McGill
Centre for
Intelligent Machines

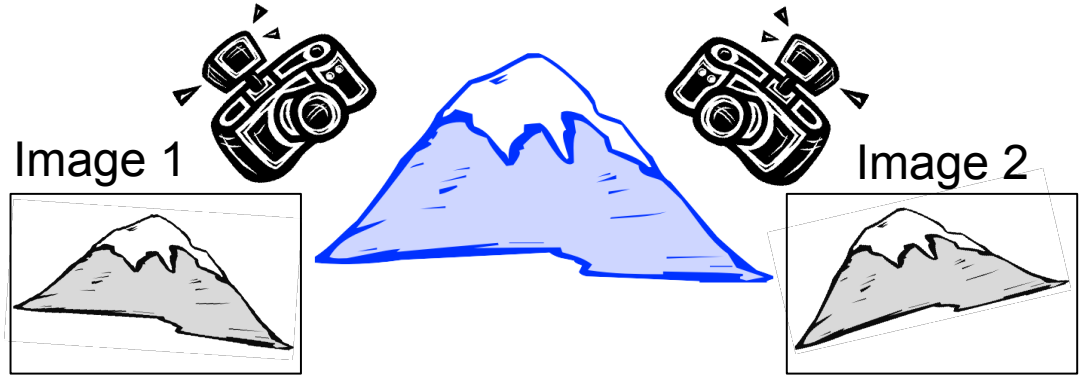
Recovering Scene Geometry via TIP: the Transfer of Invariant Parameters

Frank Riggi, Matthew Toews and Tal Arbel
Centre for Intelligent Machines, McGill University
Montreal, Quebec, Canada

McGill
Centre for
Intelligent Machines

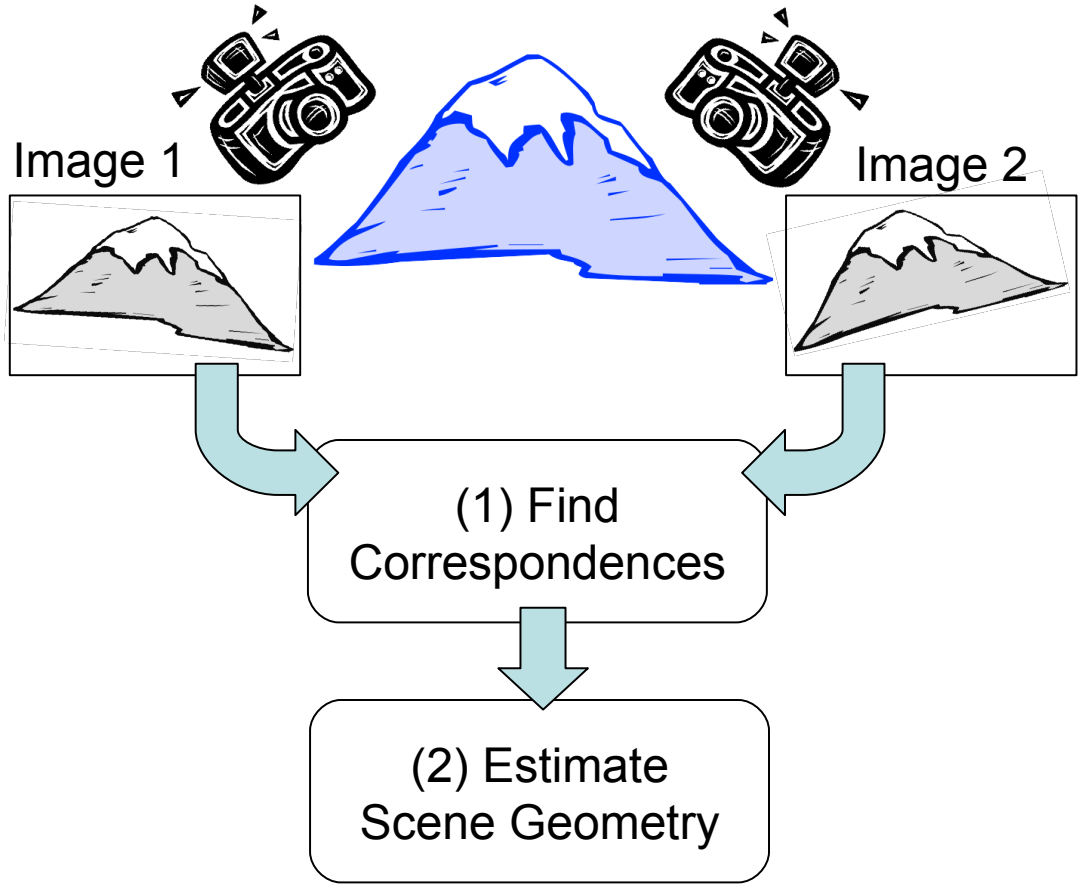
Context: 3D Scene Geometry from 2D Images

- The 3D geometry of a scene can be recovered from two or more images of the scene.
- Applications: 3D modeling, ego motion, …



