Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Title — Classical Machine Learning applied to surgical knot analysis

Supervisor — Dr. Gerard Lacey

Khushboo Goyal
MS Computer Science- Future Networked Systems

12/08/2019

Topics

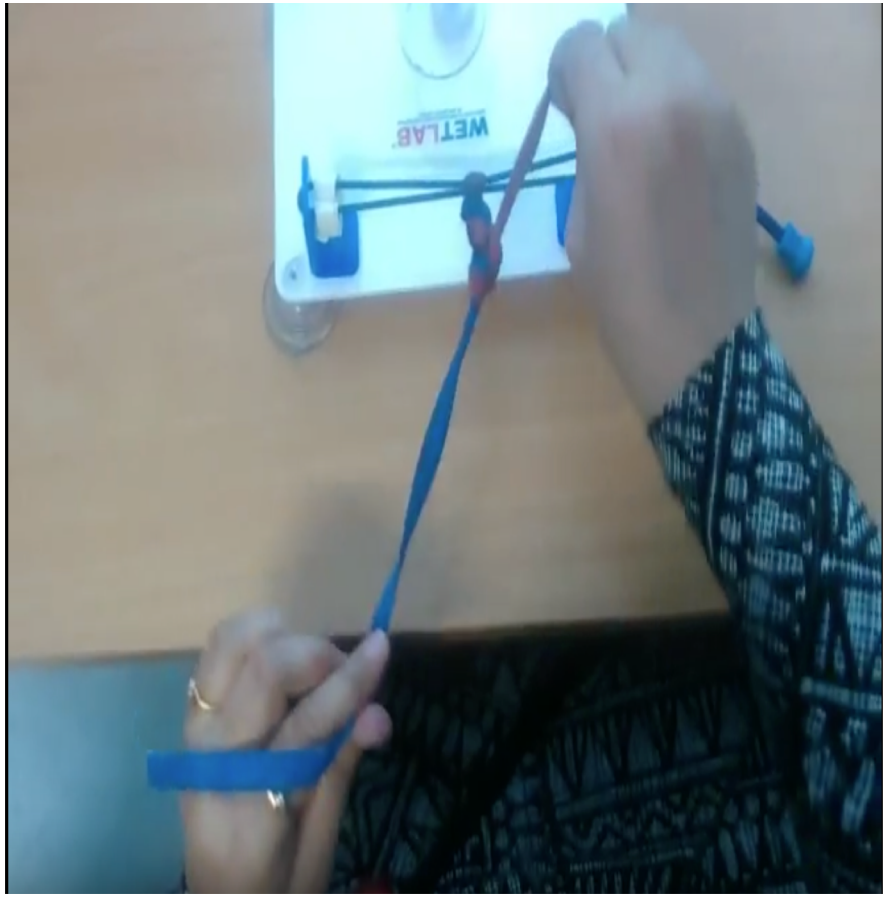
- Introduction
- Motivation
- Data Collection
- Implementation
- Result
- Conclusion
- References

Introduction

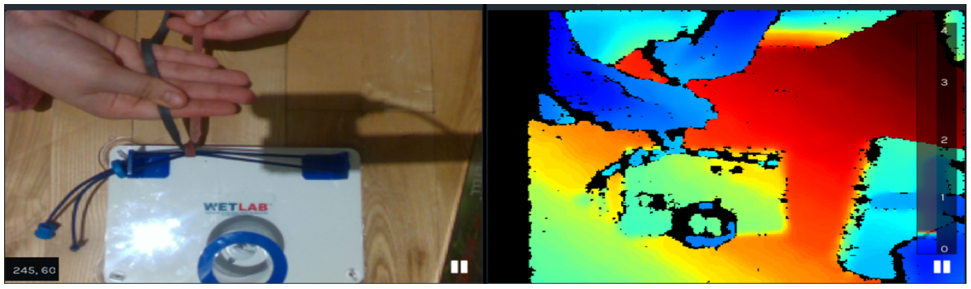
- This dissertation helps to explore hand gestures to identify surgical knot.
- 3 methods to recognize gesture
  - Sensor gloves (requires user to wear cables and additional devices)
  - 2D model based recognition(In 2D methods, a hand is represented by its geometric features and appearance features such as contours, fingertips, and skin color.)
  - 3D model based hand method(3D hand tracking method in depth space using a 3D depth sensor)
- Many application domains-
  - Games
  - Sign Language
  - Robotics
- **This thesis take reference of paper Video Based Assessment of OSATS Using Sequential Motion Textures by Irfan Essa.**
  - Irfan did analysis on the basis of OSATS(Objective Structured Assessment of Technical Skills) score with different techniques like Motion texture, Bag-of-words, SMT(Sequential Motion Texture) and many more.

