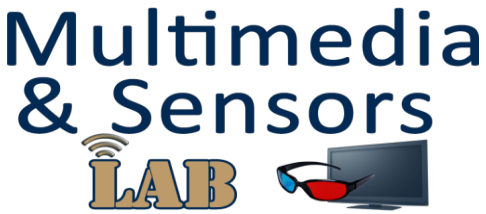


Unsupervised Estimation of Uncertainty for Video Saliency Detection Using Temporal Cues

Tariq Alshawi*, Zhiling Long, and Ghassan AlRegib

Multimedia and Sensors Lab (MSL)

Center for Signal and Information Processing (CSIP)

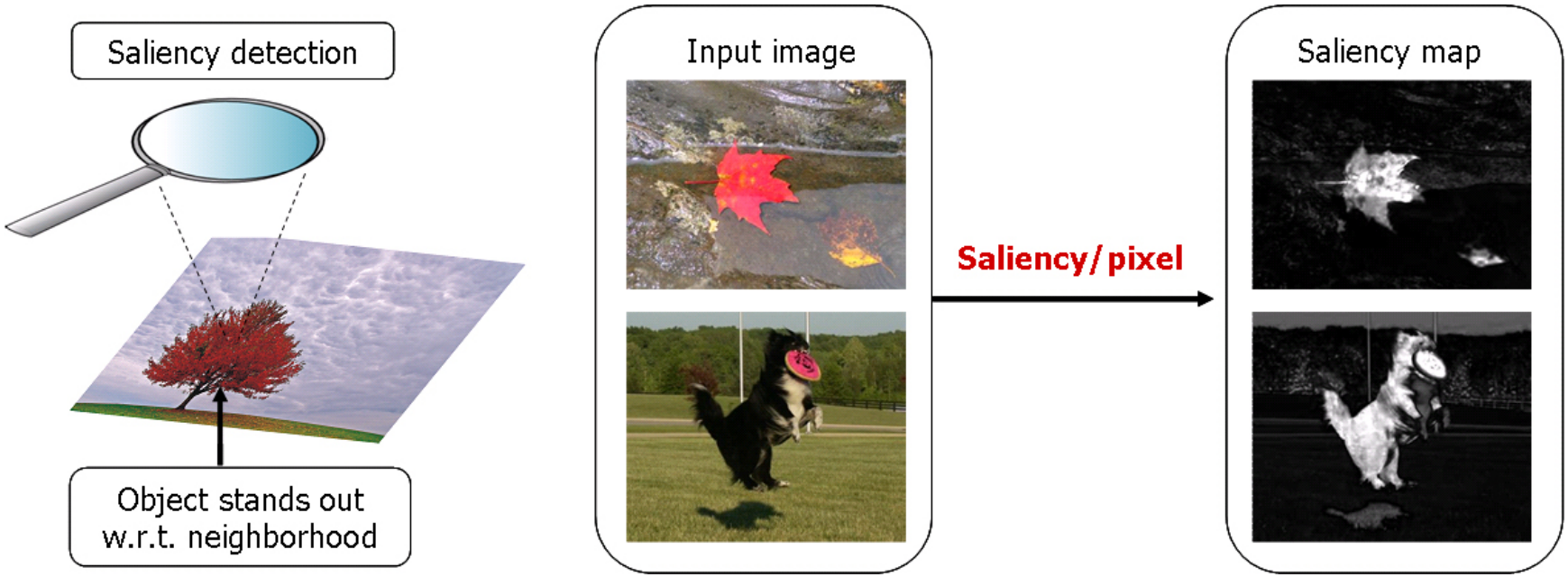
School of Electrical and Computer Engineering

Georgia Institute of Technology

Outline

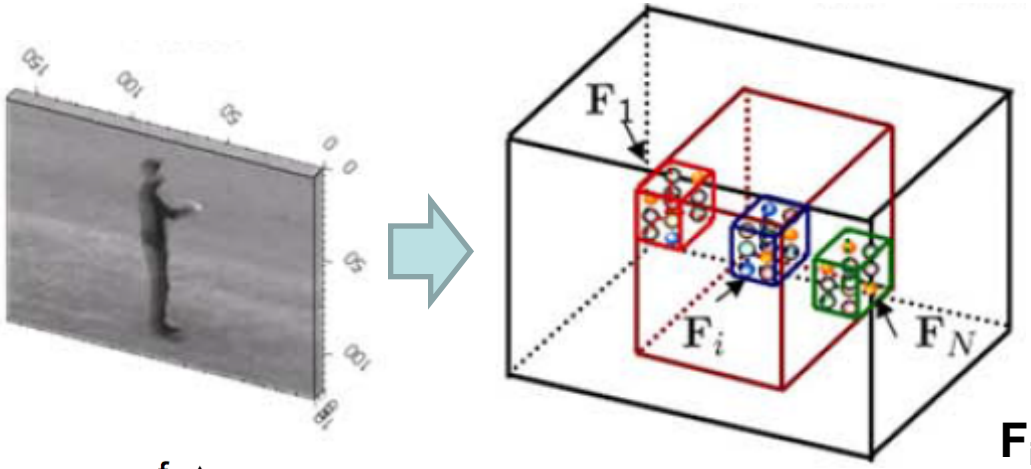
1. **Introduction to Saliency**
   - Motivation
   - 3D FFT
2. **Saliency Fusion Technique**
   - Classical Methods
   - Uncertainty-based Fusion
3. **Proposed Method**
   - Algorithm
   - Evaluation
   - Results
4. **Conclusions**