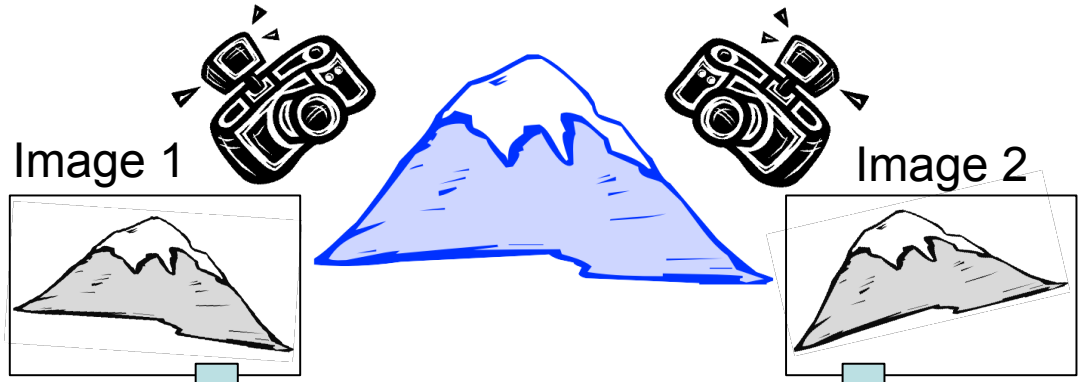
McGill
Centre for
Intelligent Machines

Context: 3D Scene Geometry from 2D Images



McGill
Centre for
Intelligent Machines

Context: 3D Scene Geometry from 2D Images

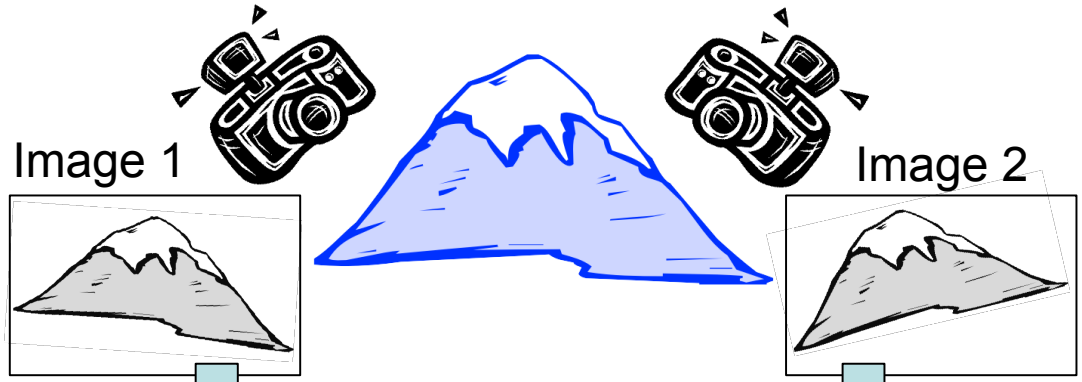


(1) Find Correspondences

Feature Matching
Point-to-point
Invariant features
i.e. SIFT

(2) Estimate Scene Geometry

Context: 3D Scene Geometry from 2D Images



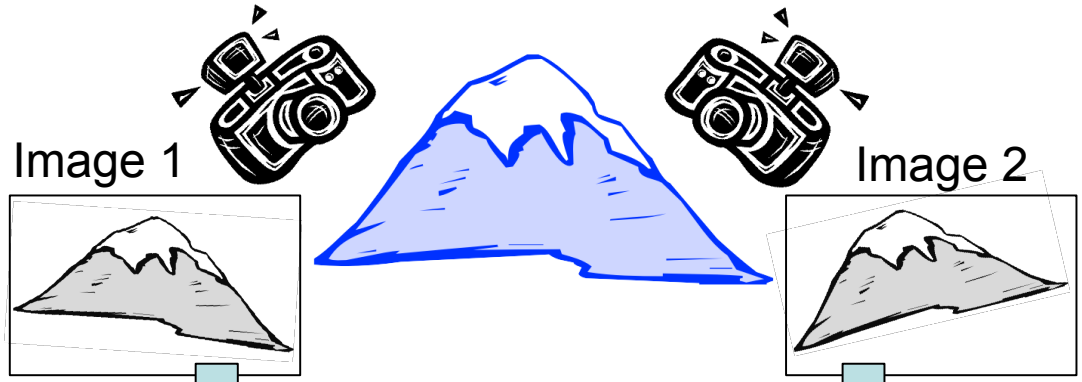
(1) Find Correspondences

(2) Estimate Scene Geometry

Epipolar geometry
Fundamental matrix
Point-based algorithms
i.e. the 8-point algorithm

McGill
Centre for
Intelligent Machines

Context: 3D Scene Geometry from 2D Images

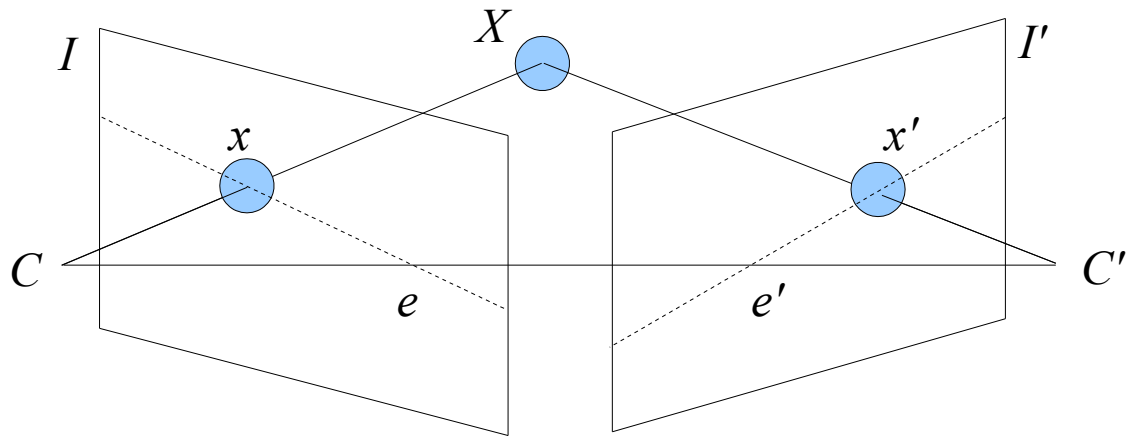


(1) Find Correspondences

(2) Estimate Scene Geometry

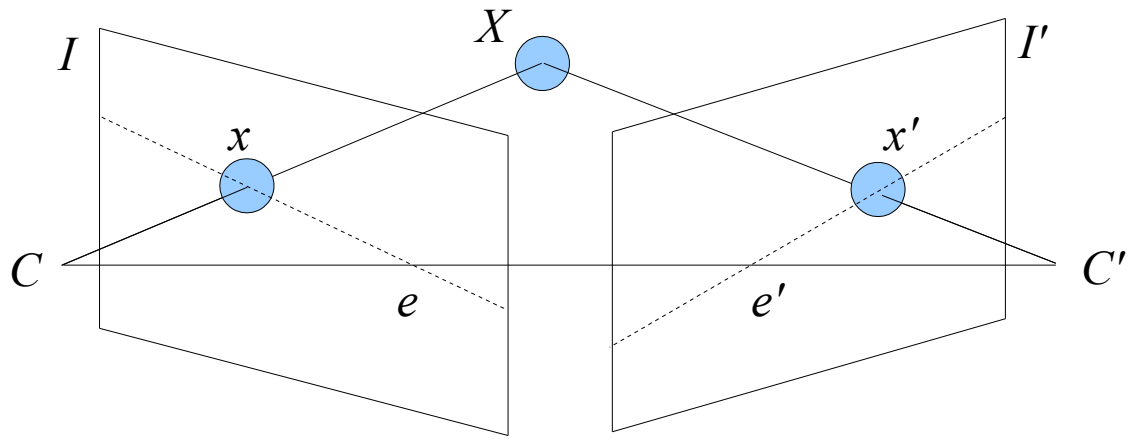
(3) **The TIP method:**
- Invariant feature correspondences
- Point-based geometry estimation

Scene Geometry



- Epipolar geometry
  - Relates a point X in a 3D scene to it's projections x, x' in different images I, I' of the scene.
  - Recovering the 3D scene geometry requires estimating the epipolar geometry.

Scene Geometry



- Fundamental matrix (FM)
  - Describes the epipolar geometry of an uncalibrated 2-image system.
  - A *3*x*3*, rank *2* matrix F satisfying:

    $$x'^{T}Fx = 0 \text{ and } \det(F) = 0,$$

    where x and x' are corresponding points in I and I'.

Fundamental Matrix Estimation

- Point-based methods:
  - Point-to-point correspondences.
  - 8-point algorithm: linear solution, 8+ correspondences.
  - 7-point algorithm: cubic solution, 7 correspondences.
  - Widely used: Intel OpenCV, others…
- Difficulties
  - Quality of estimation dependent on number and accuracy of correspondences.
  - 8 point-to-point correspondences required for linear estimation.
  - Sparse or noisy correspondence = unstable estimation.

