Motion Synthesis for Virtual Characters

Motion Capturing

Koushik Chowdhury

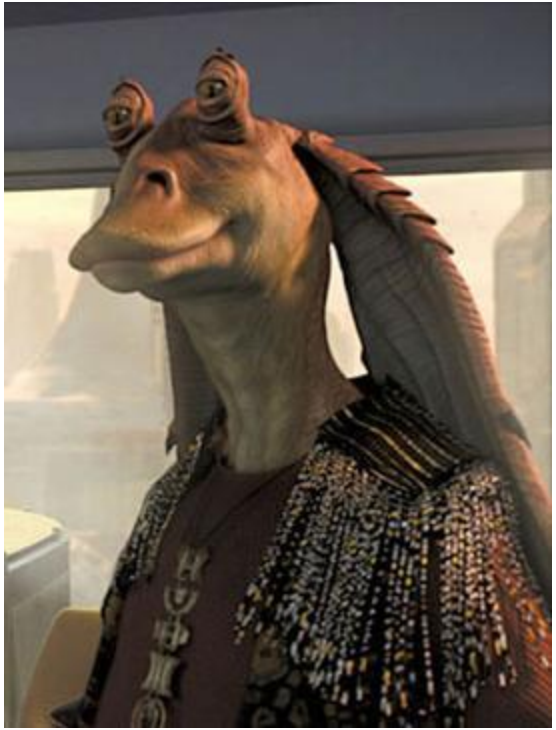
A survey of depth and inertial sensor fusion for human action recognition

Chen Chen, Roozbeh Jafari, Nasser Kehtarnavaz

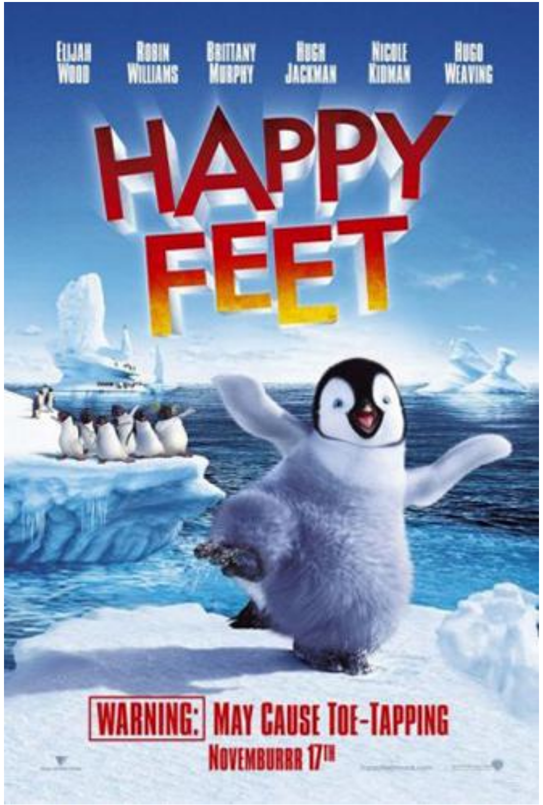
Contents

Introduction: *Motion Capture*

- Process of recording human movement.
- First Developed
- Application
- First Feature Length Film
- Motion Capture Technique

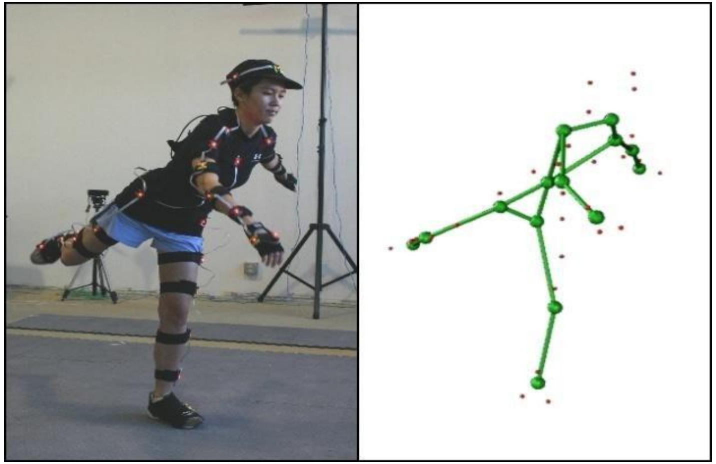


Introduction: *Motion Capture*



Introduction: *Motion Capture Technique*

- Optical
- Electro Magnetic
- Mechanical



Optical

Electro Magnetic

Mechanical

Introduction:

Objective

- Investigation: Vision and Inertial Sensor work together
- Overview of Dataset based on Vision and inertial

Human Action Recognition

- Detecting and Analyzing Human Action
- Type of Sensors, e.g. RGB Camera, Depth Camera, Range Camera etc.
- Two Approaches: Vision and Inertial

Contents

Vision-based action recognition

- What is VBAR?
- Challenges.
- Categories: gestures, actions, interactions, group activities.
- Vision Sensor such as conventional RGB Camera
- Survey

Vision-based action recognition (Cont.)

- Depth Sensor
- Provide 3D Data
- Three existing approach
- Cost-effective 3D data
- Compared to conventional RGB images
- Limitation

Depth Sensor

- Microsoft’s kinect
- Capture 3D data
- Color camera, IR emitter, IR depth sensor, monitor microphone, LED.
- 16 bit depth images, 320x240 pixels
- 8 bit depth images, 640x480

Depth Sensor: An example

a b

Depth Sensor (Cont.)

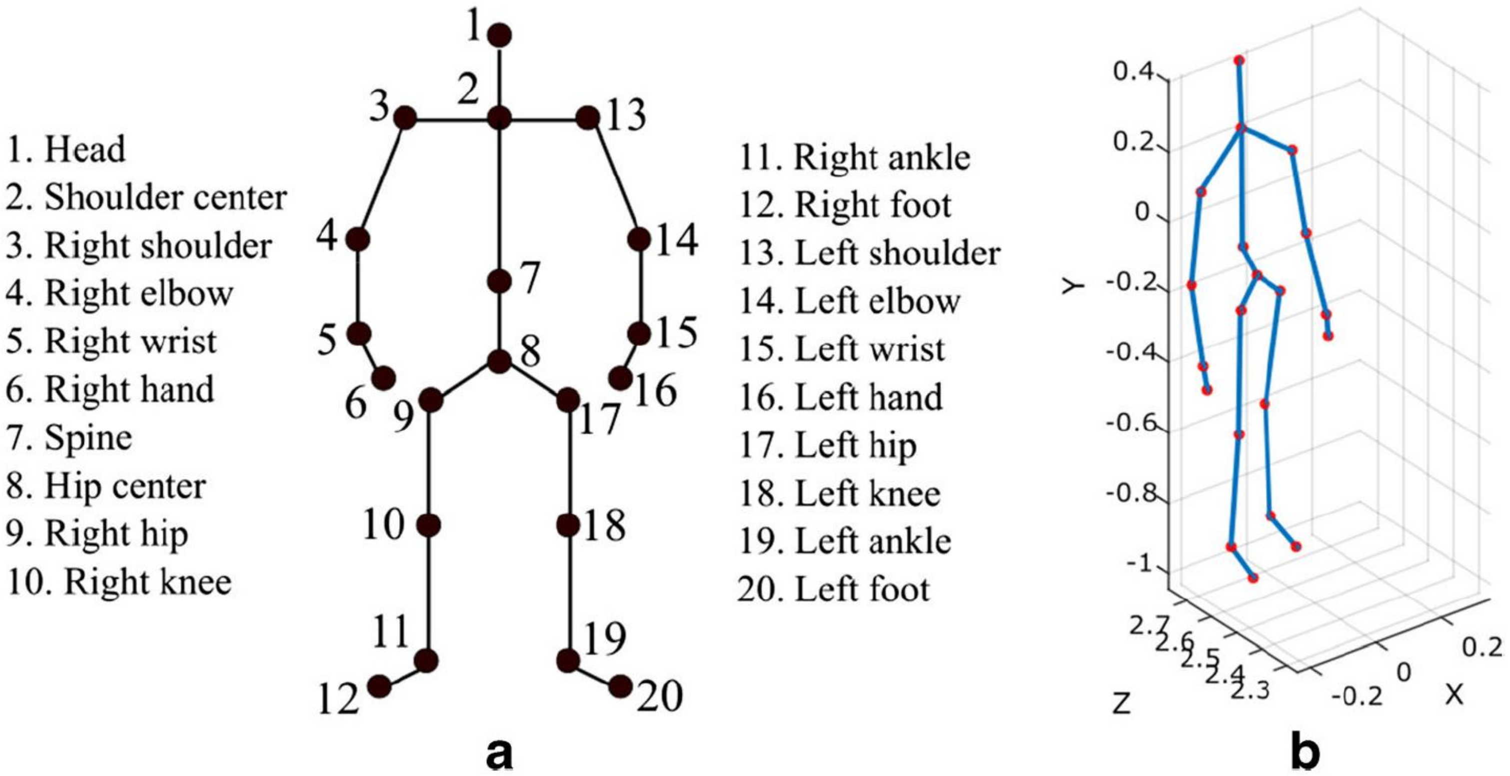
- Projected depth maps.
- Space-time occupancy
- Spatio-temporal depth cuboid
- Surface Normals
- Skeleton Joints

