LECTURE 20

Multi View Geometry of Moving Cameras

Alper Yilmaz and Mubarak Shah
Computer Vision Lab.
Univ. of Central Florida
http://www.cs.ucf.edu/~vision

University of Central Florida VISION

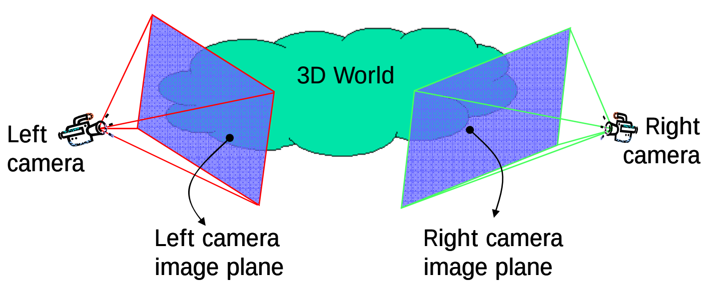
Alper Yilmaz and Mubarak Shah, Recognizing Human Actions in Videos Acquired by Uncalibrated Moving Cameras, IEEE ICCV 2005, Beijing, China, October 15-21.

http://www.cs.ucf.edu/~vision/papers/yilmaz_iccv_2005.pdf

University of Central Florida VISION

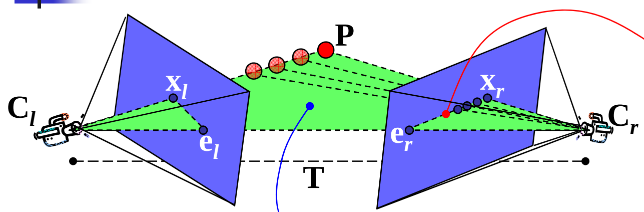
Multi View Geometry

- Defined for two or more static cameras



University of Central Florida VISION

Epipolar Geometry



Epipolar line: set of world points that project to the same point in left image, when projected to right image forms a line.

$\mathbf{X}_l$ maps to line $\mathbf{X}_r\mathbf{e}_r$.

Epipolar plane: plane defined by the camera centers and world point.

- $\mathbf{P}$: world point
- $\mathbf{C}_l$: left camera center
- $\mathbf{C}_r$: right camera center
- $\mathbf{e}_l$: left epipole
- $\mathbf{e}_r$: right epipole

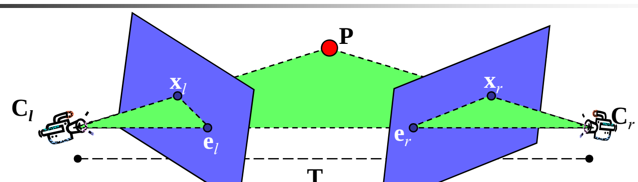
$$\mathbf{P}_l = \mathbf{C}_l\mathbf{P}$$

$$\mathbf{P}_r = \mathbf{C}_r\mathbf{P}$$

$$\mathbf{T} = \mathbf{C}_r - \mathbf{C}_l$$

University of Central Florida VISION

Essential Matrix



Coplanarity constraint between vectors $(\mathbf{P}_l\text{-}\mathbf{T})$, $\mathbf{T}$, $\mathbf{P}_l$. $(\mathbf{P}_l - \mathbf{T})^{\mathrm{T}}\mathbf{T} \times \mathbf{P}_l = 0$

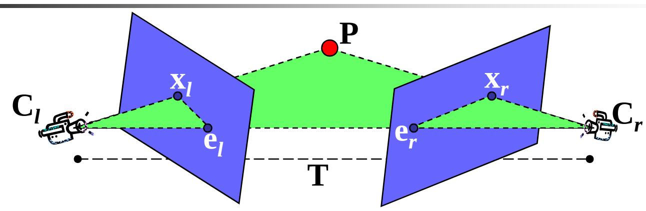
$$\mathbf{P}_r = R(\mathbf{P}_l - \mathbf{T}) \Rightarrow (R^{\mathrm{T}}\mathbf{P}_r)^{\mathrm{T}}\mathbf{T} \times \mathbf{P}_l = 0 \Rightarrow (\mathbf{P}_r{}^{\mathrm{T}}R)\mathbf{T} \times \mathbf{P}_l = 0$$

