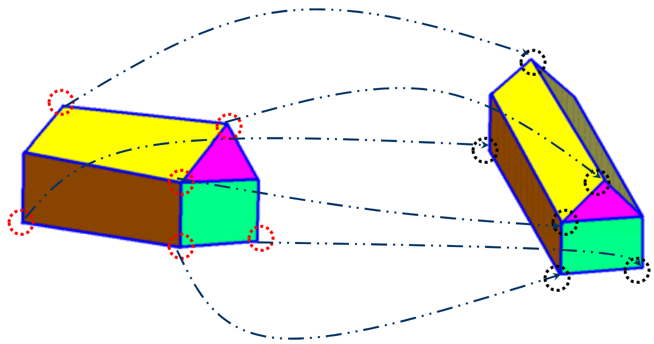
Wide Baseline Matching



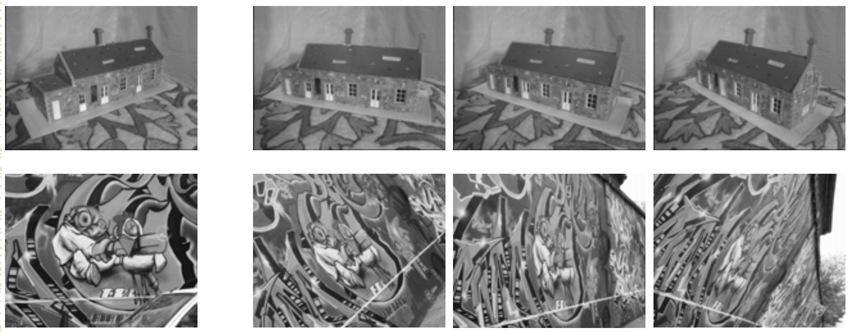
Jiangjian Xiao and Mubarak Shah, “Two-FrameWide Baseline Matching”, The Ninth ICCV , 2003.
http://www.cs.ucf.edu/~vision/papers/xiao_iccv03.pdf

Motivation

- Determine correspondences between two wide baseline images.
- Establish the epipolar geometry between these two images.
- Applications:
  - 3D reconstruction
  - image-based view morphing.

What Are Wide Baseline Images?

- A large translation, scaling, or rotation (20~70º) between two frames.
- Maximum pixel flow may be more than half of the size of the image.
- Some parts of the scene may be occluded in the images.



0° 50° 60° 70°

Main Steps

- Determine edge-corners.
- Two-stage matching by using affine matrix decomposition.
- Refine matches by epipolar geometry.

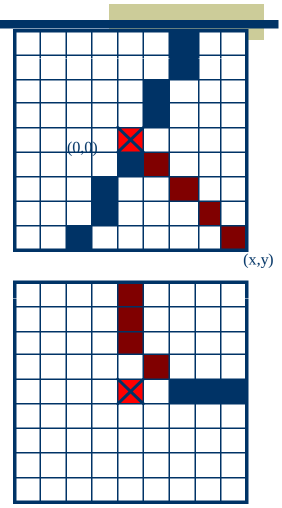
Edge-corners

- Corners are affine invariant.
- The corner should be located at the intersection of multiple edges.
- Combine three methods:
  - Canny edge detector.
  - Local Hough transform.
  - Harris corner operator (compute Eigen values of C matrix).

$$C = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

Local Hough Transform

- Fast, only considers a local area around potential corner.
- To avoid missed corners by canny edge detector:
  - Neighboring pixels around each edge pixel are also selected as corner candidates.

