3D Vision

CSc I6716
Fall 2010

Topic 4 of Part II
Visual Motion

Zhigang Zhu, City College of New York zhu@cs.ccny.cuny.edu

Cover Image/video credits: Rick Szeliski, MSR

Outline of Motion

- Problems and Applications
  - The importance of visual motion
  - Problem Statement
- The Motion Field of Rigid Motion
  - Basics – Notations and Equations
  - Three Important Special Cases: Translation, Rotation and Moving Plane
  - Motion Parallax
- Optical Flow
  - Optical flow equation and the aperture problem
  - Estimating optical flow
  - 3D motion & structure from optical flow
- Feature-based Approach
  - Two-frame algorithm
  - Multi-frame algorithm
  - Structure from motion – Factorization method
- Advanced Topics
  - Spatio-Temporal Image and Epipolar Plane Image
  - Video Mosaicing and Panorama Generation
  - Motion-based Segmentation and Layered Representation

The Importance of Visual Motion

- Structure from Motion
  - Apparent motion is a strong visual clue for 3D reconstruction
    - More than a multi-camera stereo system
- Recognition by motion (only)
  - Biological visual systems use visual motion to infer properties of 3D world with little a priori knowledge of it
    - Blurred image sequence
- Visual Motion = Video ! **[Go to CVPR 2004-2010 Sites for Workshops]**
  - Video Coding and Compression: MPEG 1, 2, 4, 7…
  - Video Mosaicing and Layered Representation for IBR
  - Surveillance (Human Tracking and Traffic Monitoring)
  - HCI using Human Gesture (video camera)
  - Image-based Rendering
  - …

3D Computer Vision
and Video Computing

Blurred Sequence

Recognition by Actions: Recognize object from motion even if we cannot distinguish it in any images …

An up-sampling from images of resolution 15x20 pixels

From: James W. Davis. MIT Media Lab

3D Computer Vision
and Video Computing

Problem Statement

- Two Subproblems
  - Correspondence: Which elements of a frame correspond to which elements in the next frame?
  - Reconstruction :Given a number of correspondences, and possibly the knowledge of the camera's intrinsic parameters, how to recovery the 3-D motion and structure of the observed world
- Main Difference between Motion and Stereo
  - Correspondence: the disparities between consecutive frames are much smaller due to dense temporal sampling
  - Reconstruction: the visual motion could be caused by multiple motions (instead of a single 3D rigid transformation)
- The Third Subproblem, and Fourth….
  - Motion Segmentation: what are the regions the the image plane corresponding to different moving objects?
  - Motion Understanding: lip reading, gesture, expression, event…

3D Computer Vision
and Video Computing

Approaches

- Two Subproblems
  - Correspondence:
    - Differential Methods - >dense measure (optical flow)
    - Matching Methods -> sparse measure
  - Reconstruction : More difficult than stereo since
    - Motion (3D transformation betw. Frames) as well as structure needs to be recovered
    - Small baseline causes large errors
- The Third Subproblem
  - Motion Segmentation: Chicken and Egg problem
    - Which should be solved first? Matching or Segmentation
      - Segmentation for matching elements
      - Matching for Segmentation

3D Computer Vision and Video Computing

The Motion Field of Rigid Objects

- Motion:
  - **3D Motion (R, T):**
    - camera motion (static scene)
    - or single object motion
    - Only one rigid, relative motion between the camera and the scene (object)
  - **Image motion field:**
    - 2D vector field of velocities of the image points induced by the relative motion.
- Data: Image sequence
  - **Many frames**
    - captured at time t=0, 1, 2, …
  - **Basics: only consider two consecutive frames**
    - We consider a reference frame and its consecutive frame
  - **Image motion field**
    - can be viewed disparity map of the two frames captured at two consecutive camera locations (assuming we have a moving camera)

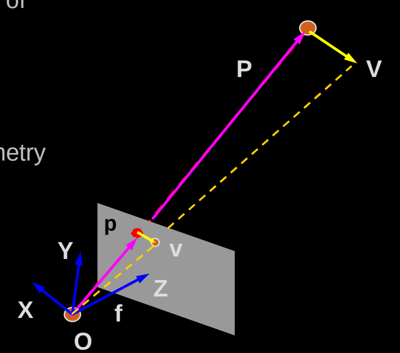
3D Computer Vision and Video Computing

The Motion Field of Rigid Objects

- Notations
  - $P = (X,Y,Z)^T$: 3-D point in the camera reference frame
  - $p = (x,y,f)^T$: the projection of the scene point in the pinhole camera
- Relative motion between P and the camera
  - $T = (T_x,T_y,T_z)^T$: translation component of the motion
  - $\omega=(\omega_x, \omega_y,\omega_z)^T$: the angular velocity
- Note:
  - How to connect this with stereo geometry (with R, T)?
  - Image velocity v= ?

