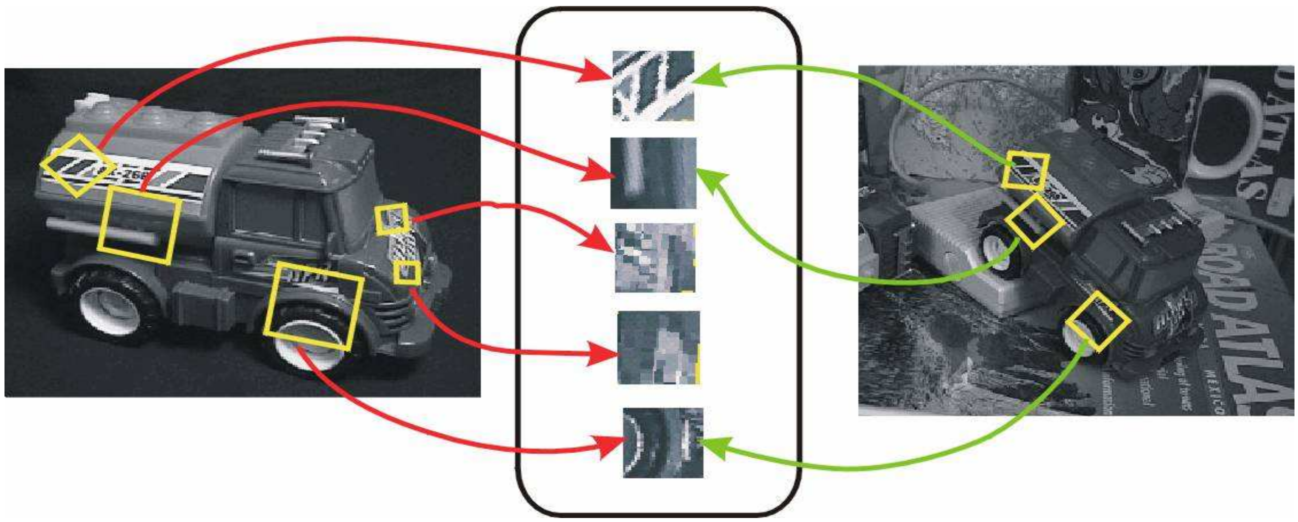
Local Image Features

Basic Idea: Encode the image structure in a spatial neighbourhood (i.e. "what" the image patch looks like) around each of a set of interest points chosen at selected scales (i.e. "where" to do this encoding).



(image from D. Lowe, CVPR03 tutorial).

Advantages:

- **Locality:** Small spatial neighbourhoods are less sensitive to image deformations. Other parts of the object can be occluded.
- **Pose Invariance:** The interest point detector can select a canonical position, scale and orientation. Matches are then made with respect to these canonical coordinates.
- **Distinctiveness:** Image neighbourhoods can be encoded to achieve a low false matching probability (e.g. 10^{-3}) with a 0.7-0.9 detection rate.
- **Applicability:** One view of a textured object is sufficient for training in order to recognize an object from nearby 3D viewing directions (but possibly different scales, image locations, and image orientations).

Interest Point Detectors

Harris Corner Points: Compute the 2×2 orientation tensor $T(\vec{x}, \sigma)$

$$T(\vec{x}, \sigma) \equiv G(\vec{x}, 2\sigma) * \left[\begin{pmatrix} I_x(\vec{x}, \sigma) \\ I_y(\vec{x}, \sigma) \end{pmatrix} \begin{pmatrix} I_x(\vec{x}, \sigma) & I_y(\vec{x}, \sigma) \end{pmatrix} \right].$$

Here $I_x(\vec{x}, \sigma) = G_x(\vec{x}, \sigma) * I(\vec{x})$ and similarly for I_y.