Contents

Inertial-based action recognition

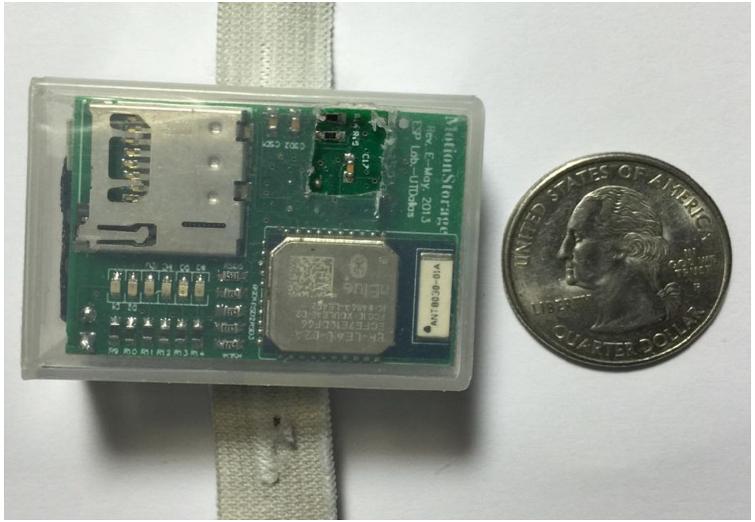
- Inertial Sensor
- Provide 3D action data
- Example: Xsens 3D motion tracking