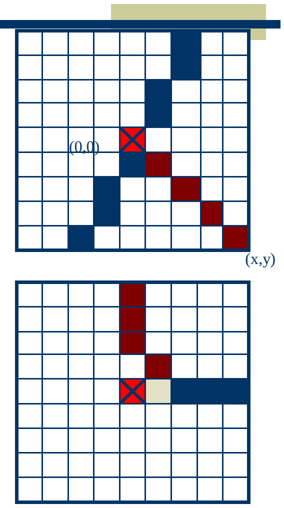


Local Hough Transform

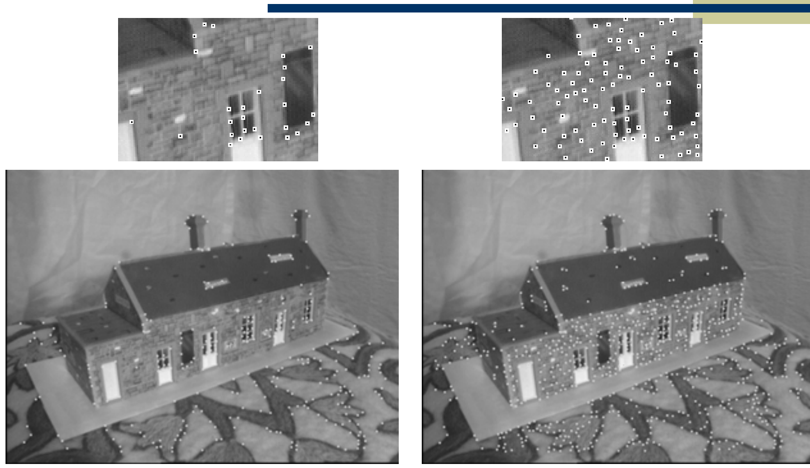
- Select a n×n window around the corner candidate, set origin to the center of the window.
- Quantize the parameter space θ in interval: [-π/2,π/2] using 2(n-1) bins for nxn window.
- Find lines going through the candidate corner point using

$$\theta = \tan^{-1}(y/x)$$

- If the number of edges passing the origin is more than 1, this point is a potential corner
- Use Harris corner operator to evaluate the goodness of this corner.
- Remove other weak neighboring corners using non-maxima suppression.



Initial Edge-corners(left)



(Our Results) (Harris Corners)

Affine Invariant Matching

- Simple Affine equation:

$$I_2(Ax + d) = I_1(x)$$

- Affine equation with illumination terms:

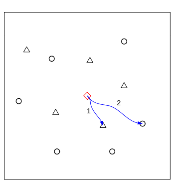
$$\mu I_2(Ax + d) + \delta = I_1(x)$$

- Image residual:

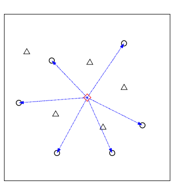
$$\varepsilon = \sum_W \left[(\mu I_2(Ax + d) + \delta) - I_1(x) \right]^2$$

This equation can be solved by using *the first order Taylor expansion* and *Newton-Raphson iteration* after assuming: A = I (identity matrix), $d = 0$, $\mu = 1$, and $\delta = 0$.

Matching Scheme



Traditional search
(Lucas-Kanade-Tomasi algorithm)

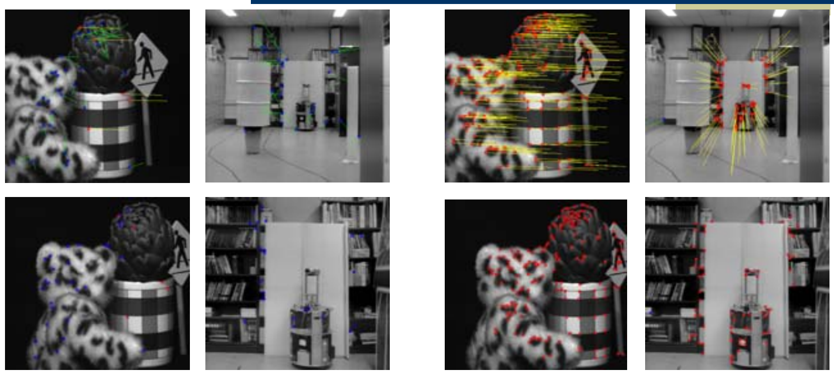


Our matching method

- For each potential corner, p_{1i} , in the first frame $I_{1.}$
  - Consider a search window in the second frame around $p_{1i.}$ For each potential corner, p_{2j} ,
    - Initialize d=0, set the origin of the corner's window at p_{1i} in I_1 and p_{2j} in I_2 respectively.
    - Perform Newton-Raphson iteration to minimize ε_{1i2j}.
  - The best match for corner p_{1i} is the p_{2j} with the minimal ε_{1i2j}.

Red diamond is a corner in first frame, triangles and circles are local minima in the second frame.
Circles also represent the corners in the second frame.

Comparison of Two Schemes

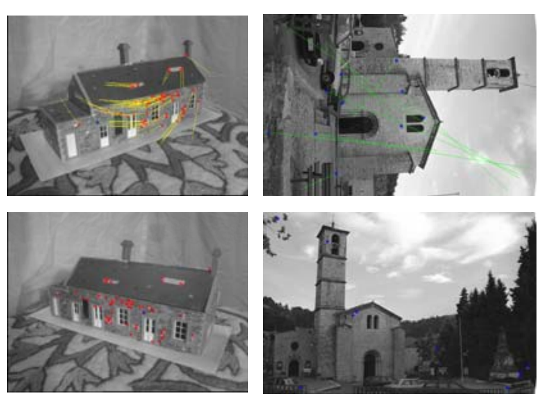


KLT algorithm | Our method

Conclusion: our method works very well if the major motion of camera is translation or looming (<2× scaling).

