Visual Odometry for Ground Vehicle Applications

David Nister, Oleg Naroditsky, James Bergen

Presented by: Chaoran Yu, Chayut Thanapirom

Some slides derived from authors' presentation
(http://faculty.cs.tamu.edu/dzsong/teaching/spring2009/cpsc643/JiPresentation%204.ppt)

Abstract

- A system that estimates the motion of a stereo head or a single camera based on video input
- Real-time navigation for ground vehicles

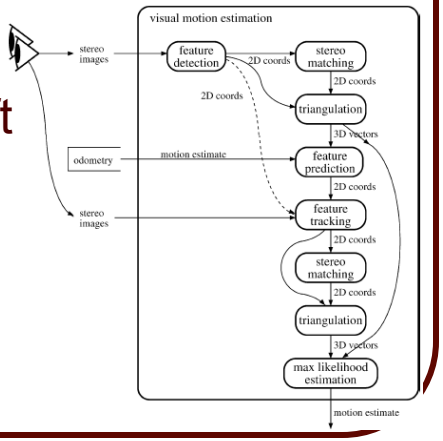
Related Work

- A Visual Odometry System – Olson 2003
- Previous Work of this Paper – Nister 2004

Related Work

- A Visual Odometry System – Olson 2003
- Previous Work of this Paper – Nister 2004

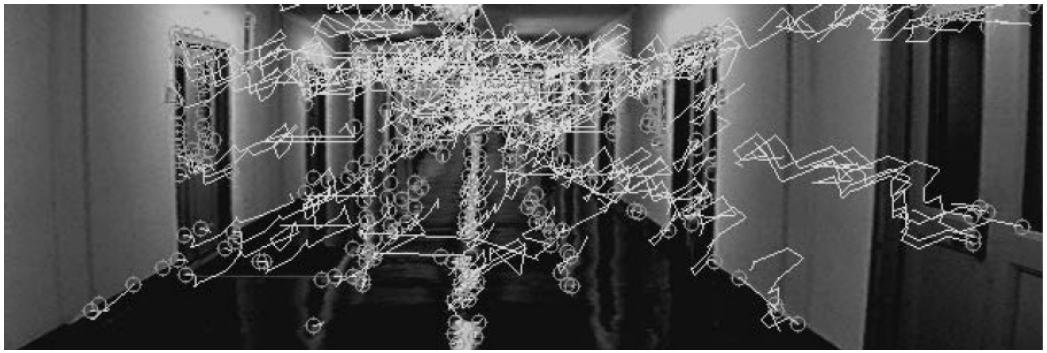
- With absolute orientation sensor
- Forstner interest operator in the left Image, matches from left to right
- Use approximate prior knowledge
- Iteratively select landmark points



Related Work

- A Visual Odometry System – Olson 2003
- Previous Work of this Paper – Nister 2004

- Estimates ego-motion using a hand-held Camera
- Real-time algorithm based on RANSAC



Motivation

Olson:	Nister:
With absolute orientation sensor	Use pure visual information
Forstner interest operator	Use Harris corner detection in all images, track feature to feature
Use approximate prior knowledge	No prior knowledge
Iteratively select landmark points	RANSAC based estimation in real-time

Feature Detection

Harris Corner Detection

Search for the local maxima of the corner strength $s(x, y)$.
d determinant, t trance, k constant, a, b window area,
I_x, I_y derivatives of input image, w weight function.

$$\max \quad s(x,y) = d(G_\alpha(x,y)) - kt(G_\alpha(x,y))^2$$

$$G_\alpha(x,y) = 2\sum_a \sum_b w(a,b)$$

$$\times \begin{bmatrix} I_x^2(x+a,y+b) & I_x I_y(x+a,y+b) \\ I_x I_y(x+a,y+b) & I_y^2(x+a,y+b) \end{bmatrix}$$

Feature Detection

- Non-max Suppression

  A feature point is declared at each pixel where the response is stronger than all other pixels in a 5*5 neighborhood.

- Local Saturation

  Limit the number of features in a local region of the image to bound processing time.

Feature Detection

Detected Feature Points

Superimposed feature tracks through images

Feature Matching

- Disparity Limit

1) A feature in one image is matched to every feature within a fixed distance from it in the next image.