$$\mathbf{T} \times \mathbf{P}_l = \begin{pmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{pmatrix}\mathbf{P}_l = S\mathbf{P}_l \Rightarrow \mathbf{P}_r{}^{\mathrm{T}}RS\mathbf{P}_l = 0 \Rightarrow \boxed{\mathbf{P}_r{}^{\mathrm{T}}E\mathbf{P}_l = 0}$$

essential matrix

University of Central Florida VISION

Fundamental Matrix

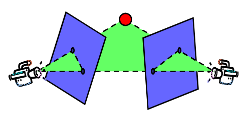


$$\left.\begin{aligned} \mathbf{P}_r &= R(\mathbf{P}_l - \mathbf{T}) \\ \mathbf{P}_l &= R_l\mathbf{P} + T_l \\ \mathbf{P}_r &= R_r\mathbf{P} + T_r \end{aligned}\right\} \quad \begin{aligned} \mathbf{R} &= R_r R_l^{\mathrm{T}} && \text{(A)} \\ \mathbf{T} &= T_l - \mathbf{R}^{\mathrm{T}} T_r && \text{(B)} \end{aligned}$$

$$\left.\begin{aligned} \mathbf{x} &= M_l \mathbf{P}_l \\ \mathbf{x}' &= M_r \mathbf{P}_r \\ \mathbf{P}_r^{\mathrm{T}} E \mathbf{P}_l &= 0 \end{aligned}\right\} \quad \begin{aligned} &\mathbf{x}'^{\mathrm{T}} M_r^{-\mathrm{T}} E M_l^{-1} \mathbf{x} = 0 \\ &\boxed{\mathbf{x}'^{\mathrm{T}} F \mathbf{x} = 0} \end{aligned}$$

fundamental matrix

Fundamental Matrix



$$\mathbf{x}'^{\mathrm{T}} F \mathbf{x} = \mathbf{x}'^{\mathrm{T}} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & m \end{pmatrix} \mathbf{x} = 0$$

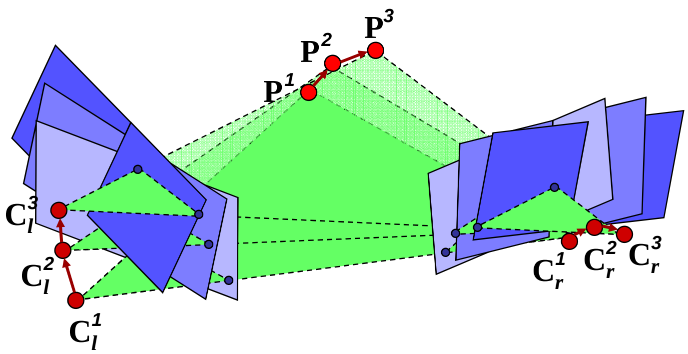
- Rank 2 matrix (due to S)
- 3x3 matrix with 9 components
- 7 degrees of freedom
- Given a point in left camera $\mathbf{x}$, epipolar line in right camera is: $\mathbf{u}_r = F\mathbf{x}$

What Happens to Fundamental Matrix When Cameras Move?

Observations

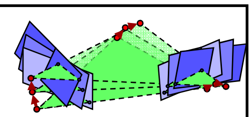
- At each time instant
  - Different epipolar geometries
  - Different epipoles
  - Different **R** and **T** (equations A and B)
  - Different fundamental matrices
- Is there any relation between consecutive epipolar geometries?