Introduction to Saliency: Motivation



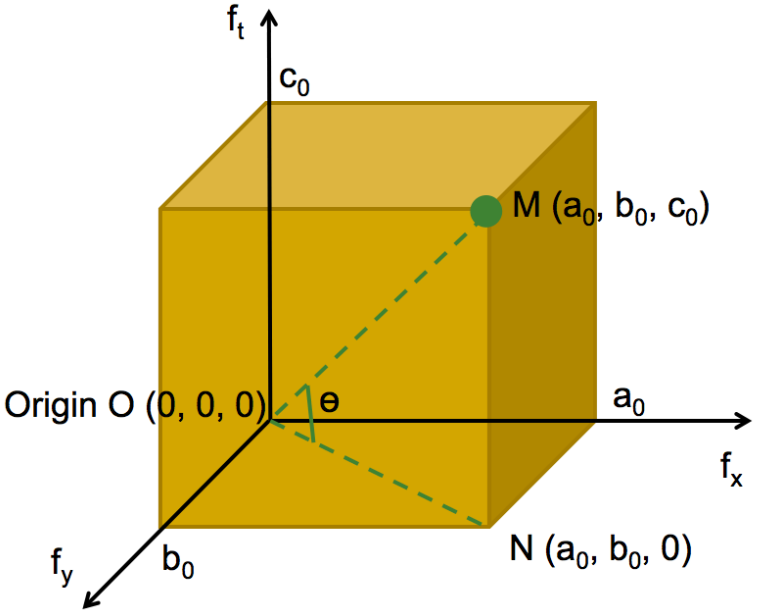
(Diagram from http://ivrgwww.epfl.ch/supplementary_material/RK_CVPR09)

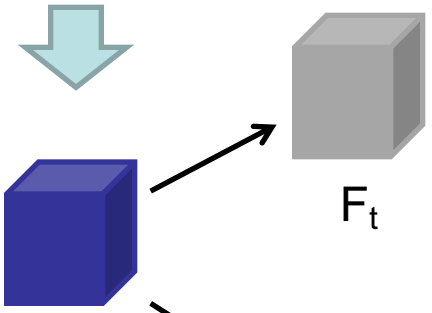
Introduction to Saliency: 3D FFT method



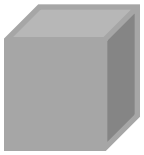
Work published in in Human Vision and Electronic Imaging XX, SPIE Electronic Imaging SPIE, 2015.