Fundamental Matrix Estimation

- Other methods:
  - Higher order correspondences, i.e. lines, curves, conics.
  - Typically involve finding roots of Kruppa's equations.
  - Less popular.

Image Correspondence

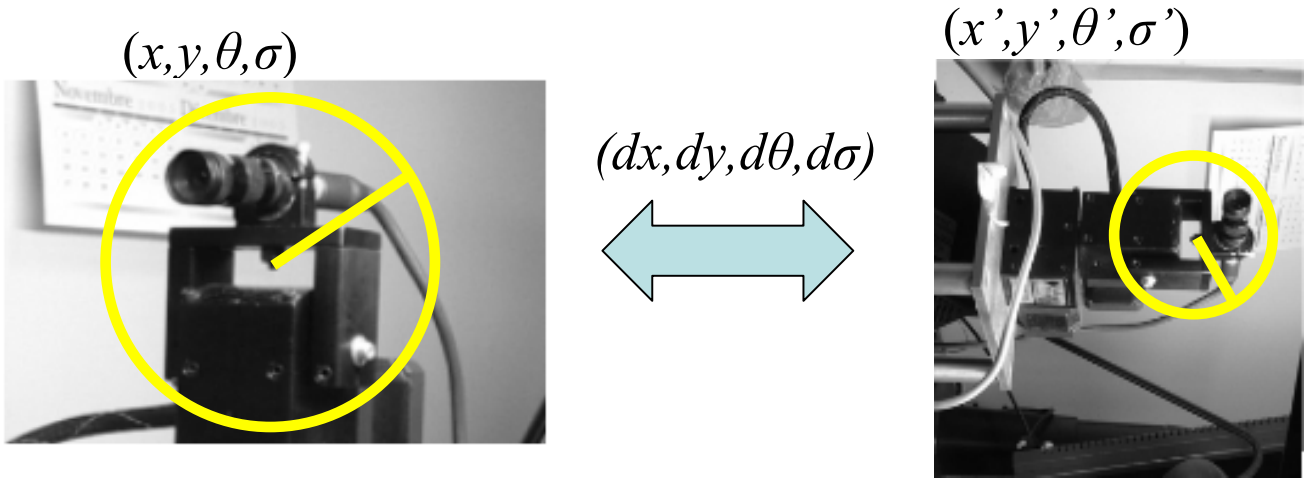
- Image correspondence
  - Identify points in different images arising from the same 3D world point.
- General approach
  - Consider informative image points that can be readily detected, and matched between images, e.g. corners.
- Difficulty
  - Detecting and matching points in the presence of image transformation arising from viewpoint change.

Invariant Feature Correspondence

- Invariant features
  - Image regions that are invariant under a transformation group, e.g. similarity transform, affine transform.
- Benefits
  - Robust to match over large changes in viewpoint.
  - Efficient to detect using image pyramids.
  - Generally applicable to many types of scenes.
  - Model a variety of image characteristics, i.e. edges, blobs, phase, entropy.
- Examples:
  - Similarity transform: scale-invariant features (SIFT, Lowe 2004)
  - Affine transform: affine-invariant features (Mikolajczyk, 2004).

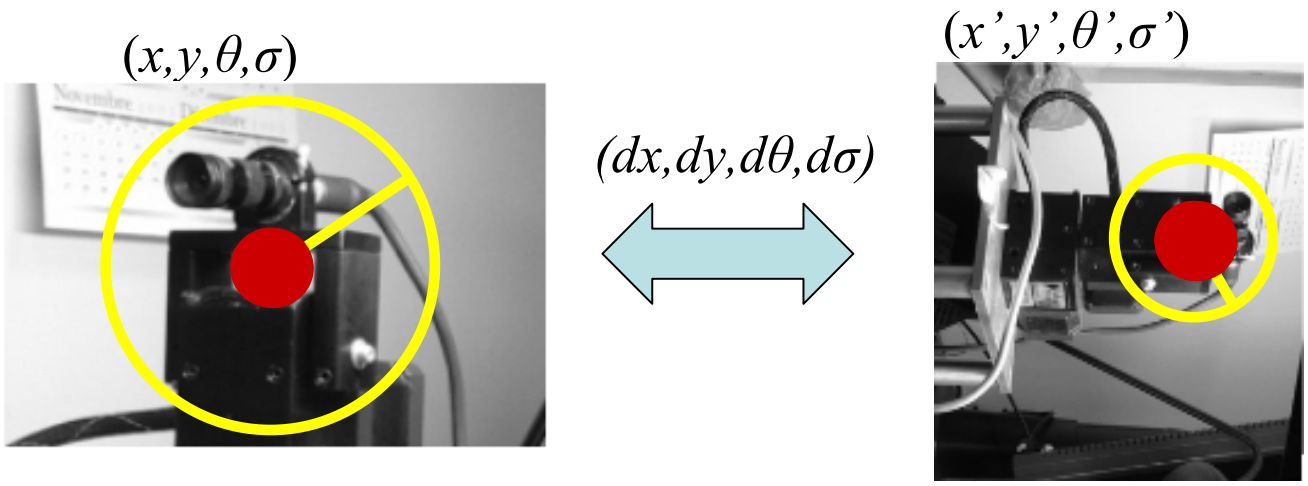
Invariant Feature Correspondence

- Invariant feature region:
  - Described by K parameters, e.g. location, orientation and scale (x,y,θ,σ).
- Invariant feature correspondence:
  - A K-parameter transform between two regions, i.e. $(dx,dy,d\theta,d\sigma)$.



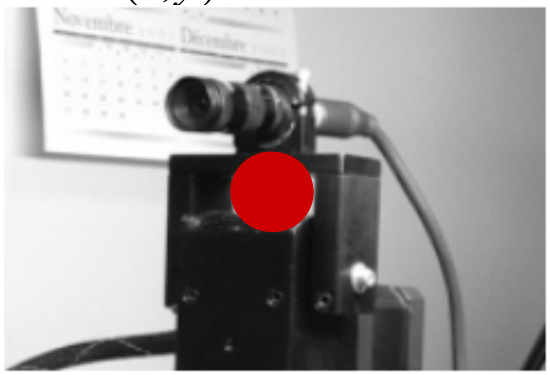
Invariant Features and FM Estimation

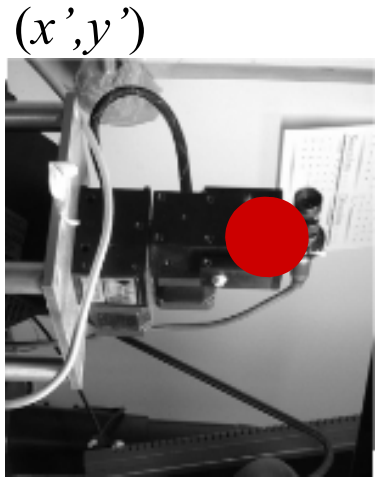
- Standard approach
  1) Use invariant region centers as point-to-point correspondences.
  2) Perform point-based FM estimation.