Epipolar Geometry of Moving Cameras



University of Central Florida VISION

Theorems Governing Temporal Epipolar Geometry

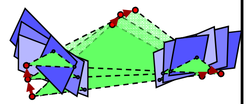


- Theorem 1 (temporal fundamental matrix):
  - Corresponding points in two sequences captured by independently moving cameras are related to each other through a temporal fundamental matrix of the form:

$$\boxed{\mathbf{x}_r^{\mathrm{T}}(t)\hat{F}(t)\mathbf{x}_l(t) = 0}$$

University of Central Florida VISION

Theorems Governing Temporal Epipolar Geometry



- Theorem 2 (on the order of polynomials in the temporal fundamental matrix):
  - Assume that motion of cameras are approximated by polynomials in time variable. Then, the temporal fundamental matrix is a 3x3 matrix whose components are polynomials of order:

$$\boxed{\deg \hat{F}_{i,j}(t) = \max(n_l, n_r, m_l, m_r) + 1}$$

where i,j refers to i^{th} row j^{th} column, and m_α and n_α refers to degree of polynomials for translational and rotational velocities.

University of Central Florida VISION

Computing TFM

- Find corresponding points in first frames
- Normalize all trajectories in both views
  - Mean normalize
  - Isotropically scale: on the average a point is (1,1,1)

University of Central Florida VISION

Computing TFM (for order 2)

- Select the order of polynomials

$$[x_r(t) \quad y_r(t) \quad 1]\begin{bmatrix} a_1+a_2t+a_3t^2 & b_1+b_2t+b_3t^2 & c_1+c_2t+c_3t^2 \\ d_1+d_2t+d_3t^2 & e_1+e_2t+e_3t^2 & f_1+f_2t+f_3t^2 \\ g_1+g_2t+g_3t^2 & h_1+h_2t+h_3t^2 & m_1+m_2t+m_3t^2 \end{bmatrix}\begin{bmatrix} x_l(t) \\ y_l(t) \\ 1 \end{bmatrix}=0$$

- Construct a linear system to solve TFM unknowns: $\mathbf{A}_{N\text{x}27}.f_{27\text{x}1}=\mathbf{0}_{N\text{x}1}$

$$\mathbf{A}_i = (\quad x_rx_l, x_rx_lt, x_rx_lt^2, x_ry_l, x_ry_lt, x_ry_lt^2, x_r, x_rt, x_rt^2, y_rx_l, y_rx_lt, y_rx_lt^2, \\ y_ry_l,\ y_ry_lt,\ y_ry_lt^2,\ y_r,\ y_rt,\ y_rt^2,\ x_l,\ x_lt,\ x_lt^2,\ y_l,\ y_lt,\ y_lt^2, 1, t,\ t^2 \quad)$$

$$f = (a_1,a_2,a_3,b_1,b_2,b_3,c_1,c_2,c_3,d_1,d_2,d_3,e_1,e_2,e_3,f_1,f_2,f_3,g_1,g_2,g_3,h_1,h_2,h_3,m_1,m_2,m_3)^{\mathrm{T}}$$

University of Central Florida VISION

Computing TFM (for order 2)

$$\mathbf{A}f = 0$$

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}f = 0$$

- Compute Singular Value Decomposition of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$
- Select minimum eigenvalued eigenvector.

University of Central Florida VISION

Quality of Recovered Geometry

- Two measures
  - Condition Number of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$: How well-conditioned is the homogenous linear system?
  - Symmetric Epipolar Distance: How correct is the estimated TFM?

Condition Number

- Select rank of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ of from singular values σ_I

$$rank(\mathbf{A}^{\mathrm{T}}\mathbf{A}) = i \ni \frac{\sum_{j=1}^{i}\sigma_j}{\sum_{j=i+1}^{N}\sigma_j} > th$$

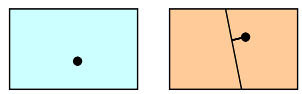
- Compute condition number of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ **by**

$$\boxed{c = \frac{\sigma_1}{\sigma_{rank(\mathbf{A}^{\mathrm{T}}\mathbf{A})}}}$$

Geometric Error in TFM

- Compute TFM unknowns
- Construct TFM by enforcing constraints
- For each point on left and right cameras
  - Compute epipolar line

$\boxed{u_r = \hat{F}(t)x_l}$ $\boxed{u_l = \hat{F}^{\mathrm{T}}(t)x_r}$



  - Compute distance of point from epipolar line

$$\boxed{g = \frac{1}{2}\left(\frac{\left|u_l^{\mathrm{T}} x_l\right|}{\left|u_l\right|} + \frac{\left|u_r^{\mathrm{T}} x_r\right|}{\left|u_r\right|}\right)}$$