F_i: FFT Local Spectrum



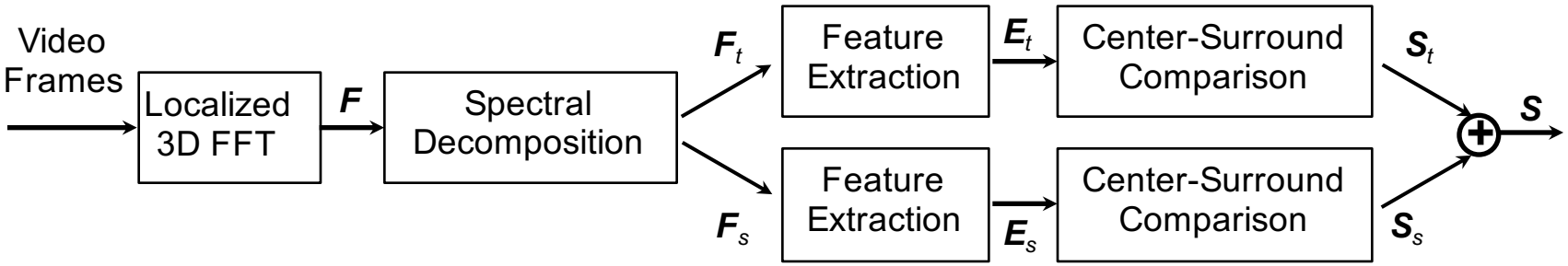


$F_t(a_0, b_0, c_0) =$
$F(a_0, b_0, c_0)$ x MN / OM

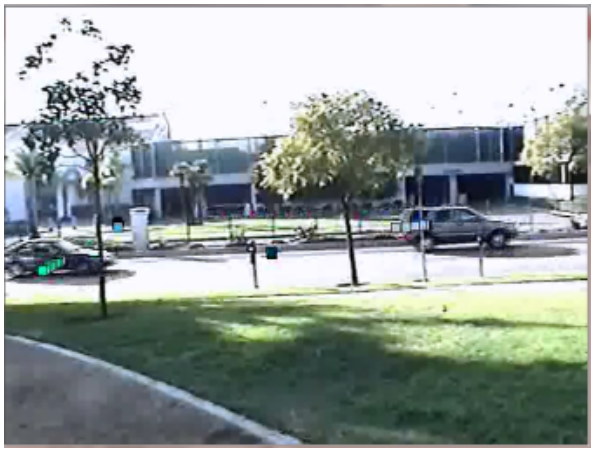


$F_s(a_0, b_0, c_0) =$
$F(a_0, b_0, c_0)$ x ON / OM

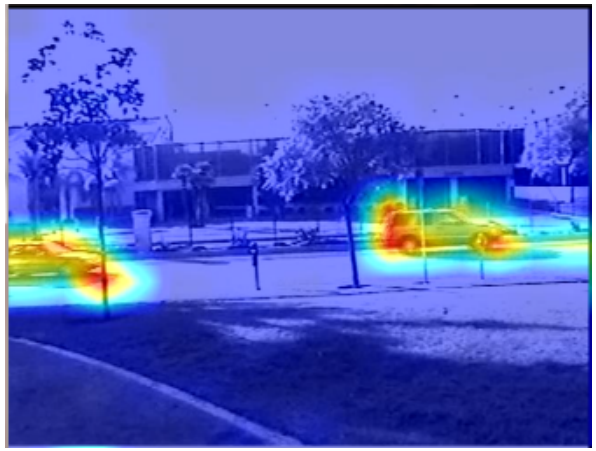
Introduction to Saliency: 3D FFT method



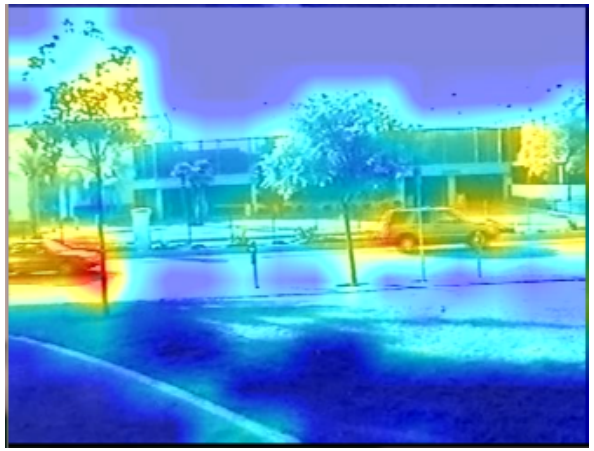
original Video



F_t energy distribution



F_s energy distribution

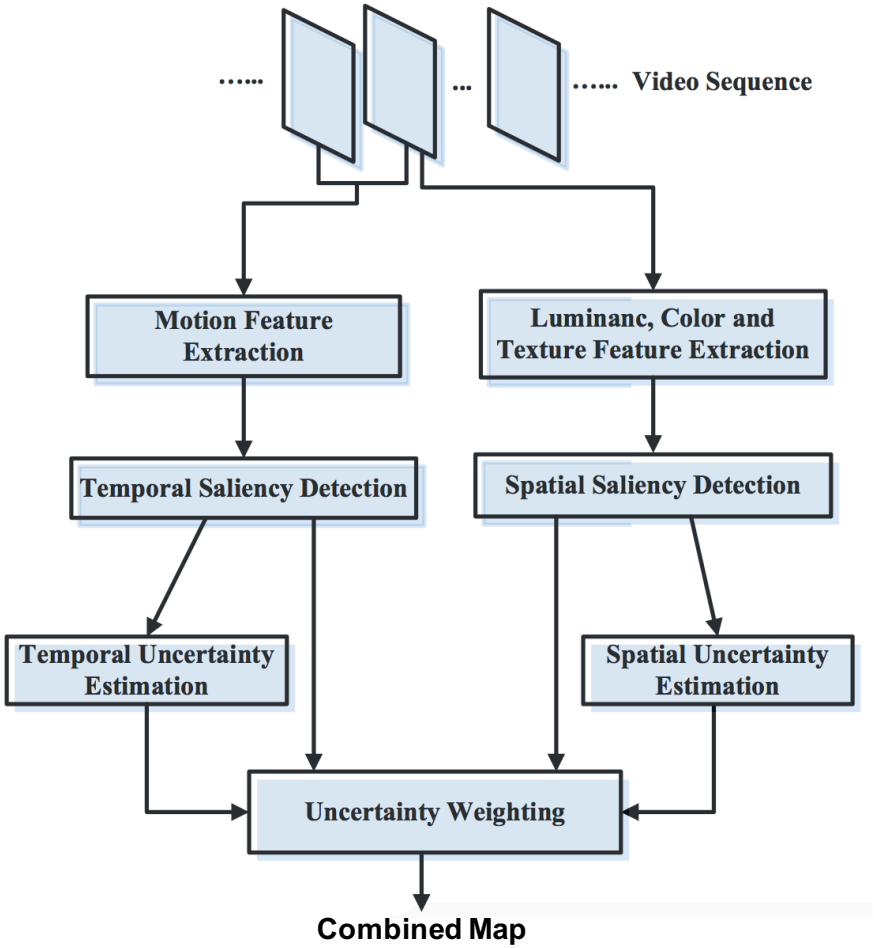


Saliency Fusion Techniques: Classical Methods

Mean Fusion	$M_F = (M_S + M_T)/2$
Max Fusion	$M_F = max(M_S, M_T)$
Multiplication Fusion	$M_F = M_S \times M_T$
Maximum Skewness Fusion	$M_F = \alpha M_S + \beta M_T + \gamma(M_S \times M_T)$
Binary Threshold Fusion	$M_F = max(M_S, M_T \cap M_B)$
Motion Priority Fusion	$M_F = (1 - \alpha)M_S + \alpha M_T$

Muddamsetty et al. "A Performance Evaluation of Fusion Techniques for Spatio-Temporal Saliency Detection in Dynamic Scenes," in ICIP 2013, pp. 1-5

Saliency Fusion Techniques: Uncertainty-based Fusion



$$M_F = \frac{U_T M_S + U_S M_T}{U_S + U_T}$$

Distance from center of mass

$$U^d = H_b(p(s|d))$$

$$p(s|d) = \alpha_1 \exp\left[-\left(\frac{d}{\beta_1}\right)^{\gamma_1}\right]$$