Limitations

- Problems when there are:
  - Large rotations.
  - Large scalings.
- Correct initial estimate is required.
  - $A = \mathrm{I}$, $d = 0$, $\mu = 1$, and $\delta = 0$ usually is ***not a correct initial state*** for wide baseline images.



59 correct inliers. | 0 correct inliers.

Affine Matrix Decomposition

$$A = UDV^T = U(V^TV)DV = (UV^T)(VDV^T)$$
$$= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} v_1 & v_h \\ v_h & v_2 \end{bmatrix}$$
$$= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \kappa & 0 \\ 0 & \kappa \end{bmatrix} \begin{bmatrix} 1/\kappa & 0 \\ 0 & 1/\kappa \end{bmatrix} \begin{bmatrix} v_1 & v_h \\ v_h & v_2 \end{bmatrix}$$
$$= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \kappa & 0 \\ 0 & \kappa \end{bmatrix} \begin{bmatrix} v_1/\kappa & v_h \\ v_h & v_2/\kappa \end{bmatrix}$$
$$= R(\alpha)S(\kappa)E$$

High weight component

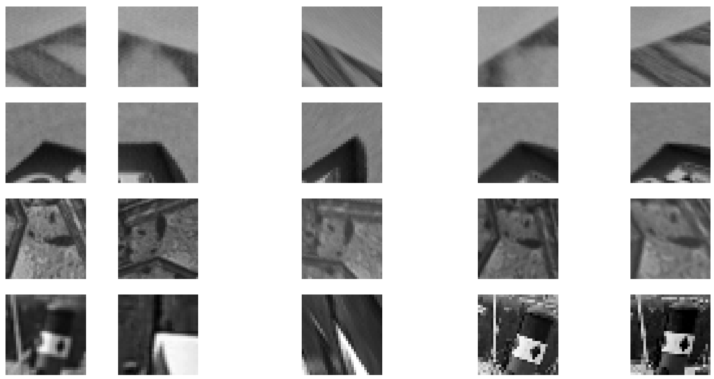
Low weight component

Two-stage Matching

- Stage one — determining a reasonable initial state.
  - For α from α_{low} to α_{high} (large step of 20° and small step of 4°) , for κ from 4, 2, 1, 0.5 and 0.25.
    - A rotation-scaling warping $R(\alpha)S(\kappa)$ is applied to a window around p_{2j} and SSD (residual) between this warped window and a window around p_{1i} is computed.
    - Select α and κ corresponding to minimum residual as the initial state.
- Stage two — minimizing residual between two corners from this initial state.
  - Initialize $A = R(\alpha)S(\kappa)$, which provides a reasonable gradient descent direction for the Newton-Raphson iteration.
  - Following this gradient descent direction, A is refined by Newton-Raphson iteration, and the residual is minimized.

Matching Procedure

The warped image patches of the second frames



Original windows around a corner in two images

Results with an initial state $A = I$, $d = 0$, $\mu = 1$, $\delta = 0$.

Stage one

Stage two

Benefits of Two-stage Matching

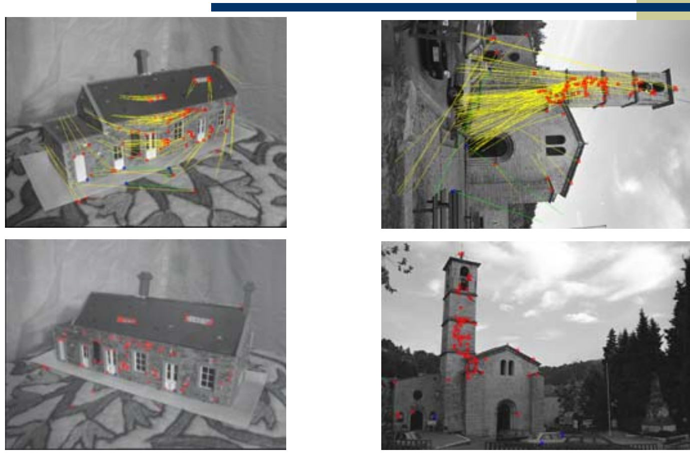
- Corner-to-corner matching avoids *exhaustive search* for every pixel in a large area.
- The first stage quantizes the affine space by using the rotation angle and scaling factor.
  - The computation of *non-linear components* $R(\alpha)$ *and* $S(\kappa)$ is avoided by computation of gradient I_x to I_y.
  - An approximate state of the geometric deformation due to the large rotation and scaling is recovered.
- A reasonable initial state provides a *correct gradient descent direction* to quickly determine an optimized solution of *A*.