2) DL chosen based on speed requirements and smoothness of the input

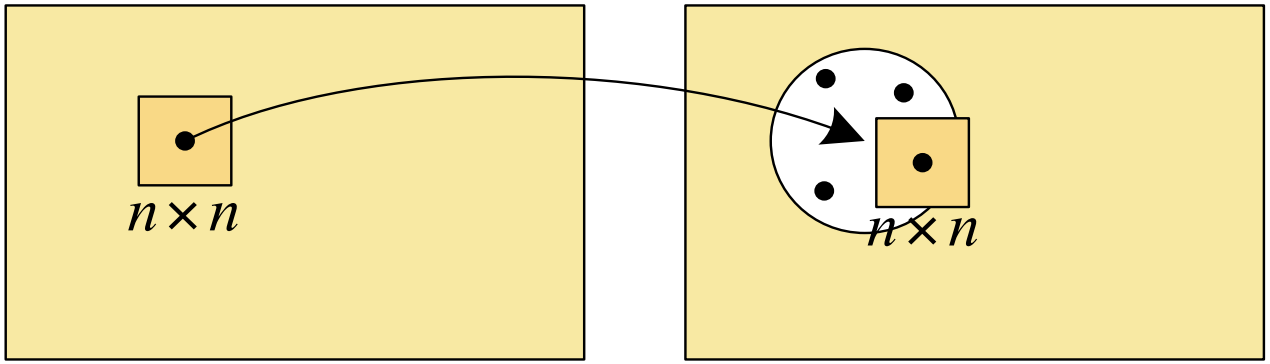
Feature Matching

Two Directional Matching (Mutual Consistency Check)

- Calculate the normalized correlation in $n \times n$ boxes centered around each detected feature, where ,I_1, I_2 are two input image patches.

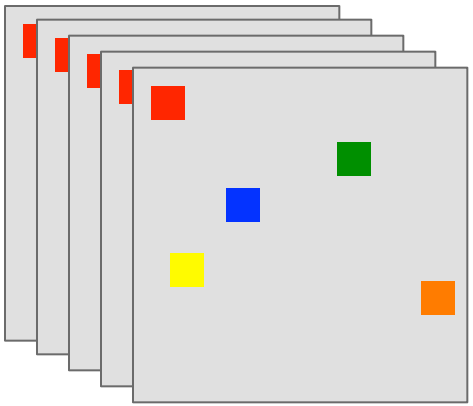
$$\max \frac{n^2 \sum I_1 I_2 - \sum I_1 \sum I_2}{\sqrt{n \sum I_1^2 - (\sum I_1)^2} \sqrt{n \sum I_2^2 - (\sum I_2)^2}}$$

- Match the feature points in the circular area that have the maximum correlation in two directions.



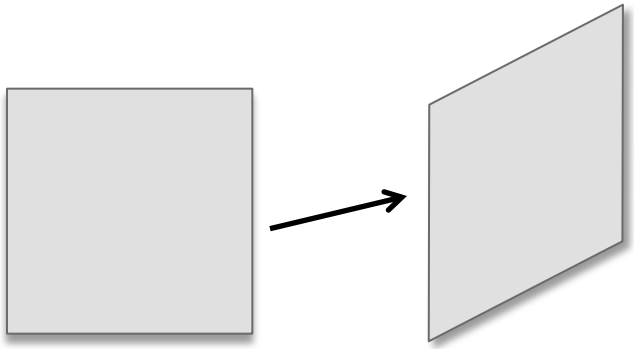
Pose Estimation Problem

Input:



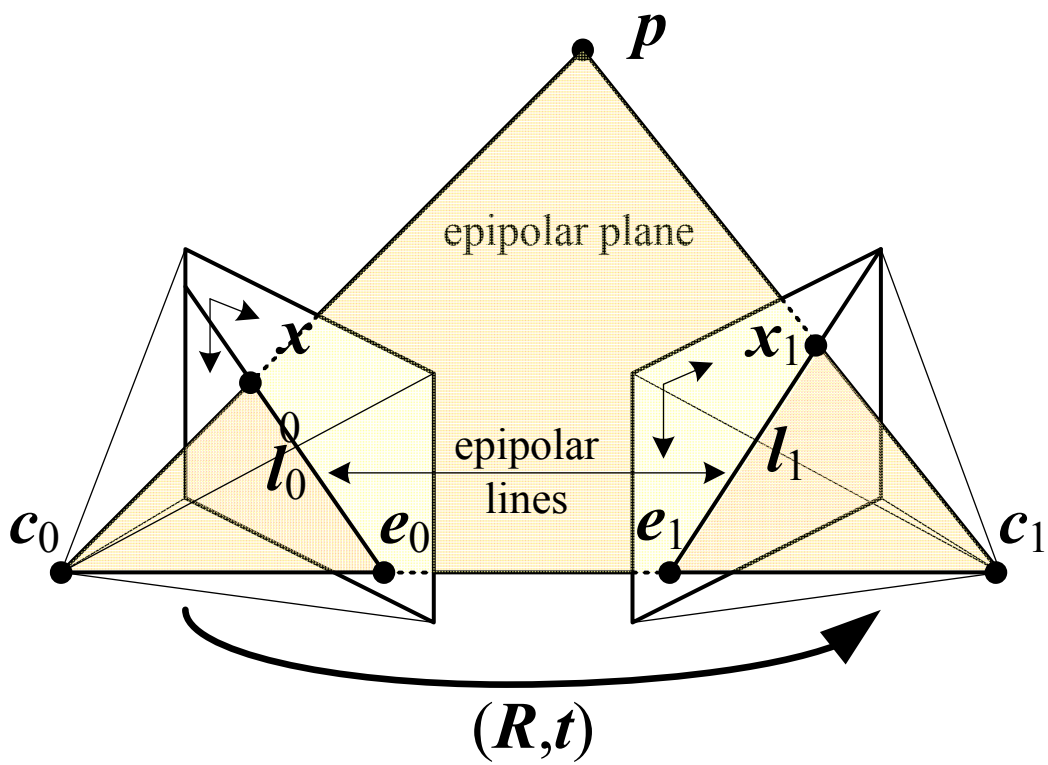
Frame(t_0,t_1,t_2,…)

Output:



Rotation
Translation

Epipolar Geometry



Epipolar constraint equation: $x_1^T E x_0 = 0$

x_i is a vector in projective space representing a 2D point in camera *i*

Rotation and translation information *(R,t)* can be extracted from the matrix *E*

Naïve vs Robust Method

Naïve method – least square error

$$x_1^T E x_0 = 0$$

$$\text{Find } E \text{ that minimize} \left[\sum_{\text{all points}} \left(x_1^T E x_0 \right)^2 \right]$$

Very bad if we have a lot of outliers

Robust method – RANSAC

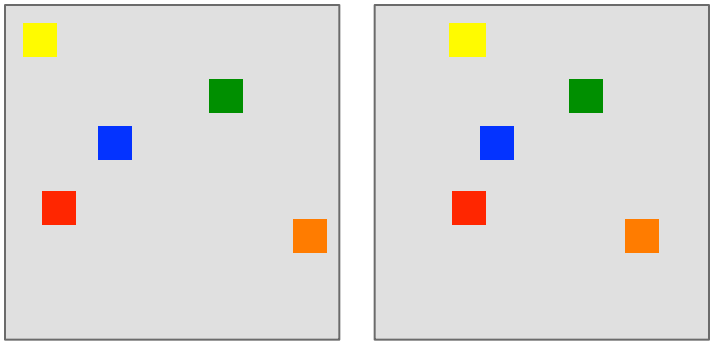
(RANdom Sample Consensus)

Pick few random points to generate pose hypothesis

Evaluate and pick the best one

Monocular Scheme: Step 1

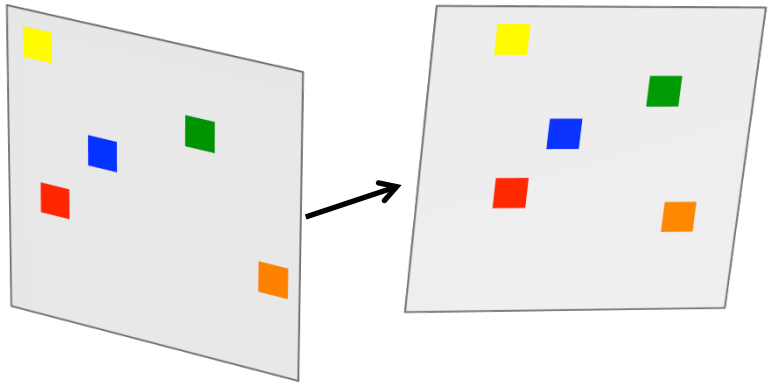
Input:



Frame(t_0) Frame(t_1)

Use **5-point algorithm** to solve epipolar equation and generate pose hypothesis in RANSAC method

Output:



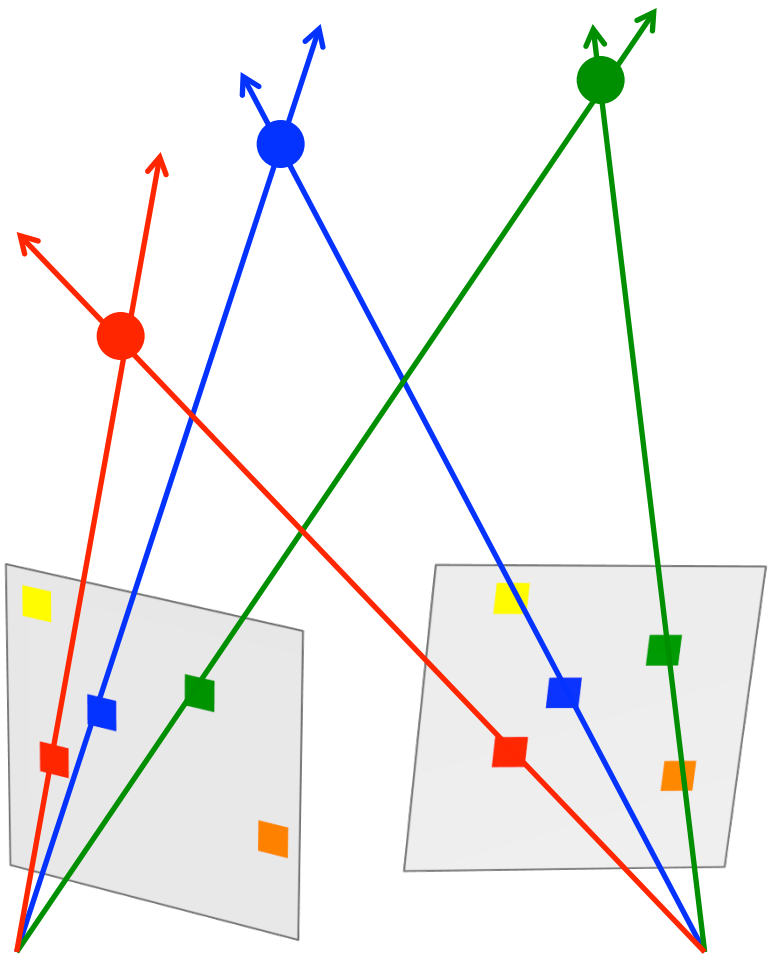
(R,t)

Monocular Scheme: Step 2

Output:

Triangulate to obtain
3D points

Input:

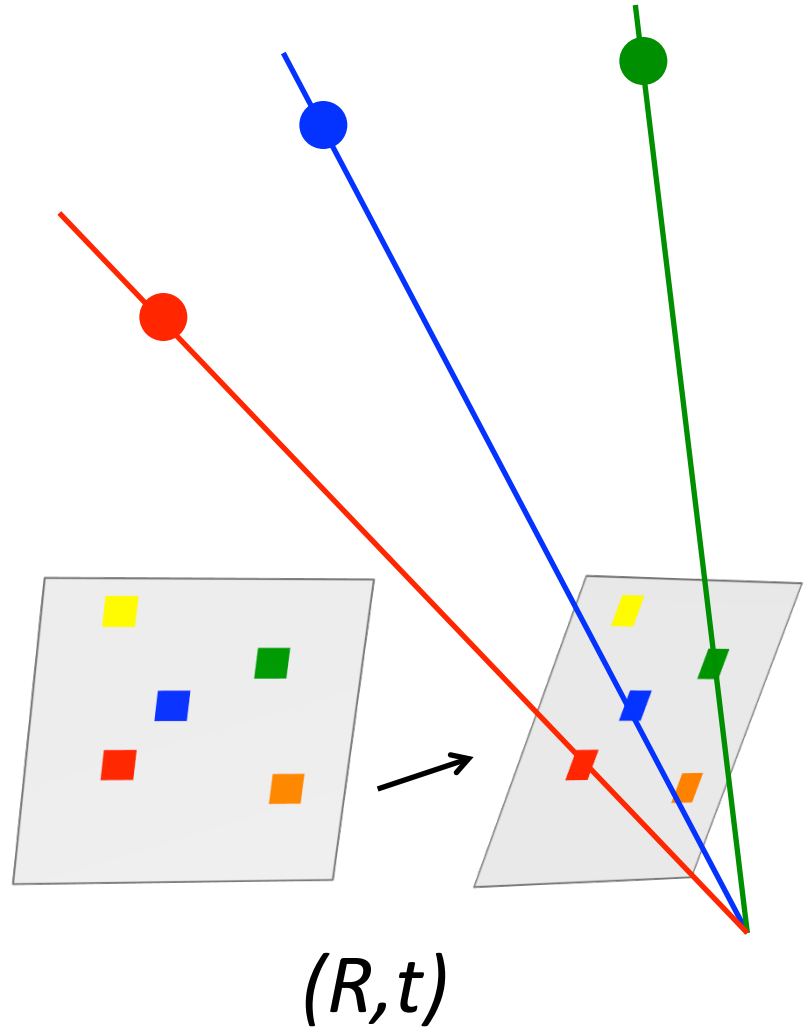


Monocular Scheme: Step 3

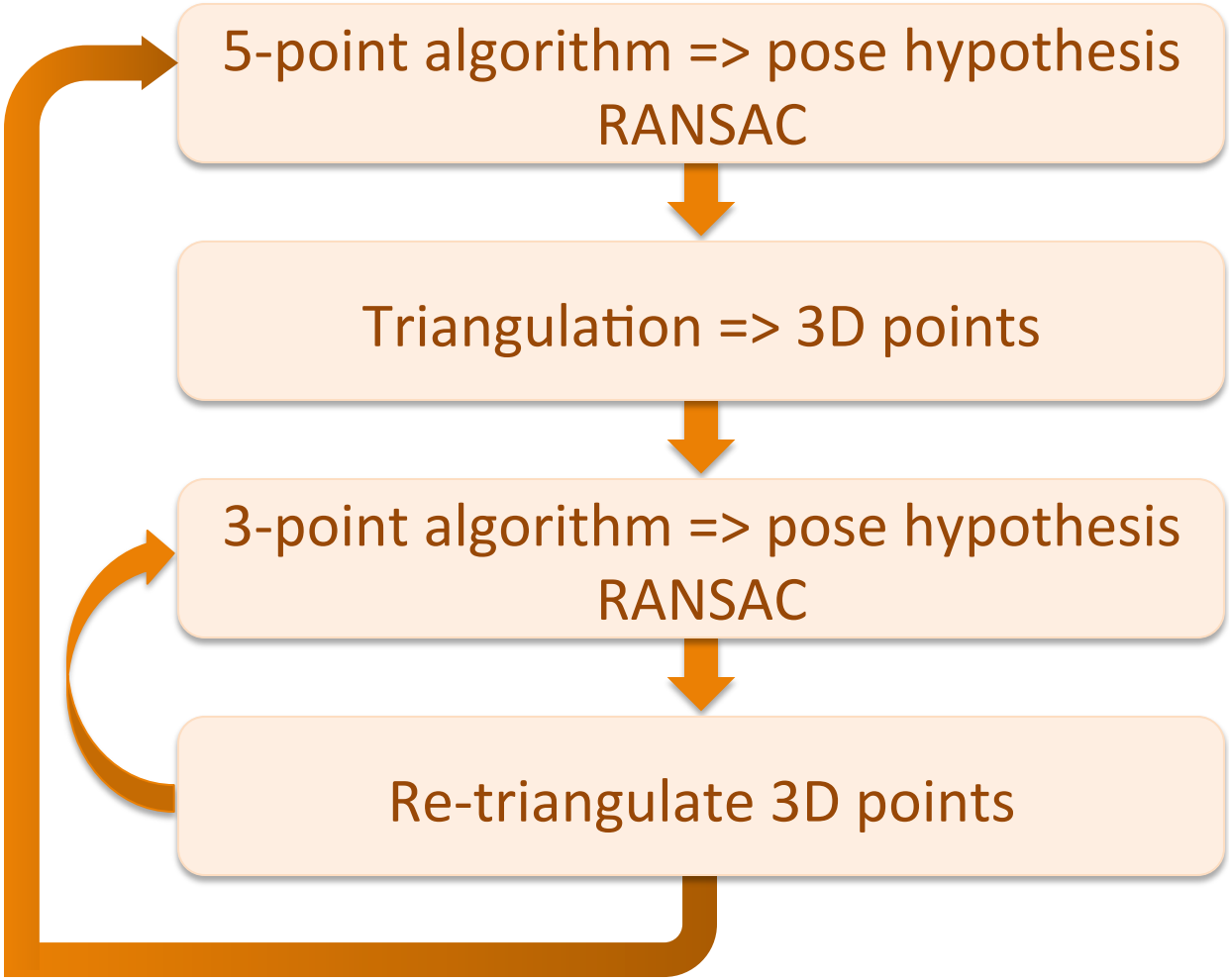
Use **3-point algorithm** to generate pose hypothesis in RANSAC method