Measure of Connectedness

$$U^c = H_b(p(s|c))$$

$$p(s|c) = 1 - \exp\left[-\left(\frac{c}{\beta_2}\right)^{\gamma_2}\right]$$

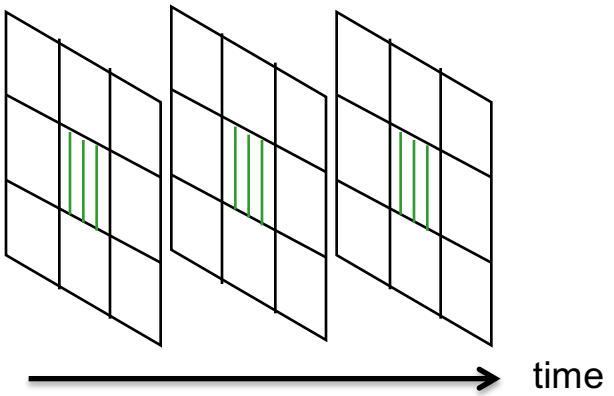
Fang et al. "Video Saliency Incorporating spatiotemporal Cues and Uncertainty Weighting," in ICME 2013, pp. 1-6

Proposed Method: Algorithm

$$\boldsymbol{S} = \begin{bmatrix} S_{11}[k] & S_{12}[k] & \dots & S_{1N}[k] \\ S_{21}[k] & S_{22}[k] & \dots & S_{2N}[k] \\ \vdots & \vdots & \ddots & \vdots \\ S_{M1}[k] & S_{M2}[k] & \dots & S_{MN}[k] \end{bmatrix}$$

$$\boldsymbol{U} = \begin{bmatrix} U_{11}[k] & U_{12}[k] & \dots & U_{1N}[k] \\ U_{21}[k] & U_{22}[k] & \dots & U_{2N}[k] \\ \vdots & \vdots & \ddots & \vdots \\ U_{M1}[k] & U_{M2}[k] & \dots & U_{MN}[k] \end{bmatrix}$$

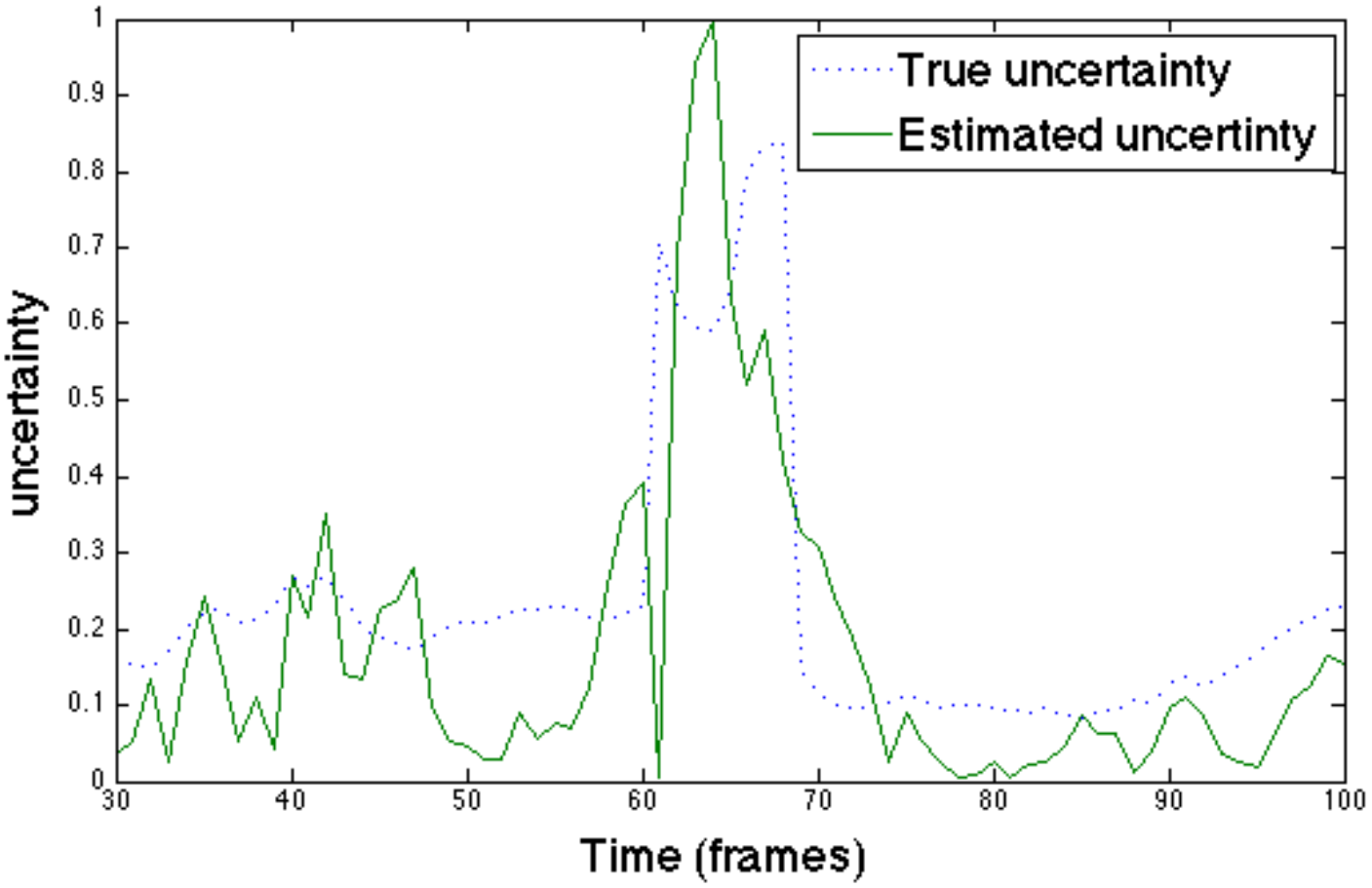
$S_{mn}[k]$



$$U_{mn}[k] = |S_{mn}[k] - W_{mn}^{(L)}[k]|$$

where $$W_{mn}^{(L)}[k] = \frac{1}{L} \sum_{i=k-\frac{L}{2}}^{k+\frac{L}{2}} S_{mn}[i]$$