Invariant Features and FM Estimation

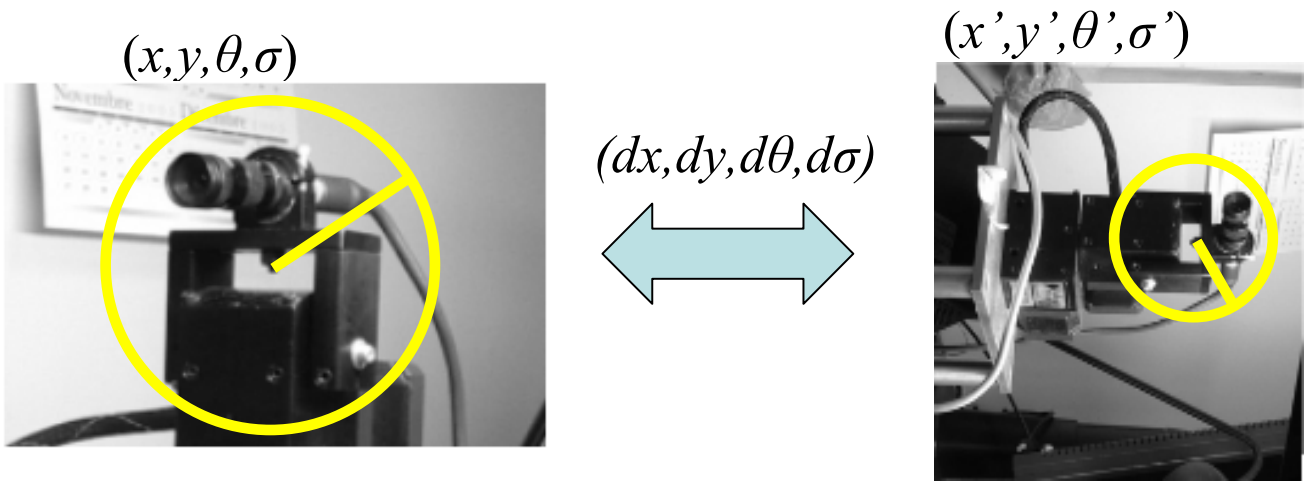
- Standard approach
  1) Use invariant region centers as point-to-point correspondences.
  2) Perform point-based FM estimation.
- Problem
  - Important geometrical information is thrown away!





Question

- Can additional geometrical information inherent to invariant feature correspondences be used to improve the quality of FM estimation?



Motivation

- 3D scene viewed as points

Motivation

- 3D scene viewed as invariant regions
  - More informative?

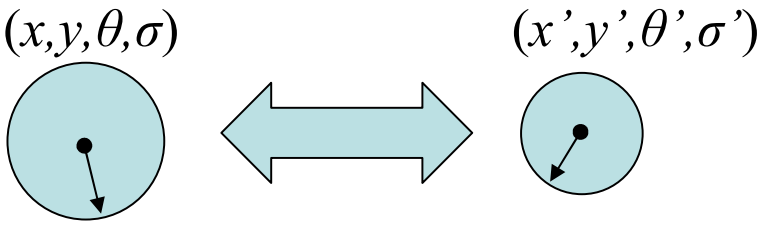
McGill
Centre for
Intelligent Machines

Motivation

- Original 3D scene

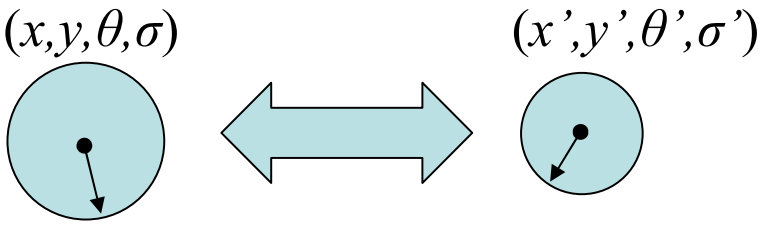
Transfer of Invariant Parameters (TIP)

- Each invariant feature correspondence represents a local geometric transform from one image to the next.



Transfer of Invariant Parameters (TIP)

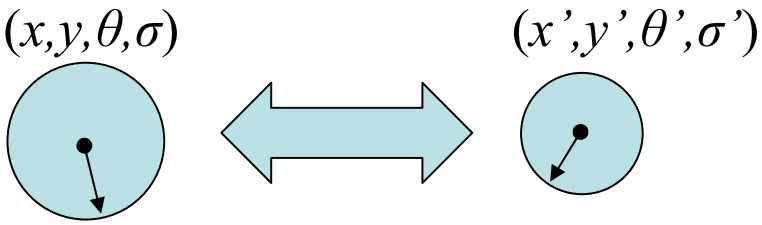
- Each invariant feature correspondence represents a local geometric transform from one image to the next.



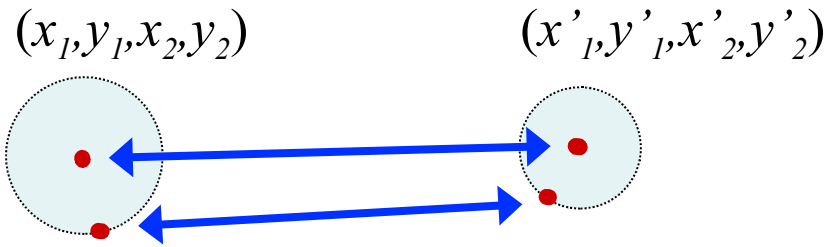
- This transform can be equivalently parameterized in terms of point-to-point matches: ***TIP matches***.

Transfer of Invariant Parameters (TIP)

- Each invariant feature correspondence represents a local geometric transform from one image to the next.



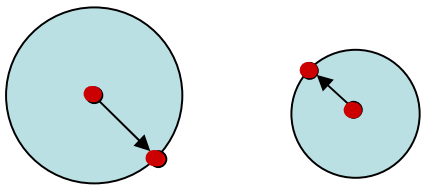
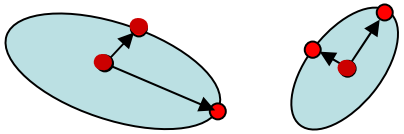
- This transform can be equivalently parameterized in terms of point-to-point matches: ***TIP matches***.