University of Central Florida VISION

Applications of This New Geometry

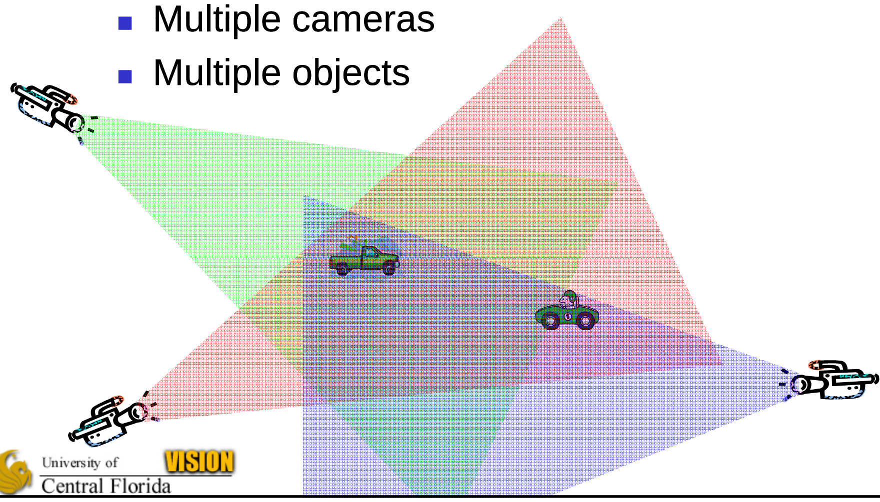
- Tracking Across Multiple Moving Cameras
- Action Recognition in Video Captured Using Uncalibrated Moving Cameras

University of Central Florida VISION

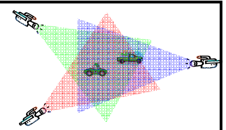
Tracking Objects Across Multiple Moving Cameras

Problem Definition

- Multiple cameras
- Multiple objects

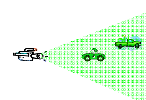


University of Central Florida VISION



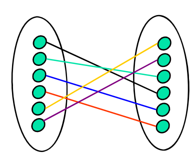
Problem Definition

- Find corresponding objects
  - Cameras move independently
  - Objects move independently



Tracking in Single Camera

- Point based
  - Object detection (background subtraction)
  - Point (centroid) correspondence
- Region based
  - Rigid motion models (mean-shift tracker, Eigentracking, etc.)
- Contour based
  - Non-rigid object deformations



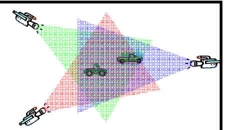
Tracking Across Multiple Cameras

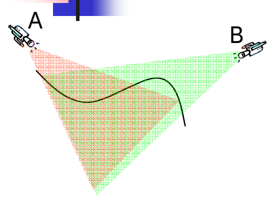
- Should be more than one object in scene
- Object tracking in single camera is performed first
- Which object in one camera associates with the objects in the second camera?
  - Object correspondence
  - Trajectory correspondence

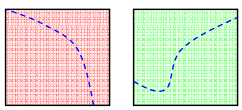
Tracking Across Multiple Cameras
Object Correspondence

- Appearance matching
  - Same scene appears different in different cameras (different camera gain)
  - Different object views

Tracking Across Multiple Cameras
Trajectory Correspondence



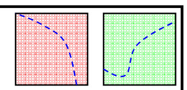




- Trajectory correspondence
  - Advantage: Cameras can be of different modalities
  - Registration based: Compute transformation between two camera views (affine, projective)
  - Epipolar geometry based

University of Central Florida VISION

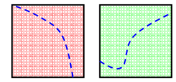
Registration Based Trajectory Matching



- If you have stationary cameras
  - (1)
    - Register one camera view onto the other
    - Label closest trajectories
  - OR…
  - (2)
    - Use epipolar geometry constraints (applied in context of action recognition)
  - OR…
  - (3)
    - Treat trajectories as 3D objects (x,y,t), compute 3D transformation between them and compute reconstruction error.