$$\mathbf{p} = \frac{f}{Z}\mathbf{P}$$

$$\mathbf{V} = -\mathbf{T} - \boldsymbol{\omega} \times \mathbf{P}$$



3D Computer Vision
and Video Computing

The Motion Field of Rigid Objects

- Notations
  - $\mathbf{P} = (X,Y,Z)^T$: 3-D point in the camera reference frame
  - $\mathbf{p} = (x,y,f)^T$: the projection of the scene point in the pinhole camera
- Relative motion between P and the camera
  - $\mathbf{T} = (T_x,T_y,T_z)^T$: translation component of the motion
  - $\omega=(\omega_x, \omega_y,\omega_z)^T$: the angular velocity
- Note:
  - How to connect this with stereo geometry (with R, T)?

$$\mathbf{p} = \frac{f}{Z}\mathbf{P}$$

$$\mathbf{V} = -\mathbf{T} - \boldsymbol{\omega} \times \mathbf{P}$$

$$\mathbf{P} - \mathbf{P}' = \mathbf{V} = -\mathbf{T} - \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}\mathbf{P}$$

$$\mathbf{P}' = \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_x & 1 \end{bmatrix}\mathbf{P} + \mathbf{T}$$

$$\mathbf{R} = \begin{bmatrix} \cos\beta\cos\gamma & -\cos\beta\sin\gamma & \sin\beta \\ \sin\alpha\sin\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & -\sin\alpha\cos\beta \\ -\cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & \cos\alpha\sin\beta\sin\gamma + \sin\alpha\cos\gamma & \cos\alpha\cos\gamma \end{bmatrix}$$

3D Computer Vision
and Video Computing

Basic Equations of Motion Field

- Notes:
  - Take the time derivative of both sides of the projection equation
  - The motion field is the sum of two components
    - Translational part
    - Rotational part
  - Assume known intrinsic parameters

$$\mathbf{v} = \frac{f}{Z^2}(Z\mathbf{V} - V_z\mathbf{P})$$

$$\mathbf{V} = -\mathbf{T} - \boldsymbol{\omega} \times \mathbf{P} \qquad \mathbf{p} = \frac{f}{Z}\mathbf{P}$$

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \underbrace{\frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}}_{\text{Rotation part: no depth information}} + \underbrace{\frac{1}{Z}\begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}\begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}}_{\text{Translation part: depth Z}}$$

Motion Field vs. Disparity

- Correspondence and Point Displacements

Stereo	Motion
Disparity	Motion field
Displacement – (dx, dy)	Differential concept – velocity (v_x, v_y), i.e. time derivative (dx/dt, dy/dt)
No such constraint	Consecutive frame close to guarantee good discrete approximation

- Next lecture

Special Case 1: Pure Translation

- Pure Translation (ω =0)
- Radial Motion Field (Tz <> 0)
  - Vanishing point p0 =$(x_0, y_0)^T$:
    - motion direction
  - FOE (focus of expansion)
    - Vectors away from p0 if Tz < 0
  - FOC (focus of contraction)
    - Vectors towards p0 if Tz > 0
  - Depth estimation
    - depth inversely proportional to magnitude of motion vector v, and also proportional to distance from p to p_0
- Parallel Motion Field (Tz= 0)
  - Depth estimation:
    - depth inversely proportional to magnitude of motion vector v

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{Z}\begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}\begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$$

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \frac{f}{T_z}\begin{pmatrix} T_x \\ T_y \end{pmatrix}$$

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{T_z}{Z}\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix}$$

$$Z = \frac{T_z}{|\mathbf{v}|}\sqrt{(x-x_0)^2+(y-y_0)^2}$$

Tz =0

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = -\frac{f}{Z}\begin{pmatrix} T_x \\ T_y \end{pmatrix}$$

$$Z = \frac{f}{|\mathbf{v}|}\sqrt{T_x^2+T_y^2}$$

3D Computer Vision
and Video Computing

Special Case 2: Pure Rotation

- Pure Rotation (T =0)
  - Does not carry 3D information
- Motion Field (approximation)
  - Small motion
  - A quadratic polynomial in image coordinates $(x,y,f)^T$
- Image Transformation between two frames (accurate)
  - Motion can be large
  - Homography (3x3 matrix) for all points
- Image mosaicing from a rotating camera
  - 360 degree panorama

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}$$

$$\mathbf{P}' = \mathbf{R}\mathbf{P}$$

$$\mathbf{p}' = \frac{f'}{Z'}\mathbf{P}' \qquad \mathbf{p} = \frac{f}{Z}\mathbf{P}$$

$$\mathbf{p}' \cong \mathbf{R}\mathbf{p}$$

3D Computer Vision and Video Computing

Special Case 3: Moving Plane

- Planes are common in the man-made world

$$\mathbf{n}^{\mathbf{T}}\mathbf{P} = d \quad \Rightarrow \quad \frac{(n_x x + n_y y + n_z f)}{f} Z = d$$

- Motion Field (approximation)
  - Given small motion

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} + \frac{1}{Z}\begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}\begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$$

  - a quadratic polynomial in image

Only has 8 independent parameters (write it out!)

