# Improved CNN-based Marker Labeling for Optical Hand Tracking

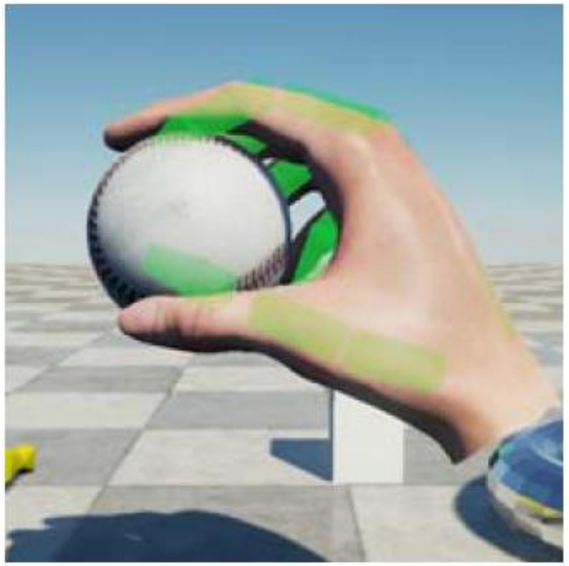
Janis Rosskamp[1], Rene Weller[1], Thorsten Kluss[2], Jaime L. Maldonado C[2]., Gabriel Zachmann[1]

[1]University of Bremen, Computer Graphics and Virtual Reality, Germany
[2]University of Bremen, Cognitive Neuroinformatics, Germany

*EuroVR 2020 – Scientific track*
*25-27 November - Valencia, Spain*

# Motivation

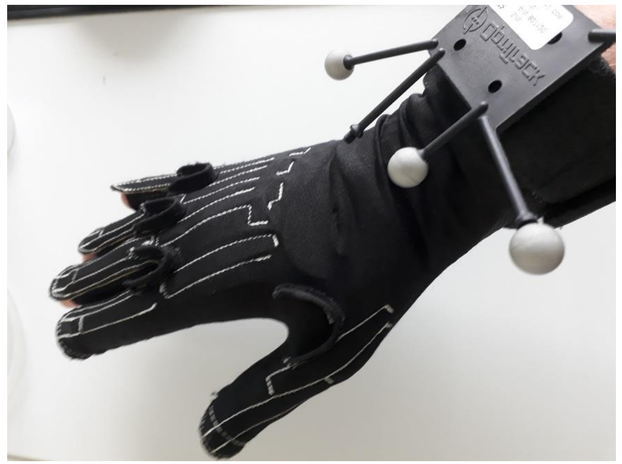
- Accurate hand tracking for
  - Physically-based grasping
  - Medical training
  - Human-to-robot transfer
    - Trajectories of finger motion
    - Heatmaps of contact points



