Interpretation of Airborne Video Sequences

Rama Chellappa
Dept. of Electrical and Computer Engineering and Center for Automation Research
University of Maryland

Supported by ARO MURI

Applications -1

- Instantaneous situational awareness
- Scene Interpretation
  - Terrain segmentation and classification
    - Where is the lake?
  - Terrain modeling, map updating and path planning
    - Can vehicles go through the terrain?
  - Video measurements
    - How tall is the building?
  - Damage assessment

Applications - 2

- Dynamic interpretation
  - Is anything moving in the scene?
  - Humans, vehicles, animals?
  - Combatants/non-combatants?
    - Civilians/insurgents/military
  - How many?
    - Counting
  - What are they up to?
    - Activity modeling and recognition

UMD’s Involvements

- Early Nineties
  - RSTA for UGVs (DARPA)
- Mid Nineties
  - Visual Surveillance and Monitoring (VSAM) effort (DARPA)
- Late Nineties
  - Airborne Visual Surveillance (AVS) effort (DARPA)
- Recent Years
  - FedLab, CTA efforts (ARL)
  - Video Verification and Identification (VIVID) effort (DARPA)
  - MURI on MAVs (ARO)

Challenges in Processing MAV Videos

- No texture over a large portion of the image
- Large inter-frame displacements
- Low-resolution and poor-quality video
- Limited onboard processing capability
- Low signal to noise ratio
    - Tons of data
    Unreliable/unsynchronized/unavailable meta data.
    - Absence of compound eyes
        Unable to do insect-style processing

Ongoing Work - 1

- **MAV Video Stabilization**
  - Earlier efforts focused on optic flow and discrete features.
  - Recent efforts have looked at the horizon, features at infinity and close by.
  - From distant points and horizon get the full rotation vector.
  - Refine rotation and estimate translation using close by features.