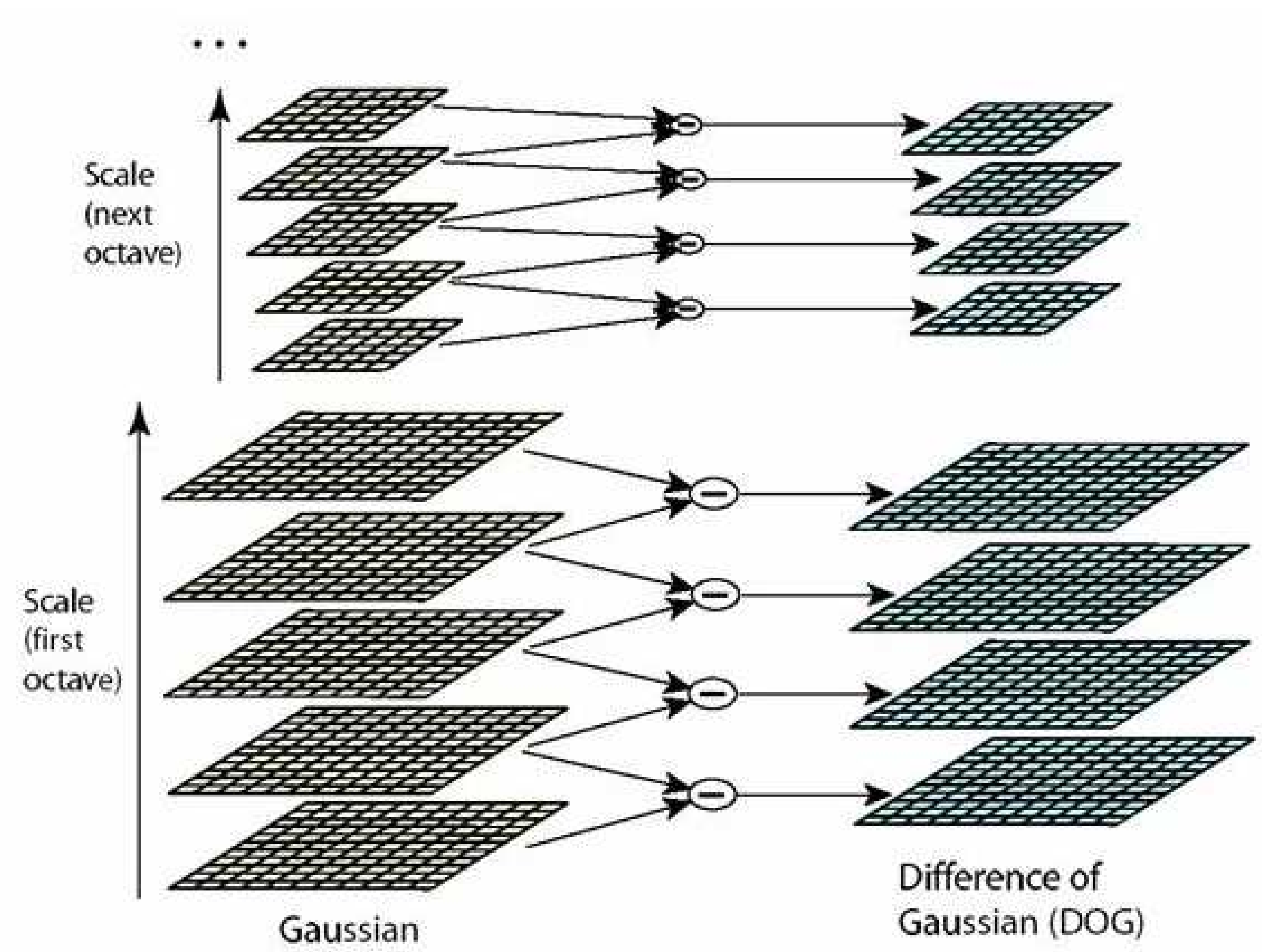
"Corner points" are identified when the eigenvalues of $T(\vec{x}, \sigma)$ are both relatively large. This indicates image gradients in the neighbourhood of $\vec{x}$ come in more than one direction (see the orientation tensor in the Canny tutorial). Such neighbourhoods are more likely to be distinctive.