Transfer of Invariant Parameters (TIP)

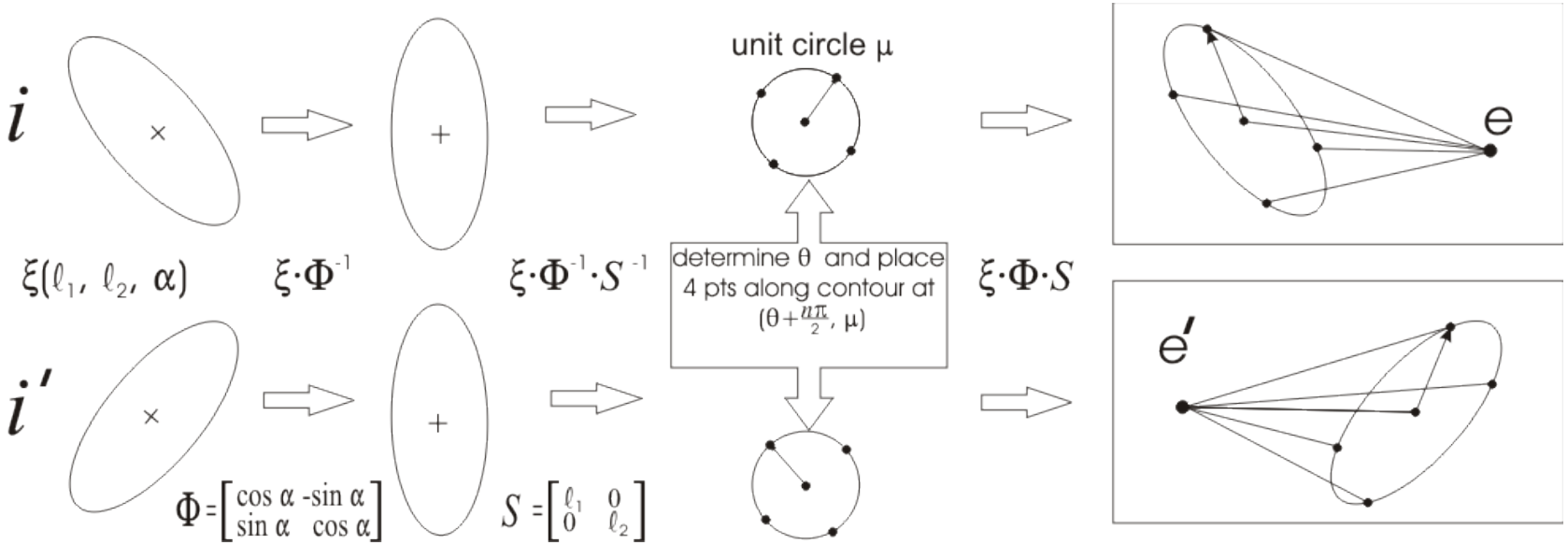
- TIP matches
  - Augment the number of point correspondences passed to FM estimation routines.
  - Improve robustness in the case of sparse correspondences.
  - Allow estimation from fewer than 8 invariant feature correspondences.
- How many additional TIP matches?
  - Determined by the number of parameters describing the invariant region K.

Transfer of Invariant Parameters (TIP)

Invariance	Parameters	K	Tip Matches	Invariant Corr. Req'd with TIP	Example TIP Matches
Translation	(x,y)	2	1	8 (*or* 7)	
Similarity (Scale)	(x,y,σ,θ)	4	2	4	
Affinity (Affine)	$(x,y,\theta,\alpha,\ell_1,\ell_2)$	6	3	3	

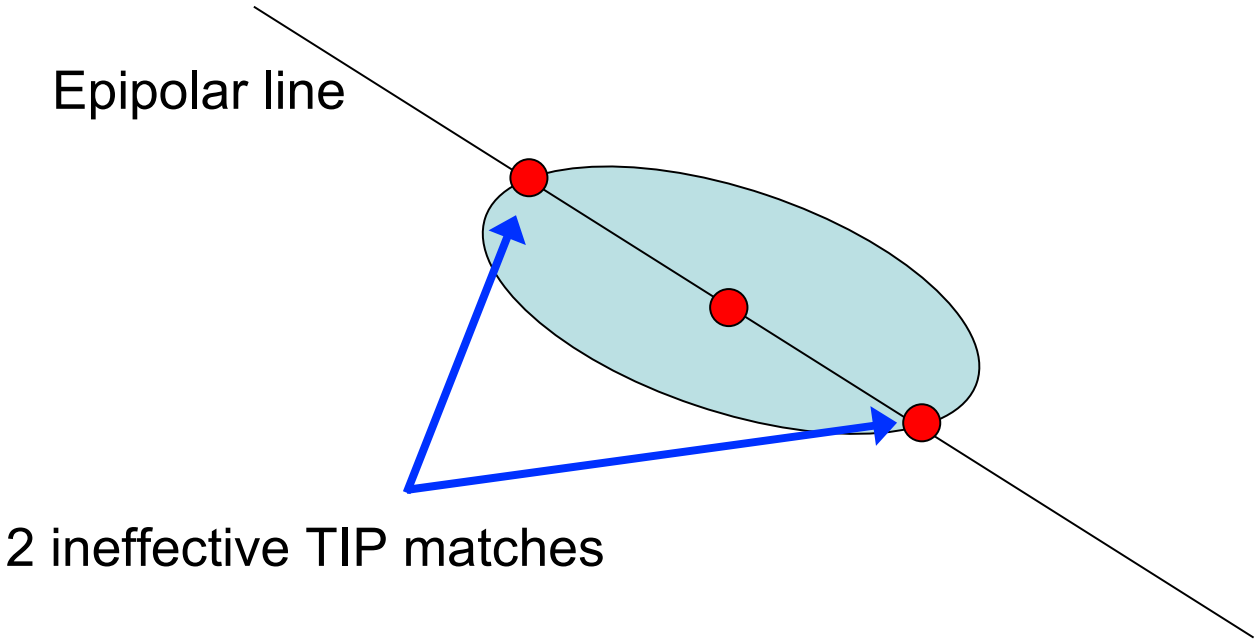
TIP with Affine Invariance

- Affine invariant correspondences
  - Elliptical invariant regions.
  - Good representation for locally planar surfaces, distant camera.
  - Add 5 TIP matches: 1 center + 4 along region boundaries.