Motivation

- Developing high quality skill(surgical knot) is time consuming for a novice surgeon.
- Requires expert surgeon's supervision and evaluation for manual assessments like OSATS(Objective Structured Assessment of Technical Skills).

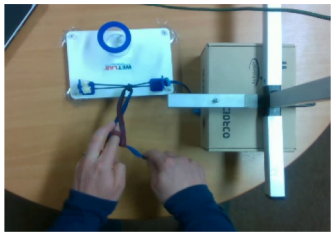


Data Collection



- Hardware – Real Sense Camera
  - Real Sense D435 with depth resolutions, which is upto 1280x720 at 30 frames per second [1]
- Software – Depth Sense SDK
  - It is the SDK which allows depth and color streaming, stored videos in .bag format
- Dataset- Divided into 2 parts
  - camera position is at 90◦ (mounted on stand)
  - camera position is at 60◦ of angle (egocentric)

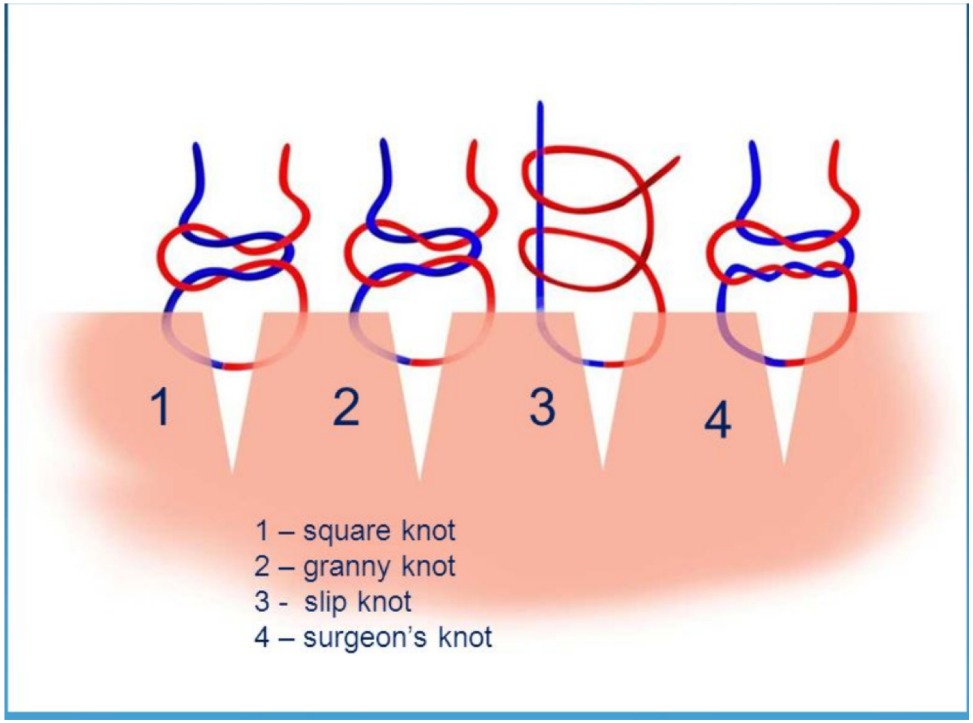
Egocenteric data (headmount)



Camera mounted on stand

Surgical Knot

- Surgical Ties are commonly described as "One-Handed" or "Two-Handed" – slightly confusing because both techniques involve two hands.
- There are many type of surgical knot, few are shown in figure.



