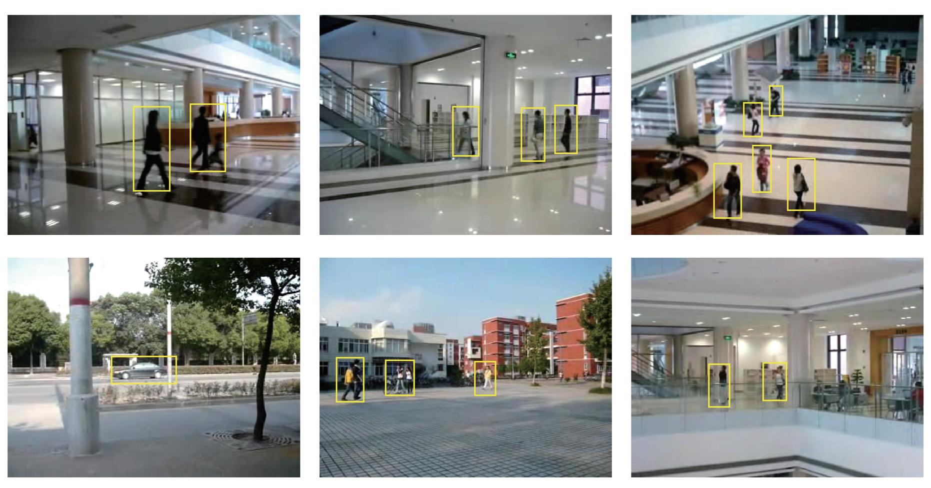
A Phase Discrephancy Analysis of Object Motion

Bolei Zhou Xiaodi Hou Liqing Zhang

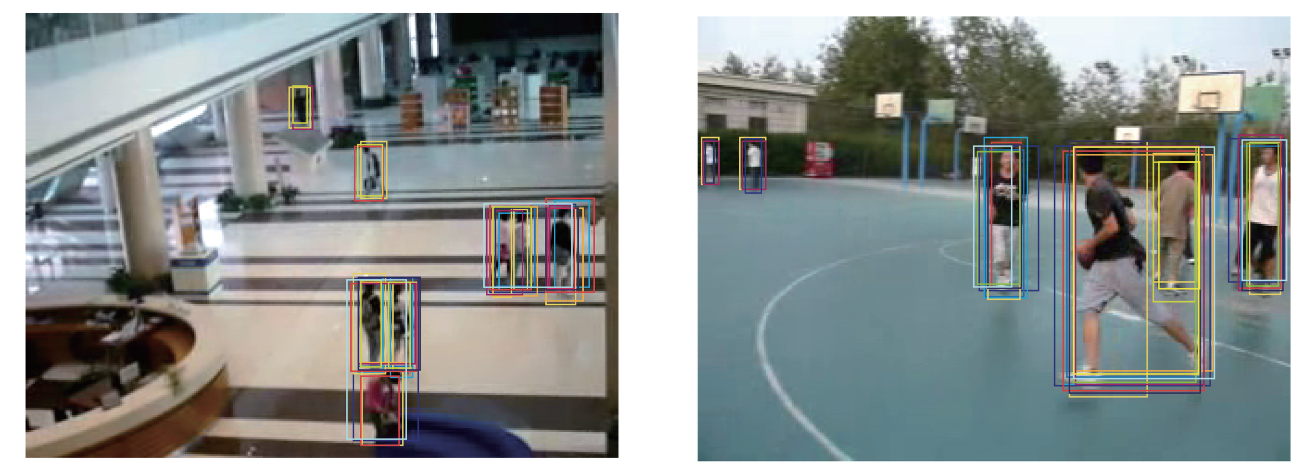
Center for Brain-Like Computing and Machine Intelligence,
Shanghai Jiao Tong University, CUHK and Caltech

Psychophysical experiments: human motion perception in natural scenes

11 subjects is instructed to annotate 'moving objects' in 20 clips of video



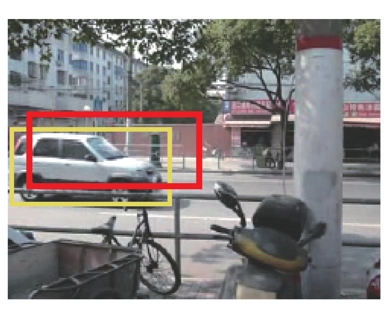
Different interpretations of moving objects by different subjects.