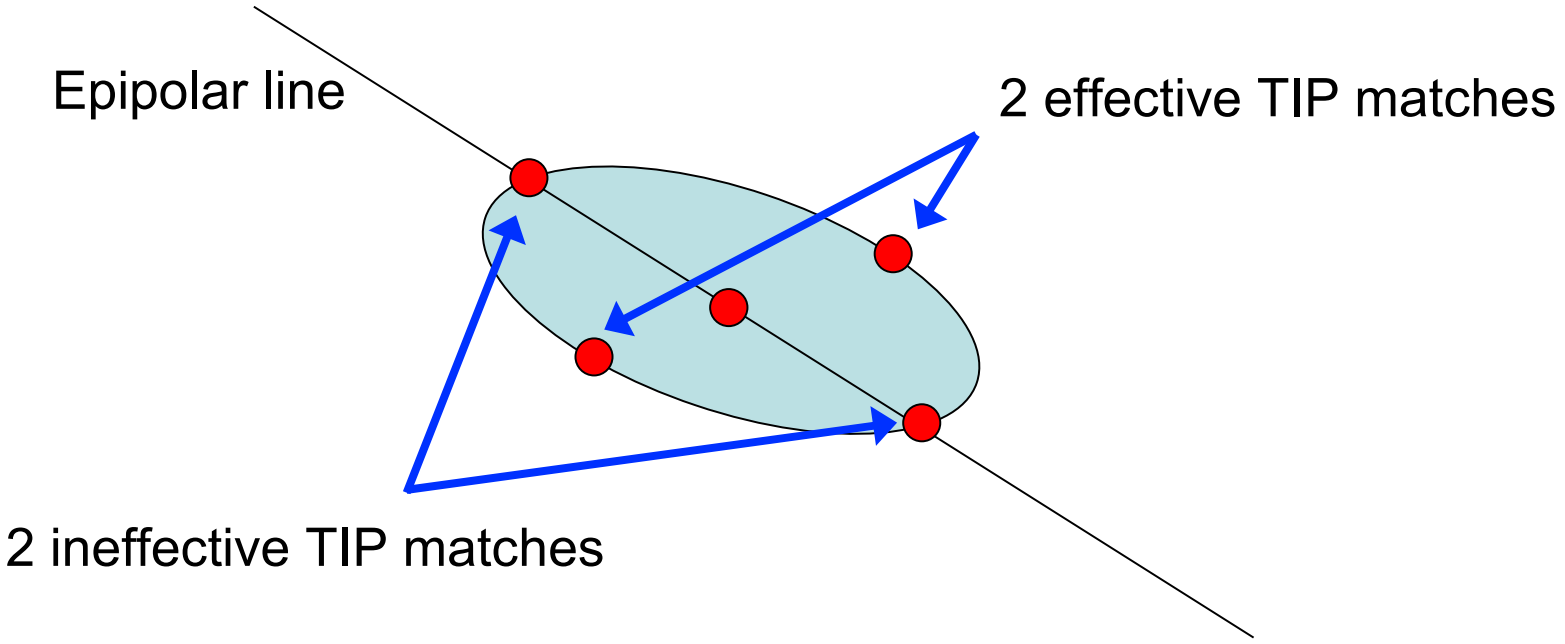
TIP with Affine Invariance

- Why 5 TIP matches instead of 3?
  - Redundancy helps to avoid ineffective TIP matches.
  - For example, TIP matches lying along the same epipolar line.



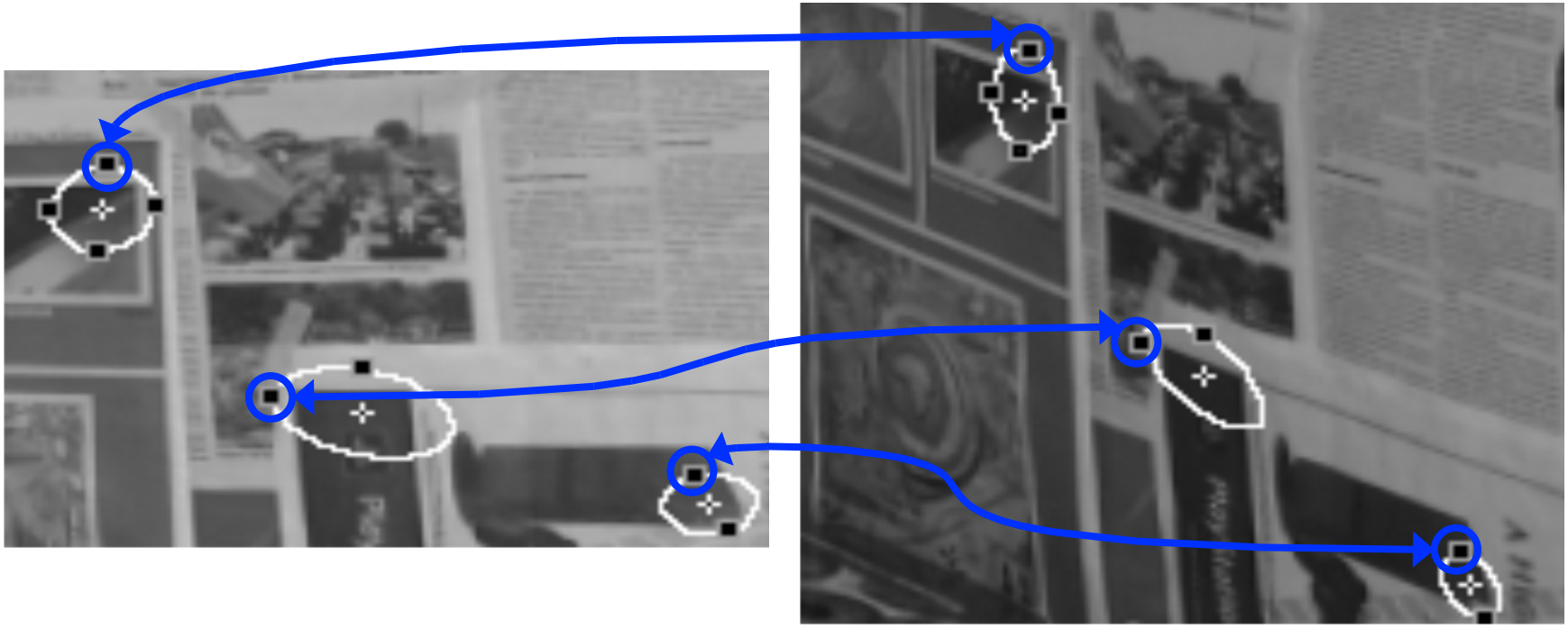
TIP with Affine Invariance

- Why 5 TIP matches instead of 3?
  - Redundancy helps to avoid ineffective TIP matches.
  - For example, TIP matches lying along the same epipolar line.



TIP with Affine Invariance

- Example: affine-invariant TIP matches



Experimentation

- Sample a set of *n* affine correspondences (inliers) between an image pair.
- For each correspondence set, estimate two FMs:

  F: estimated from feature centers for n = {20,...,7} using the 8-point algorithm, n = 7 using 7-point algorithm.

  F^{TIP}: estimated from TIP matches for n ={20,...,3} using the 8-point algorithm.

Note: F cannot be estimated for n ={6,...,3}

- Perform 100 random samples for each *n*.

