IS1-15

行動認識とデータマイニングを適用した行動意図推定
Human Intention Inference using Activity Analysis and Data Mining

○Hirokatsu Kataoka† Yoshimitsu Aoki† † Keio University

Outline

►In this presentation, we propose

(i) the "Graphical Model" integrating three types of feature descriptor

(ii) the method to predict "Human's Next Activities"

(iii) In experiment, our approach performs
- 99.1% on Weizmann Dataset
- 99.9% on Videoweb Activity Dataset
- 17.1 fps processing time (tracking, action, intention)
- "Intention Inference" in daily scene

1. Background

►Action recog is the most important topic in CV

The Applications of interest them
Surveillance, Sports, Robotics, HCI, game…

►Related works

Action Recognition

- HOOF : Histograms of Oriented Optical Flow
R.Chaudhry et al ,"Histograms of oriented optical flow and binet caushy kernels on nonlinear dynamical systems for the recognition of human actions",CVPR2009.

- 3D-HOG : Spatio-temporal representation of shape feature
A.Klaser et al ,"A spatio-temporal descriptor based on 3D-gradients ", BMVC2008.

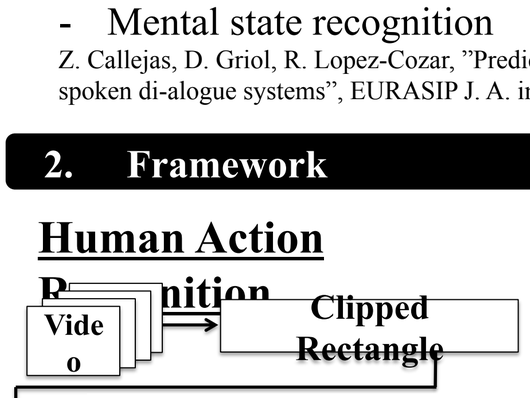
Intention Inference

- Predicting human's location
S. Pellegrini, A. Ess, K. Schindler, L. V. Gool, "You'll Never Walk Alone: Mod- eling Social Behavior for Multi-target Tracking", ICCV2009, 2009

- Mental state recognition
Z. Callejas, D. Griol, R. Lopez-Cozar, "Predicting user mental states in spoken di-alogue systems", EURASIP J. A. in Signal Processing 2011

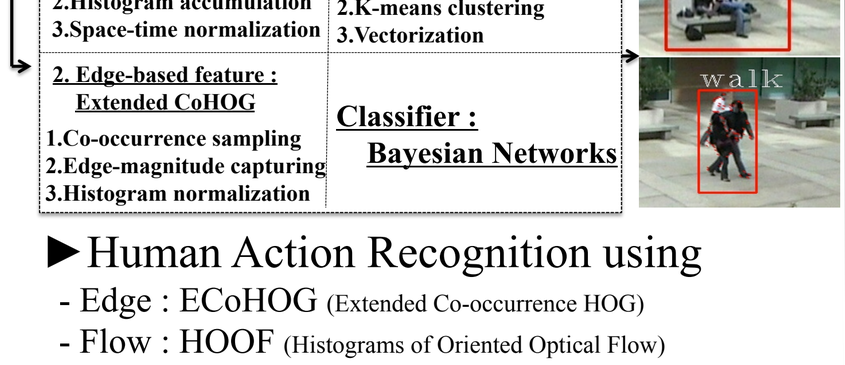
2. Framework

Human Action Recognition



►Human Action Recognition using
- Edge : ECoHOG (Extended Co-occurrence HOG)
- Flow : HOOF (Histograms of Oriented Optical Flow)
- Keypoint : BoF + SURF (Bag-of-Features)

Intention Inference



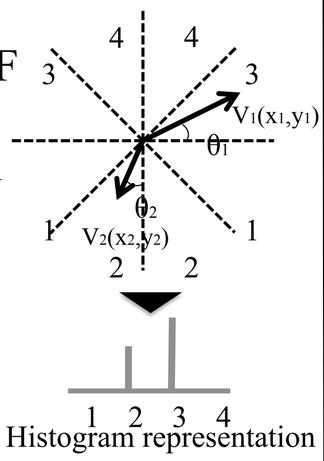
►Intention Inference
The Next Action Estimation