University of Central Florida VISION

Registration Based Trajectory Matching



- For moving cameras
  - Compensate camera motion
    - Generates trajectory as if camera is stationary
  - (1) Apply method 2 or 3 for stationary camera case
    - Yaser's work
  - OR…
  - It is like magic ☺
    - (2)
    - Temporal epipolar geometry
    - Complete perspective geometry

University of Central Florida VISION

Fundamental Matrix in Context of Trajectory Matching

- Let there be N corresponding points in left and right cameras
- Unknowns of F can be found by least squares

$$\underbrace{\begin{bmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_N x_N & x'_N y_N & x'_N & y'_N x_N & y'_N y_N & y'_N & x_N & y_N & 1 \end{bmatrix}}_{\mathbf{A}} \overbrace{\begin{bmatrix} a \\ \vdots \\ g \\ m \end{bmatrix}}^{f} = \mathbf{0} \quad \Longrightarrow \quad \mathbf{A}^{T}\mathbf{A}f = \mathbf{0}$$

- Solution is found by SVD: Eigenvector with smallest eigenvalue
- Ideally 27th singular value is 0.
- Use 27th singular value as distance measure between two trajectories.

University of Central Florida VISION

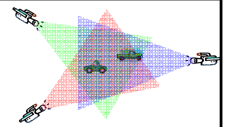
Finding Corresponding Objects Across Cameras

- Given N_r and N_l trajectories in right and left cameras
- Find correct correspondences that satisfy temporal fundamental matrix (TFM).
- What is TFM?
  - We know neither TFM nor correspondences ☹
  - We know the constraints used to compute TFM ☺
    - .

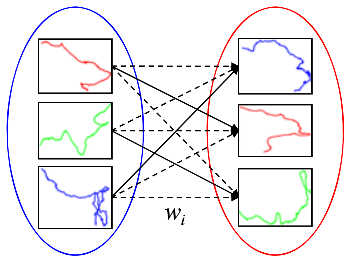
Finding Corresponding Objects Across Cameras

- Select a correspondence hypothesis and check its validity
  - Total of N_rxN_l hypotheses
- How do we check validity?
  - Algebraic error
  - Geometric reconstruction error

Computing Correspondences



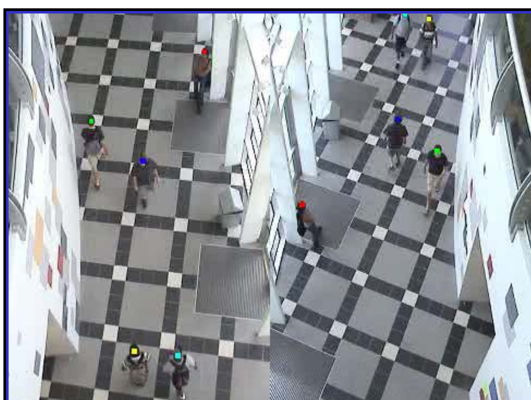
- Graph theoretic matching

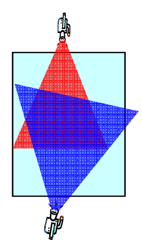


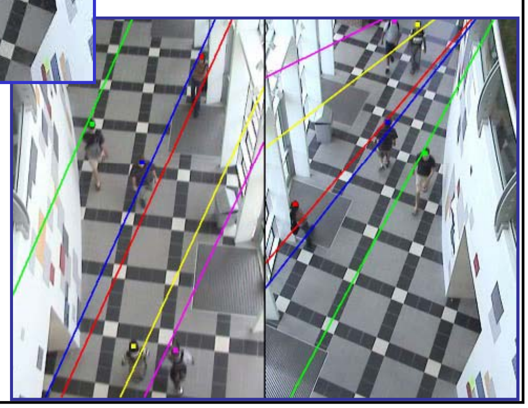
Bipartite graph: No edges between trajectories in the same camera view.

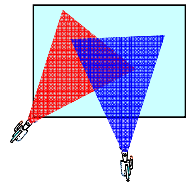
Edge weights: w_i are computed using either geometric or algebraic error.