Statistical analysis on subjects' labelings

*i*th subject's labeling as Ground-Truth (yellow box)

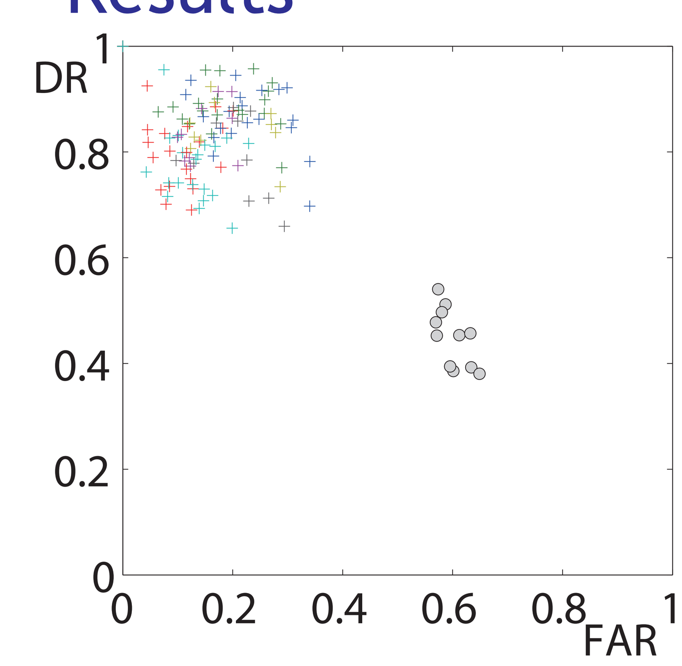
*j*th subject's labeling as test (red box)

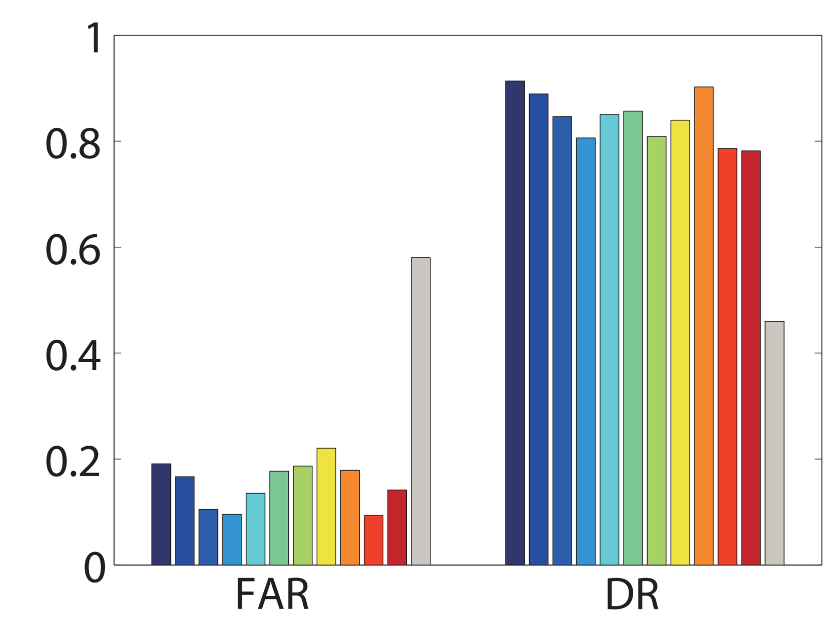


$\frac{\text{area(GT)} \cap \text{area(test)}}{\text{area(GT)} \cup \text{area(test)}} > 0.5$ as TP(true positive), otherwise as FP(false positive)

DR(detection rate)=TP/GT
FAR(false alarm rate)=FP/(TP+FP)

Results





DR=0.84±0.08 FAR=0.15±0.08

THEN, the labelings are used to construct a new benchmark for evaluation of motion detection algorithm

A simple model for motion detection

Aim:
to detect moving objects in natural scene under moving camera condition.