Input: 3D points, 2D points on camera
Output: pose *(R,t)*

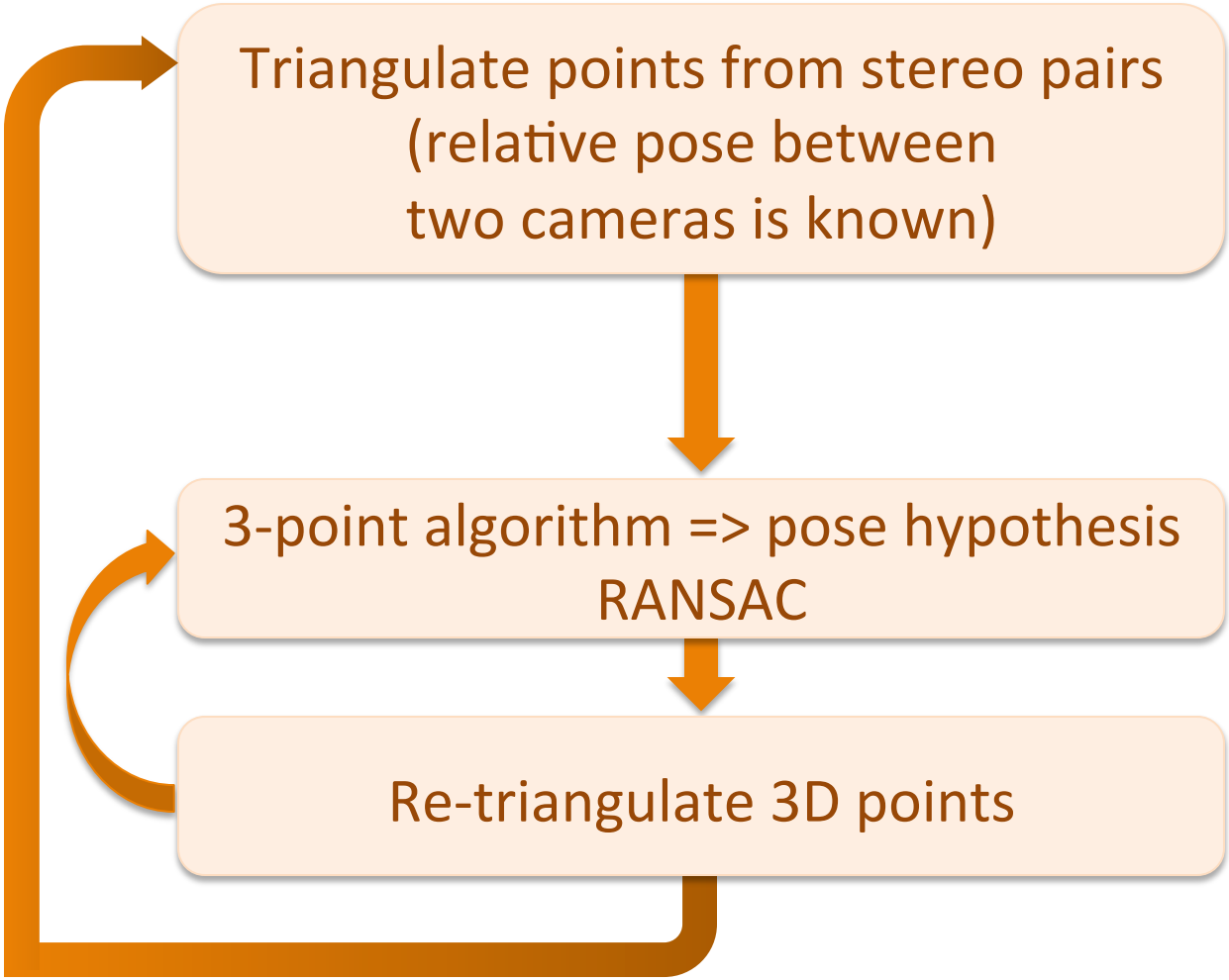
Repeat Step 3



Monocular Scheme



Stereo Scheme

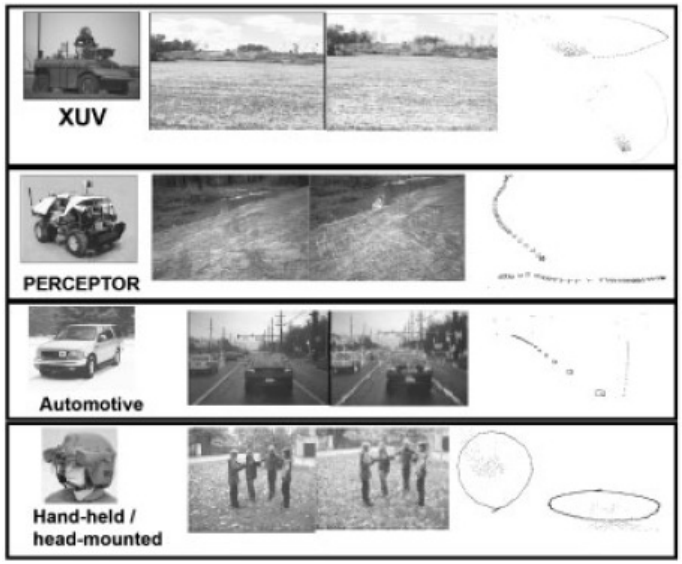


Stereo vs. Monocular

- More information
- No scale ambiguity
- More stable when motion is small

Experiments

Different Platforms



Experiments

- Evaluate performance of visual odometry system

  Ground truth: Integrated differential GPS (DGPS) and high-precision inertial navigation system (INS) – VNS.
- Align coordinate systems of visual odometry and VNS by a least square fit of initial 20 poses.

Experiments

Speed and Accuracy

Table I. Approximate average timings per 720 × 240 frame of video for the monocular system components on a modest 550 MHz machine. Disparity range for the matching is given in percent of the image dimensions. The average timings for the stereo version are very similar, the reason being that both systems are most of the SaM processing time performing RANSAC estimations of the pose with respect to known 3D points.

