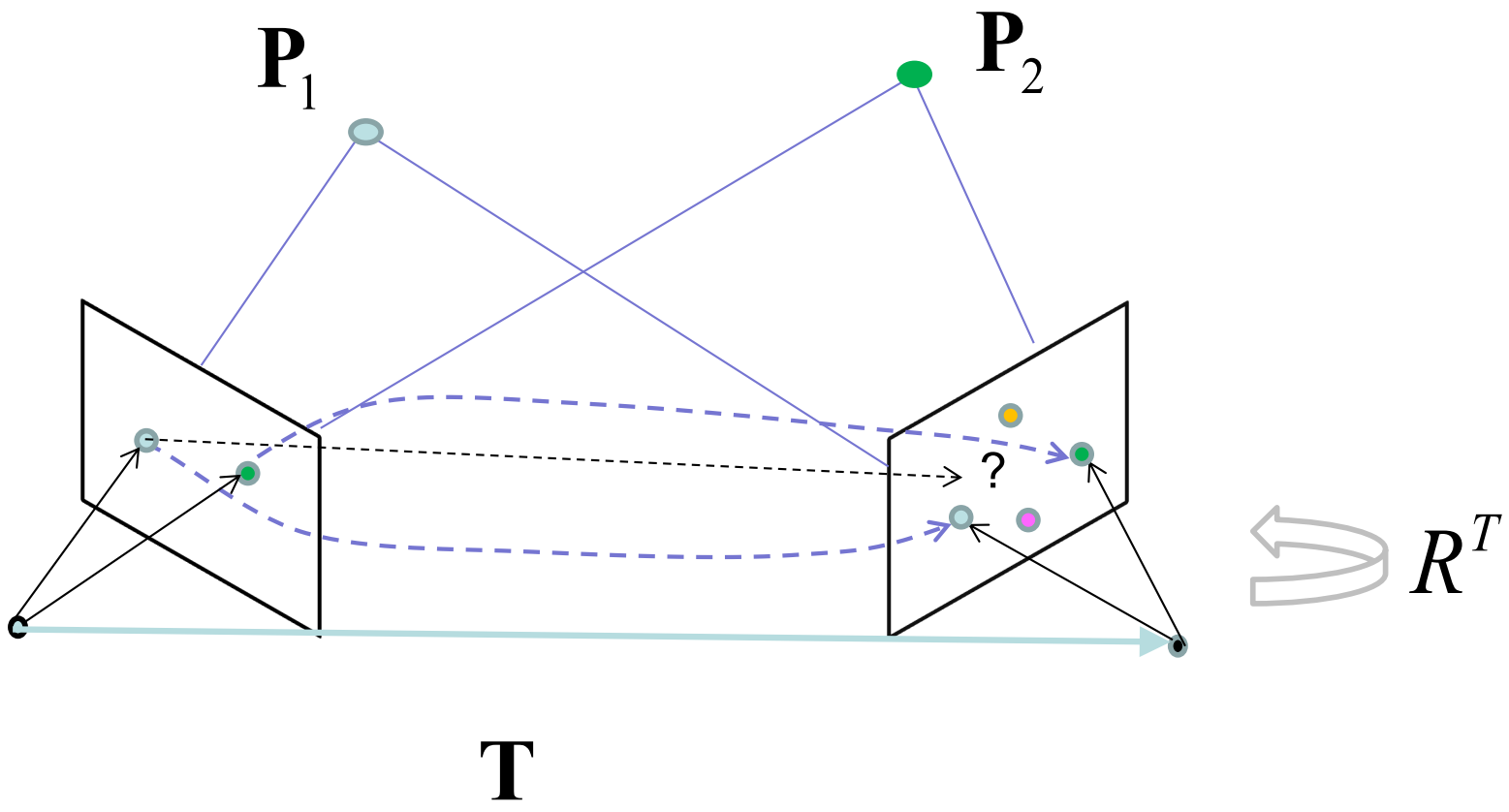
COMS30030
Image Processing and Computer Vision

Stereo – Correspondence Matching

Andrew Calway

andrew@cs.bris.ac.uk

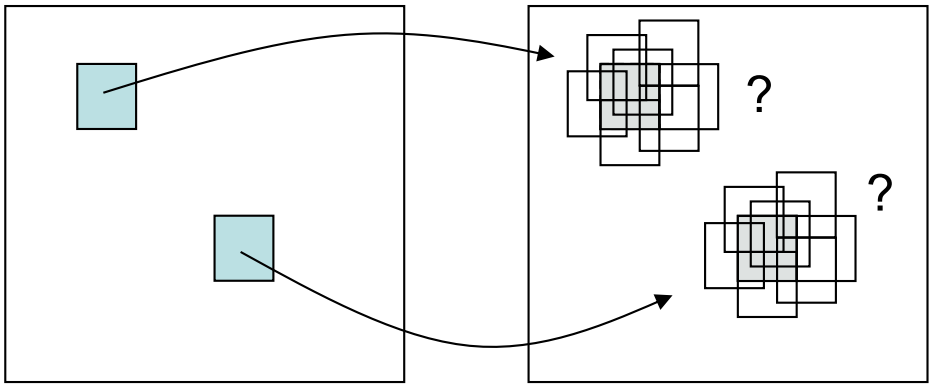
Stereo Correspondence

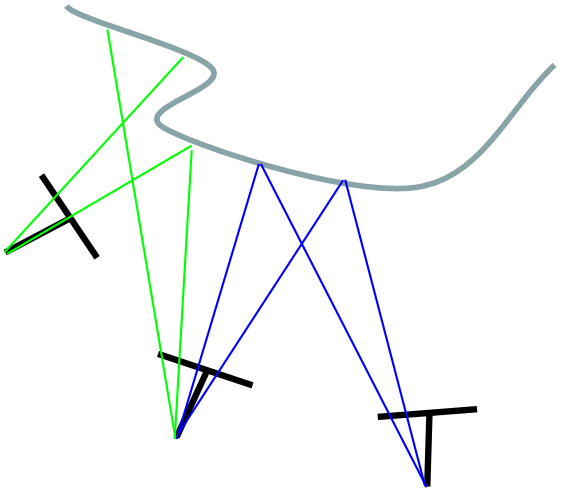


Stereo Matching is Hard

Region-Based Methods

- Compare pixel values within regions in two views.
- For region in left image, compute similarity with regions of same size in right image.
- Corresponding point - centre of most similar region





