Scale
Scale

PK’s half formed thoughts...
Scale

All computer vision algorithms require parameters.

Scale is often a hidden parameter that is implicitly set by the choice of image sizes, image resolution, filter sizes, window sizes etc.

What is it that we mean by scale?
What is it that we mean by ‘scale’?

The term is used in many contexts

- Gaussian smoothing scale
- Magnification/Zoom scale
- Bandpass frequency scale
- Resolution scale
- Window size
- Structuring element size
- Spatial frequencies present
What do we want to use knowledge of scale for?

• Rank features on the basis of their ‘salience’ or ‘importance’.

• Match features in images having different magnification

• Automatically determine what filter sizes/window sizes should be used to analyse an image.

• Other reasons...?
Classical Gaussian Scale-Space

A set of Gaussian smoothed images

Sequence of Gaussian smoothings of a waveform

Contours of $F_{xx} = 0$ in a scale-space image

Marr, Witkin, Koenderink, Lindeberg

Figures from Witkin

Justification: Gaussian smoothing ensures that no new features are created when going from a fine scale to any coarser scale.
Classical Gaussian Scale-Space

Marr, Witkin, Koenderink, Lindeberg

A set of Gaussian smoothed images

Sequence of Gaussian smoothings of a waveform

Contours of $F_{xx} = 0$ in a scale-space image

Justification: Gaussian smoothing ensures that no new features are created when going from a fine scale to any coarser scale.

... but this assumes that a ‘feature’ is a point of maximum gradient
2.2 Zero-Crossings and the Raw Primal Sketch
Lindeberg’s Scale-Space Theory  IJCV 1998

Defines a normalized derivative for Gaussian smoothed signals

\[ \partial_{\varepsilon, \gamma-norm} = t^{\gamma/2} \partial_x \]

\( t = \sigma^2 \) the variance of the Gaussian smoothing applied. 
\( \gamma \) is a constant that depends on the feature you are seeking to detect.

When applied to a sinusoidal signal the normalized derivative assumes a maximum at a scale that is proportional to the wavelength of the signal.

...but you need to know what you are looking for in order to apply the right normalized derivative.

- Edge \( \gamma \)-value \( \frac{1}{2} \)
- Ridge \( \gamma \)-value \( \frac{3}{4} \)
- Corner \( \gamma \)-value 1
- Blob \( \gamma \)-value 1
Scale

• The traditional approach to scale is to consider different low-pass or band-pass versions of an image.

• With this approach the number of features present, and their locations, vary with the scale used. This is very unsatisfactory.
• Phase Congruency depends on how the feature is built up from local frequency components.

• The number of local frequency components considered present is set by the size of the analysis window.

• For Phase Congruency the natural scale parameter to use is the size of the analysis window (the largest filter in the wavelet bank).

• This corresponds to high-pass filtering.

• Under high-pass filtering feature positions do not change with scale, only their relative significance changes.
(a) low-pass  
(b) band-pass  
(c) high-pass  
(d) phase congruency
Scale via high-pass filtering?
Using Scale to match features in images having different magnification

Lowe’s SIFT descriptor attempts to automatically determine the local scale at image locations

- Local extrema within the Difference of Gaussian image pyramid define keypoint locations.
- Difference of Gaussian filters are band-pass filters.
- Thus the keypoint scale is defined by the frequency band giving maximal local response.
Keypoints are in the middle of blobs
In humans it has been found that face recognition is tuned to a set of spatial frequencies ranging from about 20 cycles per face width down to about 5 cycles per face width.

Maximum sensitivity is centred around 8 to 13 cycles/face width.

To recognize with confidence you need to be able to resolve down to 20 cycles/face width.

(Hayes, Morrone and Burr 1986)  
(Costen, Parker and Craw 1996)  
(Nasanen 1999)
Human Face Recognition

In humans it has been found that face recognition is tuned to a set of spatial frequencies ranging from about 20 cycles per face width down to about 5 cycles per face width.

Maximum sensitivity is centred around 8 to 13 cycles/face width.

To recognize with confidence you need to be able to resolve down to 20 cycles/face width

(Hayes, Morrone and Burr 1986)
(Costen, Parker and Craw 1996)
(Nasanen 1999)
Faces do not survive compression well

JPEG (14kB)  
JPEG (24kB)  
Original
What Does Compression Do?

JPEG and MPEG

• Image is divided into 8x8 blocks.

• Discrete Cosine Transform is applied to each block.

• The transform coefficients are quantized, many will be rounded to zero.

• When reconstructed, the amplitude and phase of the spatial frequencies within each 8x8 block will be altered.

The 64 basis functions of the 8x8 Discrete Cosine Transform
12.5mm lens at 6m

~ 40 pixels

No compression
12.5mm lens at 6m

18:1 compression
12.5mm lens at 6m

18:1 compression
12.5mm lens at 6m

31:1 compression
12.5mm lens at 6m

40 pixels across face
= 5 DCT blocks

Spatial frequencies from
5 cycles/face width upwards
are all corrupted

This is exactly the range that is most important for face recognition!
Scale

- Classical Scale-Space theory assumes features are defined by gradient.
- Not clear that a value of scale can necessarily be assigned to all points in an image.
- Objects (things with two sides) can have scale but it is not clear that ‘features’ have scale.
- Objects may have multiple scales and scales that vary with orientation.