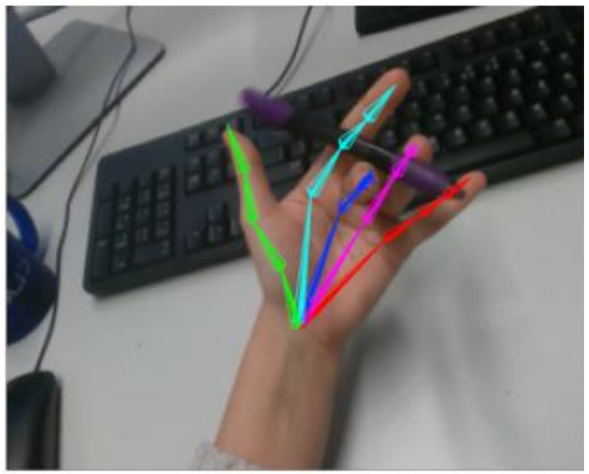
Verschoor et al. 2018

# Hand Tracking Methods

## Cyberglove



## RGB



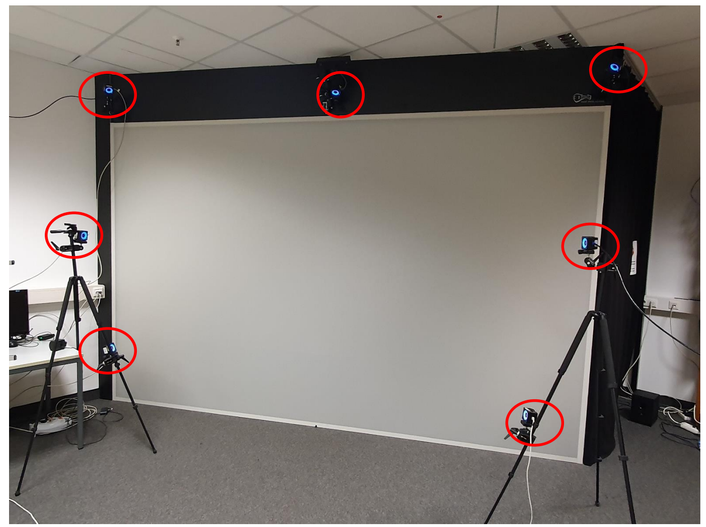
Mueller et al., 2018

## Active Markers

Pavllo et al., 2018

# Optical Marker-based Tracking

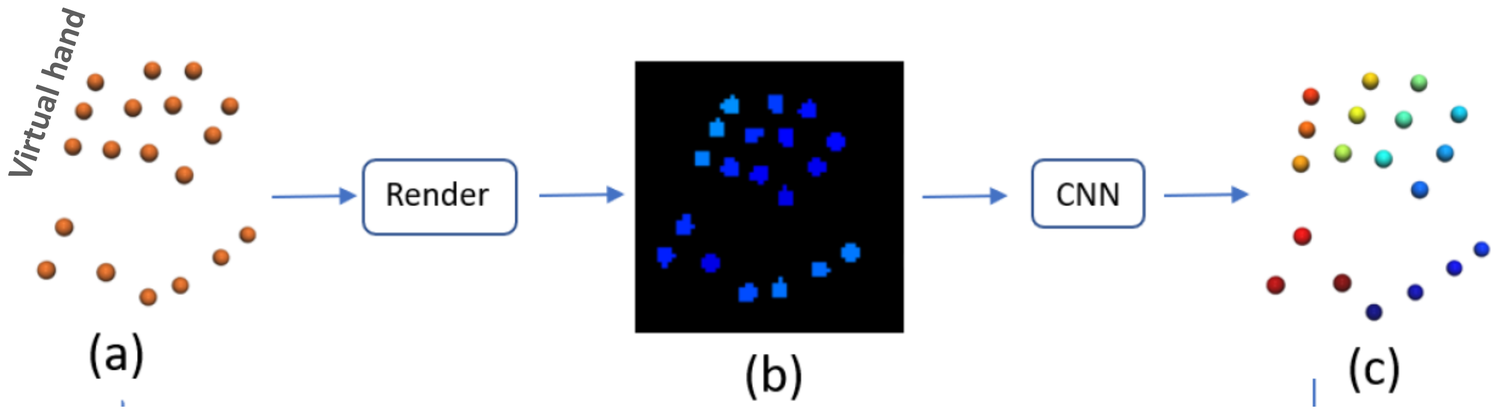
- Sub-millimeter accuracy
- Less invasive than active markers
- Challenges: Occlusions require relabeling



```
PathFileType    4   (X/Y/Z) C:/Users/stud10/golubev-hand-trackin
DataRate    CameraRate  NumFrames   NumMarkers  Units   OrigData
30  30  450 28  cm  30  0   450
Frame#  Time    Unlabeled 1660          Unlabeled 1661
        X1  Y1  Z1  X2  Y2  Z2  X3  Y3  Z3  X4  Y4  Z4  X5  Y5
0   0   48.8657 155.632 111.994 47.9975 156.13  114.748 49.8799
1   0.0333333   48.8892 155.763 111.924 48.0074 156.278 114.656
2   0.0666667   48.9287 155.87  111.869 48.0506 156.394 114.595
3   0.1 48.9924 155.977 111.833 48.1249 156.477 114.585 49.9951
4   0.133333    49.0044 156.068 111.782 48.1287 156.596 114.506
5   0.166667    49.0279 156.179 111.722 48.1459 156.713 114.446
6   0.2 49.0587 156.252 111.691 48.1701 156.789 114.403 50.1329
7   0.233333    49.1107 156.336 111.657 48.204  156.877 114.371
8   0.266667    49.1683 156.389 111.65  48.2806 156.925 114.375
9   0.3 49.2779 156.404 111.713 48.395  156.928 114.437 50.3643
10  0.333333    49.3965 156.447 111.755 48.519  156.962 114.48
11  0.366667    49.5446 156.527 111.784 48.647  157.029 114.509
12  0.4 49.6932 156.597 111.819 48.8046 157.122 114.536 50.7336
13  0.433333    49.8329 156.703 111.838 48.9494 157.23  114.558
14  0.466667    50.0036 156.818 111.864 49.1166 157.354 114.601
15  0.5 50.1768 156.931 111.914 49.3097 157.466 114.642 51.2354
16  0.533333                    49.4926 157.609 114.675 51.3995 152.
17  0.566667                    49.6966 157.777 114.695 51.5983 152.
18  0.6                         51.7564 152.719 103 60.2852
```

# Related Work

- Sparse marker sets
  - Real-time inverse kinematics (Maycock et al., 2015)
  - Gaussian mixture models (Alexanderson et al., 2017)
- Labeling of dense marker sets (Han et al., 2018)