Test Images

Error Measure

- Re-projection Error Distance (ε):
  - Pass manually-selected ground truth point correspondences through both F and F^{TIP}.
  - Calculate the re-projected distance (d) from the point ($\mathbf{x}_i$) to it's corresponding epipolar line:

$$\varepsilon^2 = \frac{1}{N}\sum_i d\left(\mathbf{x}'_i, \mathrm{F}\mathbf{x}_i\right)^2 + d\left(\mathbf{x}_i, \mathrm{F}^{\mathrm{T}}\mathbf{x}'_i\right)^2$$

- Catastrophic failure
  - FM estimates resulting in $\varepsilon > 25$

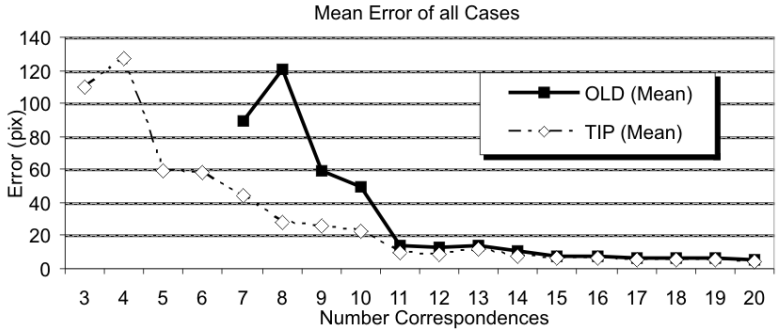
Results

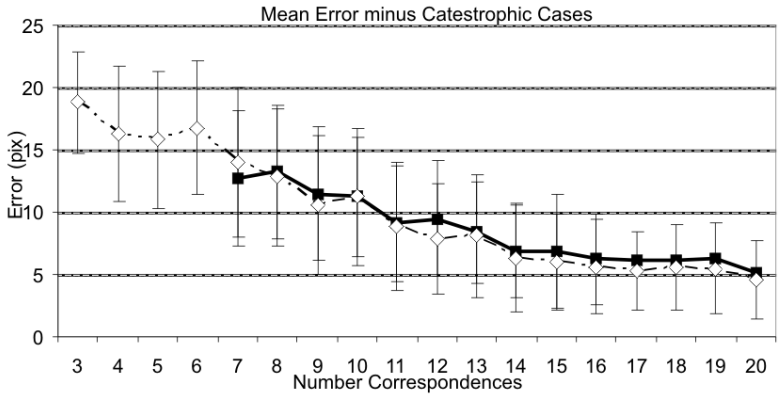
- TIP estimates result in fewer catastrophic cases.

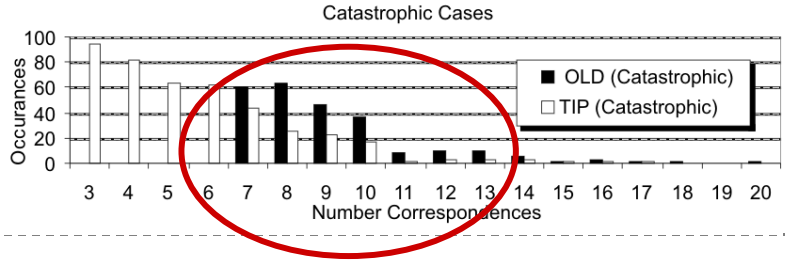
Results


MODERN ENGINEERING STUDIES
UNIX


MODERN ENGINEERING STUDIES
UNIX


Mean Error of all Cases
Error (pix)
Number Correspondences
OLD (Mean)
TIP (Mean)


Mean Error minus Catestrophic Cases
Error (pix)
Number Correspondences


Catastrophic Cases
Occurances
Number Correspondences
OLD (Catastrophic)
TIP (Catastrophic)

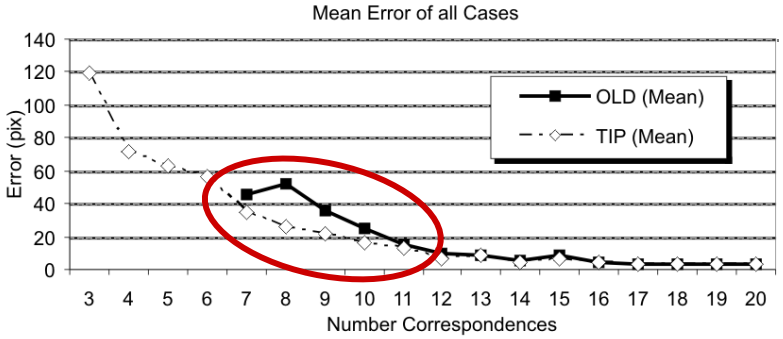
Results

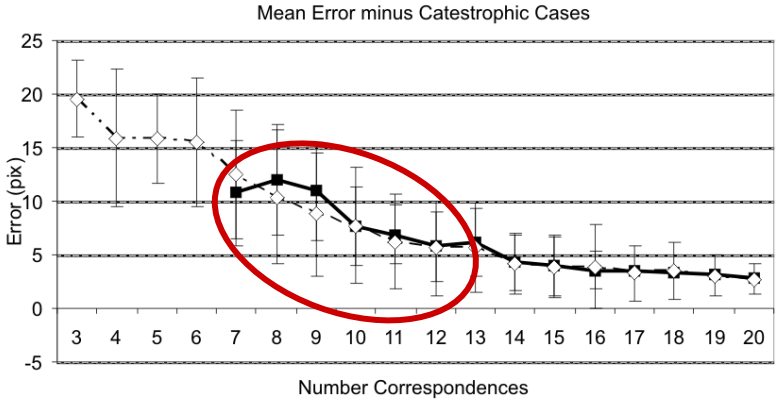
- TIP estimates result in fewer catastrophic cases.
- TIP estimates result in lower re-projection error.

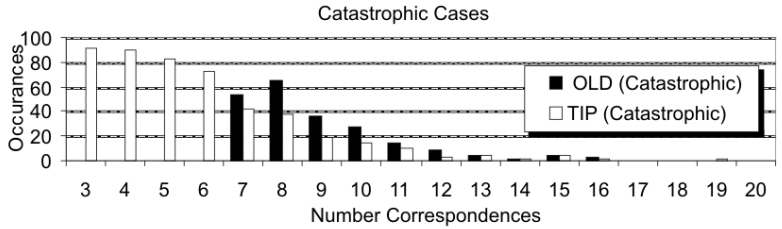
Results


MODERN
ENGINEERING
STATISTICS
UNIX


MODERN
ENGINEERING
STATISTICS
UNIX


Mean Error of all Cases
Error (pix)
Number Correspondences
OLD (Mean)
TIP (Mean)


Mean Error minus Catestrophic Cases
Error (pix)
Number Correspondences


Catastrophic Cases
Occurances
Number Correspondences
OLD (Catastrophic)
TIP (Catastrophic)