- Image Transformation between two frames (accurate)
  - Any amount of motion (arbitrary)
  - Homography (3x3 matrix) for all points
  - See Topic 5 Camera Models

$$\mathbf{p}' \cong \mathbf{A}\mathbf{p}$$

- Image Mosaicing for a planar scene
  - Aerial image sequence
  - Video of blackboard

3D Computer Vision and Video Computing

Special Cases: A Summary

- Pure Translation
  - Vanishing point and FOE (focus of expansion)
  - Only translation contributes to depth estimation
- Pure Rotation
  - Does not carry 3D information
  - Motion field: a quadratic polynomial in image, or
  - Transform: Homography (3x3 matrix R) for all points
  - Image mosaicing from a rotating camera
- Moving Plane
  - Motion field is a quadratic polynomial in image, or
  - Transform: Homography (3x3 matrix A) for all points
  - Image mosaicing for a planar scene

3D Computer Vision
and Video Computing

Motion Parallax

- [Observation 1] The relative motion field of two instantaneously coincident points
  - Does not depend on the rotational component of motion
  - Points towards (away from) the vanishing point of the translation direction
- [Observation 2] The motion field of two frames after rotation compensation
  - only includes the translation component
  - points towards (away from) the vanishing point p0 (the instantaneous epipole)
  - the length of each motion vector is inversely proportional to the depth, and also proportional to the distance from point p to the vanishing point p0 of the translation direction
  - Question: how to remove rotation?
    - Active vision : rotation known approximately?

3D Computer Vision
and Video Computing

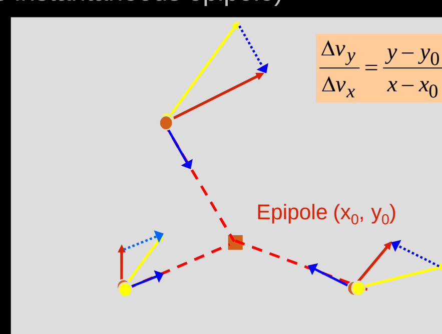
Motion Parallax

- [Observation 1] The relative motion field of two instantaneously coincident points
  - Does not depend on the rotational component of motion
  - Points towards (away from) the vanishing point of the translation direction (the instantaneous epipole)

At instant t, three pairs of points happen to be coincident

The difference of the motion vectors of each pair cancels the rotational components

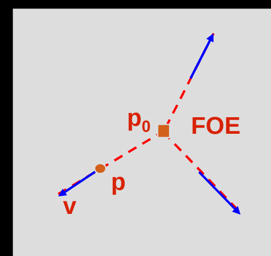
. … and the relative motion field point in (towards or away from) the VP of the translational direction (Fig 8.5 ???)



Motion Parallax

- [Observation 2] The motion field of two frames after rotation compensation
  - only includes the translation component
  - points towards (away from) the vanishing point p0 (the instantaneous epipole)
  - the length of each motion vector is inversely proportional to the depth,
  - and also proportional to the distance from point p to the vanishing point p0 of the translation direction (if Tz <> 0)

$$\frac{v_y^T}{v_x^T} = \frac{y - y_0}{x - x_0}$$



Question: how to remove rotation?
  - Active vision : rotation known approximately?
  - Rotation compensation can be done by image warping after finding three (3) pairs of coincident points

$$|\mathbf{v}| = \frac{T_z}{Z}\sqrt{(x - x_0)^2 + (y - y_0)^2}$$

3D Computer Vision
and Video Computing

Summary

- Importance of visual motion (apparent motion)
  - Many applications…
  - Problems:
    - correspondence, reconstruction, segmentation, understanding in x-y-t space
- Image motion field of rigid objects
  - Time derivative of both sides of the projection equation
- Three important special cases
  - Pure translation – FOE
  - Pure rotation – no 3D information, but lead to mosaicing
  - Moving plane – homography with arbitrary motion
- Motion parallax
  - Only depends on translational component of motion

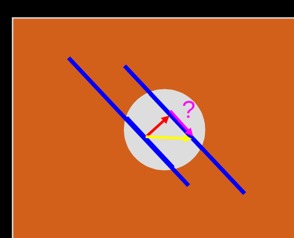
- Next lecture

Notion of Optical Flow

- The Notion of Optical Flow
  - Brightness constancy equation
    - Under most circumstance, the apparent brightness of moving objects remain constant
  - Optical Flow Equation
    - Relation of the apparent motion with the spatial and temporal derivatives of the image brightness
- Aperture problem
  - Only the component of the motion field in the direction of the spatial image gradient can be determined
  - The component in the direction perpendicular to the spatial gradient is not constrained by the optical flow equation

