
**Texture**

What defines a texture?

- Includes: more regular patterns
- Includes: more random patterns

**Includes: more regular patterns**

**Includes: more random patterns**

**Scale: objects vs. texture**

Often the same thing in the world can occur as texture or an object, depending on the scale we are considering.

**Disambiguation**

1. The nature, geometry, and wavelength composition of the illuminant(s).
2. Properties of the objects imaged, such as spectral reflectances; surface shape; surface albedo; surface texture; geometry, motion, and rotation angle.
3. Properties of the cameras (or viewer), such as (i) geometry and viewing angle; (ii) spectral sensitivity; (iii) prior knowledge, assumptions, and expectations.
Inferring surface orientation from texture

>- the assumption of uniformity constrains the problem

Dr Chris Town

Inferring surface orientation from texture

Texture

Texture is defined by the existence of certain statistical correlations across the image.
Examples:
• quasi-periodic undulations (waves, ripples, folds in clothing)
• spots, speckles
• stripes, dashes

Many natural textures can appear to be almost fractal, i.e. self-similar across different scales.

The unifying notion in all of these examples is quasi-periodicity, or repetitiveness, of some features.

So, What Scale to Choose?

• It depends on what we're looking for...
  • Too fine a scale... can't see the forest for the trees.
  • Too coarse a scale... can't tell the maple from the cherry.

Dr Chris Town
Texture

- Textures are made up of repeated sub-elements
- Representation:
  - find the sub-elements, and represent their statistics
- But what are the sub-elements, and how do we find and characterise them?

Fourier methods: capture quasi-periodicity at different scales and orientation, but have non-localised (global) response

Gabor wavelets: spatially localised, so we can analyse texture in terms of local spectral analysis

A simple texture descriptor

Magnitude of the Fourier Transform

$$A(f_x, f_y) = \sum_{x} \sum_{y} f(x,y) e^{-2\pi i (f_x x + f_y y)}$$

Magnitude of the Fourier Transform encodes unlocalised information about dominant orientations and scales in the image.

Statistics of Scene Categories

- Man-made environments
- Natural environments
- Spectral signature of man-made environments
- Spectral signature of natural environments

Gist descriptor

- Apply oriented Gabor filters over different scales
- Average filter energy in each bin

Example visual gists

Global features (I) – global features (I')

Dr Chris Town
Texture characterisation using filters

Region segmentation
> perceptual grouping

Region segmentation using Gabor Wavelets

Oriented (or “steerable”) pyramids

- Laplacian pyramid is orientation independent
- Apply an oriented filter to determine orientations at each layer
  - this represents image information at a particular scale and orientation

But we need to get rid of the corner regions before starting the recursive circular filtering

http://www.cns.nyu.edu/ftp/eero/simoncelli95b.pdf

**Steerable filters**

"Steerability"—the ability to synthesise a filter of any orientation from a linear combination of filters at fixed orientations. The basis functions of the steerable pyramid are directional derivative operators, that come in different sizes and orientations.

**Filter Set:**
- 0°
- 90°
- Synthesized 30°

**Response:**
- Raw Image
- Taking a pyramid (or equivalent set of responses to filters at different scales and orientations).
- Forming oriented pyramid (e.g. an equivalent set of responses to filters at different scales and orientations).
- Squaring the output (modulus).
- Taking statistics of responses:
  - Mean of each filter output (e.g. are there lots of spots?)
  - Standard deviation of each filter output (e.g. are the spots of similar size?)
  - Mean of one scale conditioned on another scale having a particular range of values (e.g. are the spots in straight rows?)

**Texture synthesis**

- Model texture as generated from random process.
- Discriminate by seeing whether statistics of two processes seem the same.
- Synthesize by generating image with same statistics.

**Texture synthesis**

- Model texture as generated from random process.
- Discriminate by seeing whether statistics of two processes seem the same.
- Synthesize by generating image with same statistics.