Benefits of Two-stage Matching

- The search space is dramatically reduced from the exhaustive search's ***x×y×360/4×m×n*** to our method's ***w×(360/20+20/4)×5×n***
  - In exhaustive search case: *x*, *y* are the image size, 360/4 is the rotation space's size, *m* is the scaling space's size, *n* is the stretch-shearing space's size.
  - In our approach: *w* is the number of corners in the search area, (360/20+20/4) is the size of the rotation space by using *N*-split algorithm, 5 is the quantized scaling space's size.

  Note: w« (x×y) , 5«m.

Results After Two-stage Matching



83 correct inliers. 78 correct inliers.

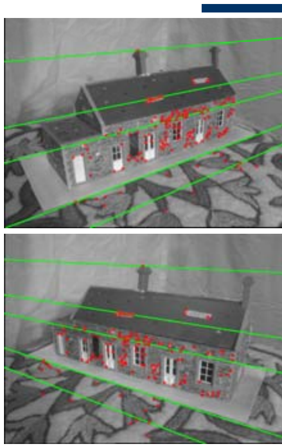
Refinement With Epipolar Geometry

- Compute distance for each pair of points p_{1i} and p_{2j}

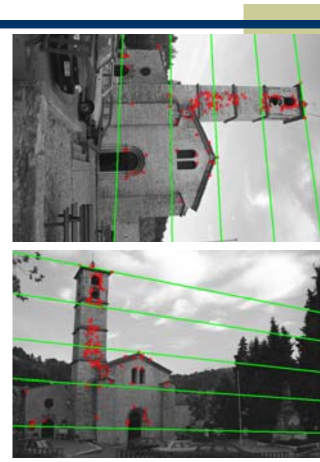
$$d_{1i2j} = \frac{(p_{2j}^T F p_{1i})^2}{\sqrt{\left(\frac{1}{(Fp_{1i})_1^2 + (Fp_{1i})_2^2}\right)}}$$

  - if distance $d_{1i2j} < \xi$ (2-4 pixels) then re-compute the residual ε_{1i2j} within a small band around the epipolar line.
- Apply common motion constraint to reduce false matches
  - Compute average motion flow ζ' in a small neighborhood of a point
  - If the direction and magnitude of point are very different then remove that point

Result After Refinement



150 correct inliers
RMS is 0.418.



102 correct inliers
RMS is 0.632.

Comparison With Other Methods

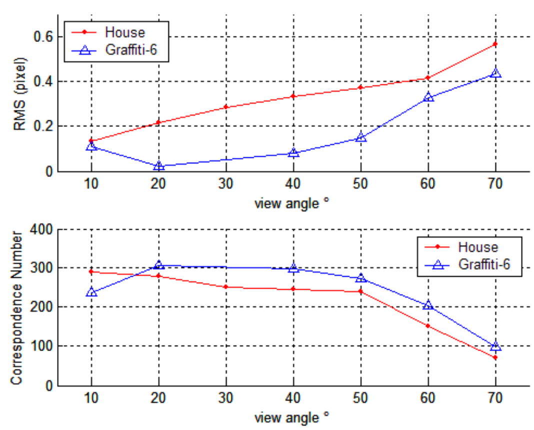
Name (Frame No)	A-H	T-S (RMS)	RF (RMS)
Graffiti-5 (4,8)	33	276 (0.110)	299 (0.127)
Graffiti-6 (1,5)	27	135 (0.341)	202 (0.327)
Boat (0,5)	22	163 (0.301)	175 (0.560)
Valbonne (5,14)	14	57 (0.665)	185 (0.658)
Valbonne (9,13)	22	155 (0.486)	194 (0.661)
UBC (8,10)	34	184 (0.266)	227 (0.242)

A-H are the results using Affine-Harris method.
T-S are the results using two-stage matching without refinement.
RF are the results after increasing correspondences using epipolar geometry.