MAV Stabilization Results - 1

MAV Stabilization Results – 2

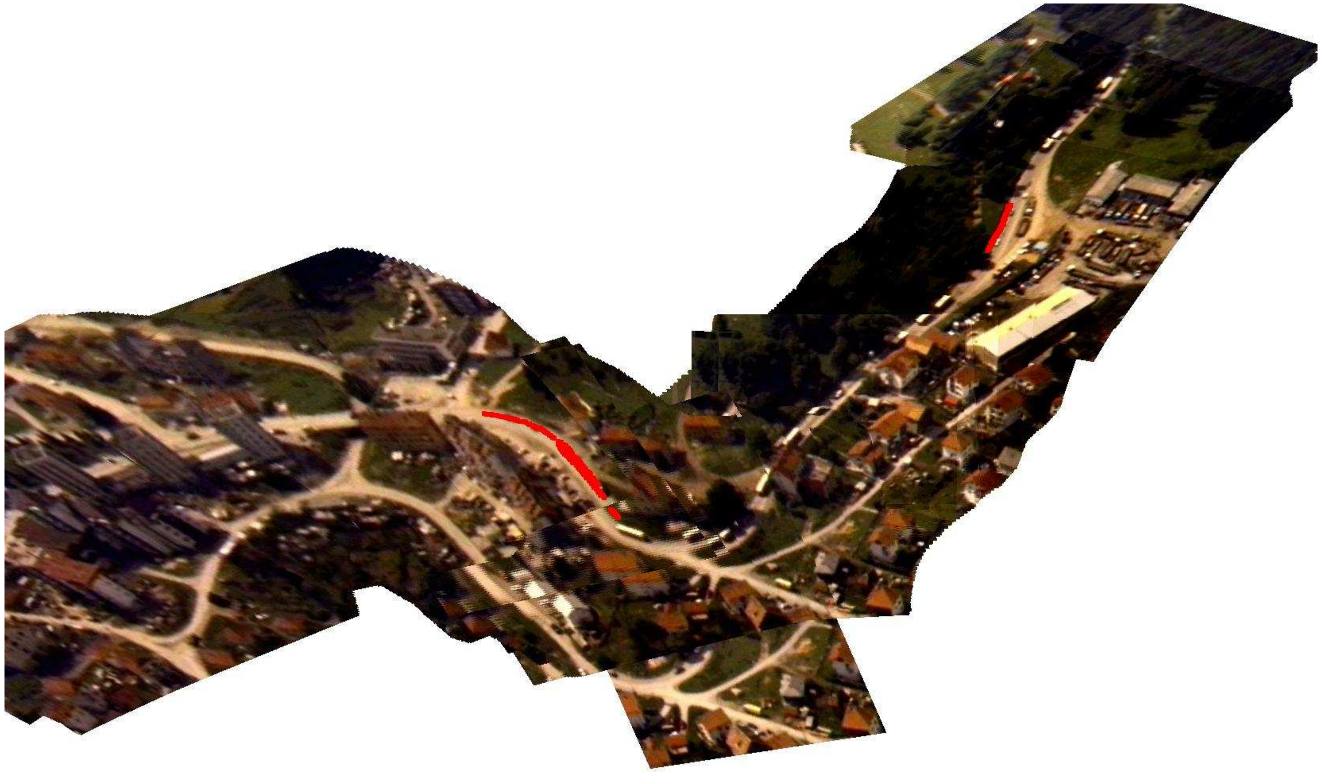
MAV Stabilization Results - 2

MAV Stabilization Results - 3

MAV Stabilization Results - 3

UAV Stabilization

Persistent Tracking in Airborne Video



Persistent tracking in a PREDATOR-generated mosaic built using 2200 frames. The red tags indicate tracking the same convoy of vehicles on the mosaic.

Ongoing Work -2

- Persistent tracking and verification of targets
  - Appearance/feature graph based
  - Maximize the probability of the target appearance given the video
  - Temporal integration of tracking and ID parameters
  - Funded by ARL/Collaborative Tech. Alliance on Advanced
  - Sensors

Detection and Tracking of Moving Objects

- Stochastic appearance tracking is a stochastic process for modeling inter-frame motion and appearance changes
  - Video frame $\{ Y_1, Y_2, \ldots, Y_t, \ldots \}$
  - Motion parameter $\{ q_1, q_2, \ldots, q_t, \ldots \}$
  - State equation (motion model): $q_t = F_t(q_{t-1}, U_t)$
  - Observation equation (model): $Y_t = G_t(q_t , V_t)$

Particle Filters

- Statistical inference
  - Computing the posterior probability $p(q_t|Y_{1:t})$
- Particle filters (PF)
  - PF approximates $p(q_t|Y_{1:t})$ using a set of weighted particles $\{q_t^{(j)}, w_t^{(j)}; j=1,\ldots,J\}$
  - Two steps: (i) propagate the particles governed by the motion model;
    (ii) update the weights using the observation model.
  - The state estimate q_t^* can be a MMSE, MAP, or other estimate based on $p(q_t|Y_{1:t})$ or $\{q_t^{(j)}, w_t^{(j)}; j=1,\ldots,J\}$.

Adaptive Visual Tracking Using PF

- Strategy: appearance-adaptive
  - State and observation models adaptive to appearances in the video
- Adaptive observation model
  - $T\{Y_t\ ;\ q_t\} \equiv Z_t = A_t + V_t$
  - A_t is a mixture appearance model (MAM) adaptive to all past observations
- Adaptive motion model
  - Time-varying Markov model: $q_t = q_{t-1} + U_t$
  - Adaptive noise variance; $U_t = n_t + r_t\, U_0;\ U_0 \sim N(0, S_0)$
  - The mean n_t and the 'variance' function r_t, both time-varying, adapt to the incoming frame Y_t

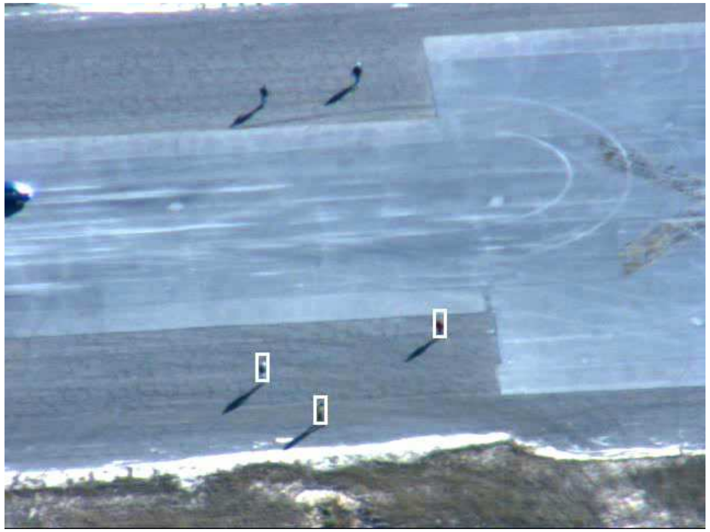
Mixture Appearance Model (MAM)