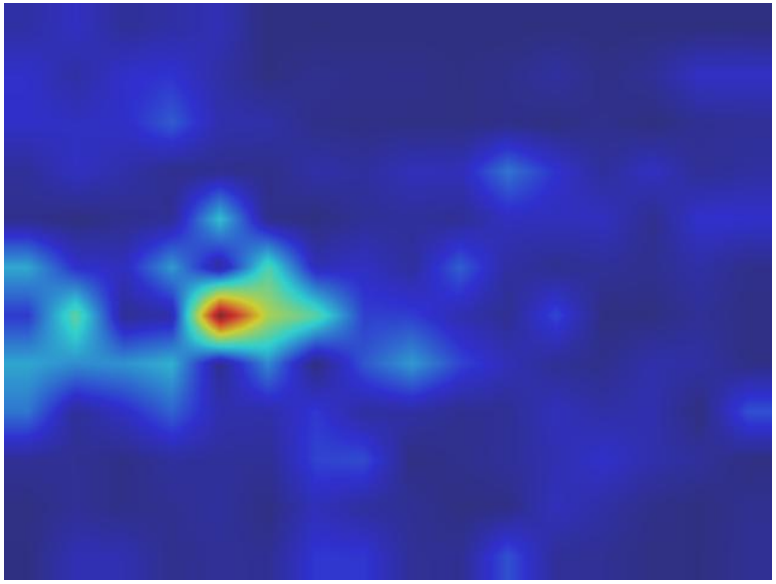
Proposed Method: Qualitative Results



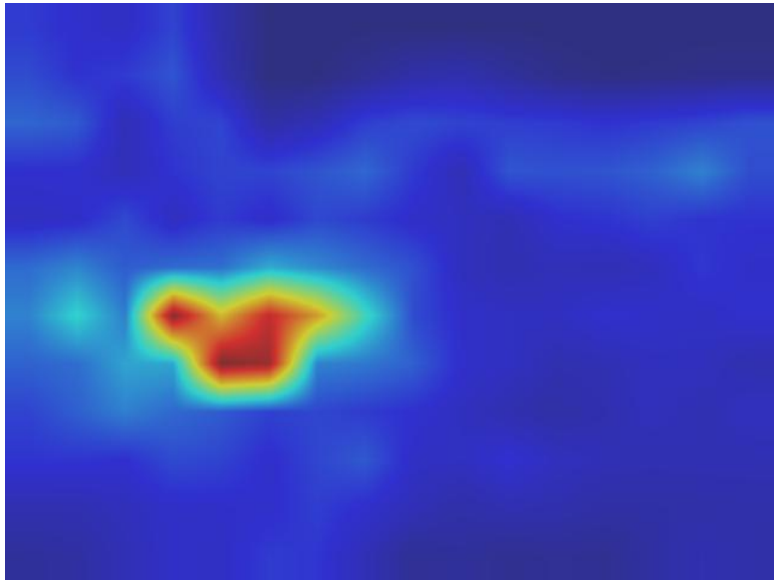
beverly01, pixel location (8;6) frame 30 to 100