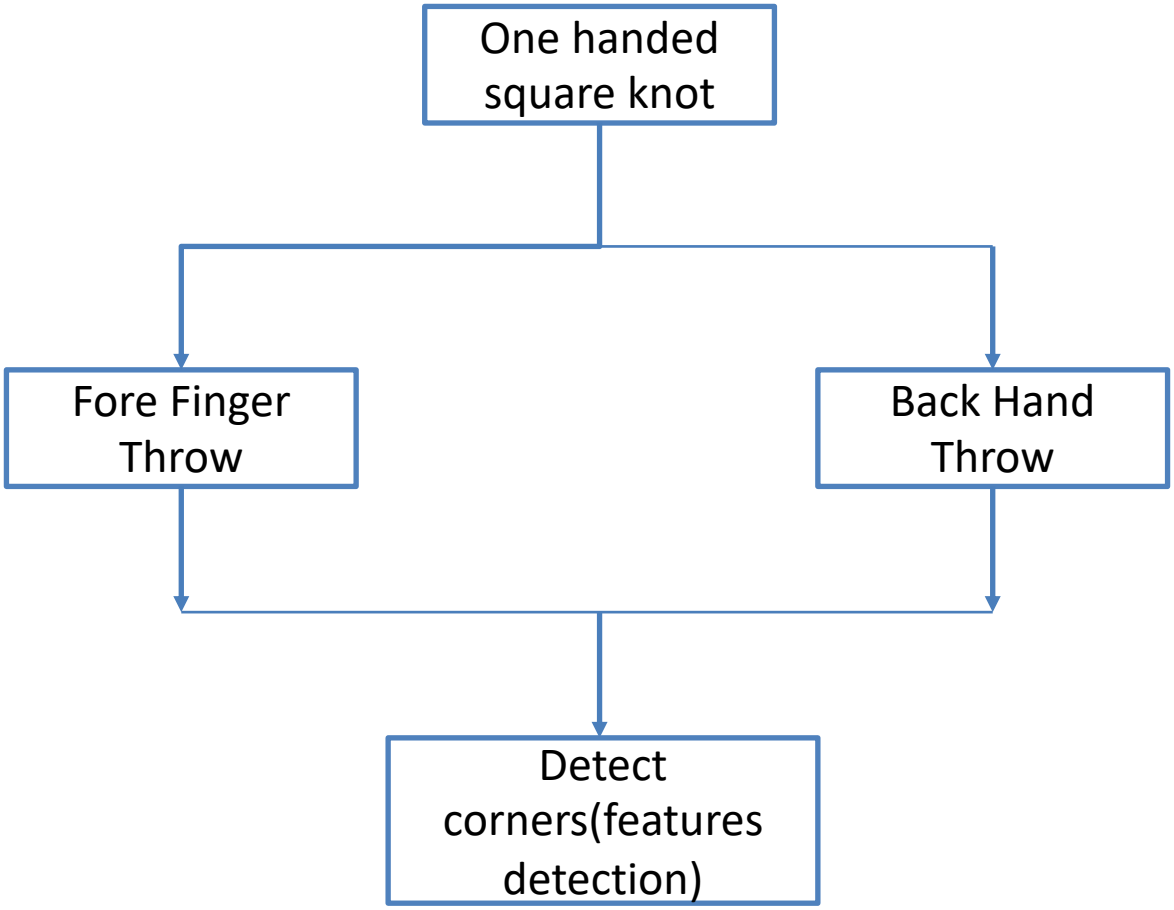
One handed Square knot

Part1- Fore Finger Throw

One handed Square knot

Part2- Back Hand Throw

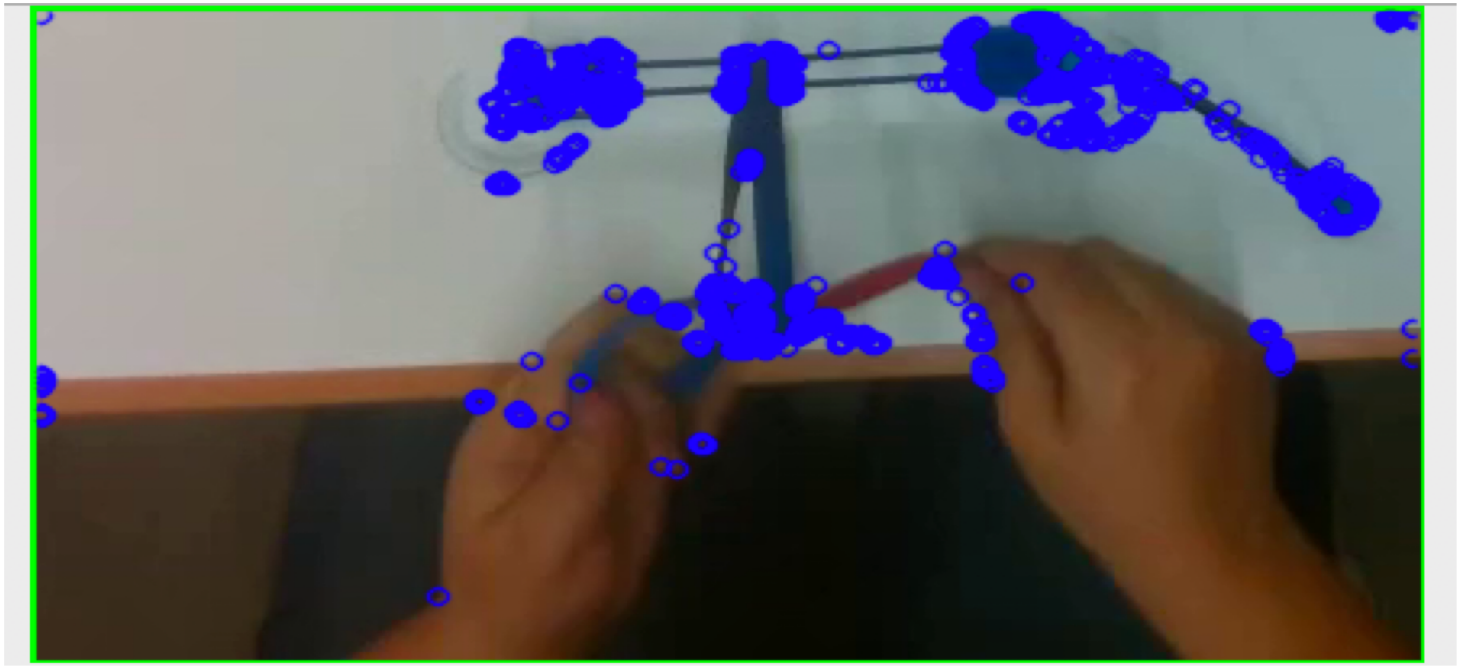
Methodology



Implementation

Feature Extraction- Harris Corner Detection per frame

- Using bounding box around the hands, extracted features but this gives wide range of features other than hand, so I applied on video in order to get better results.

