What, where and who? Classifying events by scene and object recognition

http://vision.cs.princeton.edu

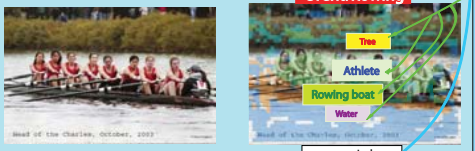
Li-Jia Li[1]
jiali3@uiuc.edu

Li Fei-Fei[2]
feifeili@cs.princeton.edu

University of Illinois Urbana-Champaign, USA[1]

Princeton University, USA[2]

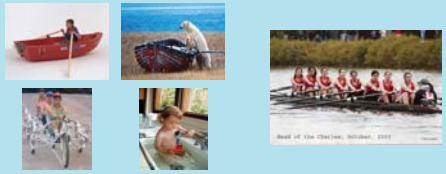
Goal



Giving an unknown event image:

- Classify the event
- Recognize the scene environment
- Provide a number of semantic labels to the objects

Definition of Event



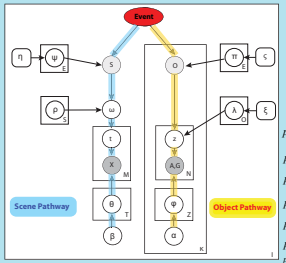
Non-Event

Event: Rowing

Event:

A semantically meaningful human activity, taking place within a selected environment and containing a number of necessary objects

Graphical Model



E - Event
O - Objects
S - Scene
A - Apperance (Object)
G - Geometry/Layout
X - Appearance (Scene)
z - Object topic
t - Scene topic

$$p(S \mid E,\Psi) = Mult(S \mid E,\Psi)$$

$$p(\omega \mid S,\rho) = Dir(\omega \mid \rho_j), S = j$$

$$p(t_m \mid \omega) = Mult(t_m \mid \omega)$$

$$p(X_m \mid t,\theta) = p(X_m \mid \theta_j), t_m = j$$

$$p(O \mid E,\pi) = Mult(O \mid E,\pi)$$

$$p(z_n \mid \lambda, O) = Mult(z_n \mid \lambda, O)$$

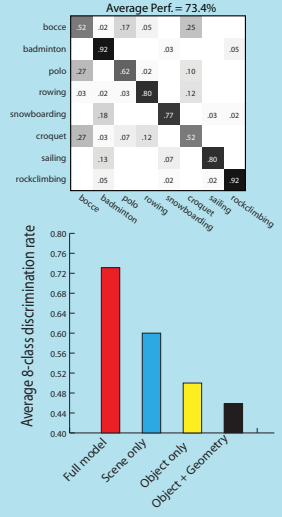
$$p(A_n, G_n \mid z,\Phi) = p(A_n, G_n \mid \Phi_j), z_n = j$$

$$p(E,S,\mathbf{O},\mathbf{X},\mathbf{A},\mathbf{G},\mathbf{t},\mathbf{z},\omega \mid \rho,\varphi,\lambda,\psi,\pi,\theta) = p(E)\cdot P(S \mid E,\psi)\cdot P(\omega \mid S,\rho)$$

$$\prod_{m=1}^{M} p(X_m \mid t_m,\theta)\, p(t_m \mid \omega)\cdot \prod_{k=1}^{K} p(O_k \mid E,\pi) \prod_{n=1}^{N} p(A_n, G_n \mid z_n,\Phi) p(z_n \mid \lambda, O_k)$$

Image Interpretation

Quantitative Result



What can you see?

PT = 500ms

Outdoor athletic event. Maybe rugby or soccer. 2 guys competing and pushing into each other. There were some players in the backgound. They were playing on grass with trees in the background and a chain link fence. (Subject: JL)

An outdoor scene. Two men in the foreground, playing some sport. ..(Subject: VT)

PT = 27ms

It looked like it was outside, maybe with one or two people running around in the front playing some kind of game; there may have been trees in the back ground. (Subject: AM)

PT = 53ms

I think I saw a farmer or a gardener in the fields? He had a rake, or some other gardening tool with a long handle. The farmer had his back turned to me, but at a slight angle. I think he had a hat (?) I think he was looking at the ground. Maybe the field was in lines/rows. There were no plants though, just soil. He was wearing long pants. (Subject: EC2)

PT = 107ms

Outside, looked like a closeup of people playing some team sport, but I'm not sure. either there were two people competing with each other in the front and center, or it was an animal, i really don't know. (Subject: LL)

Object | Scene | Event Related

Fei-Fei et al. JoV 2007

Sports Event Dataset

Labeling an Unknown Image

Object Classes - Who

$$p(I \mid O) = \prod_{n=1}^{N} \sum_{j} p(A_n, G_n \mid z_j, O) p(z_j \mid O)$$

Scene Class - Where

$$p(I \mid S,\rho,\theta) = \int p(\omega \mid \rho, S) \left(\prod_{m=1}^{M} \sum_{t_m} p(t_m \mid \omega) \cdot p(X_m \mid t_m,\theta) \right) d\omega$$

Event Class - What

$$p(I \mid E) \propto \sum_{j} p(I \mid O_j) p(O_j \mid E) p(I \mid S) p(S \mid E)$$

Conclusion

- Propose an integrative model
- Offer a rich description of an image
- Benefit multiple applications
- Aspects need to be improved:
  - Inference schemes
  - Relax the amount of supervision
  - More extensive experiment

Reference: L.-J. Li and L. Fei-Fei. What, where and who? Classifying event by scene and object recognition. IEEE International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, 2007.