1

1,The intensity constancy assumption: $f(\mathbf{x},t) = f(\mathbf{x}+\mathbf{v},t+1)$.

2,Fourier translation property: $F_{\mathbf{x}+\mathbf{v},t+1}(\boldsymbol{\omega}) = F_{\mathbf{x},t}(\boldsymbol{\omega})e^{-i\cdot\Phi(\mathbf{v})}$

Goal: To approximate the phase discrepancy of background $\tilde{\Phi}(\mathbf{v})$ as the phase discrepancy of the image $\Phi(\mathbf{v})$

Then,

$$\tilde{F}_{t+1}(\boldsymbol{\omega}) = F_t(\boldsymbol{\omega})e^{-i\cdot\tilde{\Phi}(\mathbf{v})} = |F_t(\boldsymbol{\omega})|\cdot e^{-i[\angle F_t(\boldsymbol{\omega})+\tilde{\Phi}(\mathbf{v})]} = |F_t(\boldsymbol{\omega})|\cdot e^{-i[\angle F_{t+1}(\boldsymbol{\omega})]}$$

the motion saliency map has the simple form:

$$s(\mathbf{x},t) = \left\{\mathscr{F}^{-1}[F_{t+1}(\boldsymbol{\omega})] - \mathscr{F}^{-1}[\tilde{F}_{t+1}(\boldsymbol{\omega})]\right\}^2 = \left\{\mathscr{F}^{-1}\left[(|F_{t+1}(\boldsymbol{\omega})|-|F_t(\boldsymbol{\omega})|)\cdot e^{-i\angle F_{t+1}(\boldsymbol{\omega})}\right]\right\}^2$$

The upper bound of error in $\tilde{\Phi}(\mathbf{v})$ is:

$$\max\left[\Phi(\mathbf{v}) - \tilde{\Phi}(\mathbf{v})\right] = \max\left\{\tan^{-1}\left[E(|\eta|)\right]\right\} \approx 0.31. \qquad (1)$$