Correspondences: Computed by *maximum matching of weighted bipartite graph*.

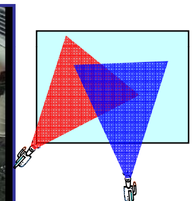


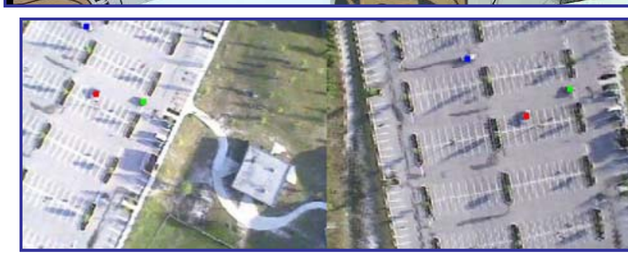


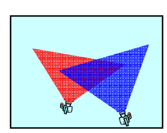




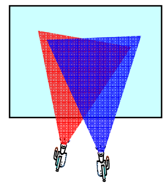
University of Central Florida VISION







University of Central Florida VISION



Algebraic & Geometric Errors in Matching

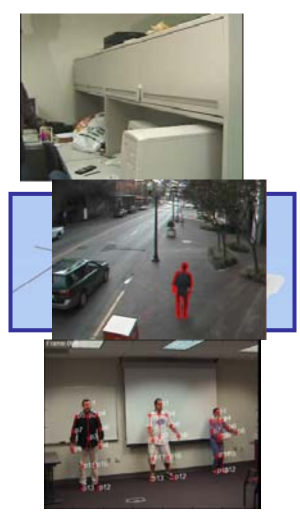
Matches	*ConNo*	*PrEr*	*ConNo*	*PrEr*	*ConNo*	*PrEr*	*ConNo*	*PrEr*	*ConNo*	*PrEr*	*ConNo*	*PrEr*
Correct	106	14	3730	9	2457	12	5940	9	2211	9	305.3	9
Closest	99	29	3712	25	1960	28	3824	55	2113	20	293	23
Worst	82	49	2238	26	1206	27	1266	81	1532	25	293.4	23

Action Recognition in Video Captured Using Uncalibrated Moving Cameras

University of Central Florida VISION

Representations for Action Recognition

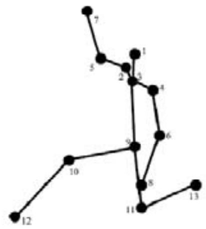
- Motion trajectories
  (Rao et al. IJCV 2002)
- Actions as Objects
  (Yilmaz&Shah CVPR 2005)
- Set of Landmarks
  (Gritai, Sheikh & Shah, ICPR 2004)



University of Central Florida VISION

Representing Human Body

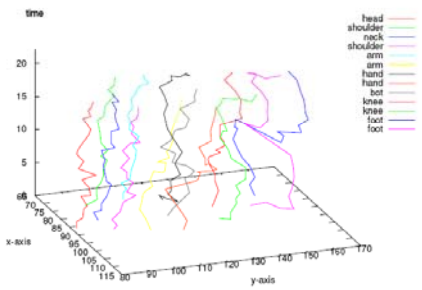
- A set of landmark points on the body



Representing Action

A collection of trajectories:

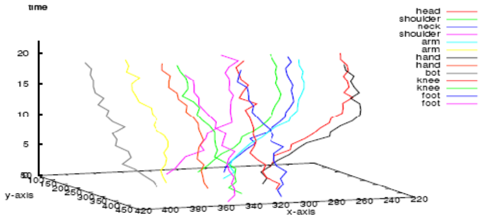
$$\mathbf{U} = \left(\Gamma_1^{T}, \Gamma_2^{T}, \ldots, \Gamma_{13}^{T}\right)$$



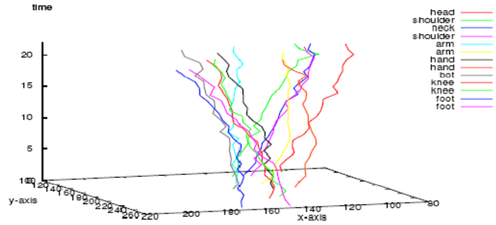
What happens to action trajectories when cameras move?