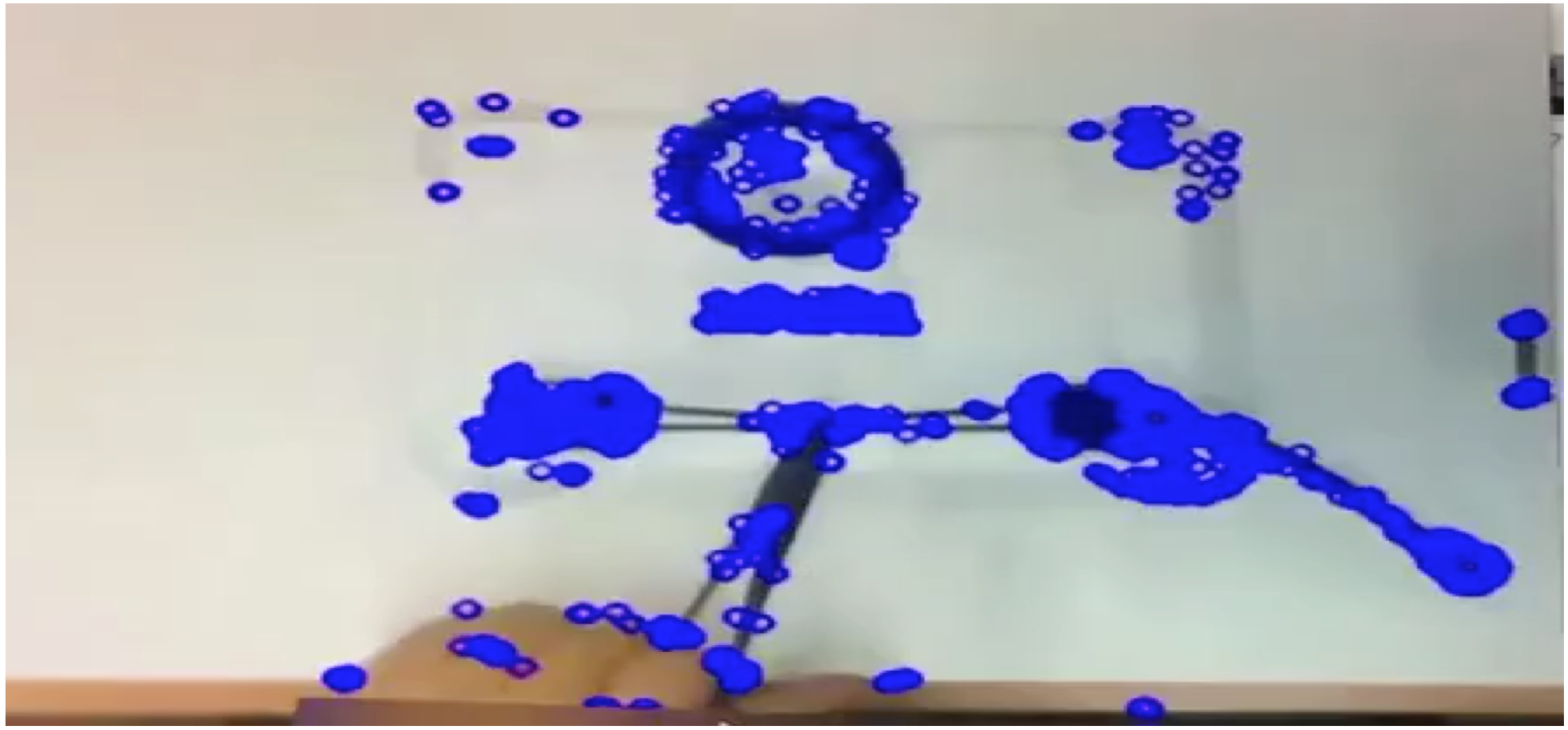


NOTE: This thesis take reference of paper Video Based Assessment of OSATS Using Sequential Motion Textures by Irfan Essa.