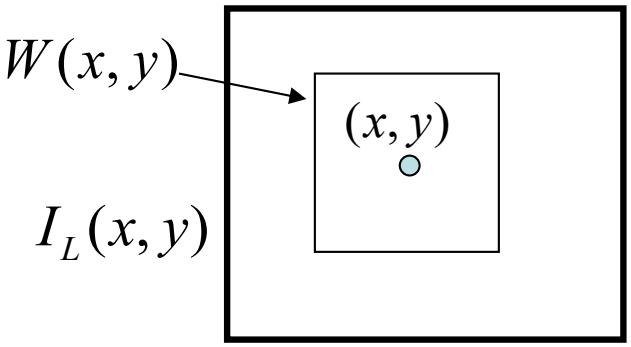
Region Matching

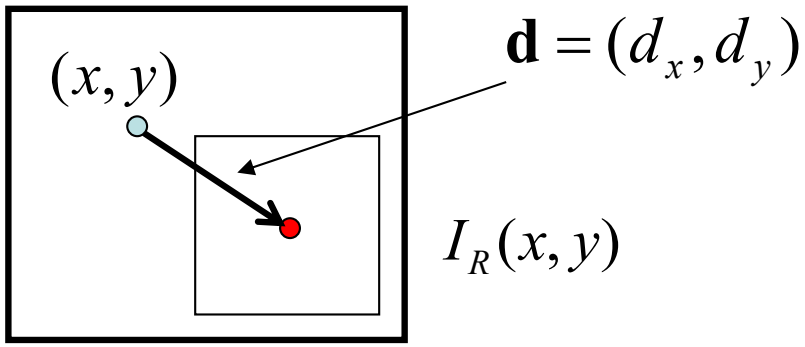
- Stereo image pair: $I_L(x,y)$ and $I_R(x,y)$
- For each pixel, find disparity $\mathbf{d}=(d_x,d_y)$ which minimises (or maximises) cost function

$$c(\mathbf{d}) = \sum_{(i,j)\in W(x,y)} s\left[I_L(i,j), I_R\left(i+d_x, j+d_y\right)\right]$$

similarity measure

$W(x,y)$ ⇨ window of pixels around (x,y)





Similarity Measures

- **Sum of squared differences:** $s(u,v) = (u-v)^2$
- **Similar pixel count:**

$$s(u,v) = \begin{cases} 1 & if\,|u-v| < T \\ 0 & else \end{cases}$$

- **Normalised cross correlation**:

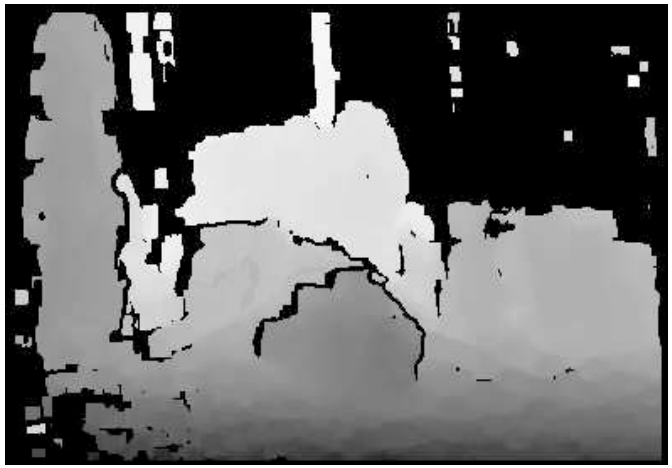
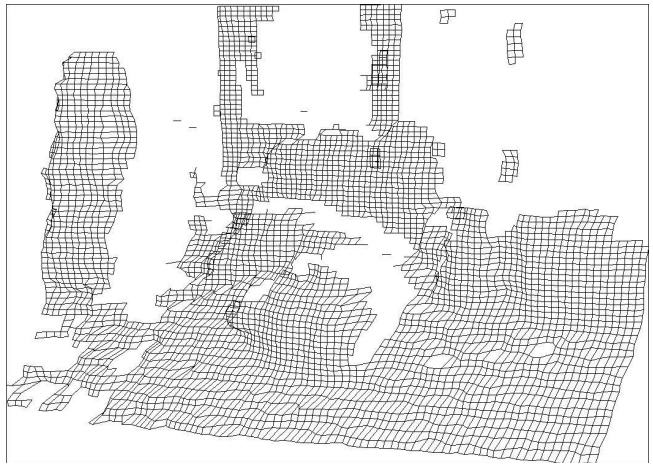
$$s(u,v) = (u-\bar{u})(v-\bar{v})/N_u N_v$$

$$\bar{u} = \frac{1}{|W|}\sum_{u \in W} u$$ mean

$$N_u = \sqrt{\sum_{u \in W} (u-\bar{u})^2}$$ Std deviation

Example

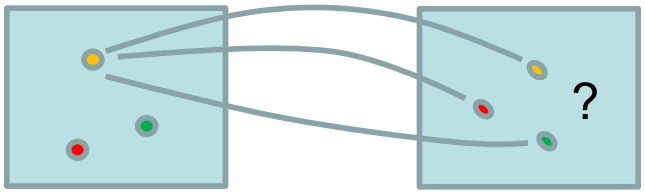


Real time correlation-based stereo, Faugeras et al, 1993

Feature-Based Methods

- Restrict search to sparse set of features
  - reduces mis-matches caused by texture-less regions
- Find salient (distinct) points in each view and match points by comparing pixels or image descriptors in local regions about each point

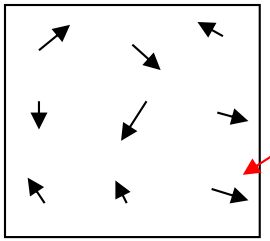


- Examples
  - Harris corner detector (salient points)
  - Scale-Invariant Feature Transform (SIFT)

