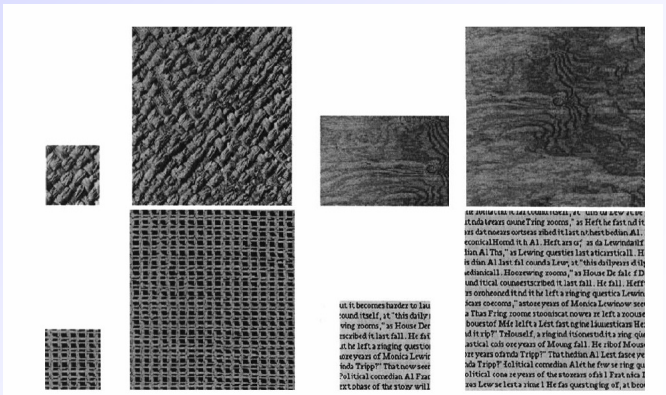


Figure from Texture Synthesis by Non-parametric Sampling, A. Efros and T.K. Leung, Proc. Int. Conf. Computer Vision, 1999 copyright 1999, IEEE

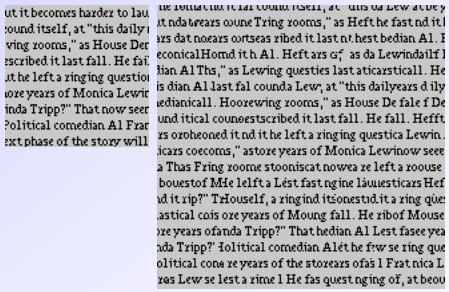
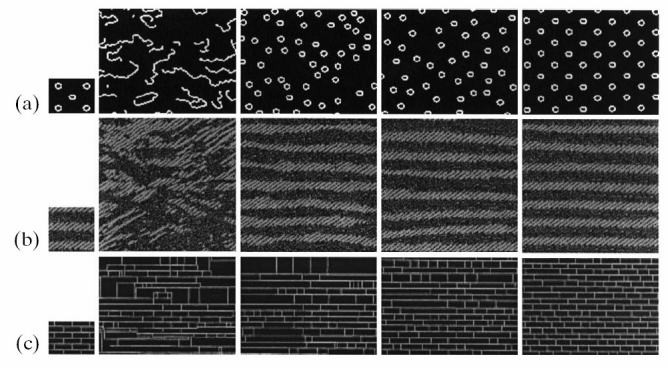


Figure from Texture Synthesis by Non-parametric Sampling, A. Efros and T.K. Leung, Proc. Int. Conf. Computer Vision, 1999 copyright 1999, IEEE

- The size of the image neighborhood to be matched makes a significant difference



Fill in holes by looking for example patches in the image If needed, rectify faces (lower images)

FIGURE 6.13: If an image contains repeated structure, we have a good chance of finding examples to fill a hole by searching for patches that are compatible with its boundaries. **Top left:** An image with a hole in it (black pixels in a rough pedestrian shape). The pixels on the region outside the hole, but inside the boundary marked on the image, match pixels near the other curve, which represents a potentially good source of hole-filling pixels. **Top right:** The hole filled by placing the patch over the hole, then using a segmentation method (Chapter 9) to choose the right boundary between patch and image. This procedure can work for apparently unpromising images, such as the one on the **bottom left**, an image of the facade of a house, seen at a significant slant. This slant means that distant parts of the facade are severely foreshortened. However, if we rectify the facade using methods from Section 1.3, then there are matching patches. On the **bottom right**, the hole has been filled in using a patch from the rectified image, that is then slanted again. *This figure was originally published as Figures 3 and 6 of "Hole Filling through Photomontage," by M. Wilczkowiak, G. Brostow, B. Tordoff, and R. Cipolla, Proc. BMVC, 2005 and is reproduced by kind permission of the authors.*