Harris 3D on video

Feature Extraction- Harris Corner Detection per video

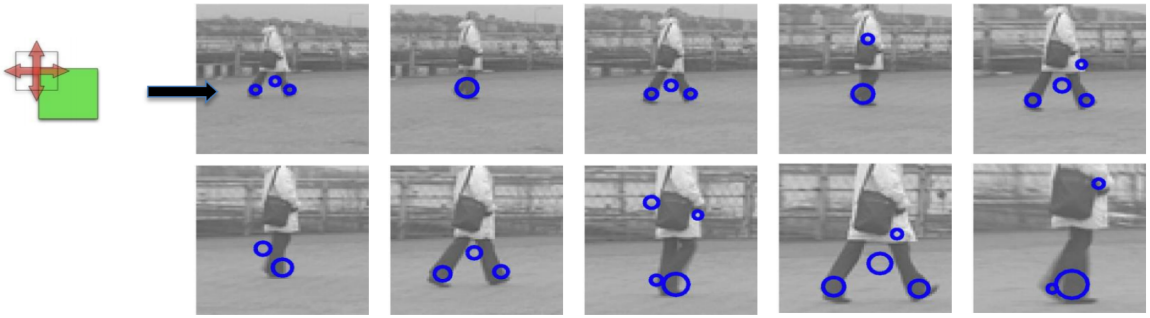
- As we got large number of extra points while doing per frame, implemented Harris Corner Detection on video to identify interest points.
- The results showed better results from previous experiment but still not efficient for analysis.