- Trajectories include camera motion
- Viewpoint constantly changes
- Action trajectories appear different

Example (Picking Up Action)



- Different actors
- Different camera motion
- Different viewpoints



Apprpach-1

- Compensate camera motion
  - High computational cost
  - Assumes planar scenes, actions are not planar
  - Compensation distorts world to image perspective projection
    - Introduces artificial deformations

Approach-II

- Compute epipolar geometry of each frame individually
  - Computational cost
    - Estimation of huge number of unknowns
  - Temporal dependency of epipolar geometries is not considered
    - Additional constraints are required

University of Central Florida VISION

Proposed Approach

- Find the geometry between two actions views using TFM
- Matching score:
  - Condition number (CN)
  - Symmetric epipolar distance (SED)

$$\mathbf{U} = \left(\Gamma_1^{\mathrm{T}}, \Gamma_2^{\mathrm{T}}, \ldots, \Gamma_{13}^{\mathrm{T}}\right)$$

$$\mathbf{U}_t^{one\,\mathrm{T}} \tilde{F}(t) \mathbf{U}_t^{two} = 0$$

Form a homogenous system of equations

University of Central Florida VISION

Similarity of Two Actions

- **Proposition:** ***Given two action videos captured by uncalibrated moving cameras, there exists a unique temporal fundamental matrix which can be computed using landmark points on the actors.***

$$\mathbf{U}_t^{\mathrm{one\,T}} \left(\sum_{i=0}^{k} F_i (t)^k \right) \mathbf{U}_t^{\mathrm{two}} = 0$$

- Similarity is defined in terms of quality of the linear system and the quality of the recovered geometry.
  - Condition number of **ATA**
  - Symmetric epipolar distance *g*

On the Matching Score

- Homogenous equation system has many solutions
  - High CN: does not mean good TFM estimate
  - Low CN: ill-conditioned equation system
  - High CN & low SED indicate correct action match

$$\mathcal{S} = \left(1 - \exp\left(-\frac{\mathcal{C}^2}{\sigma_c^2}\right)\right)\left(\exp\left(-\frac{\mathcal{G}^2}{\sigma_g^2}\right)\right)$$

University of Central Florida VISION

Database

- 19 action videos
  - Moving camera
  - Different viewpoints
  - Different actors

