STVIs for platform diving

Image and video understanding 2019 WS

C. Lenzenweger, B. Sespede

1 sentence recap

- Identify diving style from video using STVIs as a base for feature extraction

Pike

Tuck

Straight

Data (I)

- Further pruning of dataset
  - Manual labelling of best quality videos
  - Aim to remove motion blur, and low resolution as much as possible
  - Improved consistency of camera angles
  - In the end we manually labeled 100 videos per style (300 in total)

Data (II)

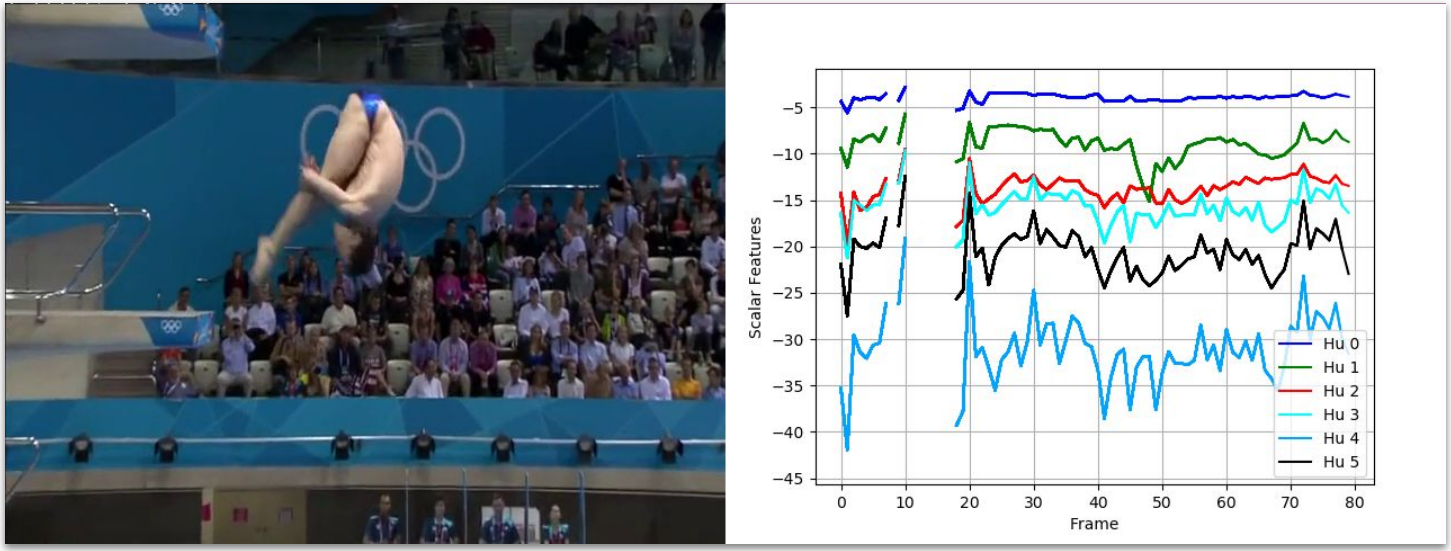
- Even though the input data improved, the STVIs were still not perfect:
  - Disjoint STVIs.
  - Empty videos.
  - Large number of small artifacts.
- As they would negatively affect our training process we had to discard some videos:
  - In the end we only used 113 videos from the original 300.

Data (III)

Feature extraction (I)

- Merge STVI-enclosing contours, if:
  - Relative area > 0.66.
  - Ratio of present STVI IDs > 0.7.
  - Both measurements were taken w.r.t. largest contour.
- Instead of using primitives to describe orientation and center of mass, use invariant Hu moments due to disjoint STVIs:
  - They superseded all other features we were planning to extract.
  - Took the absolute value and log.
- Median filter over 9 frames reduced the noise of the feature space.

Feature extraction (II)

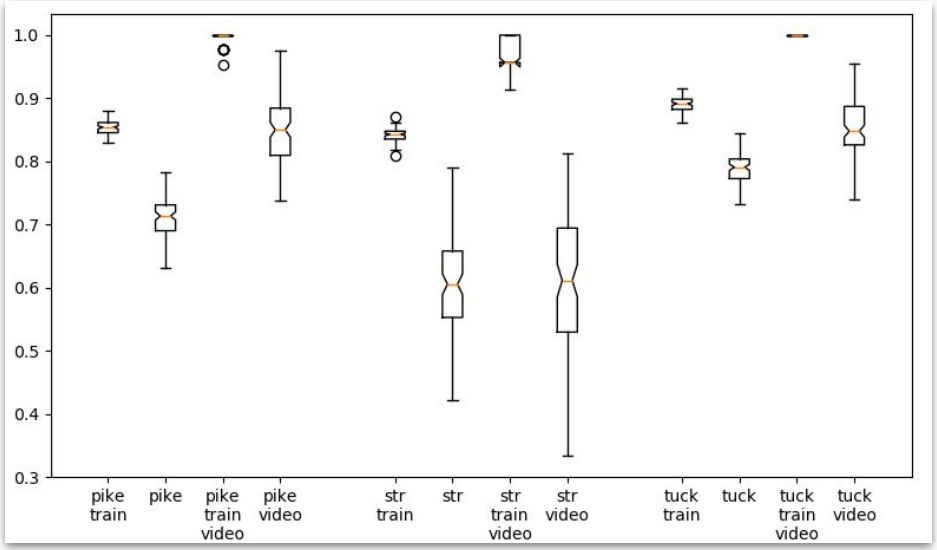


Training process

- Use a 90/10 split due to small number of training data.
- Used multi-class SVM with RBFs as Kernels.
- We decided evaluate to different classification schemes:
  - Frame-wise classification.
  - Frame-voting scheme.
- Shuffle training/test data and repeat process 100 times.

Results (I)

- Resulting accuracies for each approach:
- Why is the accuracy for straights lower?
  - Unbalanced dataset.



Results (II)

- Confusion matrices for each approach:
- When do they get confused:
  - Pike vs. tuck
  - Straight vs. pike/tuck

true \ predicted	Pike	Straight	Tuck
Pike	121.14	3.64	46.22
Straight	7.33	23.2	7.47
Tuck	53.64	3.48	207.88

Table 1. Mean (100 runs) confusion matrix for frame-wise classification. Rows represent true class. Columns represent predicted class.

true \ predicted	Pike	Straight	Tuck
Pike	33.84	0.38	5.5
Straight	4.73	11.11	2.37
Tuck	6.74	0.14	38.8

Table 2. Mean (100 runs) confusion matrix for video-wise classification. Rows represent true class. Columns represent predicted class.

Discussion and conclusions (I)

- Best training approach: frame-voting scheme, as it reduces confusion significantly.
- Filtering in feature space proved to be useful.
- Quality of STVIs.
- Hu invariant moments are a good shape descriptor for our problem.
- SVMs were sufficient considering the amount of data we had.

Discussion and conclusions (II)

- Videos (if there is time)
  - Pike: https://cloud.flipbit.eu/index.php/s/jzA78W32HdmYMzG
  - Tuck: https://cloud.flipbit.eu/index.php/s/T97QxKRRNkDkBBS
  - Straight: https://cloud.flipbit.eu/index.php/s/5SG6dpfTnXirEoH
  - Frames: https://cloud.flipbit.eu/index.php/s/k3pEXMtJhmN4bcf

Thank you for your attention!