Harris Corner Detector

- Detects salient (distinct, interesting) image points
- Covariance of spatial gradient vectors within region W

$$A = \sum_{x,y \in W} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$



spatial gradient

$$I_x \equiv I_x(x, y)$$

- Eigenvalues λ_1 and λ_2 of A indicate 'spread' of gradients in region, e.g. 2 high values → 'busy' region.
- Example saliency metric: $sal = \lambda_1 \lambda_2 / (\lambda_1 + \lambda_2)$
- Properties:
  - if eigenvalues both large → sal large
  - if either eigenvalue small → sal small

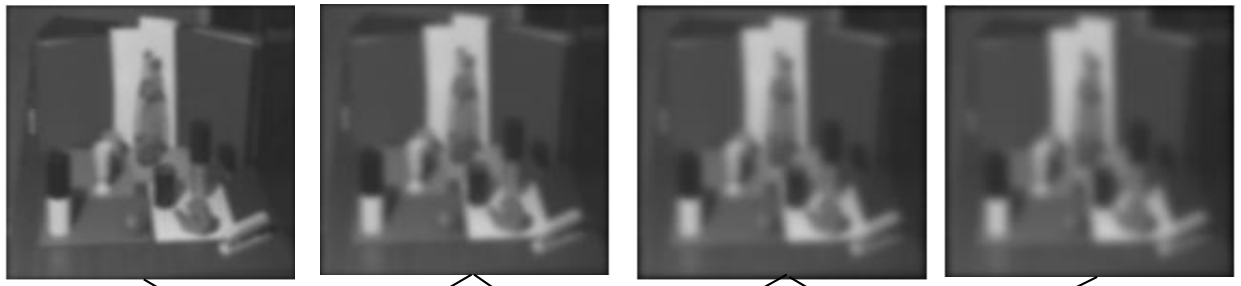
SIFT Matching

- **Two main elements**:
  - scale invariant detection of salient (key) points
  - matching by highly distinct local descriptors
- **Key point detection**:
  - extrema (max or min) in difference of Gaussian blurred versions of image → Difference-of-Gaussians (DoG) tree
  - points imaged at different resolutions appear at different levels of DoG tree ⟶ scale invariance
- **Spatial gradient descriptors:**
  - built from histograms of spatial gradients in local neighbourhood
  - invariant to rotation and perspective warp (almost)

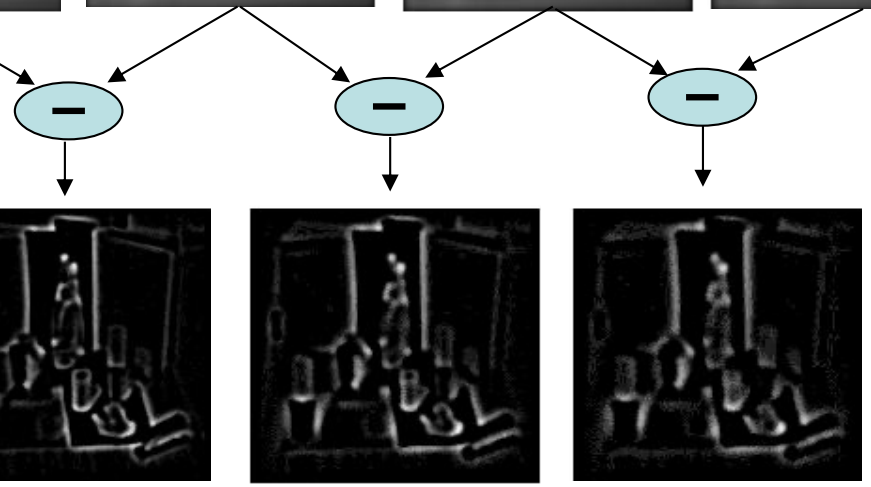
Difference of Gaussians

Most extrema present in both views possibly at different levels – **scale invariance**

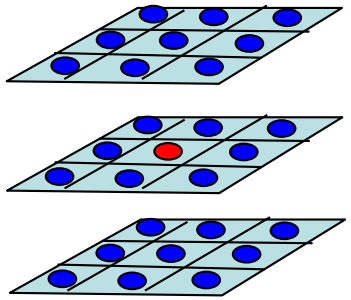
Gaussian tree



scale
('blur')