University of Central Florida VISION

Action Videos



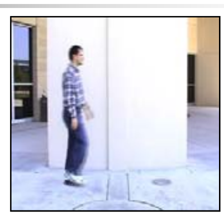

Action Videos

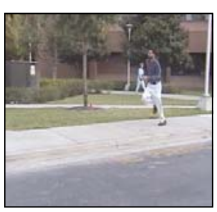


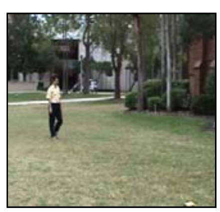
Action Videos

University of Central Florida VISION

Action Videos

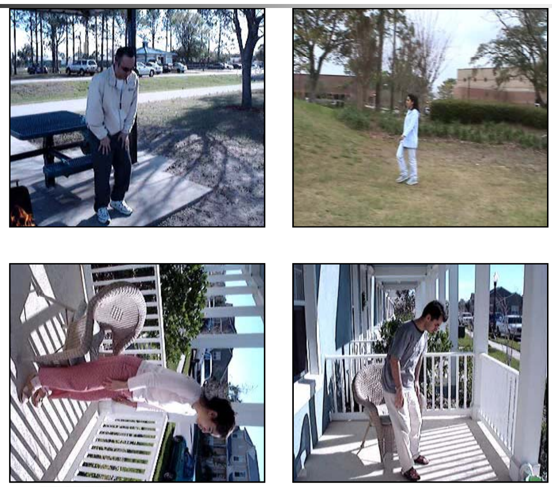






University of Central Florida VISION

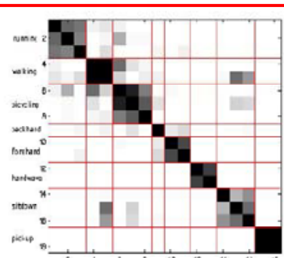
Action Videos



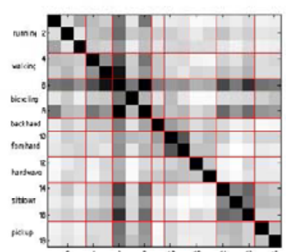
University of Central Florida VISION

Recognition Performance

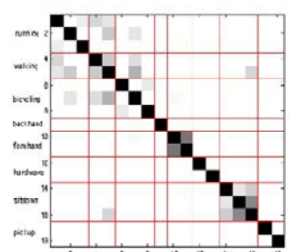
- Confusion matrices for various methods



Proposed method



Use of static epipolar geometry

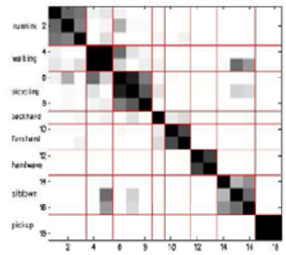


Use of static epipolar geometry and our metric

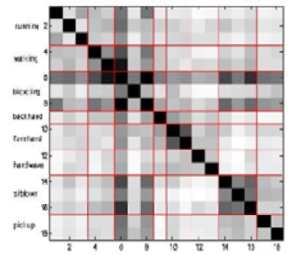
University of Central Florida VISION

Recognition Performance

- Confusion matrices for various methods



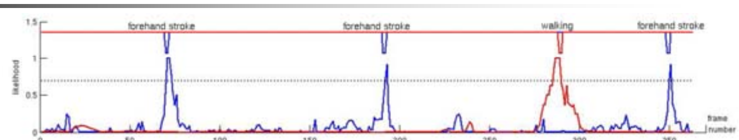
Proposed method



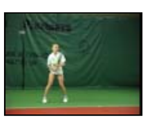
Use of static epipolar geometry

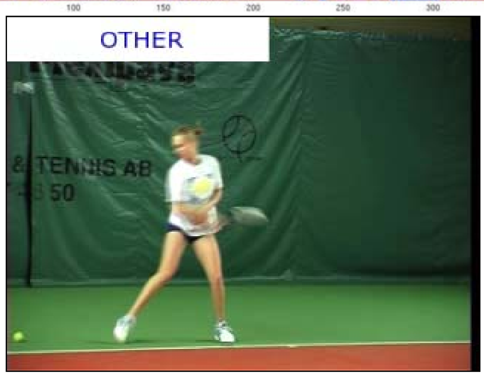
University of Central Florida VISION

Action Retrieval



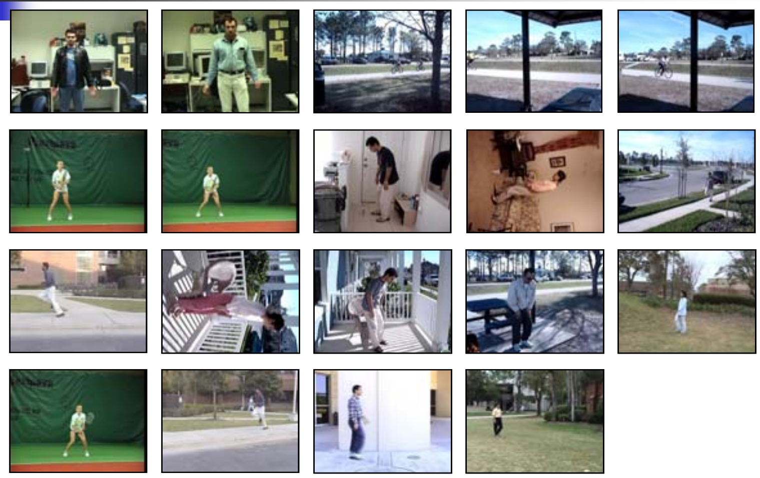
Exemplar Actions





University of Central Florida VISION

Actions



University of Central Florida VISION