Implementation

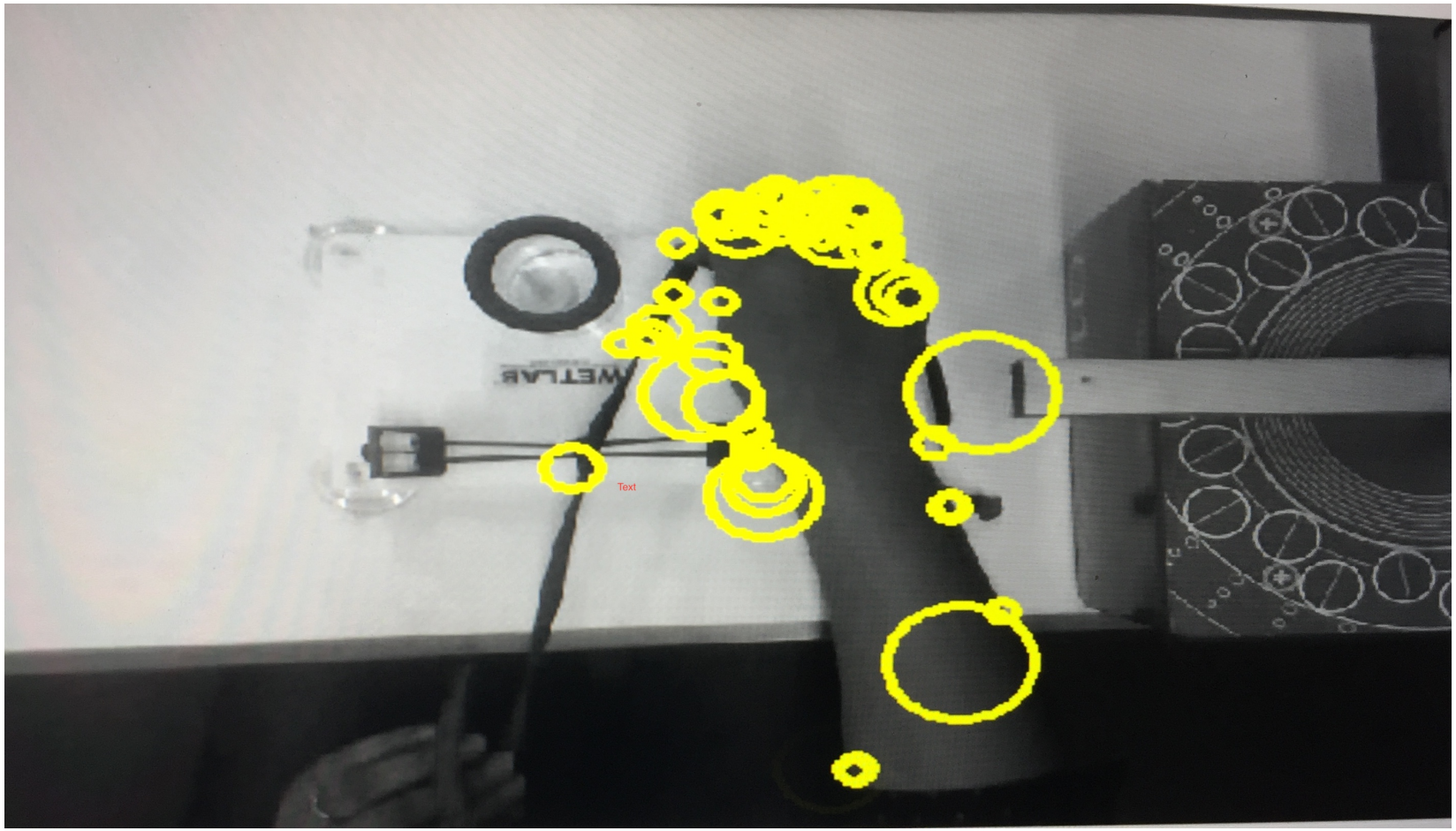
Spatio Temporal Interest Points

- Laptev developed STIP code in 2005, to identify the STIPs**(Spatio Temporal Interest Points)** and computes corresponding local space-time descriptors.
- STIPS is the extended version of spatio domain with respect to temporal domain.
- Spatio-temporal corners are located in region that exhibits a high variation of image intensity in all three directions (x, y , t). This requires that spatio-temporal corners are located at spatial corners such that they invert motion in two consecutive frames (high temporal gradient variation) [2]
- They are identified from local maxima of a cornerness function computed for all pixels across spatial and temporal scales.[2]

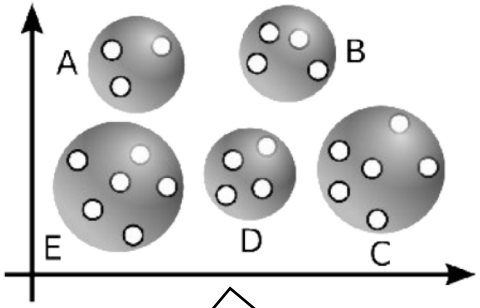


[2] http://www.micc.unifi.it/seidenari/wp-content/uploads/2010/01/A51-Spatio-temporal-features1.pdf