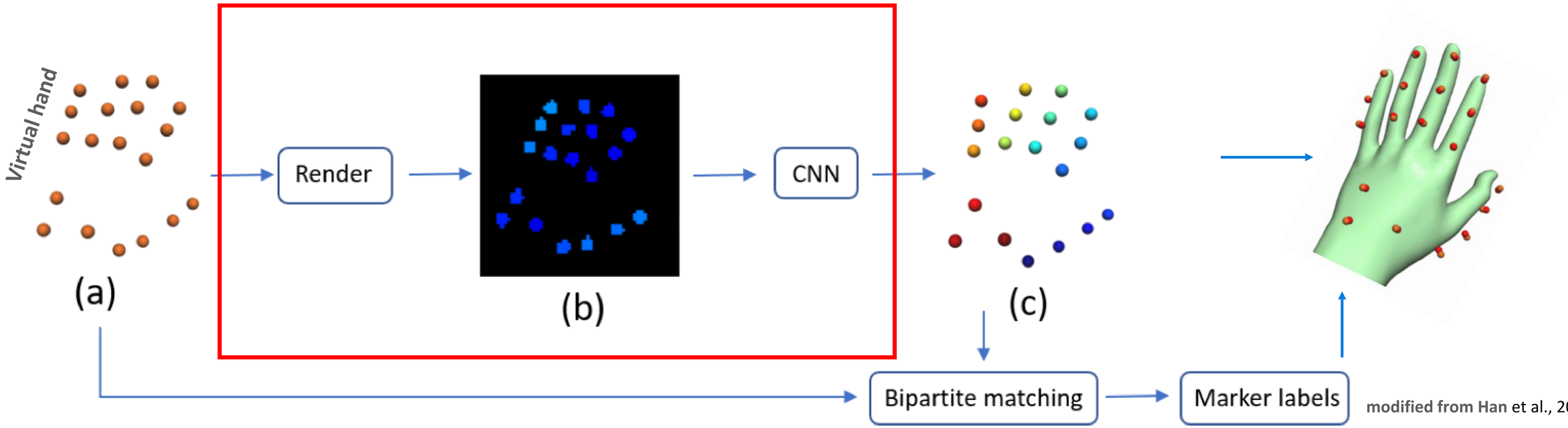



Han et al., 2018

Lecture Notes in Computer Science
LNCS LNAI LNBI
Springer

# Contribution

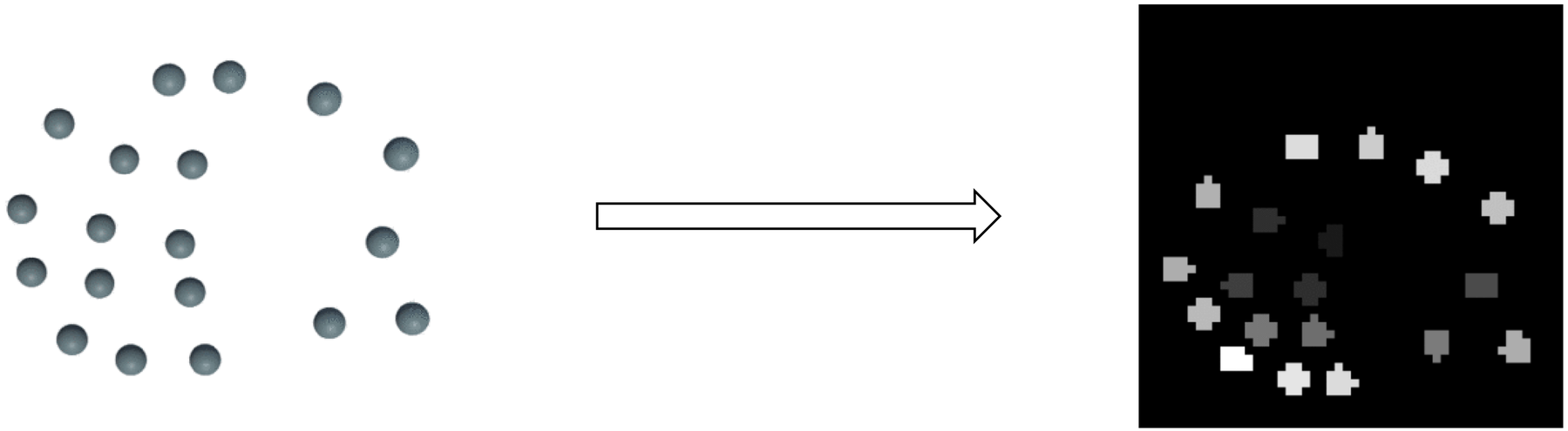
- Improvement of current state-of-the-art labeling
  - Modified depth images
  - Retraining of CNN