- **Mixture of 3 components: stable, wandering, fixed**
  - **Stable ('*S-*') component captures a slowly-varying structure in the appearance.**
  - **Wandering ('*W-*') component captures a rapidly-varying structure in the appearance.**
  - **Fixed ('*F-*') component, which is optional, captures a constant structure in the appearance.**
  - **Each component has *d* pixels, assumed to be Gaussian.**
- **IEEE Transactions on Image Processing, Nov. 2004**

Tracking for Airborne Videos

- **Background/foreground modeling**
- **Integrating intensity and motion information overcomes difficulties due to low contrast and low resolution**
- **Simultaneous tracking of background and foreground motions improves estimation of the motion parameters and segmentation.**
- **Particle weights can be adjusted using the quality of Motion and appearance cues.**

Airborne Video Examples



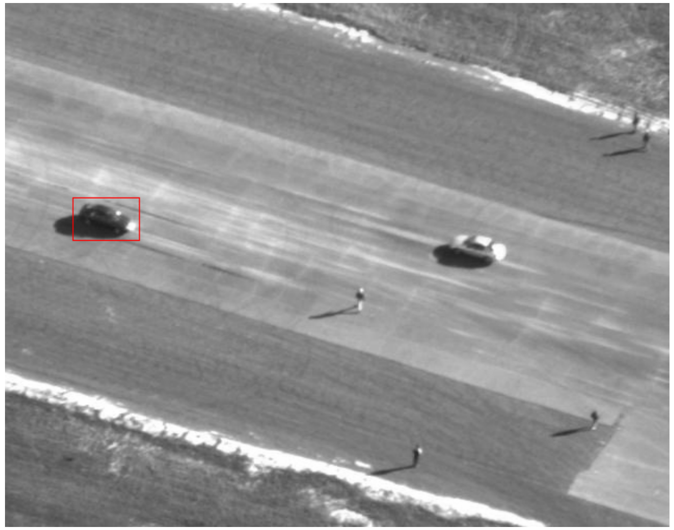
“Probing” for Human/Vehicle Classification

- **Analysis of motion signature for segmentation of humans and human/vehicle classification**
  - **The reference signal is based on periodicity and symmetry of human motion - Twin pendulam model of walking motion**
- **Assuming that the intensity at a periodic pixel (*i, j*) is the sum of a periodic signal *M*(*i, j*)(*t*) and an additive Gaussian noise *n*(*t*), perform statistical hypothesis testing.**

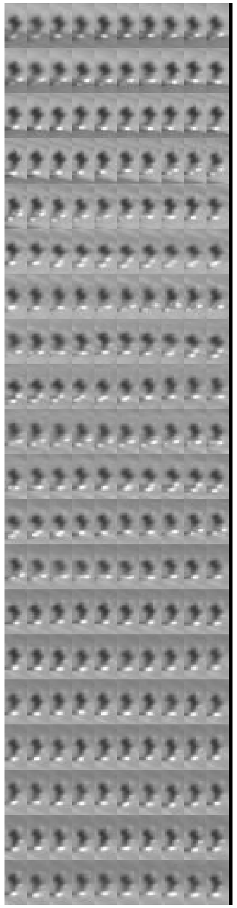
$$x_t(i,j) = M_t(i,j) + n(t)$$

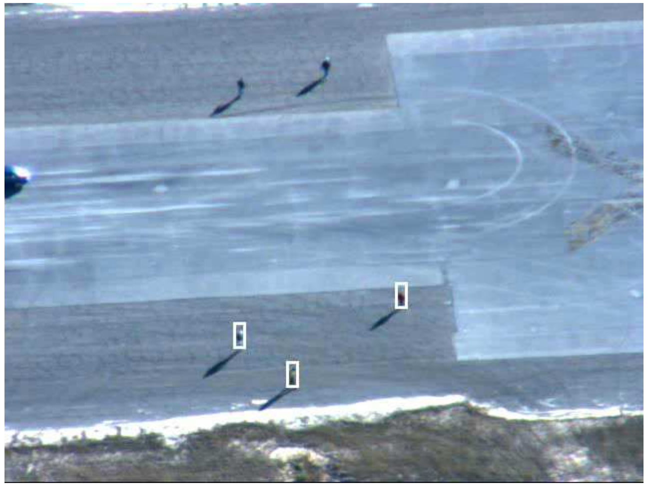
$$= \mu(i,j) + \sum_{k=1}^{\infty} [\alpha_k \cos(k\omega t) + \beta_k \sin(k\omega t)] + n(t)$$