Harris-DOG Interest Points

Position and Scale Selection: Interest point locations can be selected as local extrema in DOG (difference of Gaussian) filtered images. Here

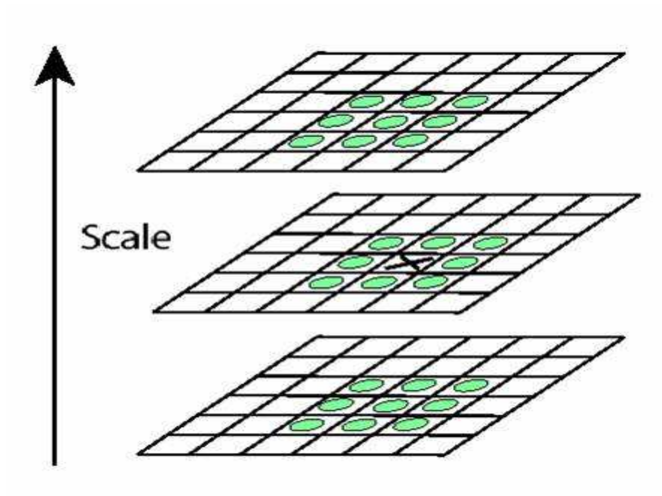
$$DOG(\vec{x}, \sigma) \equiv G(\vec{x}, \sigma) - G(\vec{x}, \rho\sigma),$$

with $\rho > 1$. Typical values for the spacing of adjacent scales, namely ρ, are $2^{\frac{1}{4}}$ or $2^{\frac{1}{3}}$.



(image from D. Lowe, CVPR03 tutorial).

The scale σ of an interest point can be selected as one at which the DOG response has a local extremum in σ as well.



(image from D. Lowe, CVPR03 tutorial).

Both the Harris corner condition and this local extrema condition (in both space and scale) must be satisfied at each interest point.

Many other forms of interest points have been developed. See the tutorial on local features presented at CVPR'03, available from:

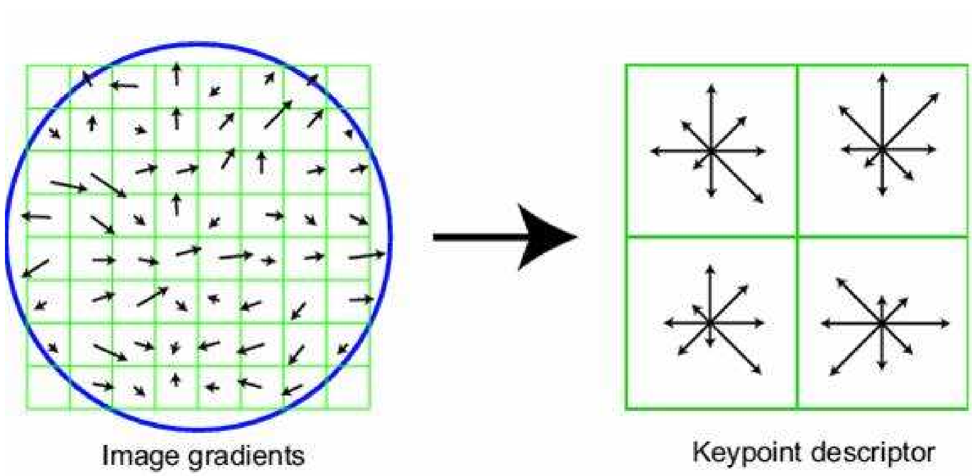
http://www.inrialpes.fr/lear/people/schmid/cvpr-tutorial03/

The critical property of an interest point detector is that it identifies image positions and scales $(\vec{x}, \sigma)$ of the **same** points on an object, despite significant changes in the imaging geometry, lighting, and noise.

Image Neighbourhood Descriptors

Once the interest points $(\vec{x}, \sigma)$ are detected at specific image locations and scales, we can then describe the image structure in the neighbourhood of $(\vec{x}, \sigma)$. The size of the neighbourhood encoded is proportional to σ.

For example, the scale invariant feature transform (SIFT) of the neighbourhood computes a 128 dimensional vector formed from histograms of the image gradients in the neighbourhood of the interest point.



(image from D. Lowe, CVPR03 tutorial).

In addition, the gradient histograms can be used to select a canonical image orientation.

Other descriptors are formed from combinations of higher order Gaussian derivative filters (see papers by Schmid et al, available from the CVPR'03 tutorial listed above), or from steerable filters (see Carneiro and Jepson, www.cs.toronto.edu/~carneiro).