modified from Han et al., 2018

# Depth Image Generation

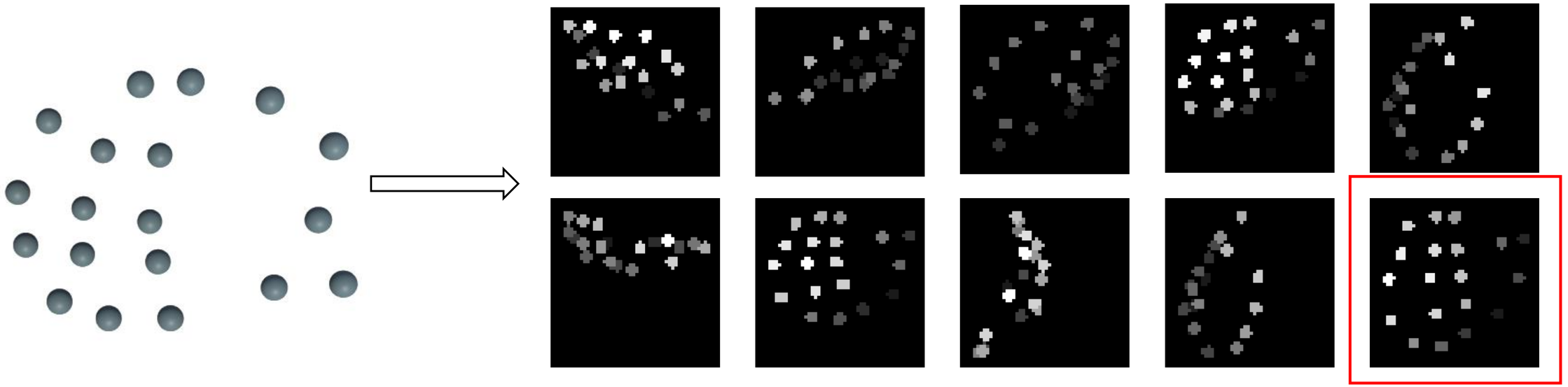
- Find projection axis for orthographic projection
- Values along the axis are normalized between [0.1,1] (depth value)
- Splatting to preserve relative depth



Lecture Notes in Computer Science
LNCS LNAI LNBI
Springer

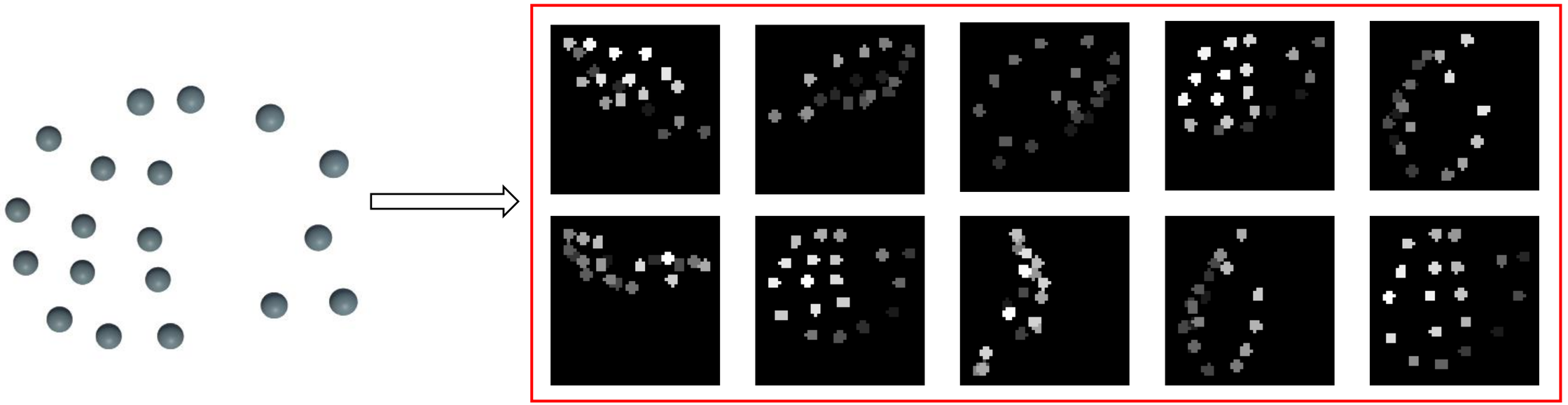
# Random Projection Axis (RPA)

- Idea: Use random projection axis (Han et. al)
  - Generate 10 random images and select the one with highest spatial spread (highest eigenvalues of covariance matrix)