Proposed Method: Qualitative Results

Estimated Uncertainty



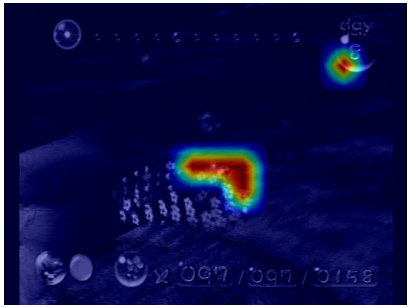
True Uncertainty



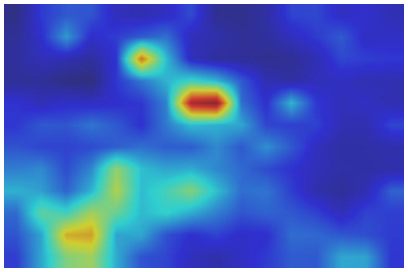
beverly05, frame 187

Proposed Performance Evaluation

Expanded Eye-fixation map

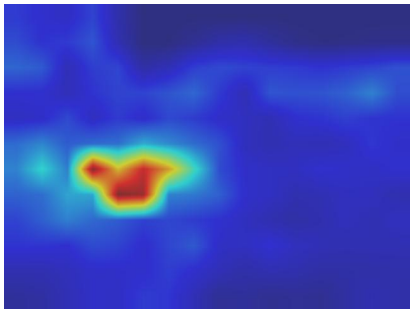


Saliency Map



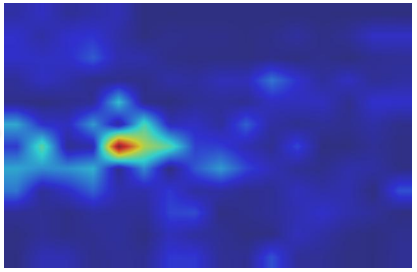
Uncertainty Estimation

True Uncertainty



Fixed Threshold

Estimated Uncertainty

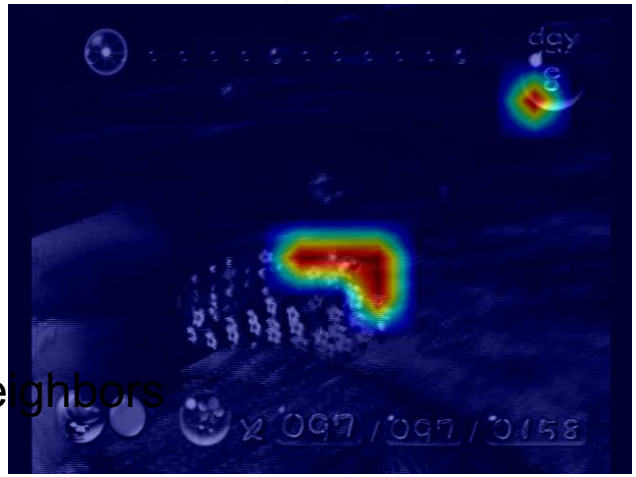


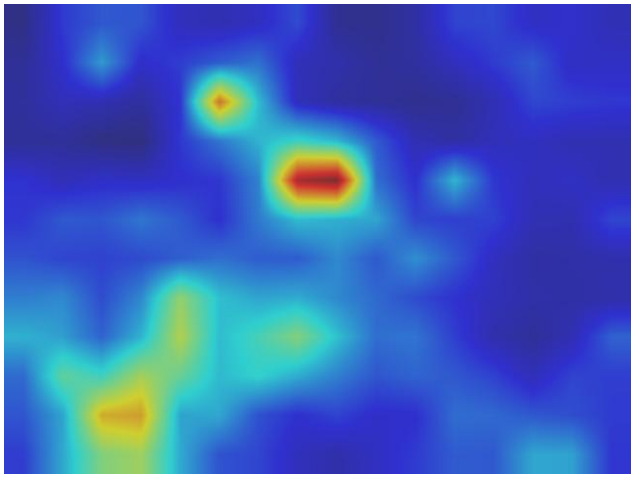
Receiver Operating Characteristics (ROC)