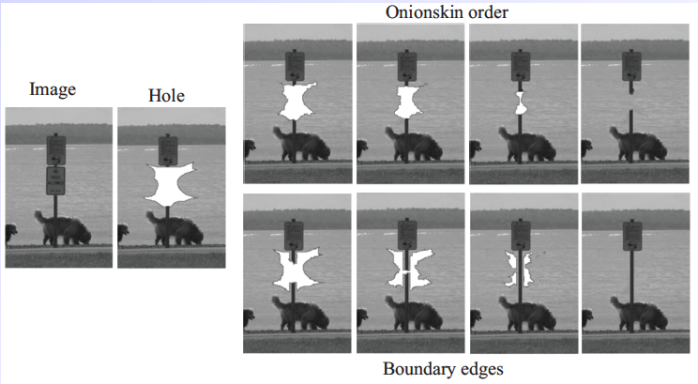
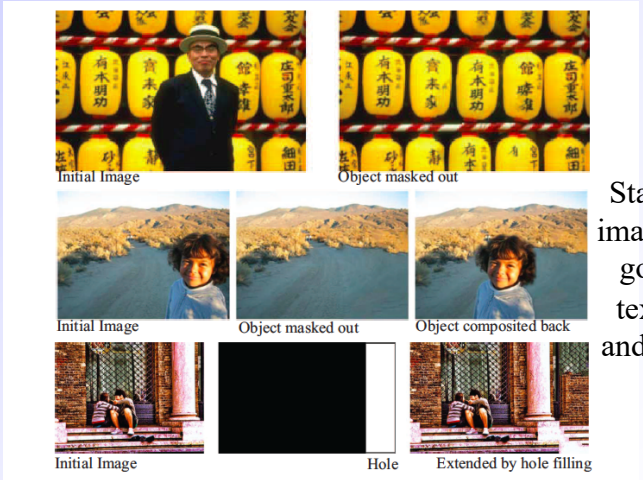


FIGURE 6.14: Texture synthesis methods can fill in holes accurately, but the order in which pixels are synthesized is important. In this figure, we wish to remove the sign, while preserving the signpost. Generally, we want to fill in pixels where most of the neighbors are known first. This yields better matching patches. One way to do so is to fill in from the boundary. However, if we simply work our way inwards (onionskin filling), long scale image structures tend to disappear. It is better to fill in patches close to edges first. *This figure was originally published as Figure 11 of "Region Filling and Object Removal by Exemplar-Based Image Inpainting," by A. Criminisi, P. Perez, and K. Toyama, IEEE Transactions on Image Processing, 2004 © IEEE, 2004.*



State of the art in image fill-in is very good. This uses texture synthesis and other methods.

FIGURE 6.15: Modern hole-filling methods get very good results using a combination of texture synthesis, coherence, and smoothing. Notice the complex, long-scale structure in the background texture for the example on the **top** row. The **center** row shows an example where a subject was removed from the image and replaced in a different place. Finally, the **bottom** row shows the use of hole-filling to resize an image. The white block in the center mask image is the "hole" (i.e., unknown pixels whose values are required to resize the image). This block is filled with a plausible texture. *This figure was originally published as Figures 9 and 15 of "A Comprehensive Framework for Image Inpainting," by A. Bugeau, M. Bertalmío, V. Caselles, and G. Sapiro, Proc. IEEE Transactions on Image Processing, 2010 © IEEE, 2010.*

Sources

- [1] Liu et al, From BoW to CNN: Two Decades of Texture Representation for Texture Classification, IJCV 2019
- Forsyth and Ponce, Computer Vision a Modern approach (2nd ed): chapter 6.
- L. G. Shapiro and G. C. Stockman "Computer Vision", Prentice Hall 2001.
- R. Gonzalez and R.E. Woods, "Digital Image Processing", 2002.
- Slides by
  - D. Forsyth @ UC Berkeley
  - G.C. Stockman @MSU