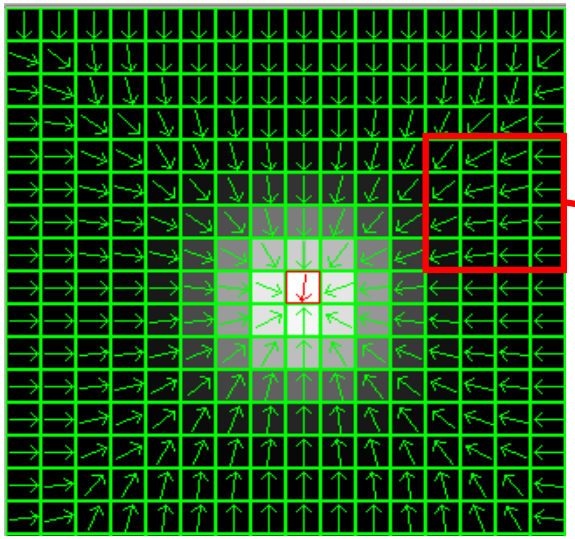


DoG tree

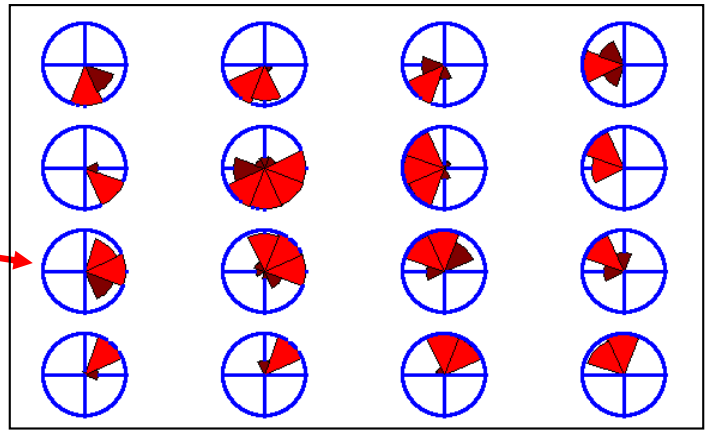


key points are extrema in DoG tree

Spatial Gradient Descriptor



16x16 gradient field



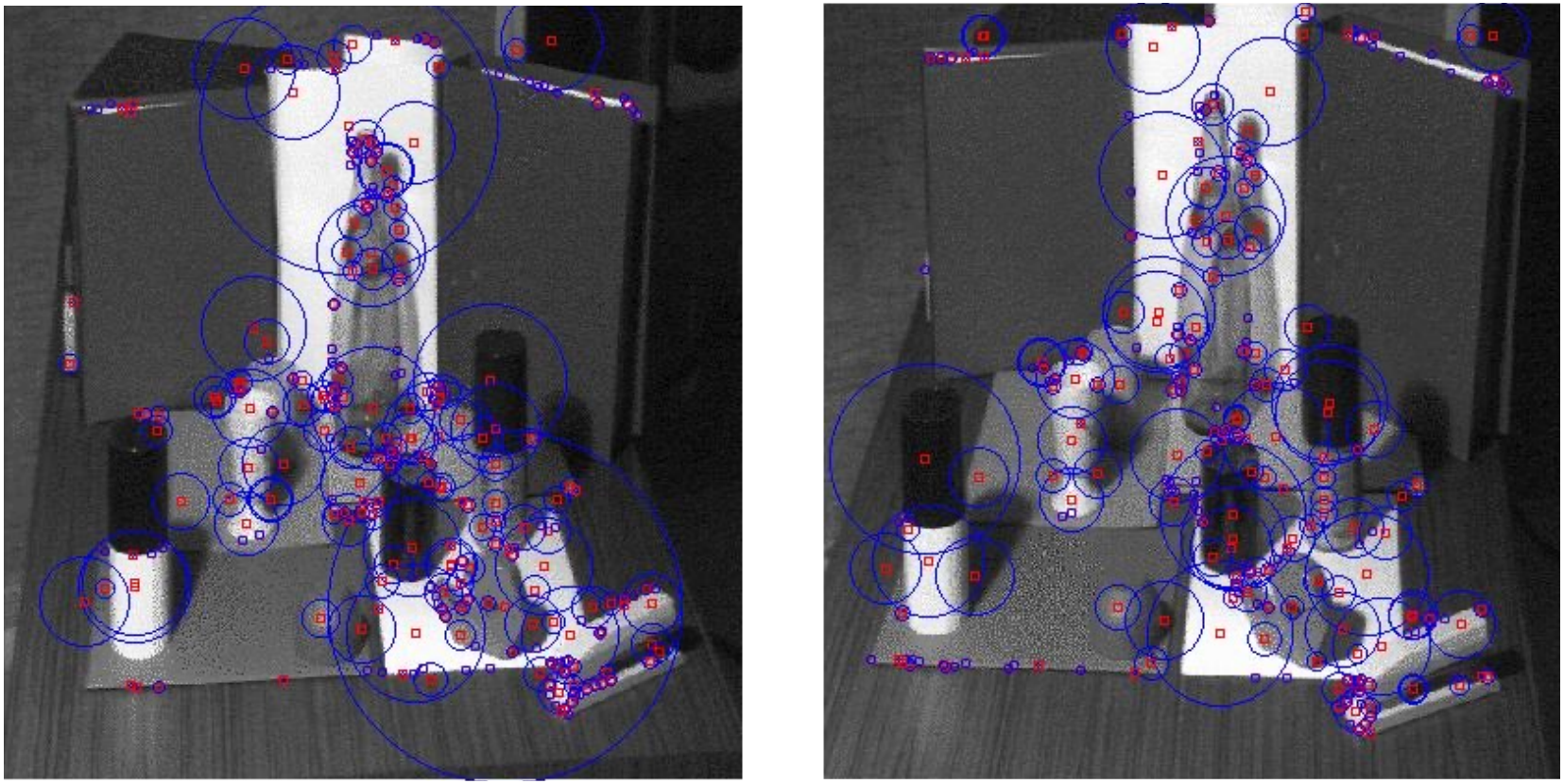
Orientation histograms (8 bins) over 4x4 blocks → 4x4 matrix of histograms, normalised → 128-D feature vector.

Histograms built w.r.t dominant orientation in region – **rotation invariance** (almost)