Airborne Video Examples



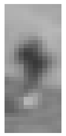
Detected object
sequence

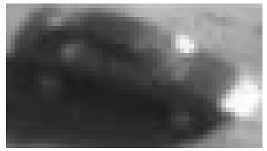




Frame: 12x24

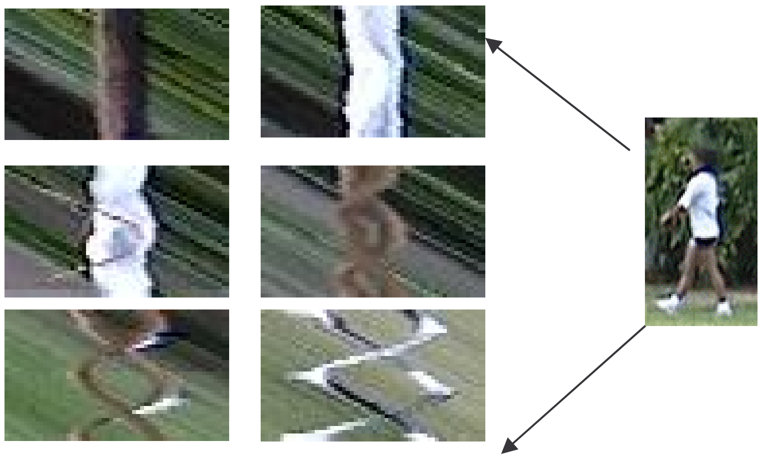
Object size: 10x15



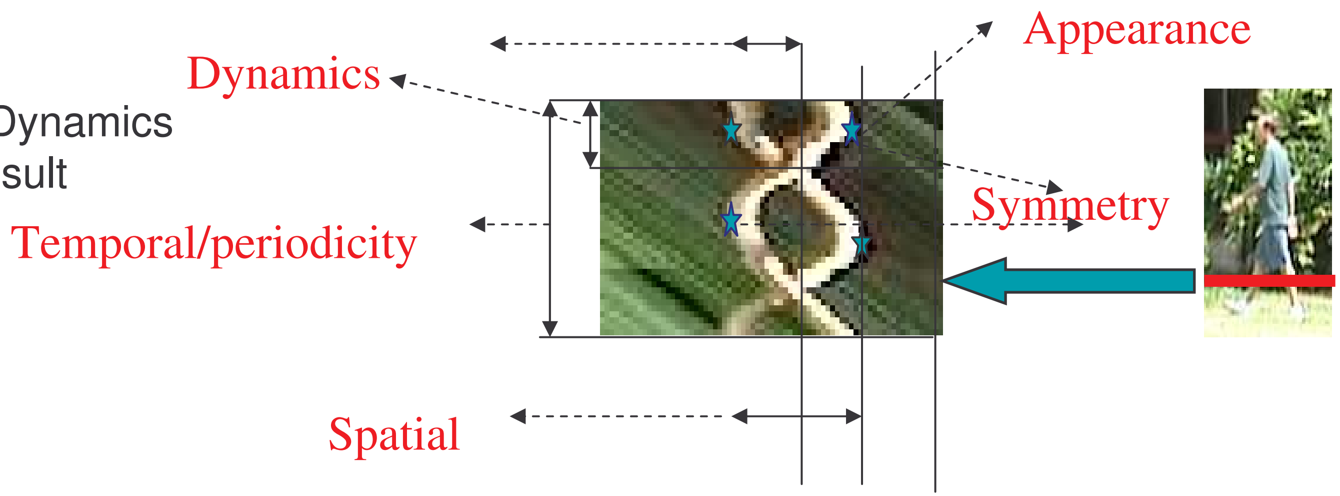


The "DNA" of Human Motion

- Look at X-t plane
  - Twin-pendulam model generates a helical structure
  - Spatio-temporal slices at various heights shown