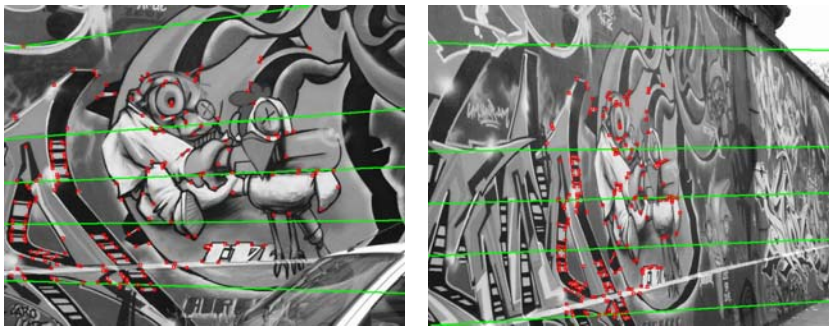
$$\varepsilon_{RMS} = \sqrt{\left(\frac{1}{2}\sum_{i=1}^{n}\frac{d(m_{2i},Fm_{1j})^2 + d(m_{1i},F^T m_{2j})^2}{2}\right)}$$

RMS Error For Different Viewing Angles

0°
50°
60°
70°

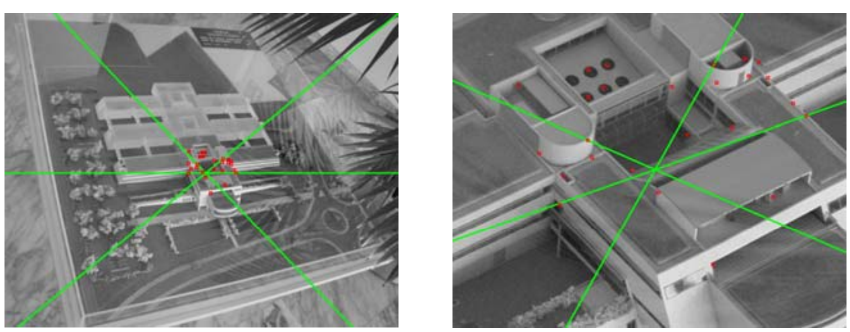


More Results



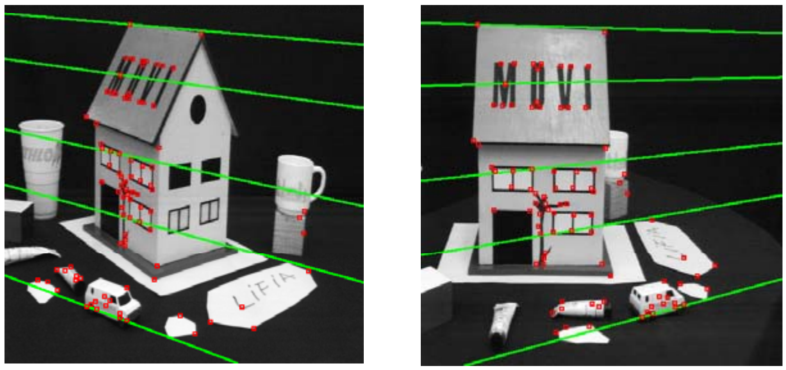
Rotation case (frame 1 and 5 of ``Graffiti-6'' from INRIA.)
202 inliers, the RMS is 0.327.

More Results



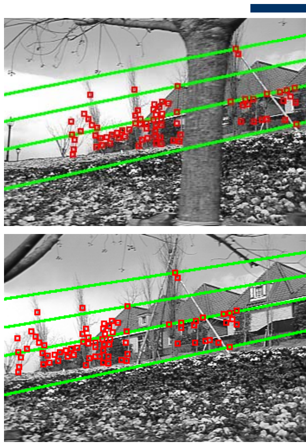
Zooming case (frame 0 and 10 of ``INRIA-Model'' from INRIA.)
25 inliers, the RMS is 0.583.

More Results

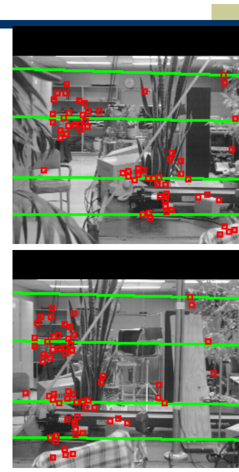


Rotation case (frame 5 and 9 of "Movie" from OSU)
89 inliers, the RMS is 0.471.