Descriptor matching based on Euclidean distance between 128-D feature vectors

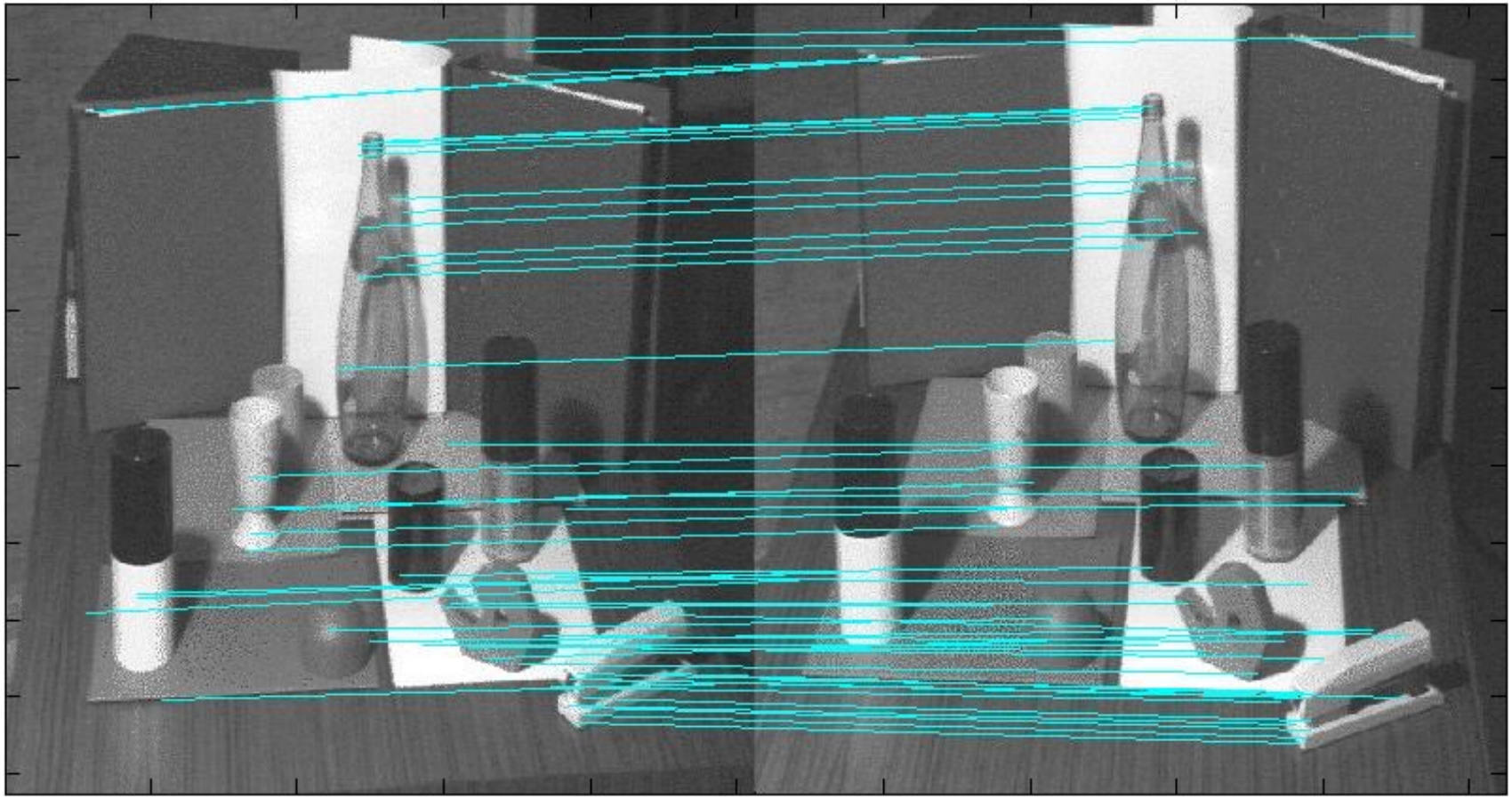
Distinctive Image Features from Scale-Invariant Keypoints, David G. Lowe, International Journal of Computer Vision, 2004.

SIFT Example – Selected Key Points



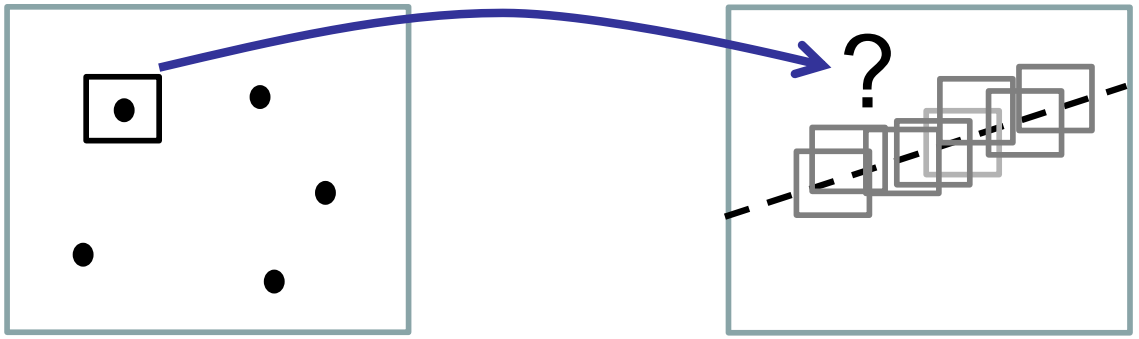