- The "**DNA**" of human motion codes:
  - Appearance
  - Symmetry
  - Kinematics/Dynamics
  - First such result



Ongoing Work - 3

- **Video verification and identification (VIVID)**
  - **Novel view synthesis of objects for improved recognition.**
  - **Use of discrete features and bundle adjustment**
  - **Homography based method (BMVC Sept. 2005)**
  - **Factorization algorithm (IEEE Motion Workshop, Jan. 2005)**

Homography Decomposition

- Get the uncalibrated homography H_π between two frames induced by the ground plane using the appearance based tracker.
- Compute the calibrated homography H by $H = K_2^{-1} H_\pi K_1$ where K_1 and K_2 are the calibration matrices obtained from metadata.
- Decompose $H = R\left(I - \varsigma t n^T\right)$, where R is the rotation between the two frames, t is the translation between the two frames, and n^T is the surface normal for the ground plane

[Bill Triggs,1998].

Multi-View Fusion Using Infinite Homography

- **For a distant plane as in airborne video, the estimated t and n^T might be unreliable but R is still accurate.**
- **The infinite homography $H_k^{\inf}$ for each pair of frames is computed from the rotation matrix R_k as** $H_k^{\inf} = K_k R_k K_1^{-1}$
- **A block matrix *W* is constructed by stacking all the transformed inter-frame homographies, and factorized into the camera center vector $[\bar{t}_k]$ and the ground plane surface normal n^T using SVD:**

$$W = \begin{pmatrix} \hat{H}_2 \\ \hat{H}_3 \\ \vdots \\ \hat{H}_n \end{pmatrix} = \begin{pmatrix} \bar{t}_2 \\ \bar{t}_3 \\ \vdots \\ \bar{t}_n \end{pmatrix} n^T$$