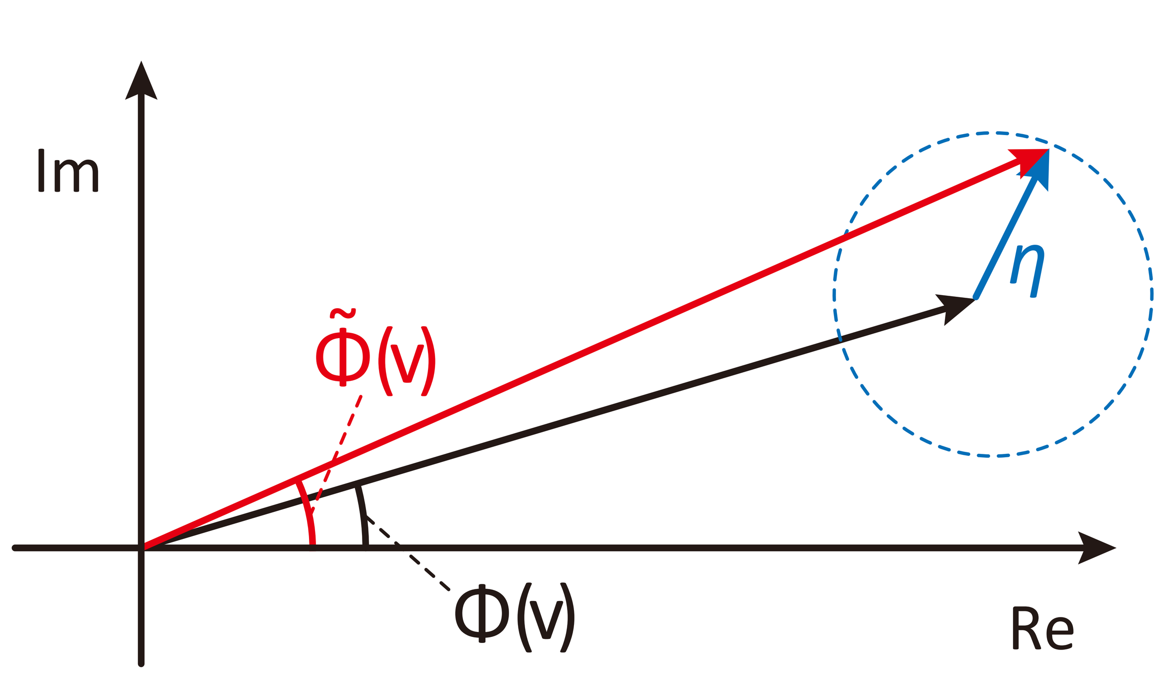
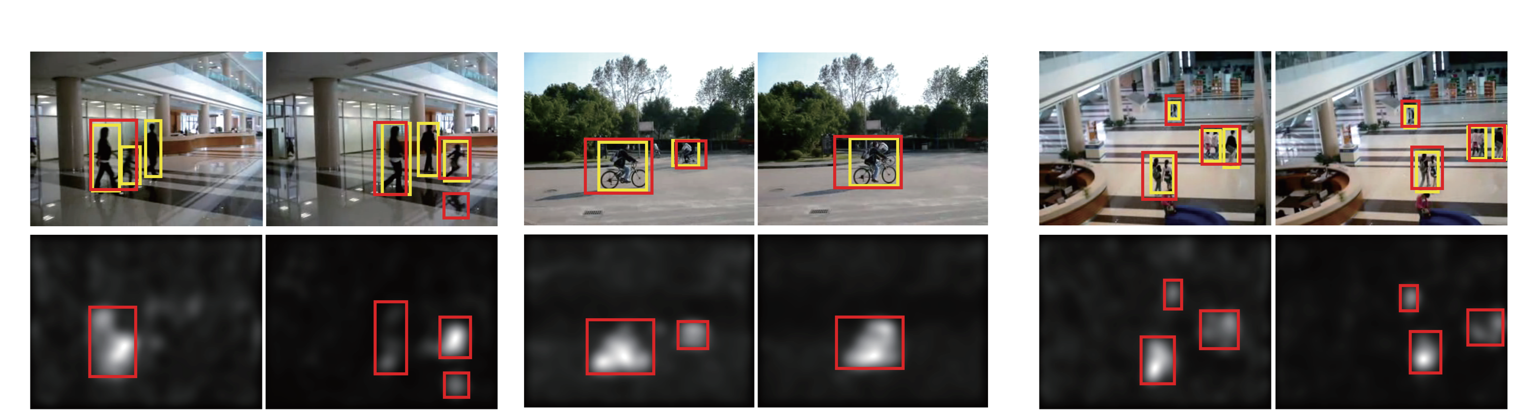


Figure 1: A diagram of the angular error calculation. Given $E(|\eta|) = \sqrt{0.1}$, the upper bound of the angular error is 0.31 (17.6°), the mean angular error is 0.21 (12.3°)

Basic implementation: 9 lines of Matlab code

```
1 FFT1=fft2(Frame1);
2 FFT2=fft2(Frame2);
3 Amp1=abs(FFT1);
4 Amp2=abs(FFT2);
5 Phase1=angle(FFT1);
6 Phase2=angle(FFT2);
7 mMap1=abs(ifft2((Amp2-Amp1).*exp(i*Phase1)));
8 mMap2=abs(ifft2((Amp2-Amp1).*exp(i*Phase2)));
9 mMap=mat2gray(mMap1.*mMap2);
```

Qualitative and quantitative result



	Detection Rate	False Alarm Rate
Human average	0.84±0.08	0.15 ± 0.08
Our model	0.46±0.14	0.58 ± 0.24
Dynamic Visual Attention	0.32±0.22	0.86±0.10
Bayesian Surprise	0.12±0.09	0.92±0.04
Saliency	0.09±0.08	0.98±0.01
Mixture of Gaussian	0.00±0.00	1.00±0.00

Full analysis is included in our submission to ACCV2010

zhoubolei@gmail.com http://bcmi.sjtu.edu.cn