Online Selection of Discriminative Tracking Features

(Robert T. Collins and Yanxi Liu)

ICCV2003

Presented by:

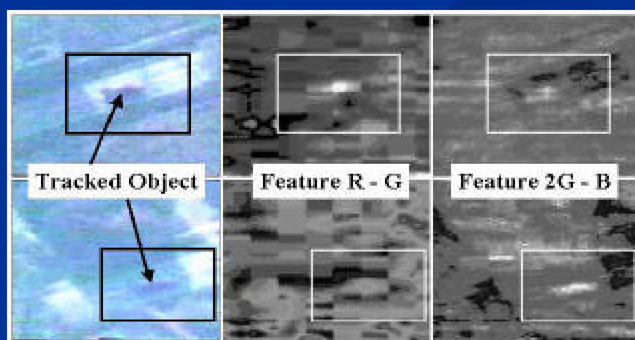
Asaad Hakeem

Main Idea

- Have a large set of features
- Select features that best discriminate object and background
- Use these features for tracking
- Re-evaluate features over time to adapt to changing appearances
- Contribution:
  Use best discriminating feature subset from a large pool of features, instead of a fixed number of features

Motivation

- Fixed set of features may not always work for tracking
- Example of tracking objects in:
  - Sunlight versus shadow



Feature Selection

- Choose *m* features from *n* candidates (m<<n)
- Rule out redundant features to improve classification
- Selection Criteria
  - Compare one feature subset against another
  - Select best feature subset that discriminates between object and background

Feature Spaces

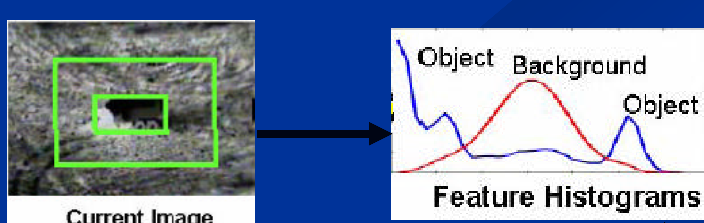
- Features for tracking include color, texture, shape, motion etc.
- Using RGB histograms for target appearance in local window
- Candidate features composed of linear combinations of RGB

$$F_1 \equiv \{w_1 R + w_2 G + w_3 B \mid w_* \in [-2,-1,0,1,2]\}$$

- Total of 5^3 initial candidates pruned (scale, zero) to 49 features
- Candidate features cover
  - Intensity R+G+B
  - Chrominance R-B
  - Excess color 2G-R-B
- Features are:
  - Normalized to 0-255 color range
  - 32-64 histogram buckets

Evaluating Feature Discriminability

- Evaluate which features yield good class separability between object and background
- Probability density calculated for:
  - Object *p(i)*: compact set of pixels in rectangle
  - Background *q(i)*: larger ring of pixels surrounding the object



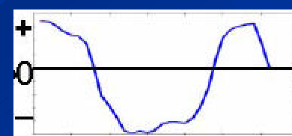
- Histograms *p(i)* and *q(i)* are normalized by number of their elements

Log Likelihood

- The log likelihood of feature *i* is calculated by:

$$L(i) = \log \frac{\max\{p(i), d\}}{\max\{q(i), d\}}, d = 0.001$$

- Non-linear log likelihood maps multimodal distributions into:
  - Positive values for objects
  - Negative values for background
  - Zeros for pixels belonging to both classes



Log Likelihood Ratio

- This log likelihood mapping becomes likelihood image
- Variance for a distribution *a(i)* is calculated from *L(i)* by:

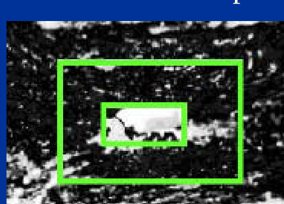
$$\mathrm{var}(L; a) = \sum_i a(i) L^2(i) - \left[\sum_i a(i) L(i) \right]^2$$

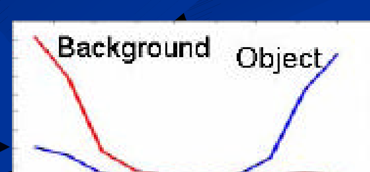
Variance Ratio

- The variance ratio of log likelihood function is given by:

$$VR(L; p, q) \equiv \frac{\mathrm{var}(L; (p+q)/2)}{[\mathrm{var}(L; p) + \mathrm{var}(L; q)]}$$

- Variance ratio will have:
  - Object and background pixels tightly clustered (denominator)
  - Both clusters spread apart (high total variance - numerator)