	Matching with disparity range			
Feature detection	3%	5%	10%	SaM
30 ms	34 ms	45 ms	160 ms	50 ms

Table II. Metric accuracy of visual odometry position estimates. The number of frames processed is given in Column 2. Total vehicle path lengths estimated by DGPS and visual odometry are given in Columns 3 and 4 with relative error in distance given in Column 4.

Run	Frames	DGPS (m)	VisOdo (m)	% error
Loops	1602	185.88	183.90	1.07
Meadow	2263	266.16	269.77	1.36
Woods	2944	365.96	372.02	1.63

Experiments

Visual Odometry vs. Differential GPS

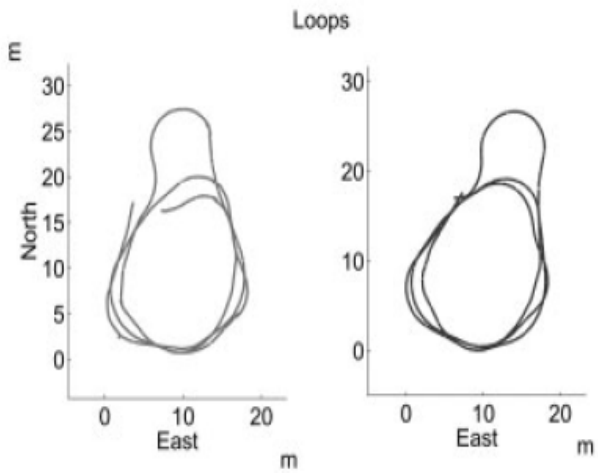


Figure 6. Vehicle positions estimated with visual odometry (left) and DGPS (right). These plots show that the vehicle path is accurately recovered by visual odometry during tight cornering as well as extended operation. In this example, the vehicle completes three tight laps of diameter about 20 m (traveling 184 m total) and returns to the same location. The error in distance between the endpoints of the trip is only 4.1 m.