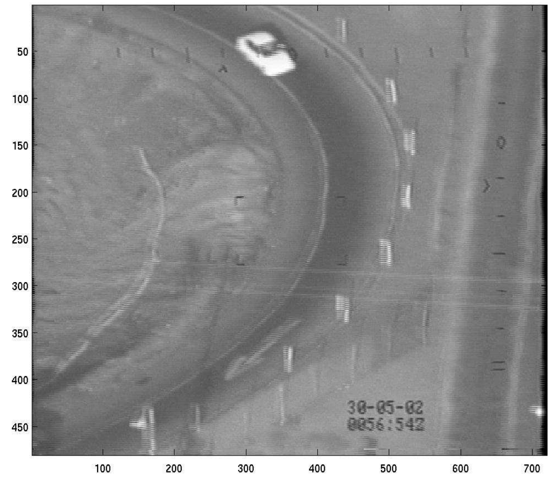
View Synthesis Using Homography

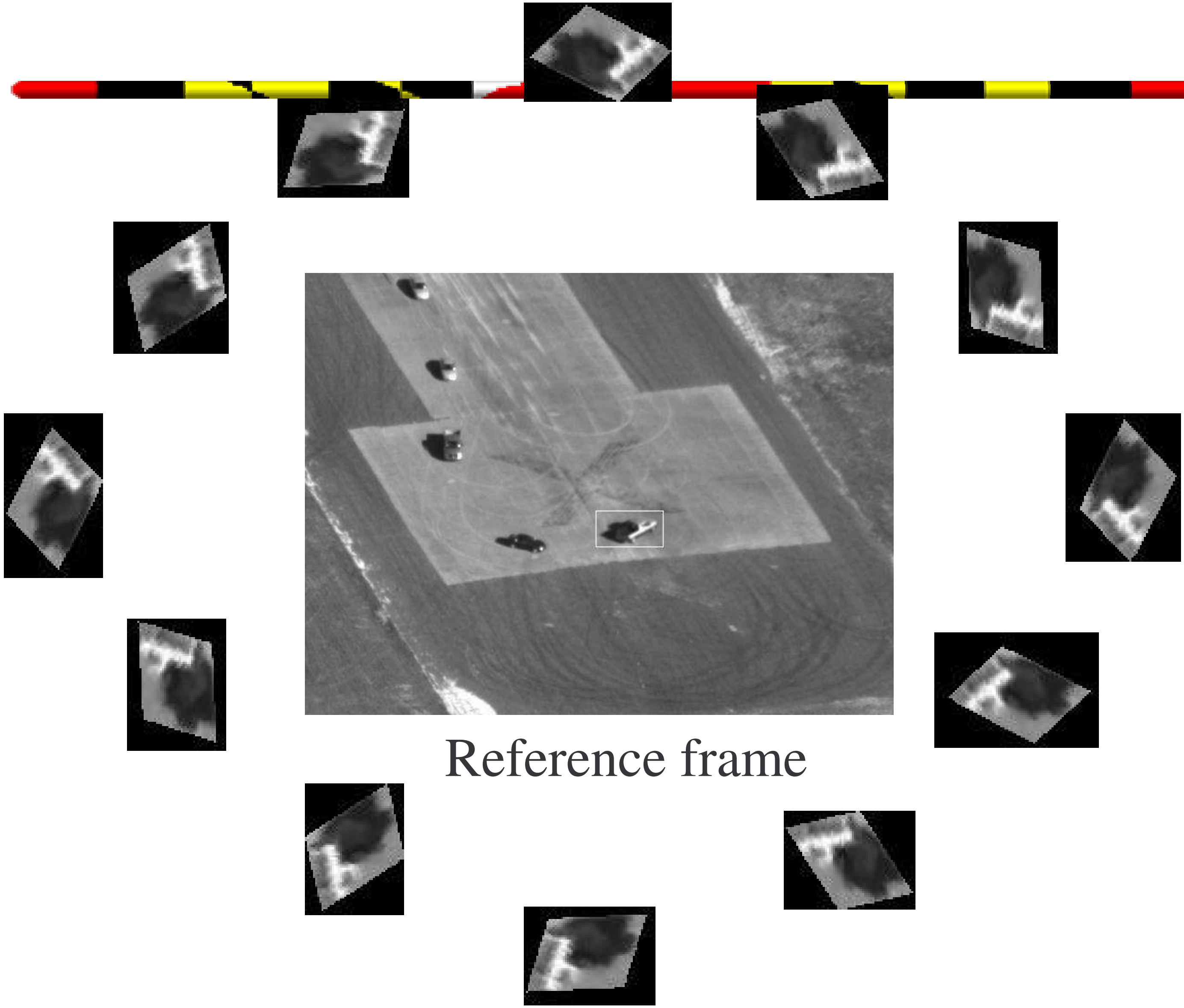
- **Given the desired viewing direction R_{new} with respect to the reference frame, generate new homography from R_{new} and n^T :**

$$H_{new} = K_{new}\left(R_{new} - tn^T\right)K_1^{-1}$$

- **A cubic interpolation is used to get the smooth synthesis result.**

Input Video Frames





Reference frame

Technical Challenge - 1

- Video stabilization, mosaicking and superresolution
  - Egomotion estimation
    - Use of IMU
  - Sub-pixel alignment
  - Background and foreground motion analysis
- Resources required
  - Reliable metadata (Time, frame, aspect angle, slant range, resolution, platform altitude, latitude, longitude)
  - Biology

Technical Challenges – 2

- Terrain modeling and navigation
  - Estimation of terrain height using optical flow
  - Landmark recognition
  - Path planning for navigation
- Video metrology
  - Measuring the height of man-made structures
  - Dynamic mensuration
- Resources required
  - DTED, Camera calibration, Biology

Technical Challenges - 3

- DCIT of humans and vehicles
  - Accurate positioning of moving objects
  - Video-based target recognition
  - Combatants/noncombatants
  - Handling Occlusion by buildings, trees etc
- Resources required
  - Fingerprinting algorithms
  - Kinematic motion models, terrain models

Technical Challenges - 4

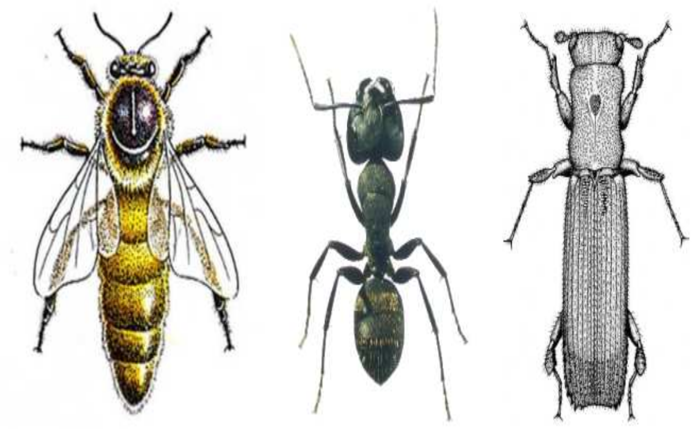
- Human/vehicle activity analysis
  - Anomaly detection
  - Interpretation of source to sink trajectories
  - Models for activities
- Motion trajectories – shapes – activities
  - Factorization theorem for activity modeling
  - Statistical shape models for activity modeling
- Resources required
  - Ontology for characterizing activities
  - Interactions with end users