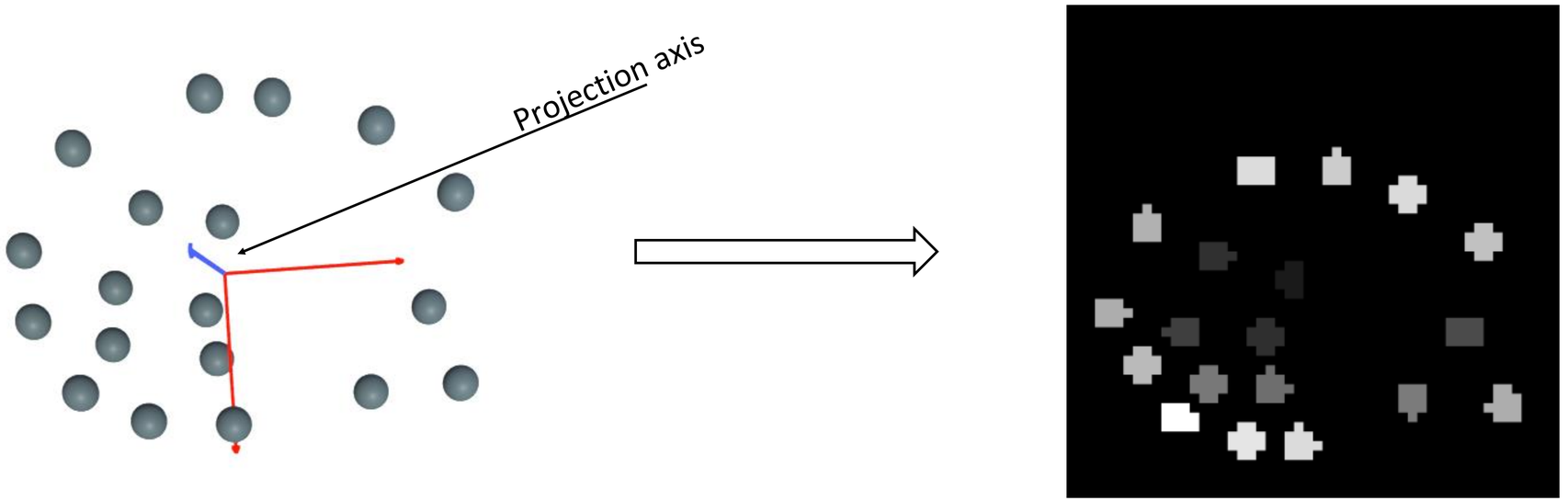
# Multi Images – Minimal Cost (Multi)

- Idea: Multiple marker matches and select best match
  - Generate multiple random projection axis and images
  - Match all images and select the one with lowest matching cost



# Principal Component Analysis (PCA)

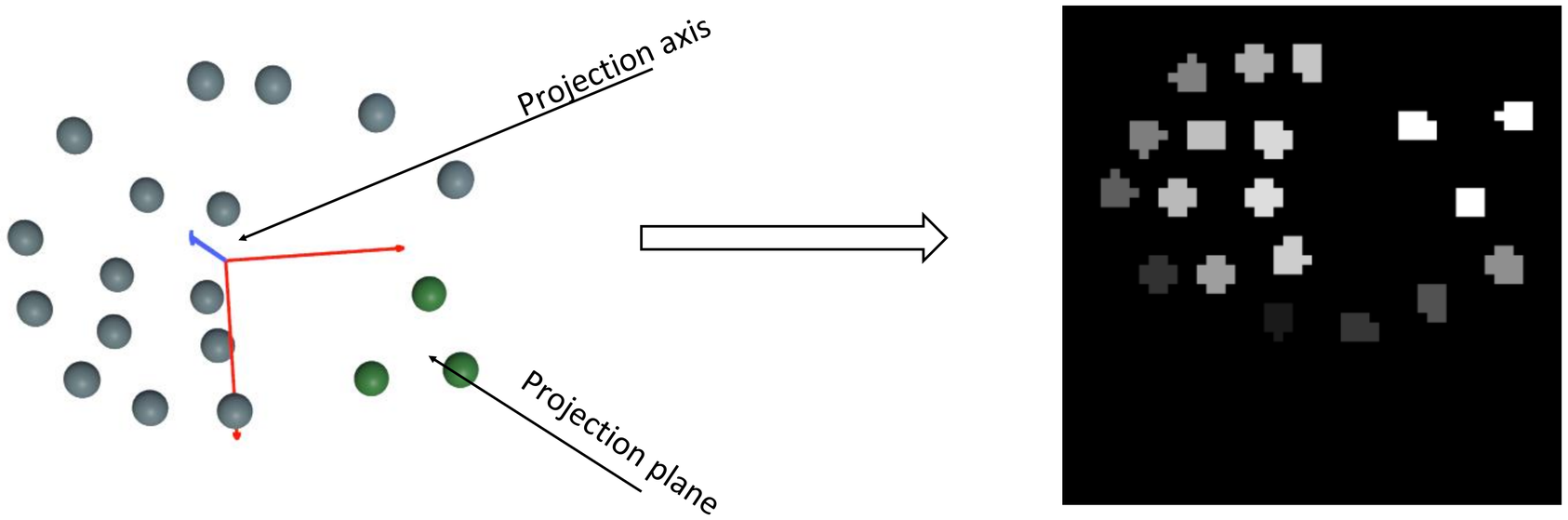
- Idea: Create an image with high spatial spread
  - Get principal axes of the 3D point cloud
  - Use principal axis with lowest eigenvalue as projection axis