Inertial Sensor

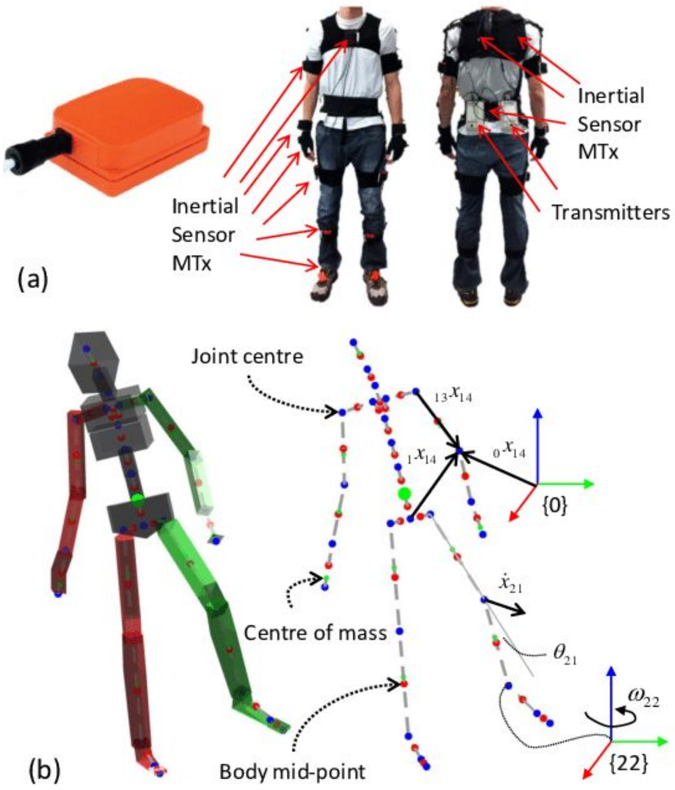


Inertial Sensor (Cont.)

- Small Size (1” x 1.5”)
- ESP at TA&MU
- 3-axis acceleration
- 3-axis angular velocity
- 3 axis magnetic strength
- Sampling rate 50Hz



Inertial Sensor



Inertial Sensor

Inertial Sensor (Cont.)

- Wearable sensor
- A fall detection system
- Human activities and different classification techniques consisting of Bayesian decision, least squares.
- Recognize the hand-twist and open hand action

Advantages

- RGB Cameras: cost effective, easy, rich texture
- Depth Cameras: cost effective, insensitive, darkness, easy, 3D structure, Color and texture.
- Inertial Sensors: cost effective, high sampling rate, darkness, unconfined environment.

Disadvantages

- RGB Cameras: subject to be in the field, sensitive, algorithm.
- Depth Cameras: subject to be in the field, noise, no color, sensitive.
- Inertial Sensors: sensitive, power consumption, multiple sensor, intrusiveness.

Contents

Combination of depth and inertial sensors

- Why?
- Human action/gesture recognition by using depth and inertial sensors

Combination of depth and inertial sensors (Cont.)

- Data Synchronous
- Action Segmentation
- Feature extraction
- Classification and fusion approach

Contents

Human action/gesture datasets

- MSR Action 3D dataset
- WARD dataset
- Datasets that contain data from depth and Inertial sensors.

Human action/gesture datasets: MSR Action 3D dataset

- Dataset that contains twenty actions.
- Each action was performed by seven subjects for three times.
- Divided the 20 actions into three subsets, each having 8 actions

Action Set 1 (AS1)	Action Set 2 (AS2)	Action Set 3 (AS3)
Horizontal arm wave	High arm wave	High throw
Hammer	Hand catch	Forward kick
Forward punch	Draw x	Side kick
High throw	Draw tick	Jogging
Hand clap	Draw circle	Tennis swing
Bend	Two hand wave	Tennis serve
Tennis serve	Forward kick	Golf swing
Pickup & throw	Side boxing	Pickup & throw

Human action/gesture datasets: WARD dataset

- http://www.eecs.berkeley.edu/~yang/software/ WAR/
- The data are sampled from 7 female and 13 male human subjects (in total 20 subjects) with age ranging from 19 to 75.
- 13 action categories: 1. Stand (ST). 2. Sit (SI). 3. Lie down (LI). 4. Walk forward (WF). 5. Walk left-circle (WL). 6. Walk rightcircle (WR). 7. Turn left (TL). 8. Turn right (TR). 9. Go upstairs (UP). 10. Go downstairs (DO). 11. Jog (JO). 12. Jump (JU). 13. Push wheelchair (PU).