Each key point shown with circle indicating scale

SIFT Example – Matched Key Points



Where to look - Calibrated

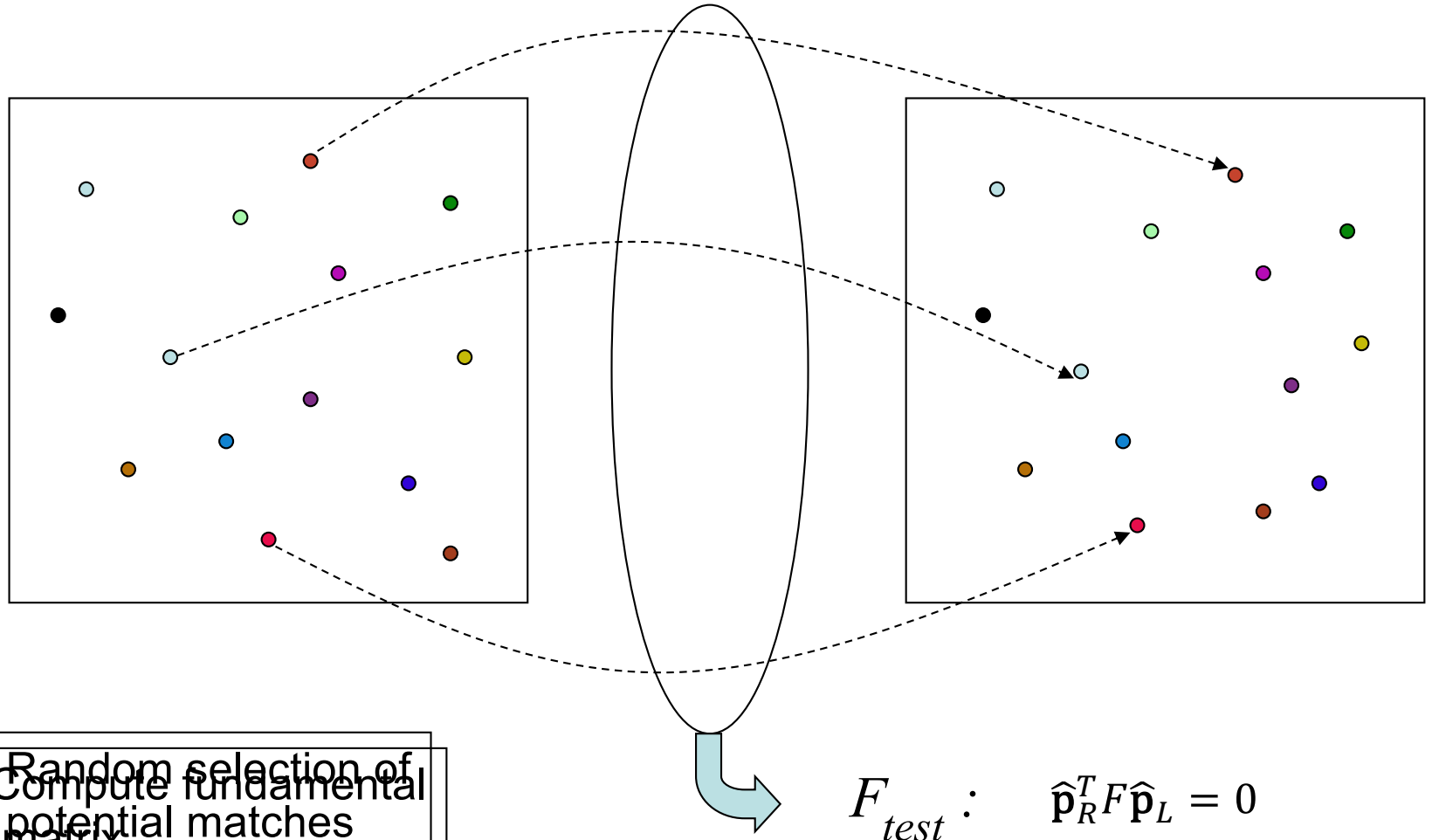
- For calibrated stereo set up, corresponding points lie on epipolar lines
- Hence, given point in left image, search for point in right image along band about epipolar line:
- Increases speed, reduces mismatches



