Fall Detection in RGB-D Videos by Combining Shape and Motion Features

Durga P. Kumar, Yixiao Yun, Irene Yu-Hua Gu

Signal Processing Group, Dept. of Signals and Systems,
Chalmers Univ. of Technology, Gothenburg, Sweden

IVMSP-L10.2, ICASSP 2016

16:20 - 16:40, March 25, 2016 | Shanghai, China

Outline

1. **Introduction**
   Addressed Problems, Motivations, . . .
2. **The Proposed Method**
3. **Experimental Results**
   Dataset, Evaluations, . . .
4. **Conclusion**

Introduction

Addressed Problem

Fall detection in RGB-D videos

Applications

Elderly care, e-Health, assisted living, ...

Observations

- Drastic pose change
- Large physical movement

Introduction

Brief Review

- **Bounding box based features**
  [Debard'12], [Charfi'13], ...
  Insufficient description of motion from using the bounding box sorely
- **Multiple cameras (3D modeling)**
  [Auvinet'11], [Mastorakis'14], [Stone'15], ...
  Computationally demanding

Introduction

Motivations

- **Reduce the risk** (bone fracture, coma, death, . . .) due to falls
- **Automatically** detect falls and trigger alarms
- Effectively detect falls from **a single camera view**
- Exploit the **spatio-temporal** features of **pose change** and **body motion**

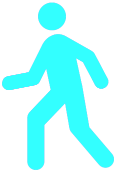
Introduction

Main Novelties

- Extract effective **time-dependent (spatio-temporal)** features:
  - **Global shape+motion** from **RGB videos**
  - **Local shape+motion** from **Depth videos**
- **Combine** different features for fall detection through classification of 2 most confusing classes **(falls vs. lie-down)**
- Study the contribution of **individual component feature** to overall performance

The Proposed Method

Foreground Human Detection



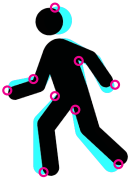
The Proposed Method

Foreground Human Detection

1. Differencing consecutive **RGB** frames

The Proposed Method

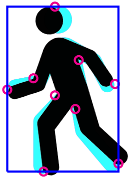
Foreground Human Detection



1. Differencing consecutive **RGB** frames
2. SURF **keypoint detection** in **difference images**

The Proposed Method

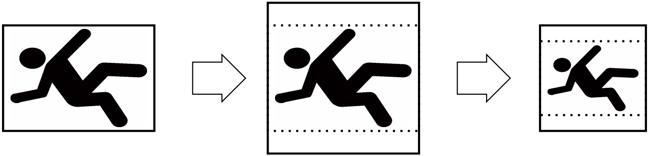
Foreground Human Detection



1. Differencing consecutive **RGB** frames
2. SURF **keypoint detection** in **difference images**
3. ROI defined by **bounding box of keypoints**

The Proposed Method

Shape Features (RGB)



- **Size normalization** of ROI:

$$(w, h) \Rightarrow (l, l) \Rightarrow (\lambda, \lambda)$$

- Shape features are implicitly represented by **HOG** (Histogram of Oriented Gradients) descriptors

The Proposed Method

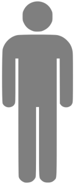
Motion Features (RGB)

Based on **HOGOF** (Histogram of Oriented Gradients of Optical Flow)

1. **Optical flow** is estimated between normalized ROIs
2. **Magnitudes** and **orientations** of optical flow are color-coded by **HSV (hue, saturation)**
3. Motion features are implicitly represented by **HOG** descriptors

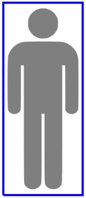
The Proposed Method

Target Contour Extraction



The Proposed Method

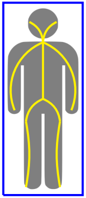
Target Contour Extraction



- Corresponding ROI in **depth** video frames

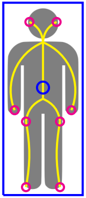
The Proposed Method

Target Contour Extraction



- Corresponding ROI in **depth** video frames
- **Morphological skeleton** estimated from ROI of depth video

Target Contour Extraction



- Corresponding ROI in **depth** video frames
- **Morphological skeleton** estimated from ROI of depth video
- **8 local extrema** obtained from the skeleton + **1 centroid**

The Proposed Method

Shape Features (Depth)

- (x, y) **coordinates** of contour centroid and local extrema
- **Distances** between centroid and local extrema
- **Orientation** and **aspect ratio** of the bounding box
- **Eccentricity** of the ellipse bounded by the rectangular box

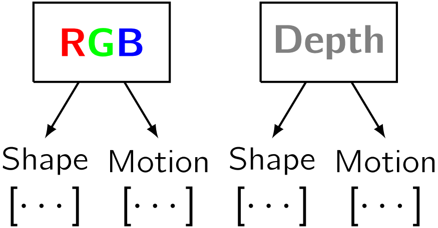
The Proposed Method

Motion Features (Depth)

- Based on **consecutive frames**
- **Gradient of distances** between centroid and local extrema
- **Inter-frame speed** of centroid and local extrema