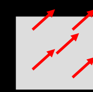
$$\frac{dE(x,y,t)}{dt}=0$$

$$E_x u + E_y v + E_t = 0$$



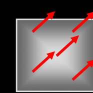
Estimating Optical Flow

- Constant Flow Method
  - Assumption: the motion field is well approximated by a constant vector within any small region of the image plane
  - Solution: Least square of two variables (u,v) from NxN Equations – NxN (=5x5) planar patch
  - Condition: A^TA is NOT singular (null or parallel gradients)



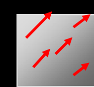
- Weighted Least Square Method
  - Assumption: the motion field is approximated by a constant vector within any small region, and the error made by the approximation increases with the distance from the center where optical flow is to be computed
  - Solution: Weighted least square of two variables (u,v) from

    NxN Equations – NxN patch



- Affine Flow Method
  - Assumption: the motion field is well approximated by a affine parametric model $u^T = Ap^T+b$ (a plane patch with arbitrary orientation)
  - Solution: Least square of 6 variables (A,b) from NxN

    Equations – NxN planar patch



Using Optical Flow

- 3D motion and structure from optical flow (p 208- 212)
  - Input:
    - Intrinsic camera parameters
    - dense motion field (optical flow) of single rigid motion
  - Algorithm
    - (good comprise between ease of implementation and quality of results)
    - Stage 1: Translation direction
      - Epipole (x0, y0) through approximate motion parallax
      - Key: Instantaneously coincident image points
      - Approximation: estimating differences for ALMOST coincident image points

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \frac{f}{T_z}\begin{pmatrix} T_x \\ T_y \end{pmatrix}$$

    - Stage 2: Rotation flow and Depth
      - Knowns: flow vector, and direction of translational component
      - One point, one equation (without depth)–
        - Least square approximation of the rotational component of flow
      - From motion field to depth

$$\frac{v_y^T}{v_x^T} = \frac{y-y_0}{x-x_0}$$

  - Output
    - Direction of translation (f Tx/Tz, f Ty/Tz, f) = (x0, y0, f)
    - Angular velocity
    - 3-D coordinates of scene points (up to a common unknown scale)

3D Computer Vision
and Video Computing

Some Details

- Step 1. Get (Tx, Ty, Tz) = s (x0,y0,f)
- Step 2. For every point (x,y,f) with known v, get one equation about ω from the motion equation (by eliminate Z since it's different from point to point)
- Step 3. Get Z (up to a scale s) given T/s and ω

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \underbrace{\frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}}_{\text{Rotation part: no depth information}} + \underbrace{\frac{1}{Z}\begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}\begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}}_{\text{Translation part: depth Z}}$$

3D Computer Vision
and Video Computing

Feature-Based Approach

- Two frame method - Feature matching
  - An Algorithm Based on the Constant Flow Method
    - Features – corners detection by observing the coefficient matrix of the spatial gradient evaluation (2x2 matrix A^TA)
    - Iteration approach: estimation – warping – comparison
- Multiple frame method - Feature tracking
  - Kalman Filter Algorithm
    - Estimating the position and uncertainty of a moving feature in the next frame
    - Two parts: prediction (from previous trajectory) and measurement from feature matching
- Using a sparse motion field
  - 3D motion and structure by feature tracking over frames
  - Factorization method
    - Orthographic projection model
    - Feature tracking over multiple frames
    - SVD

3D Computer Vision and Video Computing

Motion-Based Segmentation

- Change Detection
  - Stationary camera(s), multiple moving subjects
  - Background modeling and updating
  - Background subtraction
  - Occlusion handling
- Layered representation (I)– rotating camera
  - Rotating camera + Independent moving objects
  - Sprite - background mosaicing
  - Synopsis – foreground object sequences
- Layered representation (II)– translating (and rotating) camera
  - Arbitrary camera motion
  - Scene segmentation into layers

3D Computer Vision and Video Computing

Summary

- After learning motion, you should be able to
  - Explain the fundamental problems of motion analysis
  - Understand the relation of motion and stereo
  - Estimate optical flow from a image sequence
  - Extract and track image features over time
  - Estimate 3D motion and structure from sparse motion field
  - Extract Depth from 3D ST image formation under translational motion
  - Know some important application of motion, such as change detection, image mosaicing and motion-based segmentation

Next

- Reviews, Exam and Projects

Exam
&
Project Presentations

- Homework #4 due in Nov 29, 2010