Where to look - Uncalibrated

- When geometry unknown, can only match points using pixel values – region or feature based approach.
- Often leads to mismatches amongst true matches, known as **outliers** and **inliers**.
- We know that inliers will be related by epipolar constraint equation $\hat{\mathbf{p}}_R^T F \hat{\mathbf{p}}_L = 0$
- We can use **RAN**dom **SA**mple **C**onsensus to sort out:
  - select subset of matches at random (minimum 4)
  - compute fundamental matrix F from subset (lecture 8, slide 7)
  - assess support for F amongst other correspondences
  - repeat until best F found

RANSAC Matching

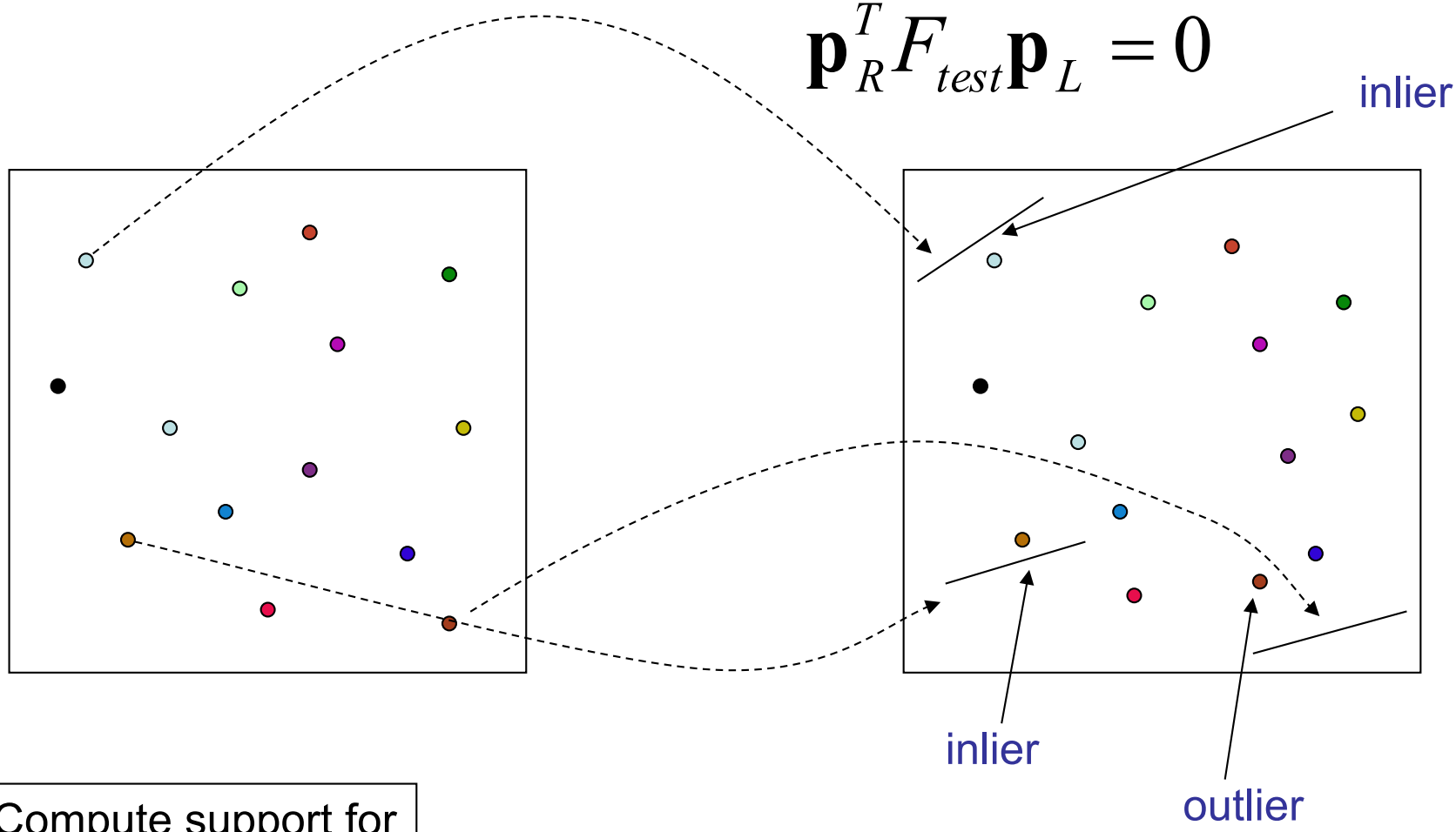


1. Random selection of potential matches
2. Compute fundamental matrix

$$F_{test}: \quad \hat{\mathbf{p}}_R^T F \hat{\mathbf{p}}_L = 0$$

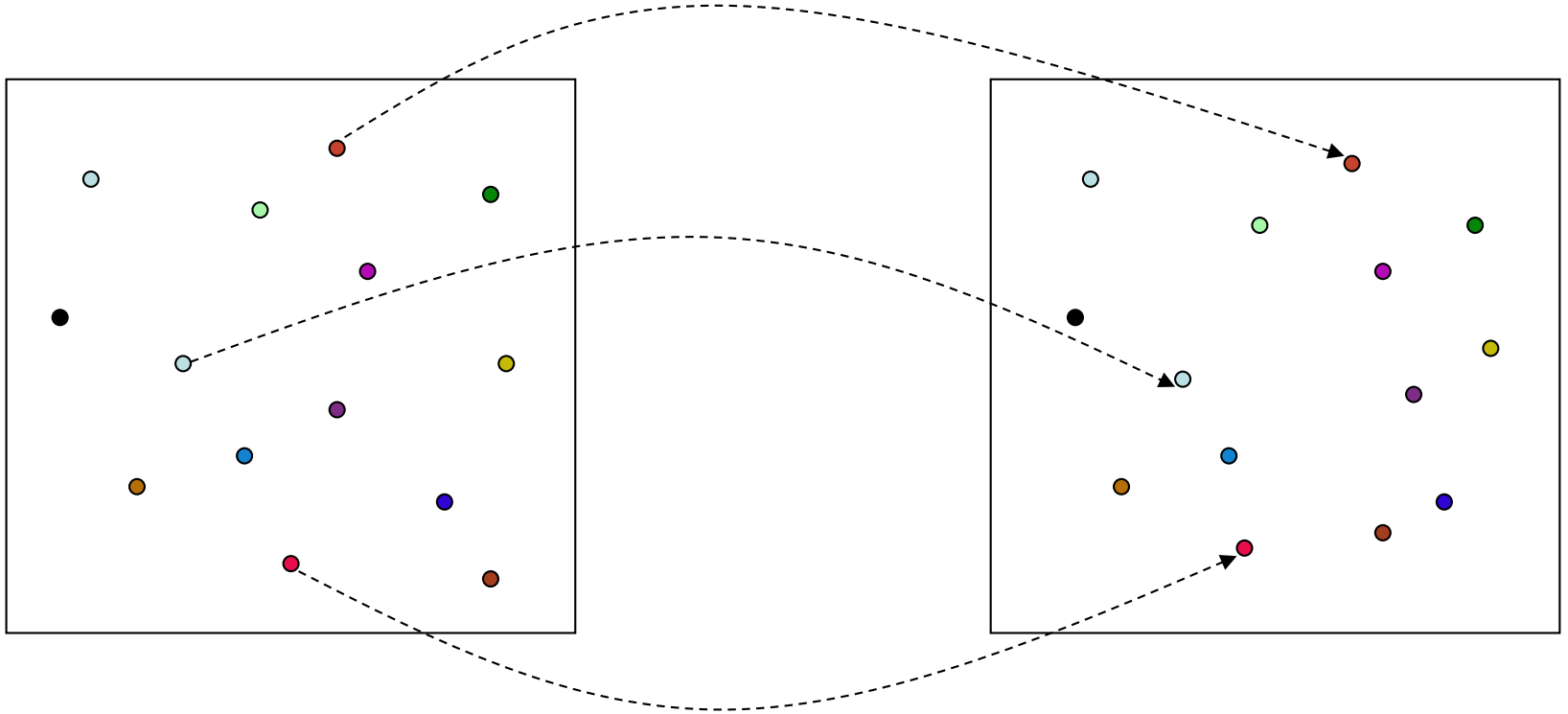
RANSAC Matching

$$\mathbf{p}_R^T F_{test} \mathbf{p}_L = 0$$



3. Compute support for fundamental matrix

RANSAC Matching

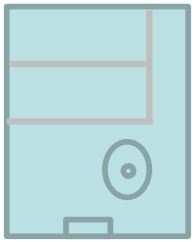


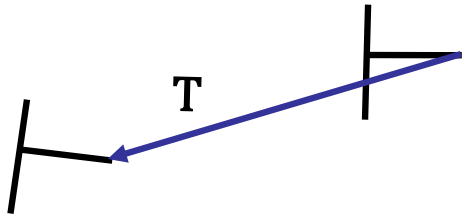
1. Random selection of potential matches