**Statistical in-fill**

- Statistical in-fill
**Texture analysis**

- Wavelet decomposition (steerable pyr)

![Image](histogram)

The texture is represented as a collection of marginal histograms.

---

**Texture synthesis**

- Input texture

![Image](histogram)

Heeger and Bergen, 1995

---

**Inferring Object Colour**

- Let \( I(\lambda) \) represent the wavelength composition of the illuminant (i.e. the amount of energy it contains as a function of wavelength \( \lambda \), across the visible spectrum from about 400 nanometers to 700 nm).
- Let \( O(\lambda) \) represent the inherent spectral reflectance of the object at a particular point: the fraction of incident light that is scattered back from its surface there, as a function of the incident light’s wavelength \( \lambda \).
- Let \( R(\lambda) \) represent the actual wavelength mixture received by the camera at the corresponding point in the image of the scene.

![Graphs](histogram)

---

**Inferring Object Colour**

- Clearly, \( R(\lambda) = I(\lambda)O(\lambda) \). The problem is that we wish to infer the “object colour” (its spectral reflectance as a function of wavelength, \( O(\lambda) \)), but we only know \( R(\lambda) \), the actual wavelength mixture received by our sensor. So unless we can measure \( I(\lambda) \) directly, how could this problem of inferring \( O(\lambda) \) from \( R(\lambda) \) possibly be solved?

![Graphs](histogram)

---

**Inferring Object Colour**

**Measuring \( f(\lambda) \)**

Search for highly specular (shiny, metallic, glassy) regions in an image. Then we could infer \( O \) by dividing \( R \) by \( I \).

Problems:
- We may not find any specular surfaces in the image
- Most materials are not purely specular (e.g. metals which have a brassy colour)
- Not robust, too dependent on highly localised measurements

---

**Colour Constancy**

These images show a bowl of fruit photographed in three lighting conditions:
- artificial light (left)
- hazy daylight (middle)
- clear blue sky (right)

Notice the marked variation in colour balance caused by the spectral properties of the illuminant. We are not normally aware of this variation because colour constancy mechanisms discount illumination effects.

http://www.psypress.co.uk/mather/resources/topics.asp?topic=ch12-tq-04

---

Dr Chris Town
Dr Chris Town

**Colour Cube Illusion**

Dr Chris Town

**Colur Constancy in Goldfish**

In David Ingle's experiment, a goldfish has been trained to swim to a patch of a given color for a reward—a piece of liver. It swims to the green patch regardless of the exact setting of the three projectors' intensities. The behavior is strikingly similar to the perceptual result in humans.

http://neuro.med.harvard.edu/site/dh/b45.htm

Dr Chris Town

**Colour Constancy**

Possible explanations:

1. **Local colour contrast**—cone excitation level of one surface relative to another remains constant when both surfaces experience the same change in illumination. → Relative cone excitation levels are invariant ratios

2. **Colour adaptation**—reduces the contribution from the source illumination by lowering activity in the most highly active cone classes.

3. **Global contrast**—global spectral changes generally represent changes in the illuminant; localised differences usually correspond to reflectance differences.

4. **Range of reflected spectrum**—gives some indication of the breadth of the illuminating spectrum.

Colour constancy is not perfect (83% accuracy), and the most powerful cue to constancy is local colour contrast.

http://www.pupress.co.uk/mather/resources/topic.asp?topic=ch12-tp-04

Dr Chris Town

**Retinex**


*Journal of the Optical Society of America*

Lightness and Retinex Theory

E.H. Land* and J.J. McCann

Optical Society of America

Volume 61, Number 11

November 1971

Dr Chris Town

**Retinex**

The key idea is that the colours of objects or areas in a scene are determined by their surrounding spatial context. A complex sequence of ratios computed across all the boundaries of objects (or areas) enables the illuminant to be algebraically discounted in the scene shown in the previous Figure, so that object spectral reflectances $O(A)$ which is what we perceive as their colour, can be inferred from the available retinal measurements $R(A)$ without explicitly knowing $R(A)$.