►Intention Inference using Data Mining from
- Time Zone
- Previous Action
- Current Action

3. Action Recognition

►Flow : Normalized HOOF
- Histogram (orientation, strength)
- Normalized by Gaussian function

$$f(x) = \frac{1}{2\pi\sigma^2}\left(-\frac{x^2}{2\sigma^2}\right)$$

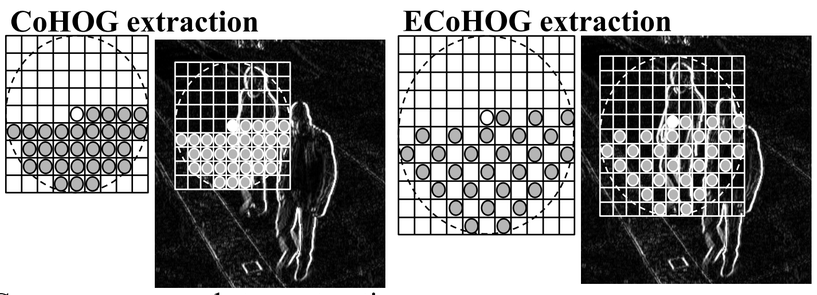


Spatio-temporal normalization using Gaussian function

Gaussian distribution	1	2	3	4	5
	1/16	1/4	3/8	1/4	1/16

NHOOF = 1/16 · H_1 + 1/4 · H_2 + 3/8 · H_3 + 1/4 · H_4 + 1/16 · H_5

►Edge : Extended CoHOG
- Co-occurrence edge orientation representation
- Edge magnitude accumulation and step extraction
- Spatio-temporal histogram from 10 frames



Co-occurrence and step extraction

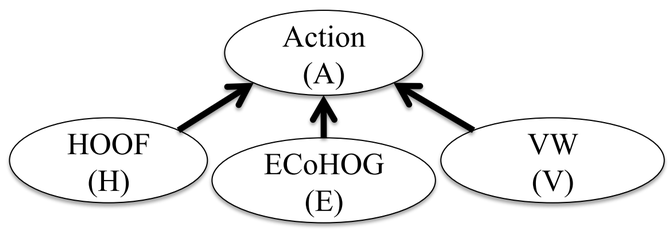
$$C_{x,y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m} \begin{cases} m_1(x_1,y_1) + m_2(x_2,y_2) \\ (if\, d(p,q) = i \\ and\ d(p+x, q+y) = j) \\ \\ 0\ (otherwise) \end{cases}$$

$$C'_{x,y}(i,j) = \frac{C_{x,y}(i,j)}{\sum_{p=1}^{8}\sum_{q=1}^{8} C_{x,y}(i',j')}$$

Edge magnitude accumulation and normalization

►Keypoint : Bag-of-Features + SURF
- K-means (20 dims / frame)
- Time-series representation

►Graphical Model
- Integrating 3 types of feature descriptor



P(A|H)P(H) · P(A|E)P(E) · P(A|V)P(V)

4. Action Recognition Experiment

►2 experiments

(i) Weizmann Action Dataset
- Single-person Action Recognition

(ii) Semantic Description of Human Activities (SDHA)
- High-level multi-person Action Recognition

►Settings

HOOF
- 100 dimensions (4 directions × 25 frames)

ECoHOG
- 10880 dimensions (1088 dimensions × 10 frames)

–> 100 dimensions (PCA : Principal Component Analysis)

BoF + SURF
- 100 dimensions (20 dimensions × 5 frames)