The Proposed Method

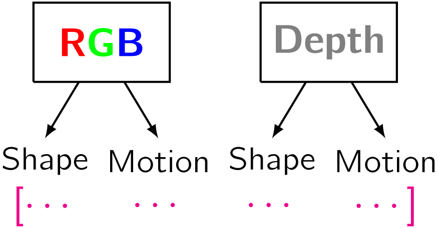
Feature Fusion (RGB+D)



- For each frame, features extracted from RGB and depth images are **concatenated**

The Proposed Method

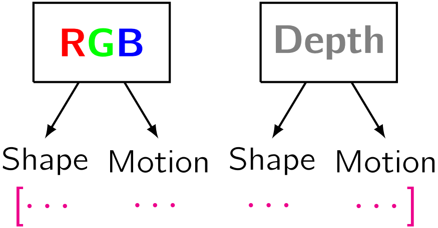
Feature Fusion (RGB+D)



- For each frame, features extracted from RGB and depth images are **concatenated**

The Proposed Method

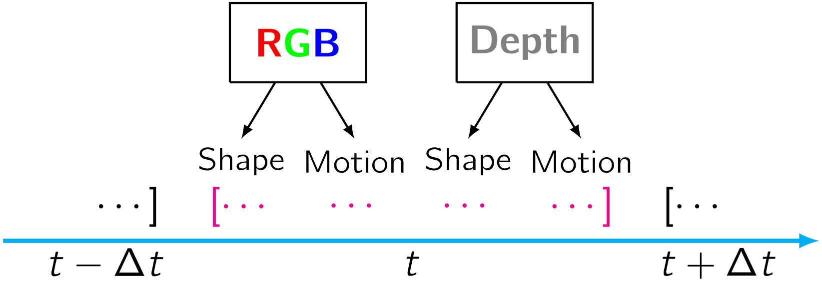
Time-Dependent Features



- For each video event, augmented features of each frame are **temporally stacked**

The Proposed Method

Time-Dependent Features



- For each video event, augmented features of each frame are **temporally stacked**

Length Norm. of Video Events

$\cdots \longrightarrow L$

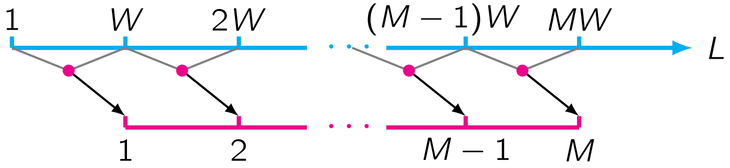
- A video event of length L (#frames)

The Proposed Method

Length Norm. of Video Events

- A video event of length L (#frames)
- Divided into M (fixed) **segments**, each of length $W = \lfloor L/M \rfloor$

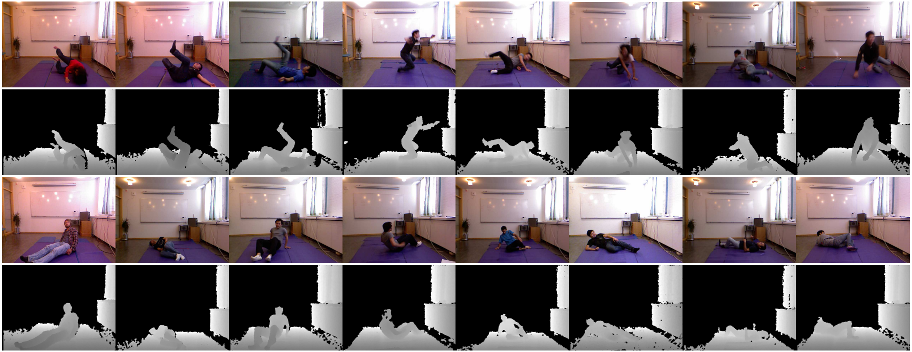
Length Norm. of Video Events



- A video event of length L (#frames)
- Divided into M (fixed) **segments**, each of length $W = \lfloor L/M \rfloor$
- In each segment, features **averaged** over W frames $\Rightarrow$ normalized length M

Experimental Results

RGB-D Dataset



Class#	Activity	#Subjects	#RGB Video	#Depth Video
1	Falling down	20	400	400
2	Lying down	20	400	400

Experimental Results

Setup

- Normalized length of video events: $M = 10$
- Binary C-SVM + RBF kernel
- **Case-1**: 50% training, 50% testing
- **Case-2**: 80% training, 20% testing

Results and Evaluations on Test Set

(a) Case-1: fusion vs. standalone features

Feature	Detection rate (%)	FNR (%)	FPR (%)
HOG	93.75	6.25	5.00
HOGOF	94.00	6.00	4.00
Contour	92.75	7.25	9.00
Fusion	95.25	4.75	5.00

(b) Case-1 vs. Case-2

Case	Detection rate (%)	FNR (%)	FPR (%)
1	95.25	4.75	5.00
2	97.50	2.50	2.50

Conclusion

- **Spatio-temporal** features of pose change and body motion are exploited
- **Time-dependent shape+motion** features from RGB and depth videos are combined
- Trained on large number of RGB-D videos, results on test set showed **high detection rate (97.5%)** and **low false alarms (2.5%)**

Future Work: Extend the method and tests on more video activities