Dr Chris Town

**Retinex**

Reflectance tends to be constant across space except for abrupt changes at the transitions between objects. Thus a reflectance change shows itself as step edge in an image, while illuminance changes gradually over space. By this argument one can separate reflectance change from illuminance change by measuring the response to spatial derivatives.

Dr Chris Town
What Is Stereo Vision?

- Generic problem formulation: given several images of the same object or scene, compute a representation of its 3D shape.

Stereo

Important information about depth can be obtained from the use of two (or more) cameras, in the same way that humans achieve stereoscopic depth vision by virtue of having two eyes. Objects in front or behind of the point in space at which the two optical axes intersect (as determined by the angle between them, which is controlled by camera movements or eye movements), will project into different relative parts of the two images. This is called stereoscopic disparity.

Pinhole Camera

Image plane:

Source: Forsyth & Ponce

\[ d = \frac{fb}{(\alpha + \beta)} \]
Camera calibration

There are many other possible intrinsic camera calibration parameters, such as:
- **skew coefficients** accounting for non-orthogonality
- **distortion coefficients** representing radial and tangential distortions of the lens

The Correspondence Problem

Features (pixels, edge responses, SIFT features etc.) in the two images need to be matched.

If each image has N features, then there are $N^2$ possible pairings.

However, the number of potential pairings is $N \times (N-1) \times (N-2) \ldots = N!$

Simplest Case: Rectified Images

- Image planes of cameras are parallel.
- Focal points are at same height.
- Focal lengths same.
- Then, epipolar lines fall along the horizontal scan lines of the images.
- We will assume images have been rectified so that epipolar lines correspond to scan lines.
  - Simplifies algorithms
  - Improves efficiency
Essential Matrix and Epipolar Lines

Epipolar constraint: if we observe point $p$ in one image, then its position $p'$ in second image must satisfy this equation.

$\mathbf{p}'^T \mathbf{E} \mathbf{p} = 0$

$\mathbf{E}' \mathbf{p}$ is the coordinate vector representing the epipolar line for point $p$

$\mathbf{E} \mathbf{p}'$ is the coordinate vector representing the epipolar line for point $p'$

Recap: Stereo Image Rectification

- In practice, it is convenient if image scanlines are the epipolar lines.

- Algorithm
  - Reproject image planes onto a common plane parallel to the line between optical centers
  - Pixel motion is horizontal after this transformation
  - Two homographies (3x3 transforms), one for each input image reprojection

Motion information

- For stereo vision, we need to solve the Correspondence Problem for two images simultaneous in time but acquired with a spatial displacement.
- For motion vision, we need to solve the Correspondence Problem for two images coincident in space but acquired with a temporal displacement.
- The object’s spatial “disparity” can be measured in the two image frames once their backgrounds have been aligned. This can be calibrated to reveal motion information when compared with the time interval, or depth information when compared with the binocular spatial interval.

The Aperture Problem

Perceived motion
• Automated motion analysis generally limited to opaque and solid objects
• Challenges: flocks of birds, clouds, vapours, waves, fire, the wind in the willows...

Motion and Perceptual Organisation
• Even “impoverished” motion data can evoke a strong percept
**Video**

- A video is a sequence of frames captured over time
- Now our image data is a function of space \( (x, y) \) and time \( t \)

---

**Motion Estimation Techniques**

- **Direct methods**
  - Directly recover image motion at each pixel from spatio-temporal image brightness variations
  - Dense motion fields, but sensitive to appearance variations
  - Suitable for video and when image motion is small

- **Feature-based methods**
  - Extract visual features (corners, textured areas) and track them over multiple frames
  - Sparse motion fields, but more robust tracking
  - Suitable when image motion is large (10s of pixels)

---

**Motion Field and Parallax**

- \( P(t) \) is a moving 3D point
- Velocity of scene point: \( V = \frac{dP}{dt} \)
- \( p(t) = (x(t), y(t)) \) is the projection of \( P \) in the image.
- Apparent velocity \( V \) in the image: given by components \( v_x = \frac{dx}{dt} \) and \( v_y = \frac{dy}{dt} \)
- These components are known as the motion field of the image.