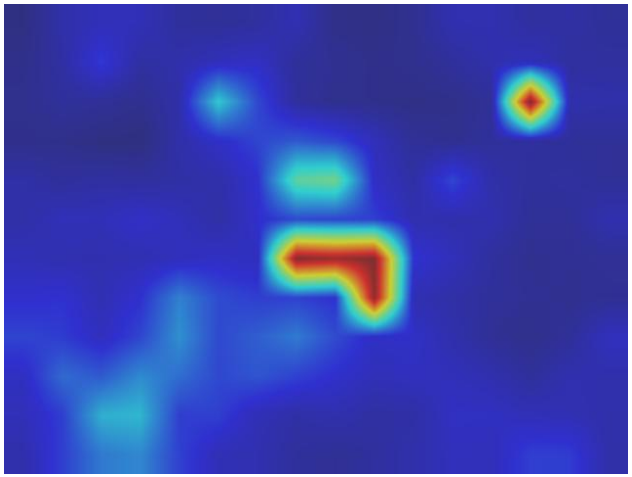
Proposed Evaluation: Example

Original Frame

Expanded Eye fixation map







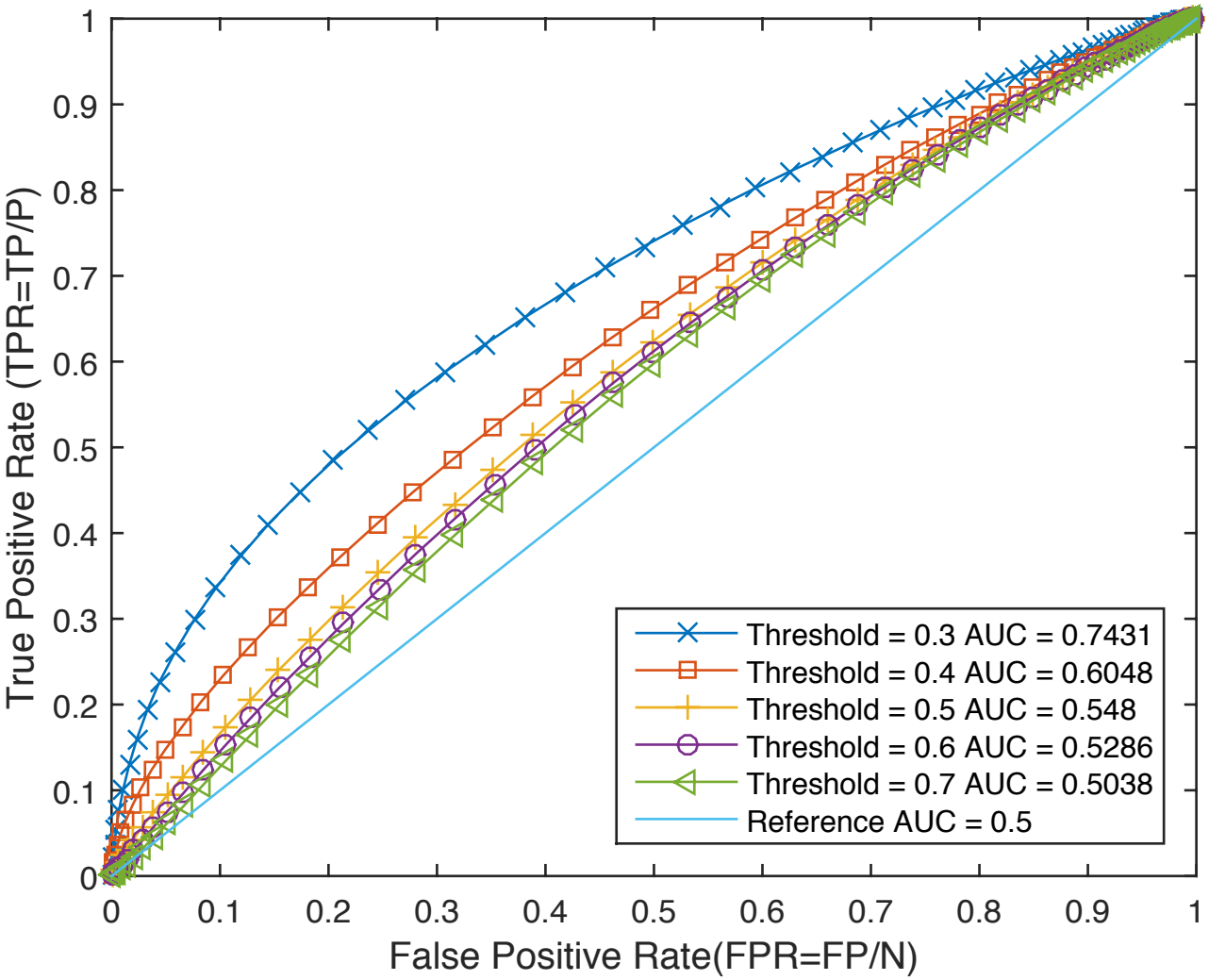
Saliency Map

True Uncertainty

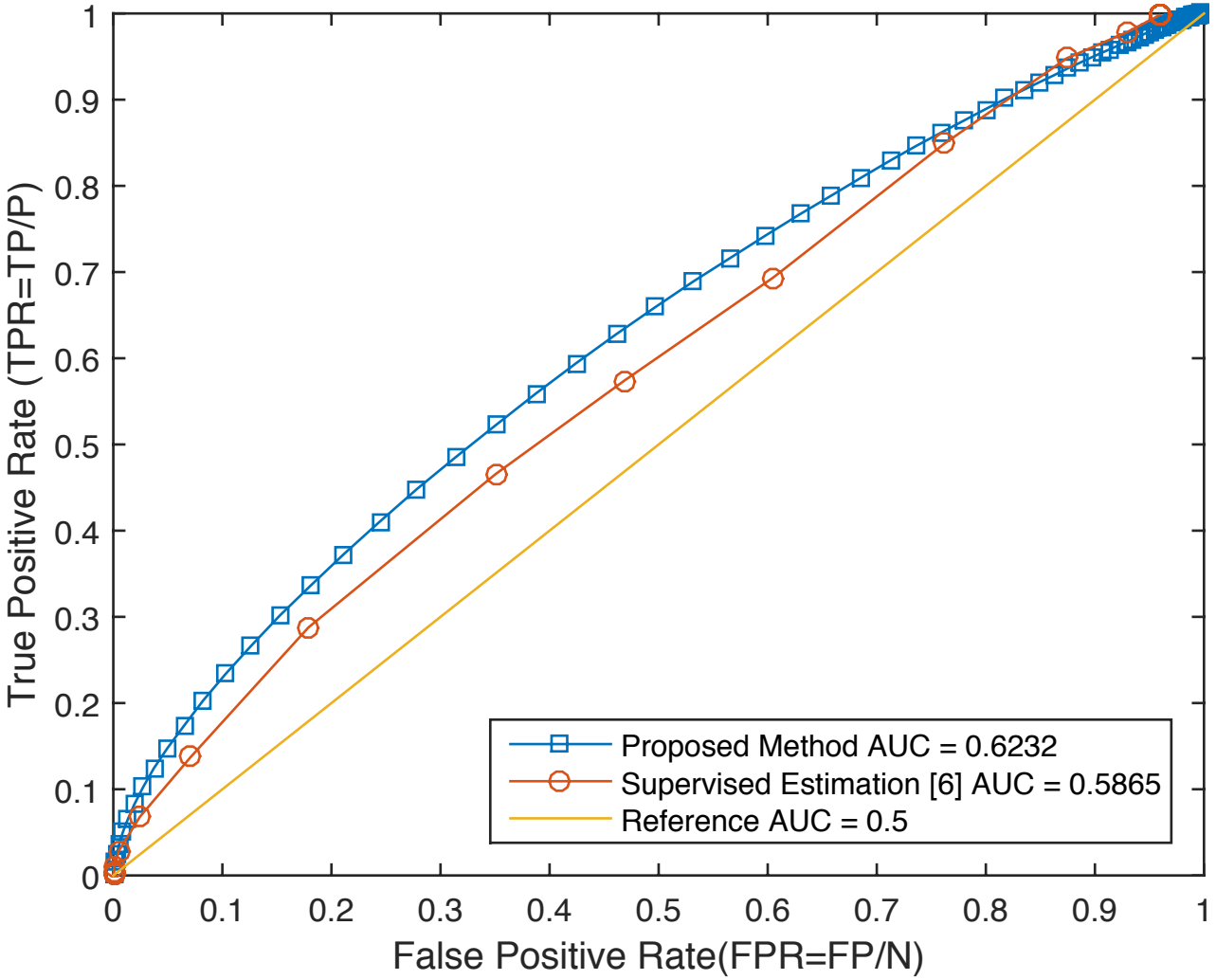
Dataset

- Collaborative Research Cognitive NeuroScience (CRCNS) dataset
- 50 video clips, 5-90 seconds
- Street scenes, TV sports, TV news, TV talks, video games, etc.
- Ground truth by human subjects (eye tracking)