Human action/gesture datasets: Berkeley MHAD

- Contains temporally synchronized data
- A motion capture system, 12 RGB cameras, 2 Microsoft Kinect depth cameras, 6 wearable accelerometers, and 4 microphones.
- 659 data sequences from 11 human action
- 12 subject
- 6 accelerometers

Human action/gesture datasets: URFD

- Fall detection application
- 70 sequences
- 5 subject
- 30 fall sequences
- 40 sequences

Human action/gesture datasets: Other dataset

- UTD-MHAD
- 50 salads dataset
- ChAirGest multimodal dataset
- TST fall detection dataset.
- Huawei/3D dataset

Human action/gesture datasets: Summary

Dataset	Modality					# Sub	# Act	# Seq
	M	V	D	A	I			
Berkeley MHAD [47]	1	12	2	4	6	12	11	660
URFD [36]	–	2	2	–	1	5	>5	70
UTD-MHAD [12]	–	1	1	–	1	8	27	861
50 salads [58]	–	1	1	–	7	25	17	966
ChAirGest [52]	–	1	1	–	4	10	10	1200
TST Fall detection database [27]	–	–	1	–	2	11	8	264
Huawei/3DLife dataset [59]	–	5	5	5	8	17	22	3740

Some Motion Capture datasets

- SFU Motion Capture Database
- MOSH Dataset

Contents

Conclusion

- Better human action recognition performance
- Do experiment of multiple type sensor instead one sensor.

Contents

Future Work

- Developing view-invariant features for depth images.
- Developing intelligent fusion approaches.
- Examining human and object interactions.

Contents

- Introduction
- Depth Sensor
- Inertial Sensor
- Combination of depth and inertial sensors
- Human action/gesture datasets
- Conclusion
- Vision-based Sensor
- **Recent works**

Recent Work

Human Action Recognition and Prediction

by Yu Kong, Member, IEEE, and Yun Fu, Senior Member, IEEE

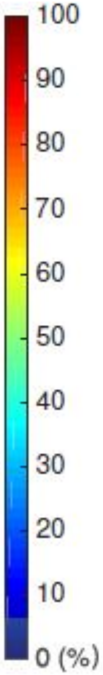
- Action recognition
- Action prediction

A recent Work: UCF101 dataset

- Videos
- 11 action categories
- UCF101 gives the largest diversity in terms of actions and with the presence of large variations in camera motion, object appearance and pose, object scale, viewpoint, cluttered background, illumination conditions, etc.

Recent Work: Action Prediction

Instantly Predictable	Early Predictable	Late Predictable
Billiards	Fencing	JavelinThrow
IceDancing	FrisbeeCatch	HighJump
RockClimbingIndoor	SoccerPenalty	FrontCrawl
PlayingPiano	VolleyballSpiking	HeadMassage
PommelHorse	HulaHoop	Haircut
Rowing	FieldHockeyPenalty	PlayingViolin
Skijet	BasketballDunk	HandstandWalking
JugglingBalls	CliffDiving	PoleVault
SoccerJuggling	Bowling	CricketBowling
TaiChi	TennisSwing	ThrowDiscus



Recent Work: Action Prediction

- presented a complete survey of state-of-the-art techniques for action recognition and prediction from videos.
- These techniques became particularly interesting in recent decades.

Recent Work:

Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera

by Timo von Marcard, Roberto Henschel, Michael J. Black, Bodo Rosenhahn, Gerard Pons-Moll

- Estimate accurate 3D poses in the wild

Recent Work: Dataset

- 60 video
- 2D pose
- 3D poses
- Camera poses for every frame in the sequences.
- 3D body scans and 3D people models.
- 18 3D models in different clothing variations.

Recent Work: Result

- Tracking Accuracy

Approach	[39]	[16]	IT	VIP-2D	VIP-Cam	VIP-IMU6	VIP-IT	**VIP**
MPJPE	70.0	(62)	55.0	15.1	25.3	39.6	28.2	**26.0**
MPJAE	-	-	16.9	10.1	12.1	15.3	12.0	**12.1**

Table 1. Mean Joint Position Error (MPJPE) in mm and Mean Per Joint Angular Error (MPJAE) in degrees evaluated on TotalCapture.

Recent Work: PhD Thesis

Vision-based Human Action Recognition using Machine Learning Techniques

By Allah Bux, Lancaster University

Thank You