More Results

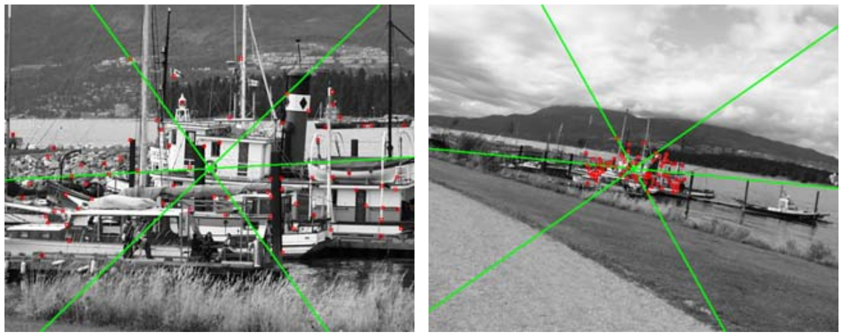


Translation case (frame 0 and 60 of "Follower garden" from OSU.)



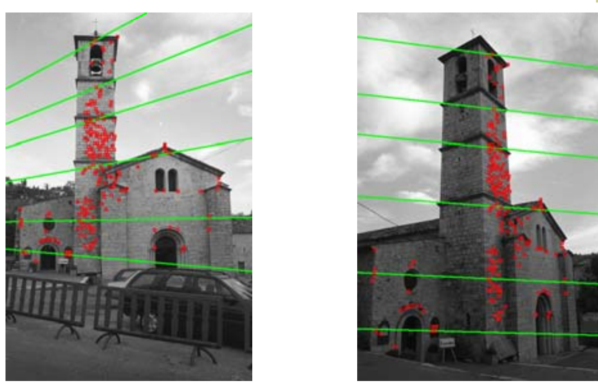
Translation case (frame 1 and 90 of "SRI" from CMU.)

More Results



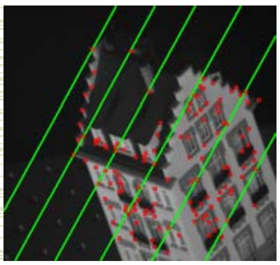
Zooming case (frame 8 and 0 of ``Boat" from INRIA.)
75 inliers, the RMS is 0.686.

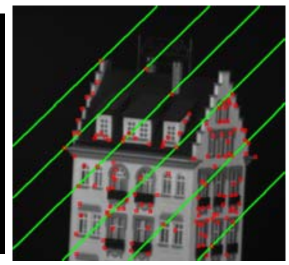
More Results



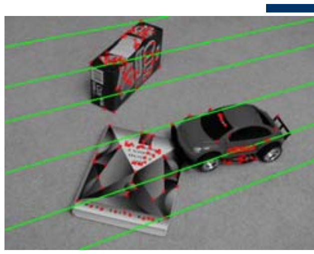
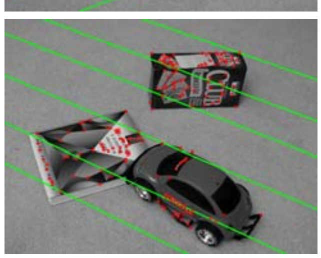
Rotation case (frame 5 and 14 of ``Valbonne" from Oxford Univ.)
185 inliers, the RMS is 0.658.

View Synthesis

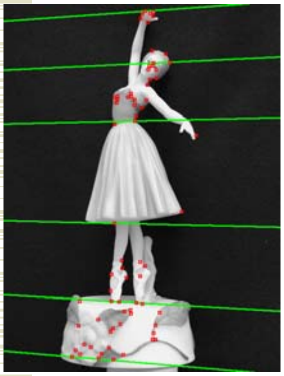


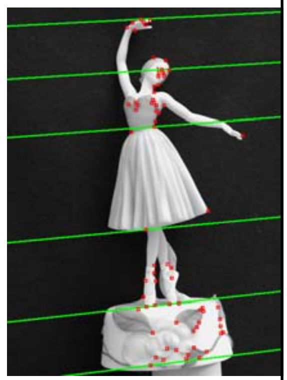


View Synthesis


View Synthesis





Summary

- Location of ***reliable*** edge-corners over wide baseline.
- Using two-stage matching by affine matrix decomposition to dramatically ***reduce*** search space.
- Only based on ***two gray level*** images, reliable appearance based corner matches are obtained by employing ***a large gap*** invariant affine transformation.