---

**Optical Flow**

- Definition: optical flow is the apparent motion of brightness patterns in the image.
- Ideally, optical flow would be the same as the motion field.
- Have to be careful: apparent motion can be caused by lighting changes without any actual motion.
  - Think of a uniform rotating sphere under fixed lighting vs. a stationary sphere under moving illumination.

---

**Optical Flow Field**

- Given two subsequent frames, estimate the apparent motion field \( u(x,y) \) and \( v(x,y) \) between them.

- Key assumptions
  - Brightness constancy: projection of the same point looks the same in every frame.
  - Small motion (temporal coherence): points do not move very rapidly.
  - Spatial coherence: points move like their neighbors.
Spatial Coherence

Assumption
- Neighboring points in the scene typically belong to the same surface and hence typically have similar motions.
- Since they also project to nearby points in the image, we expect spatial coherence in image flow.

Optical flow constraint (also known as Brightness constancy constraint)

**Brightness Constancy**

\[ I(x + u, y + v, t + 1) = I(x, y, t) \]

(assumption)

The Brightness Constancy Constraint

\[ I(x_u + u, y_v + v, t + 1) = I(x, y, t) \]

\[ I(x + u, y + v, t + 1) - I(x, y, t) = 0 \]

Divide through by \( dt \)

\[ u \frac{\partial}{\partial x} I(x, y, t) + v \frac{\partial}{\partial y} I(x, y, t) + \frac{\partial}{\partial t} I(x, y, t) = 0 \]

Notation

\[ I_u + I_v + I_t = 0 \]

\[ \nabla I^T \mathbf{u} = -I_t \]

\[ \nabla \mathbf{I} = \begin{bmatrix} I_x \\ I_y \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} u \\ v \end{bmatrix} \]
The Brightness Constancy Constraint

\[ I_x \cdot u + I_y \cdot v + I_t = 0 \]

- How many equations and unknowns per pixel?
  - One equation, two unknowns

- Intuitively, what does this constraint mean?
  \[ \nabla I \cdot (u, v) + I_t = 0 \]

- The component of the flow perpendicular to the gradient (i.e., parallel to the edge) is unknown.

\[ \text{gradient} \]

\[ (u,v) \quad (u',v') \]


Dynamic Zero-Crossing Models

Measure image velocity by first finding the edges and contours of objects (using the zero-crossings of a blurred Laplacian operator!), and then take the time-derivative of the Laplacian-Gaussian-convolved image:

\[ \frac{\partial}{\partial x} \left[ \nabla^2 G_d(x, y) * I(x, y, t) \right] \]

in the vicinity of a Laplacian-zero-crossing. The amplitude of the result is an estimate of speed, and the sign of this quantity determines the direction of motion relative to the normal to the contour.

Also known as the "Hildreth model", after Ellen Hildreth.


Spatio-Temporal Correlation Models

Image motion is detected by observing a correlation of the local image signal \( I(x, y, t) \) across an interval of space and an interval of time \( t \). Finding the pair of these intervals which maximizes the correlation between \( I(x, y, t) \) and \( I(x - \tau x, y - \tau y, t - \tau) \) determines the two components of image velocity \( \tau x \) and \( \tau y \) which we desire to know.

\[ \text{correlation} \left[ \int \int I(x, y, t) \cdot I(x - \tau x, y - \tau y, t - \tau) \, dx \, dy \, dt \right] \]

Detailed studies of fly neural mechanisms (above) for motion detection and visual tracking led to elaborated correlation-based motion models.


Intensty Gradient Models

Assume that the local time-derivative of image intensities at a point, across many images frames, is related to the local spatial gradient in image intensities because of object velocity \( \mathbf{v} \):

\[ \frac{\partial I(x, y, t)}{\partial t} = \mathbf{v} \cdot \nabla I(x, y, t) \]

Then the ratio of the local image time-derivative to the spatial gradient is an estimate of the local image velocity (in the direction of the gradient).