Weizmann Action Dataset

Framework	Accuracy (%)
Proposed method	99.1
Wang *et al.*	97.2
Satkin *et al.*	95.8
Chaudhry *et al.*	94.4
Klaser *et al.*	90.7

	be	ja	ju	pj	ru	si	sk	wa	w1	w2
be	100									
ja		100								
ju			100							
pj				100						
ru					100					
si						100				
sk						0.5	99.5			
wa							1.2	98.8		
w1								2.4	97.6	
w2		0.2							3.4	96.4

be : bend ja : jack ju : jump pj : pjump ru : run
si : side sk : skip wa : walk w1 : wave w2 : wave2

SDHA Dataset

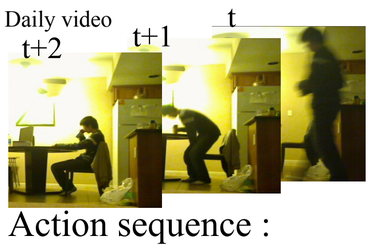
Framework	Accuracy (%)
Proposed method	81.8
Ryoo *et al.*	91.1
Niebles *et al*	81.5

	Shake	Hug	Kick	Point	Punch	Push
Shake	90.6	5.9	0.1	0.9	0.3	1.8
Hug	4.4	92.4		0.5	0.2	2
Kick	6.3	6.2	84.6		1.3	1.4
Point	1.9	1.1		96	0.4	0.2
Punch	17	17.6	3.9		55.4	5.9
Push	10	28	0.5		2	59.2

►Proposed Approach
- runs around 20 fps on these datasets
- performs high-accuracy action recognition
- is needed to improve in the context of interaction (give a posture information for complicated actions)

5. Intention Inference

►Next Activity Prediction
- We estimate "Intention" as next activity
- Using Naïve Bayes Classifier
- Information : Time Zone, Previous & Current Action



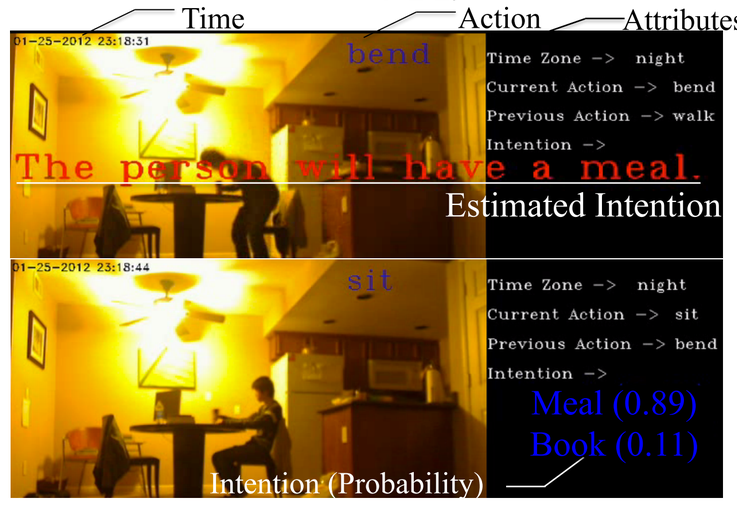
Action sequence :
Walk -> Bend -> Sit -> Meal
Time zone : Night

```
/-------------------------/
/-       Database        -/
/-------------------------/
morning walk    sit     book
morning walk    stand   book
day     walk    sit     pc
night   bend    sit     meal
night   walk    sit     pc
```

Part of daily database
(Time Zone | Previous | Current | Intention)

Naïve Bayes Classifier

$$\hat{\theta}_{Naive\ Bayes} = \arg\max_{\theta} \frac{\prod_i P(x_i|\theta)\ P(\theta)}{\int \prod_i P(x_i|\theta)\ P(\theta) d\theta}$$



Process	Time (ms)	FPS
Tracking	17.3	57.5
Action Recognition	40.8	24.4
Intention	0.000052 (52 ns)	19,230,769 (19 MIPS)
Total	58.2	17.1

6. Future Works

► Pose Estimation to improve Action Recognition

► Pose Estimation for Human-Object Interaction

► Data Mining to discover more effective info from DB