Lecture Notes in Computer Science
LNCS LNAI LNBI
Springer

# Palm Prediction Method (PalmP)

- Idea: Similar images independent of hand pose
  - Projection axis perpendicular to the palm's orientation
  - Palm orientation from rigid markers




Lecture Notes in Computer Science
LNCS LNAI LNBI
Springer

# CNN

- Training set[1] of 168691 frames provided by Han et al.
- VGG-style neural network
- Retraining of provided CNN[1] for PCA & PalmP for improved results
  - 137357 frames for training & 31.334 frames for validation
  - Improves accuracy up to 20 percent points

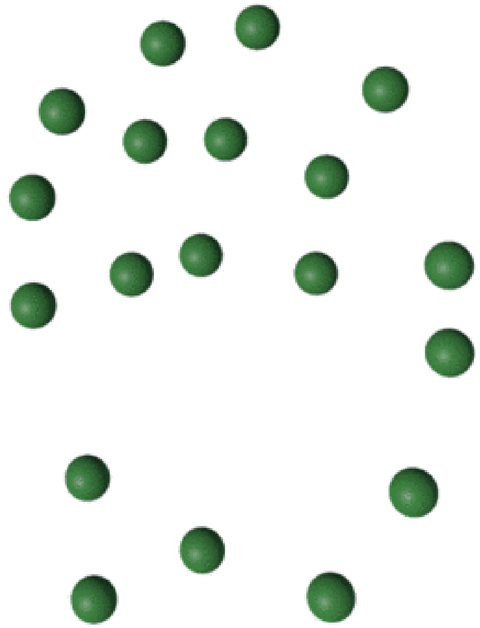
| Layer id | Type | Filter shape | Input size |
|---|---|---|---|
| 1 | Conv + BN + ReLU | $64 \times 3 \times 3$ | $1 \times 52 \times 52$ |
| 2 | Conv + BN + ReLU | $64 \times 3 \times 3$ | $64 \times 50 \times 50$ |
| 3 | Maxpool | $2 \times 2$ | $64 \times 48 \times 48$ |
| 4 | Conv + BN + ReLU | $128 \times 3 \times 3$ | $64 \times 24 \times 24$ |
| 5 | Conv + BN + ReLU | $128 \times 3 \times 3$ | $128 \times 22 \times 22$ |
| 6 | Conv + BN + ReLU | $128 \times 3 \times 3$ | $128 \times 20 \times 20$ |
| 7 | Maxpool | $2 \times 2$ | $128 \times 18 \times 18$ |
| 8 | Reshape | N/A | $128 \times 9 \times 9$ |
| 9 | FC + ReLU | $2048 \times 10368$ | 10368 |
| 10 | FC | $2048 \times 57$ | 2048 |
| 11 | Reshape | N/A | 57 |