…… and so on

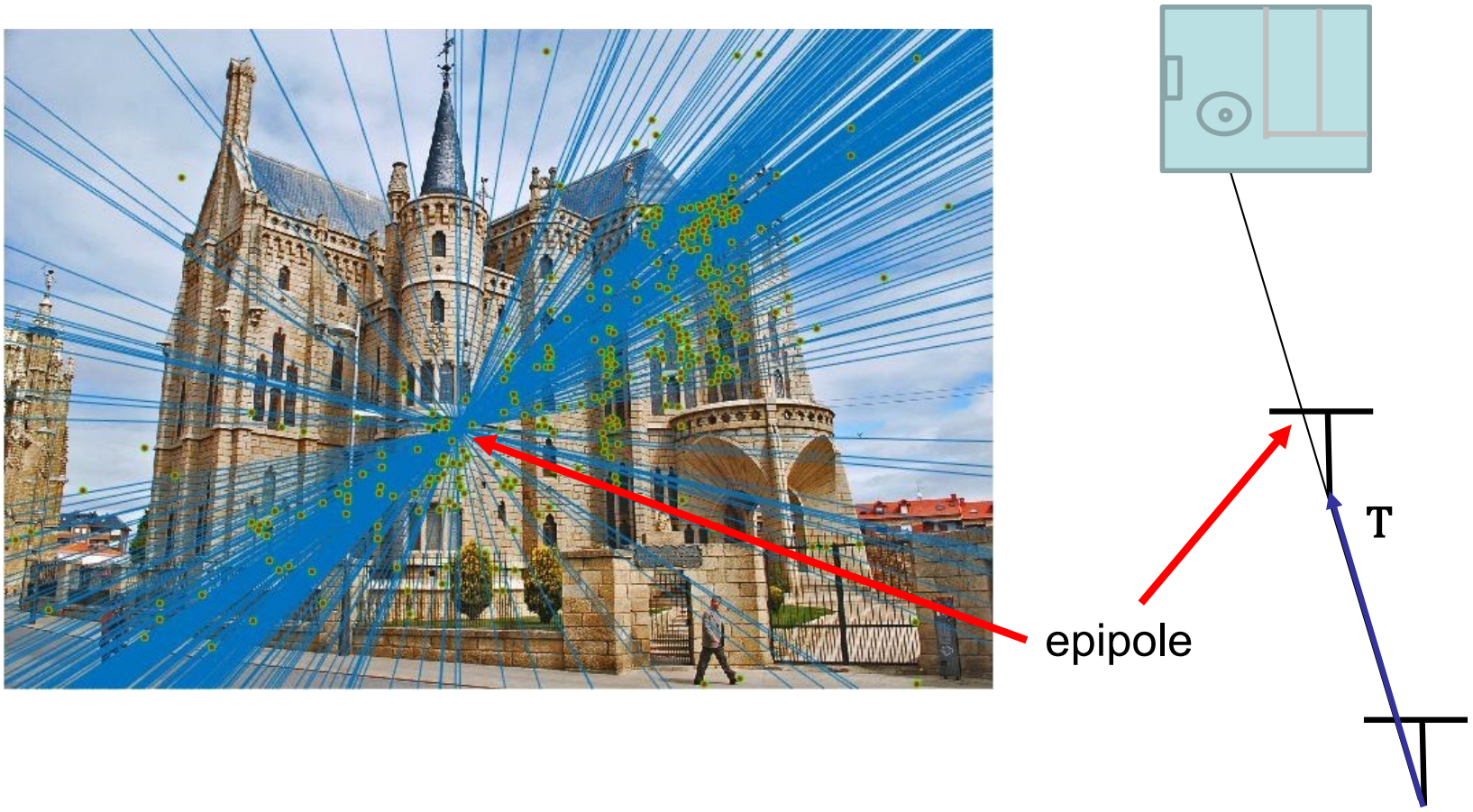
F with most support defines outliers and inliers → most likely correspondences

SIFT Matching – Keypoint Matches

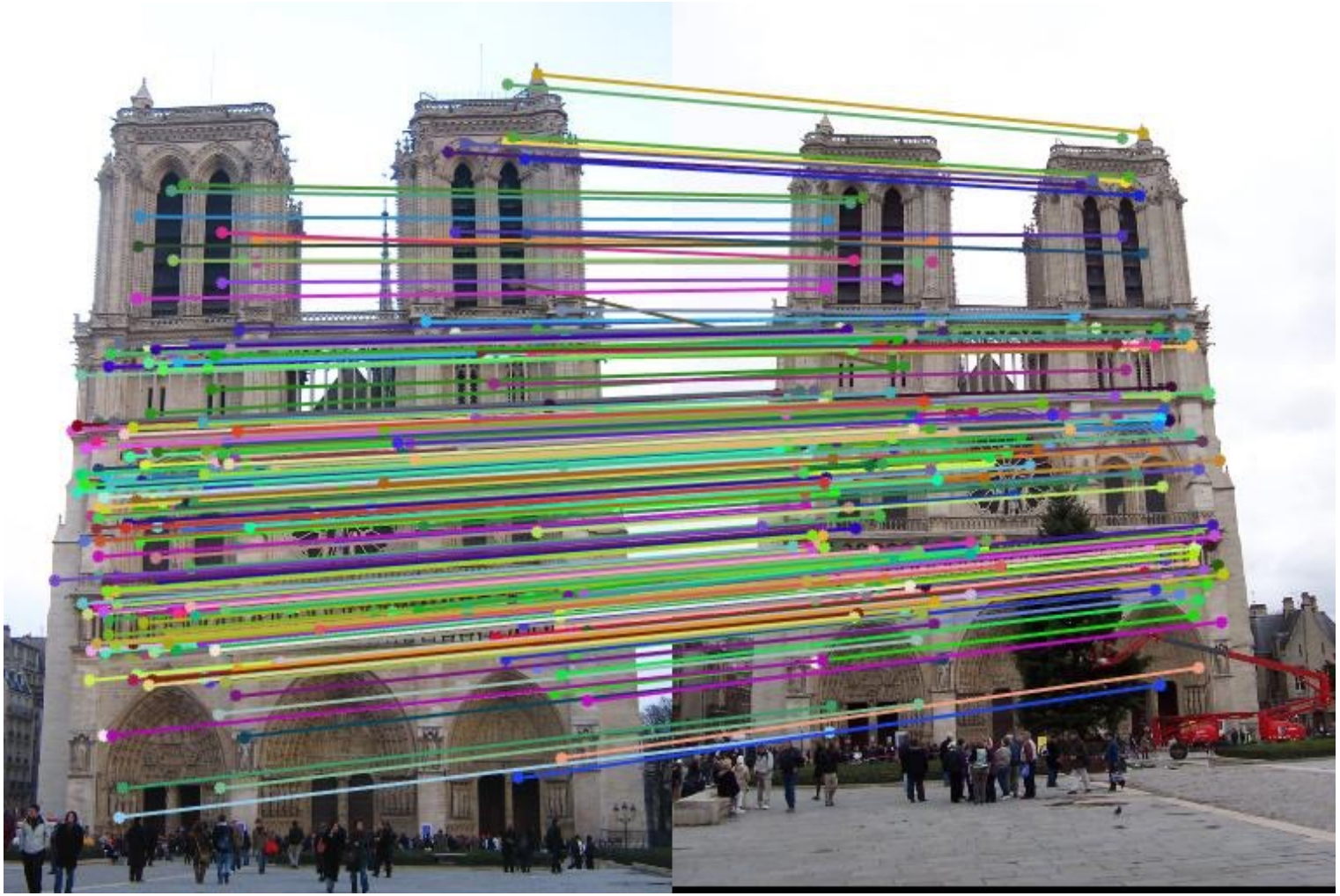




SIFT Matching – Epipolar Lines



SIFT Matching – Keypoint Matches



SIFT Matching – Epipolar Lines