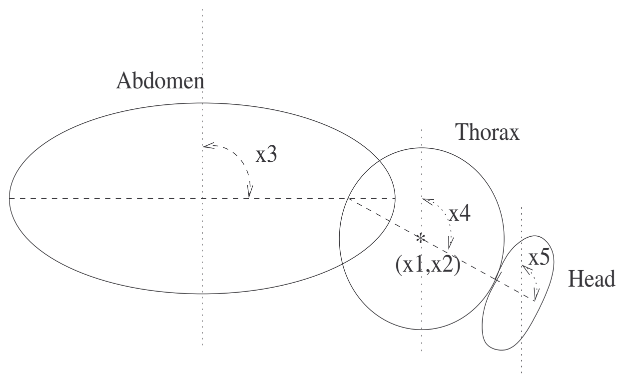
Behavior Tracking

- Behavior analysis of insects has led to advances in navigation, control systems etc.
- Goal: To automate tracking and labeling of insect motion i.e., track the position and the behavior of insects.
- Ashok Veeraraghavan spending 10 weeks in ANU.

Anatomical Modeling

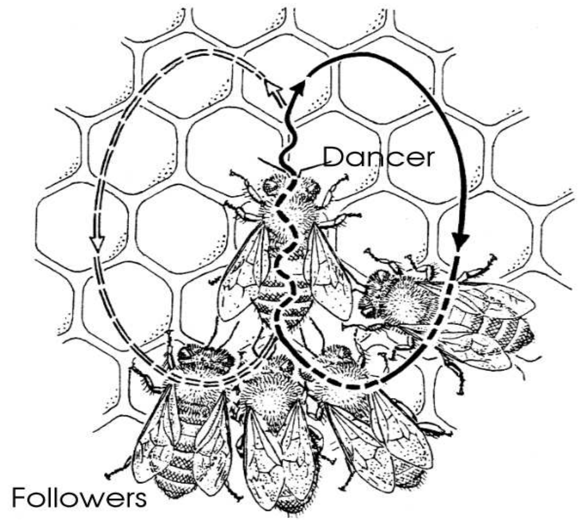
- All insects have similar anatomy.
- Hard Exoskeleton, soft interior.
- Three major body parts- Head, Thorax and abdomen.
- Each body part modeled as an ellipse.
- Anatomical modeling ensures
  - Physical limits of body parts are consistent.
  - Accounts for structural limitations.
  - Accounts for correlation among orientation of body parts
  - Insects move in the direction of their head.

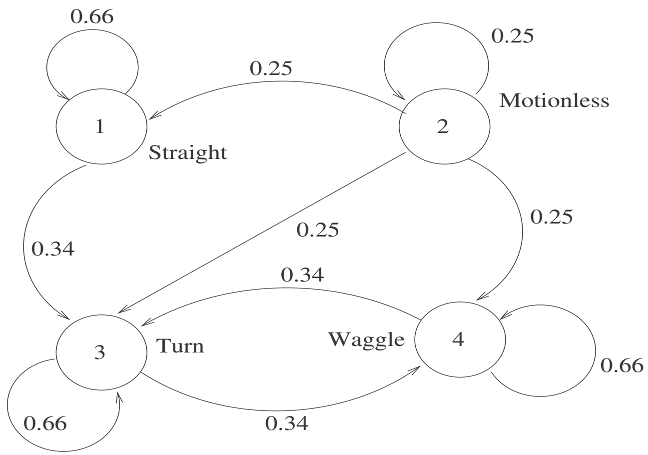




Waggle Dance

- Foragers perform waggle dance.
- Orientation of waggle axis
  Direction of Food source.
- Intensity of waggle dance
  Sweetness of food source.
- Frequency of waggle
  Distance of food source.
- Recruits follow the dancer.
- Behavior Modeling:
  Markov Model on basic motions.





Behavior-Encoded Tracking