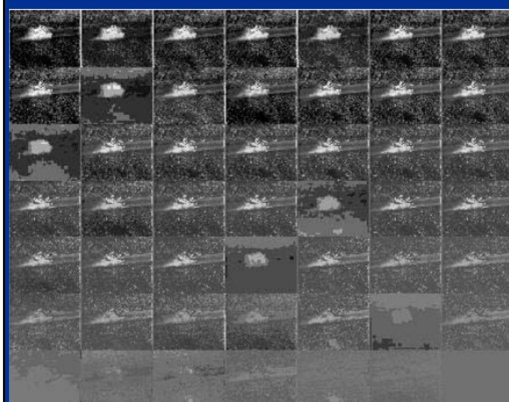

Likelihood Image

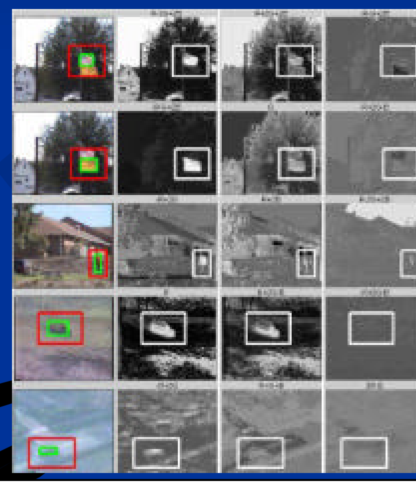


Likelihood Histograms

Ranked Likelihood Images

- Rank-order of features is based on the two-class variance ratio method
- Likelihood images for all feature spaces are shown below:





Overview of Feature Selection

- Given an appearance model from previous view:
  - Compute object and background feature value distribution
  - Candidate features are rank-ordered by measuring separability
  - Discriminative features create likelihood maps
    - High values for object pixels
    - Low values for background pixels

Tracking

- Object and background distributions are calculated in current frame, given the location of tracked object
- Top *N* discriminative features (N=5) are used to compute likelihood images for next frame
- Local mean-shift process is initialized in each *N* new likelihood images
- These perform gradient ascent to find nearest local mode
- Compute median of *N* estimates to find new object location
- This algorithm is applied at each iteration with *N* best discriminating features

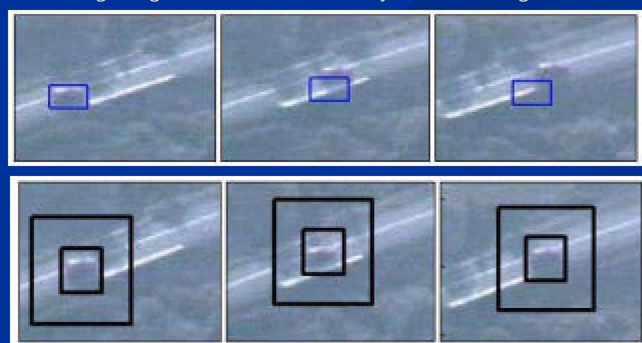
Model Drift

- The bounding box drifts from object location due to noise induced by:
  - Misclassified background pixels labeled as object
  - Adaptive updating of appearance model
- Leads to further misclassification and tracking failure

- Solution: (constrained)
  - Combines current and original object density (first frame histogram) as the object histogram
  - Assumes that the object appearance will not change drastically

Mean Shift Tracking

- In the above figure, the mean shift tracking fails as the car goes over the similar colored bridge
- The adaptive tracker solves this problem by:
  - Keeping object and background color distribution models
  - Down-weighting color common in object and background

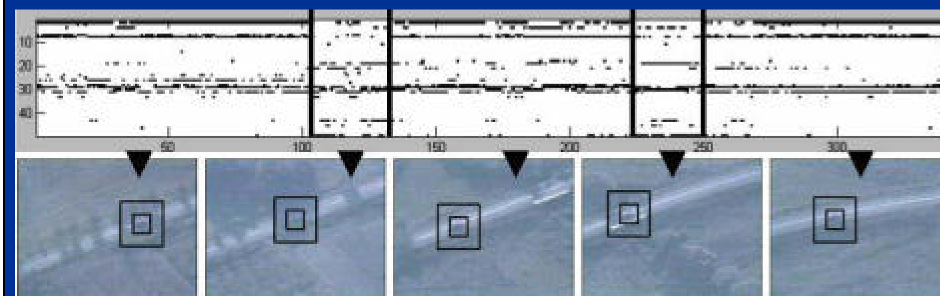


Overview of Tracking

- Continuous evaluation and updating of feature sets
- Select locally discriminative feature sets
- Use best local feature set at each time for Mean Shift tracking
- Less restrictive than trackers with fixed feature set
  - As they require global discriminative features

Experiments

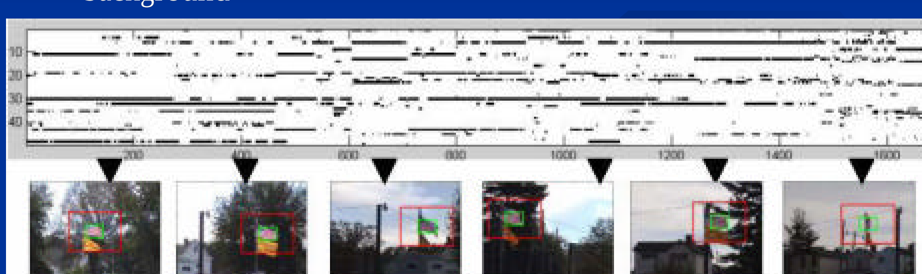
- First video is low-contrast aerial footage of car driving through sunny and shadowy patches



- Horizontal bars are features chosen repeatedly
- Discontinuity between feature space (set of features) is marked with an arrow

Experiments (contd.)

- Second video is a flag blowing non-rigidly in the wind
- Camera viewpoint continually changes, causing varying background



- Here discontinuity is difficult to mark, therefore sample frames from the sequence are shown

Conclusion

- Fixed set of features may not always work for tracking
- Have a large set of features
- Select features that best discriminate object and background
- Use these features for tracking
- Re-evaluate features over time to adapt to changing appearances