Real-time Gymnast Detection and Performance Analysis With a Portable 3D Camera

William Hoff

whoff@mines.edu
Colorado School of Mines
Golden, Colorado, USA

bill.hoff@daqri.com
DAQRI
Vienna, Austria

My co-authors:
Brian Reily, Hao Zhang
Colorado School of Mines

Need for Performance Analysis

• Need quantitative data, related to performance
• Traditional methods – force plates, motion capture systems

• Problem: can’t use force plates and markers while competing
• Need low cost system, easy to set up for non technical users

Alpine skiing [1]

Cycling [2]
Pommel Horse Gymnastics Event

Sam Mikulak - Pommel Horse - 2012 Visa Championships - Sr. Men [3]  
https://www.youtube.com/watch?v=19N6uruAyos
Portable 3D Camera

- New developments in depth 3D cameras
- New opportunities – low cost portable, accurate
- Kinect 2.0 specs:
  - Time of flight sensor
  - 512x424 depth image, 30 fps
  - 0.5-4.5 m

https://www.youtube.com/watch?v=9YVmB0Alrvw
Human Detection

- Histograms of oriented gradients (HOG)
  - Descriptor composed of HOG cells
  - Sliding window
  - SVM classifier
  - Extended to part-based models

- Problem: Trained for typical upright body poses

Human Detection using depth data

- A “random forest” classifier labels each pixel according to body part
- Used in Microsoft’s Kinect
- About 1M training images

Skeleton Estimation in Kinect

• Starting from a torso point, construct skeleton

• Problem – since it is trained on upright poses, it generates noisy and inaccurate data when applied to gymnasts
Our Approach

1. Depth of Interest Segmentation
2. Gymnast Silhouette used to Recognize Action
3. Performance Analyzed

September 25, 2017
Depth of Interest Segmentation

- Segment scene based on depth
- Steps
  - Select $n$ pixels randomly
  - Describe each with a Gaussian function, and sum these
    \[ P(x) = \sum_{i=1}^{n} \exp \left( \frac{-(x-D(x_i))^2}{2 \times \text{MAXDEPTH}} \right) \]
    - Identify peaks in this distribution
- Each peak is a proposal for a segmentation

Note: The stationary pommel horse is automatically removed from the scene
Depth of interest segmentation

- Experimentally, the human is completely contained in the segmentation corresponding to one of the top three proposals 97.8% of the time.

- Segmentation greatly reduces the amount of data that later stages of the pipeline need to process.
- On average, non-zero pixels are 37.8% of image size.
Human Detection from Silhouette

- Identified depths of interest are input to a HOG based detector trained to identify silhouettes.
- HOG features are computed on depth imagery, treating this data as a grayscale image to obtain the gradients.

- SVM sliding window classifier
  - Single class: human vs not human
  - Trained on a large variety of gymnast poses; robust to changes in body size and orientation.
Recognition of Spinning Activity

• Need to detect when the gymnast is spinning

• Compute a Silhouette Activity Descriptor:
  – Width, height of silhouette
  – Depth values at the left and right sides
  – Change in top, bottom, left, and right image coordinates compared to the previous frame
Recognition of Spinning Activity

- A support vector machine classifier was trained to recognize spin/no spin
  - Radial basis function kernel
    \[ K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \]

- Temporal smoothing
  - Classifier is applied to each frame
  - Classifications are smoothed over 5 frames
  \[ c_i = \frac{1}{5} \sum_{j=-2}^{2} c_{i+j} \]
Performance analysis of spins

• Goals:
  – Track the position of the feet
  – Track body angle

• Procedure:
  – From the body centroid, identify
    the longest vector to the body
    contour
  – Then identify the shortest
    vector – this is the waist
  – Using the bend in the body,
    identify the second longest
    vector

*Feet are assumed to the lower of the two long vectors*
Timing spins

- Find the times when the feet achieve greatest deviation in x
- Fit to a cubic spline, to interpolate the exact time of an extrema
- The duration of the spin is the amount of time between consecutive left extrema or consecutive right extrema.
Spin detection video

https://www.youtube.com/v/LRK8vK6NXfg
Spin detection (slow)

https://www.youtube.com/v/IFTE_Lna9So
Data collection

- 39 routines
- 10115 depth images
- Dataset available at http://hcr.mines.edu

- Annotated with
  - Spinning (yes/no)
  - Location of head and feet
  - Time of extrema
Evaluation

• Activity recognition
  – Data split into 5024 training frames and 5091 testing frames
  – Classified spin/no spin with 94.83% accuracy

• Spin times
  – RMS error was 12.99 ms compared to ground truth

Average spin time for a top gymnast is 960ms, with a standard deviation of only 25ms
Case study – application development

• An application was developed for use by coaches for training

• Software
  – C++, OpenCV, Qt
User Interface

Select Athlete

Record

View Frame

Analyze

See Past Statistics
User Interface

Consistency = \frac{\text{Mean} - \text{Std Deviation}}{\text{Mean}}
Conclusions

- Introduced an automated system to provide an analysis of a gymnast’s performance, using a portable 3D camera
- Steps:
  - Detect a gymnast using novel “depth of interest” method
  - Identify when a gymnast is performing circles
  - Analyze their performance
- Performance
  - Identify a depth of interest with 97.8% accuracy
  - Detect spinning with 93.8% accuracy
  - Analyze spin consistency with less than a 13ms RMSE
- Created an app for gymnastics coaches
- Dataset with ground truth
Thank you!

References