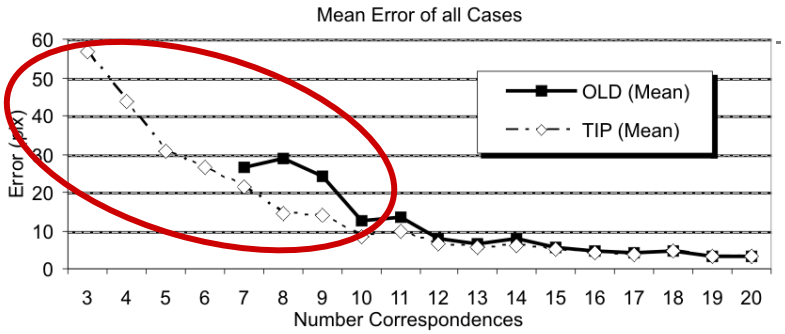
McGill
Centre for
Intelligent Machines

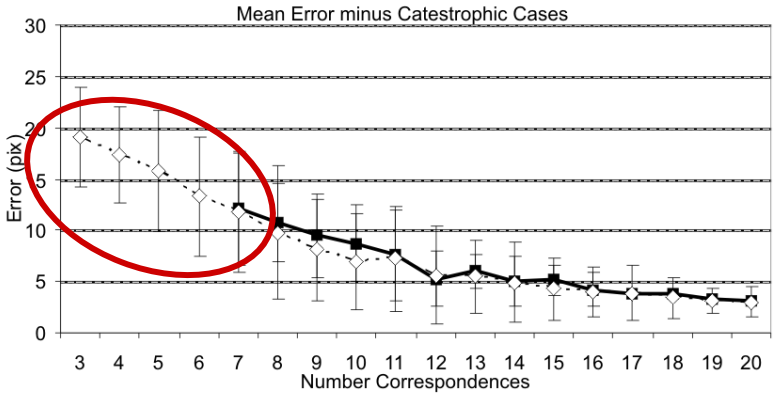
Results

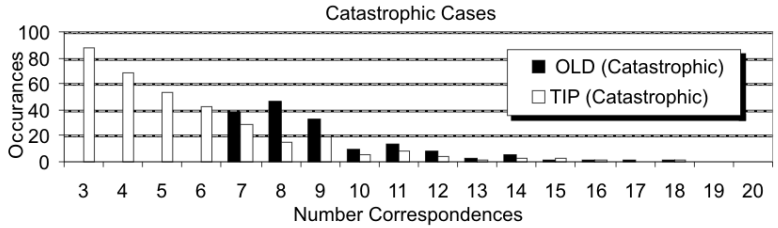
- TIP estimates result in fewer catastrophic cases.
- TIP estimates result in lower re-projection error.
- TIP allows for FM estimates with less than 7 invariant feature correspondences using standard point-based techniques.

Results


MODERN ENGINEERING STATISTICS
UNIX
MODERN ENGINEERING STATISTICS
UNIX


Mean Error of all Cases
Error (pix)
Number Correspondences
OLD (Mean)
TIP (Mean)


Mean Error minus Catestrophic Cases
Error (pix)
Number Correspondences

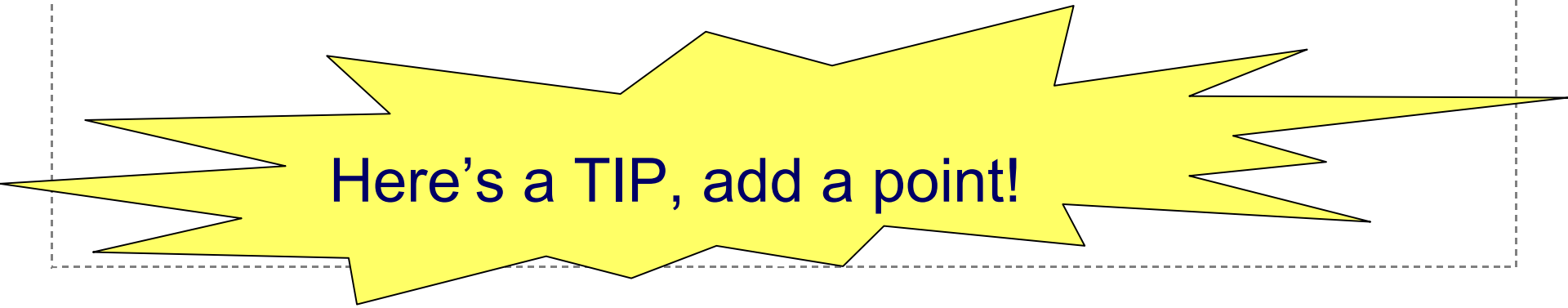

Catastrophic Cases
Occurances
Number Correspondences
OLD (Catastrophic)
TIP (Catastrophic)

Discussion

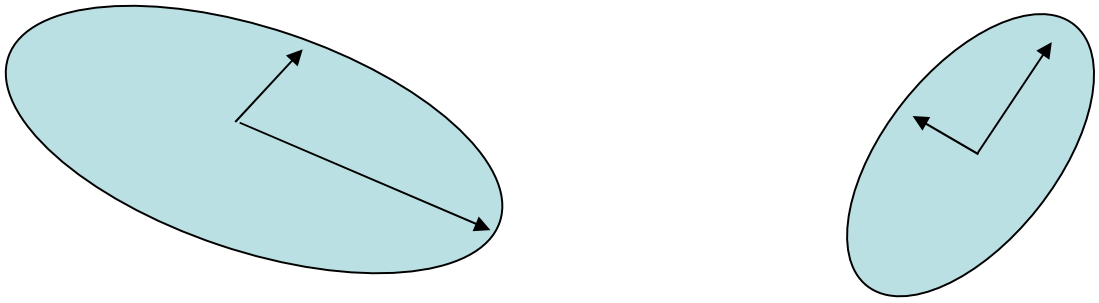
- The TIP method
  - Effective for correspondences arising from planar surfaces.
  - Less effective for complex 3D surfaces, difficult to represent accurately via invariant regions.
- Invariant feature geometrical information
  - Sufficient for matching but less effective for precise geometrical modeling.
  - Additional alignment of invariant regions after correspondence would improve the precision of additional TIP points.
  - Perspective alignment would allow estimation from as few as 2 invariant correspondences.

Summary

- The TIP method
  - Converts invariant correspondence parameters into point-to-point correspondences.
  - Incorporates local geometrical information into common point-based FM estimation routines.
  - Results in more robust FM estimation from fewer correspondences.

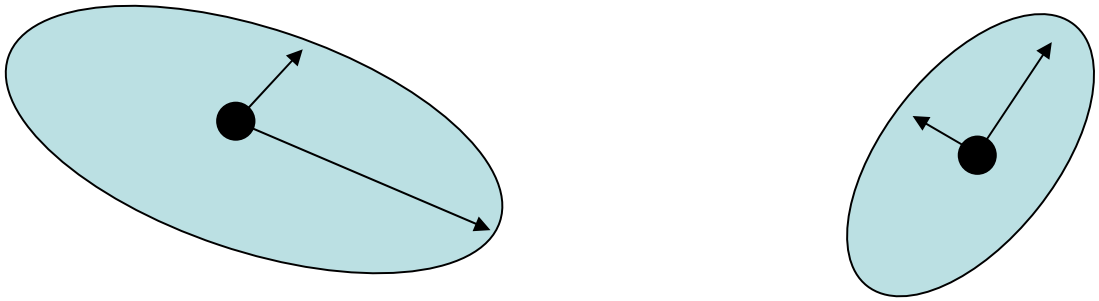


TIP Methodology



TIP Methodology

- Each invariant feature region is centered on an (x, y) point which we refer to as the *parent* point.



TIP Methodology

- Each invariant feature region is centered on an (x, y) point which we refer to as the *parent* point.
- Additional *child* points are canonically placed along the border of the invariant region of given scale, and rotation.