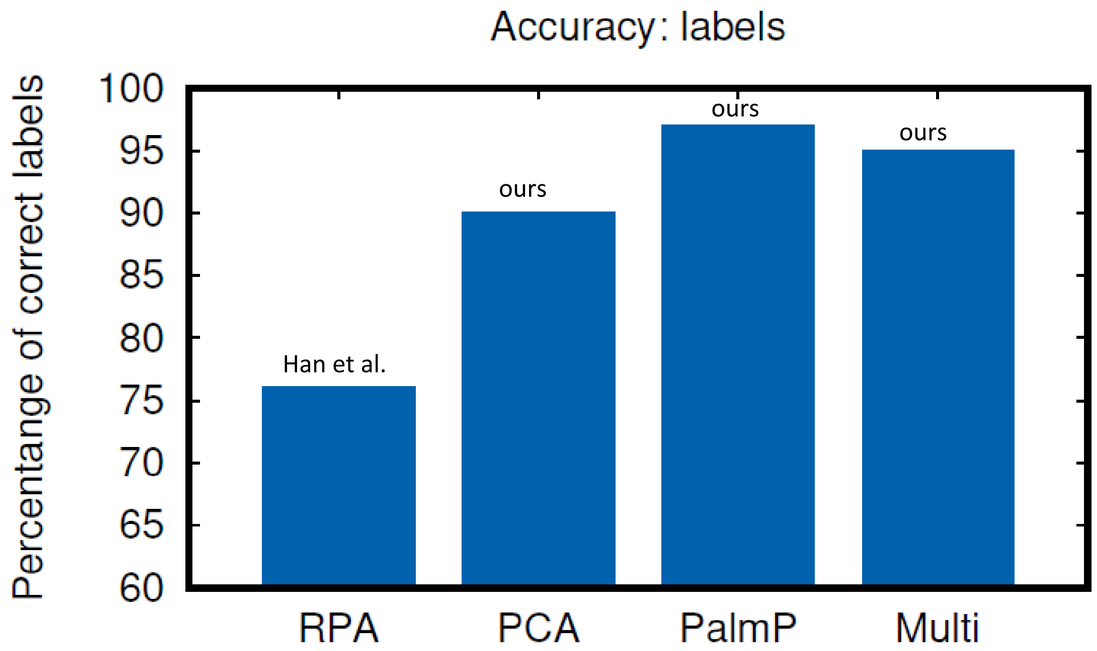
Han et al., 2018

[1] https://github.com/Beibei88/Mocap_SIG18_Data

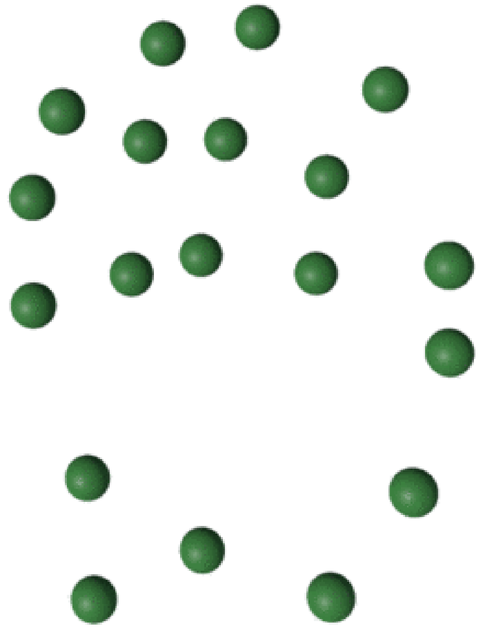
Lecture Notes in Computer Science
LNCS LNAI LNBI
Springer