Figure 7. Visual odometry vehicle position (light line) superimposed on DGPS output (dark line). No a priori knowledge of the motion was used to produce the visual odometry. A completely general 3D trajectory was estimated in all our experiments. In particular, we did not explicitly force the trajectory to stay upright or within a certain height of the ground plane.

Experiments

Visual Odometry vs. Inertial Navigation System (INS)

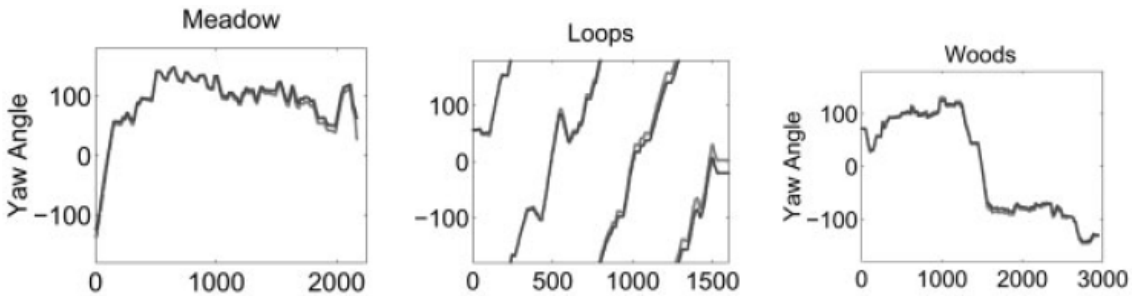


Figure 8. Yaw angle in degrees from INS and visual odometry. The correspondence is readily apparent. In most cases, visual odometry yields subdegree accuracy in vehicle heading recovery. The accumulated yaw angle is shown with respect to frame number.

Experiments

Visual Odometry vs. Wheel Recorder

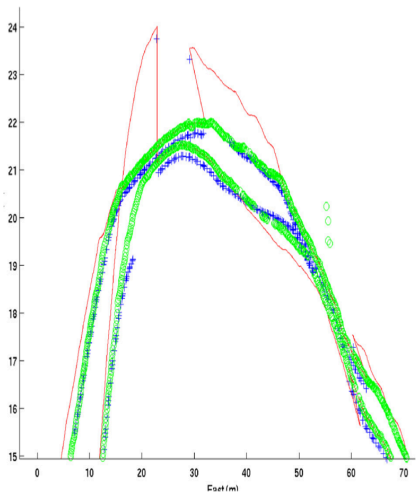


Figure 14: An example of the effect of wheel slip without visual odometry or GPS. DGPS - Dark Blue plus signs. Wheel encoders fused with IMU - Thin Red. Visual odometry - Thick Green. Note the incorrect overshoots from the wheel encoders. The motion of the vehicle was left to right in the bottom arc and right to left in the top arc.

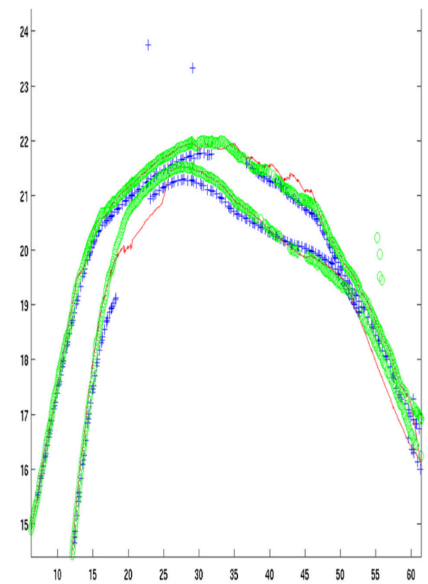


Figure 15: Results corrected by adding the visual odometry. DGPS - Dark Blue plus signs. Wheel encoders fused with visual odometry - Thin Red. Visual odometry - Thick Green.

Comparison with Existing Systems

- GPS/DGPS

  May have better accuracy but GPS signals are not always available.
- Wheel encoder

  Suffer from wheel slip.
- Visual Odometry + IMU

  Smoother path than GPS and more accurate than wheel encoders.

Conclusion

- **A real-time ego motion estimation system.**
- **Work both on monocular camera and stereo head.**
- **Results are accurate and robust.**