Spatio Temporal Interest Points on video

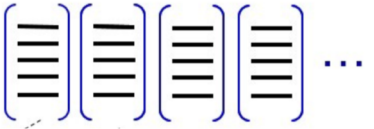


Methodology



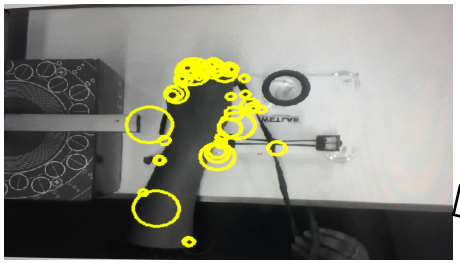
K-means

Back Hand Throw



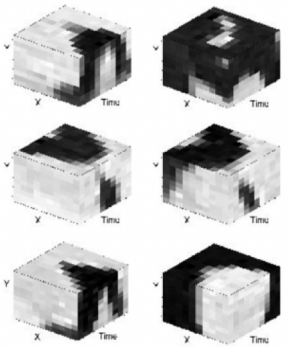
Feature Vector

Fore Finger Throw



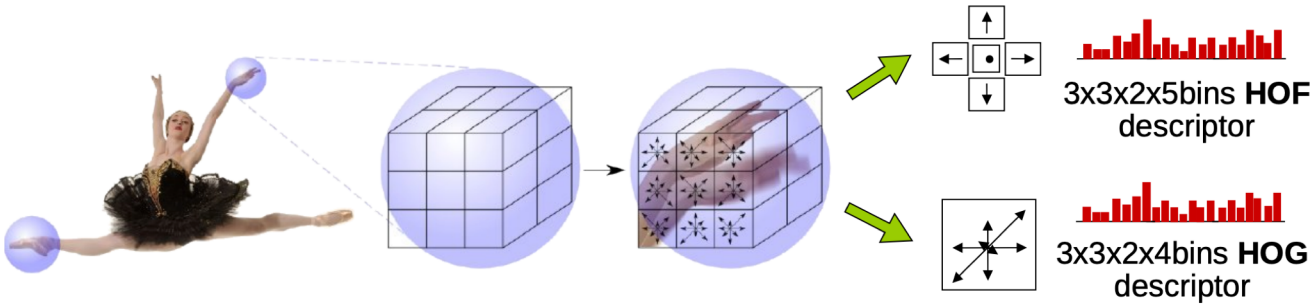
STIPs

Space-time patches



STIP- Result

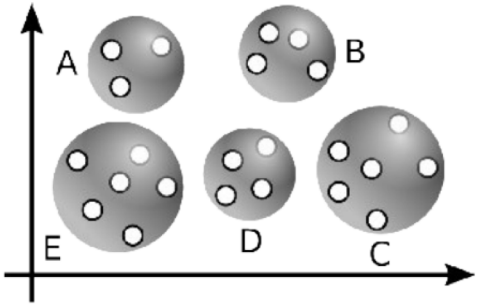
To obtain the frame kernel matrix, we detect the spatio-temporal interest points (STIP) using Laptev detector. We compute the HOG (Histogram of Oriented Gradients) and HOF (Histogram of Optical Flow) on a 3D video patch around each detected STIP to get a 162 element HOG-HOF descriptor.



Computation of HOG-HOF descriptor[3]

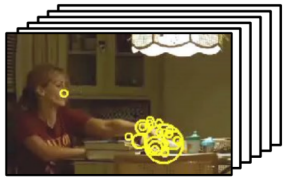
[3] https://pdfs.semanticscholar.org/4e7e/b4a7a21ed94177e65c0bea270c997d379410.pdf

Methodology

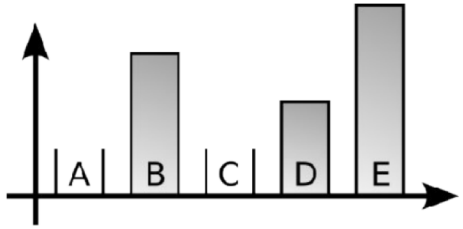


K-means

Classification with SVM with kernel = linear, poly and rbf



Each feature vector is assigned to its closest cluster center (visual word)



An entire video sequence is represented as occurrence histogram of visual words

Implementation

- We collect all the detected STIPs and their corresponding HOG-HOF descriptors from two every video. We classify the STIPs into k distinct clusters by applying k-means clustering to the HoG-HoF descriptors.
- Each cluster of points represents a distribution for a particular motion class in the data.
- We choose k=20 clusters for better results.
- Each feature vector is assigned to its closest cluster center.
- An entire video sequence is represented as occurrence histogram of visual words which is also known as BoW (Bag of Words).
- Finally, SVM is used for classification between two classes bht and fft which is labelled 0 and 1 respectively.

Results

Algorithm SVM kernel	Accuracy-Harris3D	Precision-recall Score (Harris3D)	Accuracy-Dense	Precision-recall Score (Dense)
Linear	76%	0.73	46%	.50
rbf	41%	0.59	49%	.50
Poly	71%	0.68	37%	.50

Conclusion

- Linear kernel for Harris3D detector outperforms while 49% accuracy is obtained with rbf kernel in Dense detector.
- In the paper, Video Based Assessment of OSATS Using Sequential Motion Textures by Irfan Essa accuracy is calculated by using OSATS results with different techniques like Motion texture, Bag-of-words etc from which SMT(Sequential Motion Texture) approach outperforms.
- Knot-tying is a complex task. Analysis of hand gestures is not easy task due to occlusions.
- The results obtained can be better for the different technique used by Irfan Essa in the paper. The accuracy can be significantly increased using own a good quality of training data.

Refrences

[1] https://en.wikipedia.org/wiki/Intel RealSense

[2] http://www.micc.unifi.it/seidenari/wp-content/uploads/2010/01/A51-Spatio-temporal-features1.pdf

[3]https://pdfs.semanticscholar.org/4e7e/b4a7a21ed94177e65c0bea270c997d379410.pdf

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Thank You