Spatio-Temporal Spectral Models

It is possible to detect and measure image motion purely by Fourier means. This approach exploits the fact that motion creates a covariance in the spatial and temporal spectra of the time-varying image \( I(x, y, t) \), whose 3-dimensional (spatio-temporal) Fourier transform is defined:

\[ F(\omega_x, \omega_y, \omega_t) = \int \int \int I(x, y, t) e^{-j(\omega_x x + \omega_y y + \omega_t t)} \, dx \, dy \, dt \]

In other words, rigid image motion has a 3D spectral consequence: the local 3D spatio-temporal spectrum, rather than filling up 3-space \( (\omega_x, \omega_y, \omega_t) \), collapses onto a 2D inclined plane which includes the origin. Motion detection then occurs just by filtering the image sequence in space and in time, and observing that tuned spatio-temporal filters whose centre frequencies are co-planar in this 3-space are activated together. This is a consequence of the Spectral Co-Planarity Theorem, which states that translational image motion of velocity \( \mathbf{v} \) has a 3D spatio-temporal Fourier spectrum that is non-zero only on an inclined plane through the origin of frequency-space. Spherical coordinates of the unit normal to this spectral plane correspond to the speed and direction of motion.
Theorem: Translational image motion of velocity \( \vec{v} \) has a 3D spatio-temporal Fourier spectrum that is non-zero only on an inclined plane through the origin of frequency-space. Spherical coordinates of the unit normal to this spectral plane correspond to the speed and direction of motion.

Let \( I(x, y, t) \) be a continuous image in space and time.

Let \( F(\omega_x, \omega_y, \omega_z) \) be its 3D spatio-temporal Fourier transform:

\[
F(\omega_x, \omega_y, \omega_z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y, t) e^{-j(\omega_x x + \omega_y y + \omega_z t)} dx dy dt.
\]

Let \( \vec{v} = (v_x, v_y) \) be the local image velocity.

Uniform motion \( \vec{v} \) implies that for all time shifts \( t_0 \),

\[
I(x, y, t) = I(x - v_x t_0, y - v_y t_0, t - t_0).
\]

Taking the 3D spatio-temporal Fourier transform of both sides, and applying the shift theorem, gives

\[
F(\omega_x, \omega_y, \omega_z) = e^{-j(\omega_x v_x t_0 + \omega_y v_y t_0 + \omega_z t_0)} F(\omega_x, \omega_y, \omega_z).
\]

Taking the 3D spatio-temporal Fourier transform of both sides, and applying the shift theorem, gives

\[
F(\omega_x, \omega_y, \omega_z) = e^{-j(\omega_x v_x t_0 + \omega_y v_y t_0 + \omega_z t_0)} F(\omega_x, \omega_y, \omega_z).
\]

The above equation can only be true if \( F(\omega_x, \omega_y, \omega_z) = 0 \) everywhere the exponential term doesn’t equal 1.

This means \( F(\omega_x, \omega_y, \omega_z) \) is non-zero only on the 3D spectral plane

\[
\omega_x v_x + \omega_y v_y + \omega_z = 0
\]

Q.E.D.

The spherical coordinates \((\theta, \phi, \lambda)\)

\[
\phi = \tan^{-1} \left( \frac{\omega_y}{\sqrt{\omega_x^2 + \omega_z^2}} \right)
\]

\[
\lambda = \sec^{-1} \left( \frac{\omega_z}{\omega_x} \right)
\]

of the inclined spectral plane’s unit normal are determined by \( \vec{v} \) and correspond to the speed \((\phi)\) and direction \((\theta)\) of motion:

\[
\phi = \sqrt{v_y^2 + v_z^2}
\]

\[
\lambda = \sec^{-1} \left( \frac{v_z}{v_x} \right)
\]