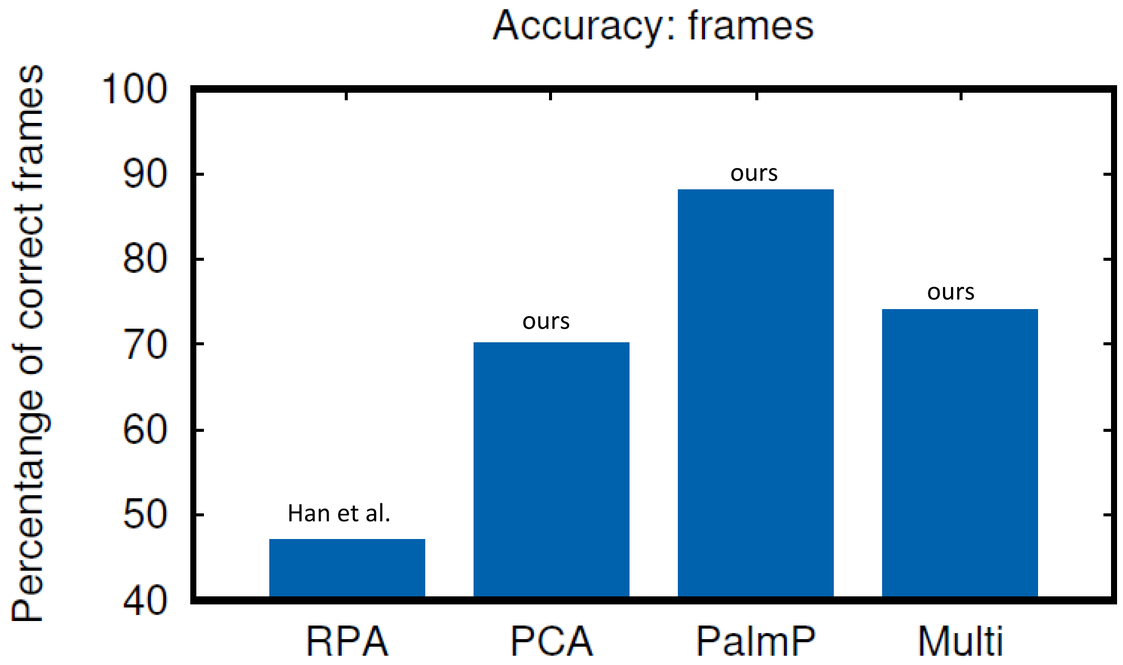
# Results – Label Prediction





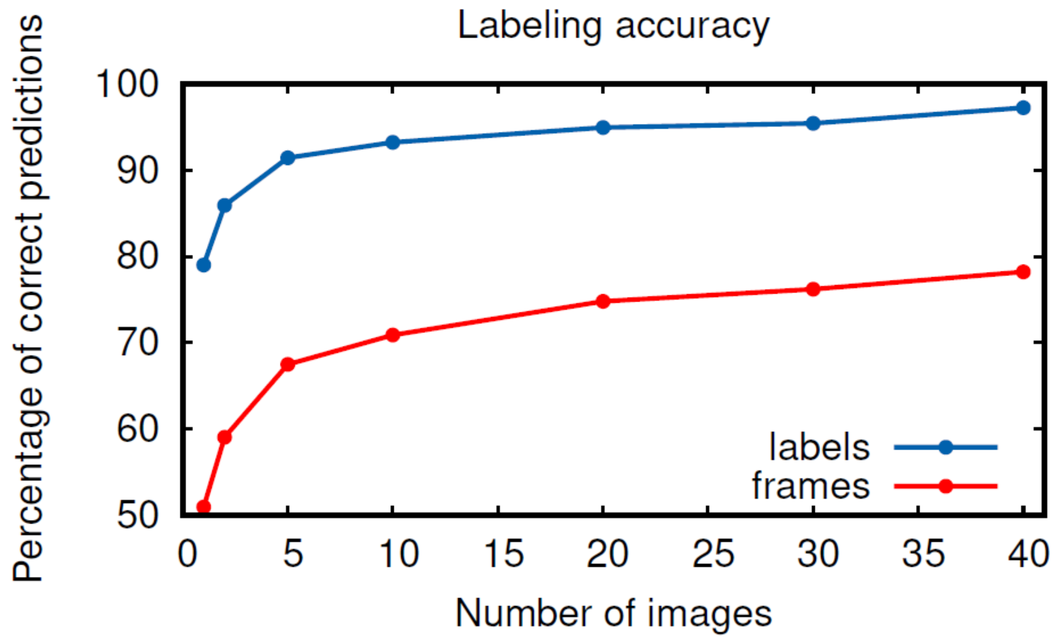
# Results – Frame Prediction

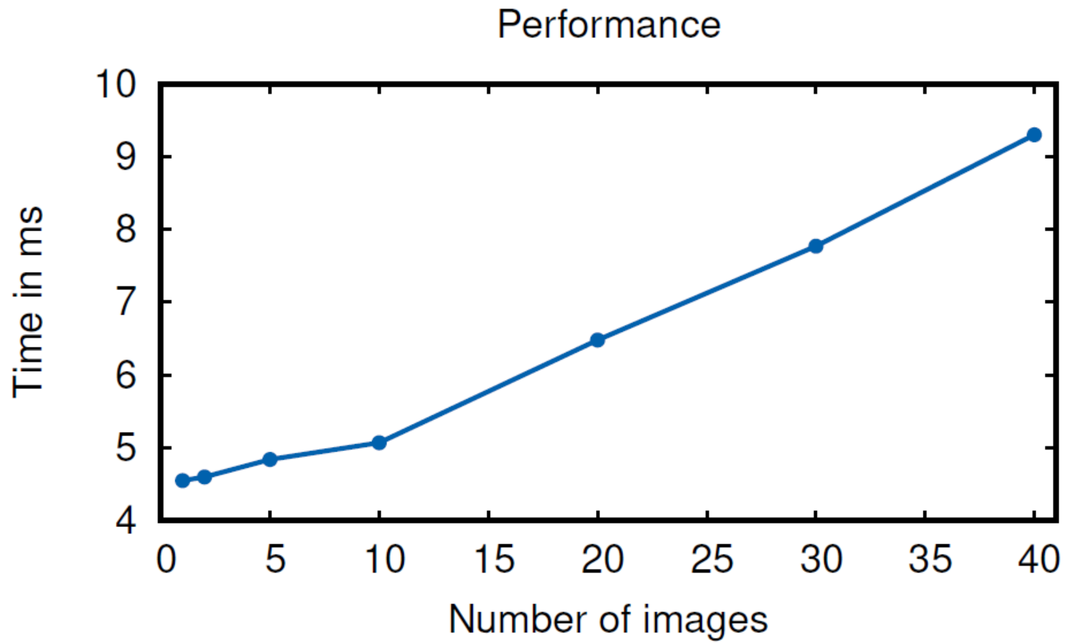




# Results – Multi

- How many images can be used for the Multi method?



# Conclusion

- Current state-of-the-art labeling up to 40 percent points improved
- Multiple methods depending on use-case
  - Multi: Independent of marker set but multiple CNN passes are necessary
  - PCA: Independent of marker set and fast
  - PalmP: Prior knowledge required (in our case the palm)