Viewpoint Insensitivity

A set of local image features is extracted from a model image. Each of these features contains a record of its:

- **Position**, that is, the pixel location $\vec{x}$;
- **Scale**, the particular value of σ;
- **Orientation**, the dominant orientation of the image structure in the neighbourhood;
- **Local Image Structure in Canonical Coordinates**, encoded in terms of gradient histograms (eg. SIFT), or local phase properties of steerable filters (eg. Carneiro), or high order image derivatives (eg Schmid). The local image structure **is encoded relative to** the position, scale and orientation determined by the interest point detector.

Given a test image, local features can be extracted in the same manner.

The features from the test image can be compared directly to the features obtained from the model image, despite changes in position, scale and orientation.

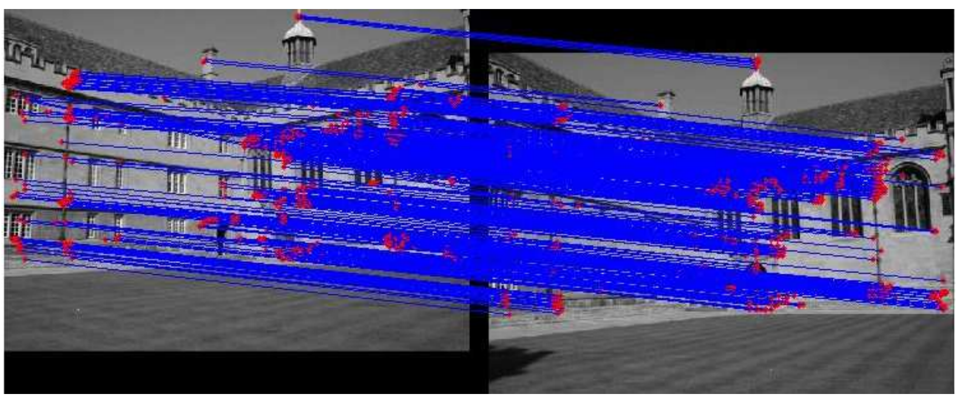
This partial invariance to the viewing geometry is a consequence of encoding the local image structure relative to the position, scale and orientation at the interest point. We are leaning on the feature point detector to get these quantities the same in both the model and test images.

Matching Wadham College

Matches using phase-based local features and their spatial layout.

Image Pair.

Corresponding Points.



Epipolar Lines.

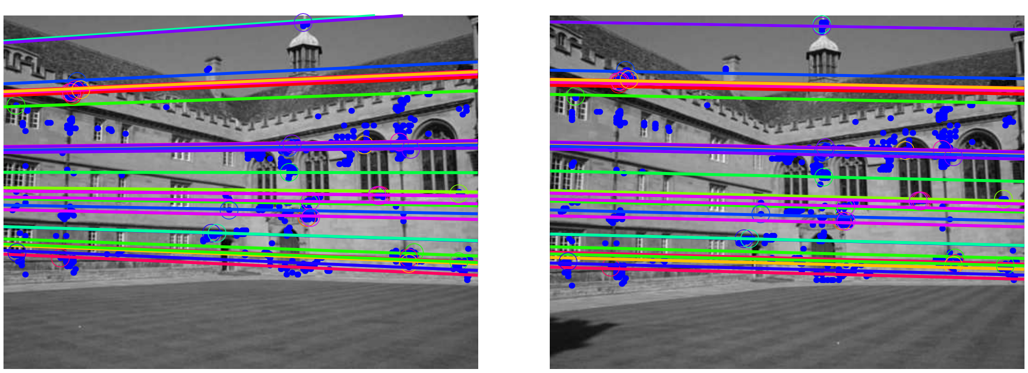


Image data available from: www.robots.ox.ac.uk/~vgg/data/

Matching the Valbonne Church

Matches using phase-based local features and their spatial layout.

Change in Scale and Position.

Change in Scale, Position, 3D Viewpoint, and Brightness.