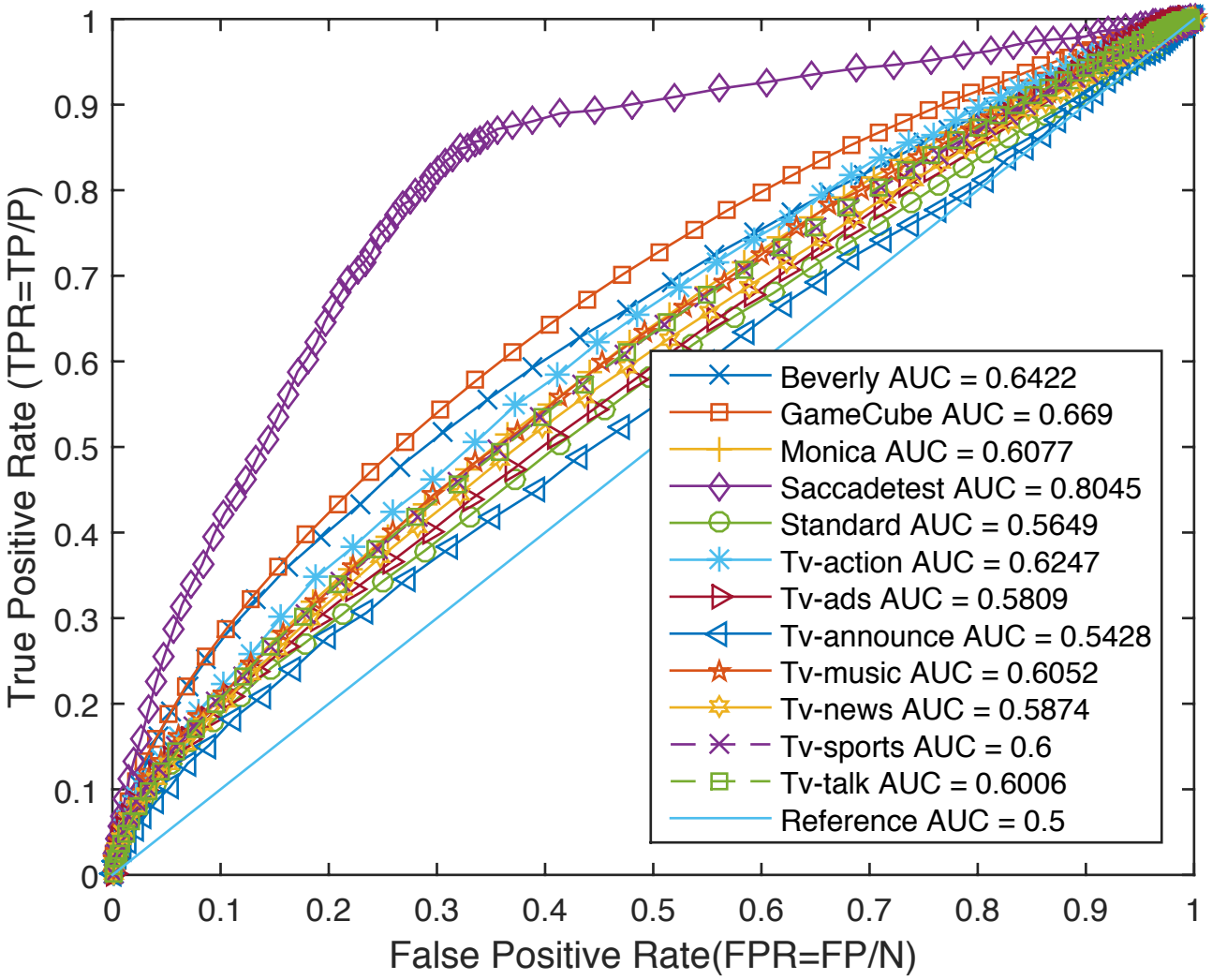
Proposed Method: Results



Proposed Method: Results

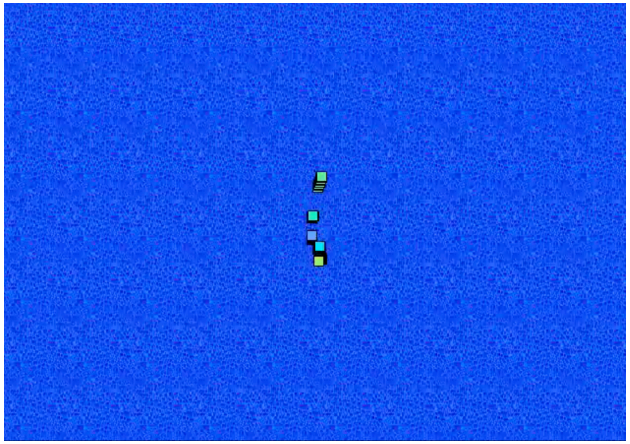


Proposed Method: Results (cont.)



Proposed Method: Results (cont.)

saccadetest



Semantically non-complex

gamecube02

Center-Bias

tv-news03

Semantically complex

Conclusions

- Exploit saliency's temporal correlation for unsupervised uncertainty estimation
- Computationally efficient; real-world applications
- True uncertainty generation by relying on discrepancy between saliency map and eye-fixation map
- Direct performance evaluation for uncertainty estimation using true uncertainty and ROC analysis
- Reasonably good results on CNCRS dataset
- Video content plays a significant role in temporal correlation of saliency maps.

Questions?