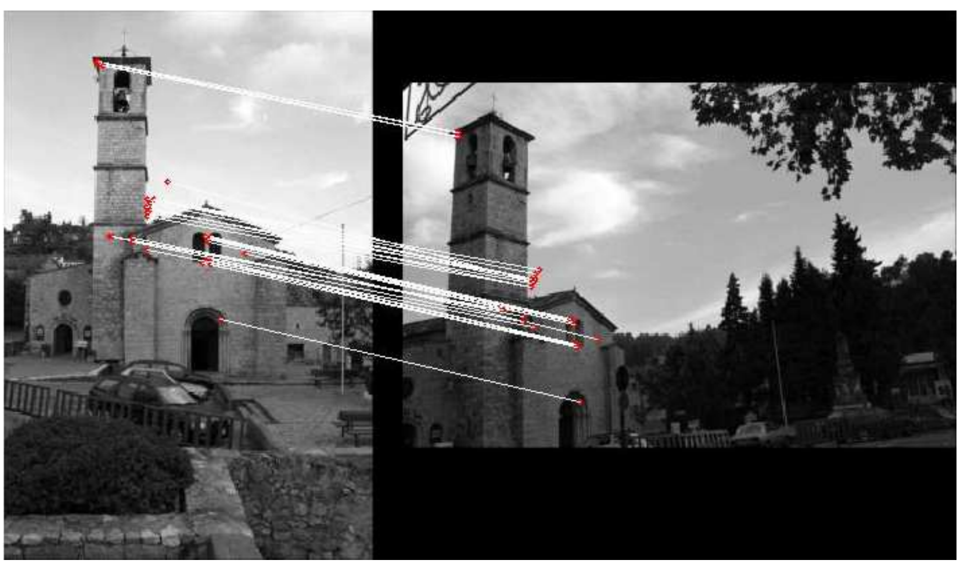


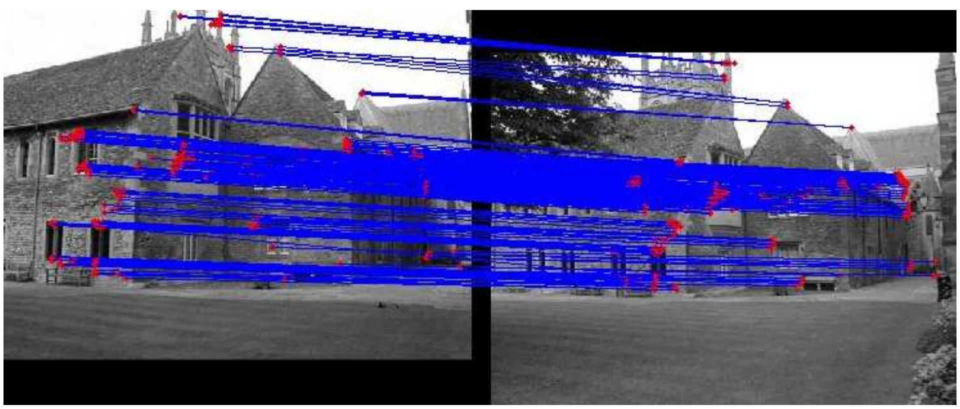
Image data available from: www.robots.ox.ac.uk/~vgg/data/

Matching Merton College

Matches using phase-based local features and their spatial layout.

Image Pair.

Corresponding Points.



Epipolar Lines.

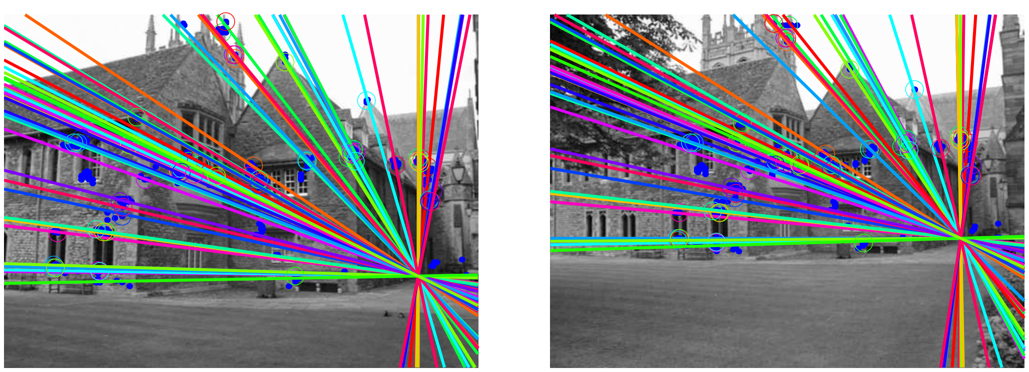


Image data available from: www.robots.ox.ac.uk/~vgg/data/