
# ChaLearn Multi-Modal Gesture Recognition 2013: Grand Challenge and Workshop Summary

http://gesture.chalearn.org/

Sergio Escalera, Jordi Gonzàlez, Xavier Baró, Miguel Reyes, Isabelle Guyon, Vassilis Athitsos, Hugo J. Escalante, Leonid Sigal, Antonis Argyros, Cristian Sminchisescu, Richard Bowden, Stan Sclaroff

# Context of the Workshop

**ChaLearn Gesture Recognition Challenges and Workshops**

**CVPR 2011** - Workshop and Challenge on Gesture Recognition

**CVPR 2012** - Workshop and Challenge on Gesture Recognition

**ICPR 2012 -** Workshop and Challenge on Gesture Recognition

- **Quantitative competition:**
- One-shot learning
- New depth-rgb data set
- Dictionaries among 5-8 gesture categories
- Leveinstein: recognizing list of sequences within each sequence

**ICMI 2013 - Workshop and Challenge on Gesture Recognition**

- **Quantitative competition:**
- User independent multiple instance learning
- New depth-rgb-mask-skeleton-audio data set
- Dictionary of 20 gesture categories
- Leveinstein: recognizing list of sequences within each sequence

**JMLR Special Topic on Gesture Recognition: Deadline 15/2/14**

Gesture Recognition 2014 - Workshop and Challenge on Gesture Recognition

- **Quantitative competition:**
- Improved Ground truth definition at frame level
- Gesture spotting
- Begin-end gesture recognition (overlapping basis)

- **To be annunced on January 2014**

# Workshop program

**9 accepted papers split into different workshop tracks:**

- Multi-modal Gesture Recognition Challenge 2013: Dataset and Results
- Multi-modal Gesture Recognition Challenge I and award ceremony
- Multi-modal Gesture Recognition Challenge II
- Challenge for Multimodal Mid-Air Gesture Recognition for close HCI (organized by Simon Ruffieux)
- Multi-modal Gesture Recognition Applications

**Four invited speakers:**

Invited speaker I: Professor Leonid Sigal, Disney Research
Title: Action Recognition and Understanding: Latest Challenges and Opportunities

Invited speaker II: Professor Cristian Sminchisescu, Lund University
Title: Human Actions and 3D Pose in the Eye: From Perceptual Evidence to Accurate Computational Models

Invited speaker III: Professor Antonis Argyros, Univ. of Crete, Institute of Computer Science
Title: Tracking the articulated motion of human hands

Invited speaker IV: Professor Richard Bowden, University of Surrey
Title: Recognising spatio-temporal events in video

Special thanks to Professor Stan Sclaroff, Boston University, Associate Editor in Chief of IEEE Transactions on Pattern Analysis and Machine Intelligence

# Challenge organization

## Multi-modal ChaLearn Gesture Recognition Challenge and Workshop

**http://gesture.chalearn.org/**
**sunai.uoc.edu/chalearn**

**Web of the competition**
**Data**

The emphasis of this edition of the competition will be on multi-modal automatic learning of a **vocabulary of 20 types of Italian anthropological/cultural gestures performed by different users.**

- **User independent continuous gesture recognition combined with audio information**.
- Multi-modal dataset recorded with **Kinect** (providing RGB images, depth images, skeleton information, joint orientation and audio sources)
- **13,858 labeled Italian gestures from near 30 users.**

# Gesture categories (1/2)

(1) *Vattene*

(2) *Viene qui*

(3) *Perfetto*

(4) *E un furbo*

(5) *Che due palle*

(6) *Che vuoi*

(7) *Vanno d'accordo*

(8) *Sei pazzo*

(9) *Cos hai combinato*

(10) *Nonme me friega niente*

# Gesture categories (2/2)

(11) *Ok*

(12) *Cosa ti farei*

(13) *Basta*

(14) *Le vuoi prendere*

(15) *Non ce ne piu*

(16) *Ho fame*

(17) *Tanto tempo fa*

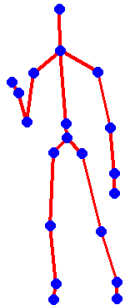
(18) *Buonissimo*

(19) *Si sono messi d'accordo*

(20) *Sono stufo*

# Data and modalities



- Framerate 20FPS
- RGB: 640x480
- Depth: 640x480
- Audio: Kinect 20 michropone array
- Users: 27
- Italians: 81%

- Total number of sequences: 956 ϵ [1,2] min.
- Total number of gestures: 13,858
- Total number of frames: 1.720.800
- Noisy gestures

**Data structure information:** *S. Escalera, J. Gonzàlez, X. Baró, M. Reyes, O. Lopes, I. Guyon, V. Athistos, H.J. Escalante, "Multi-modal Gesture Recognition Challenge 2013: Dataset and Results", ICMI 2013*.

# MMGR Challenge
## Categories and data modalities

# MMGR Challenge

Easy and challenging aspects of the data.

| Easy |
| --- |
| Fixed camera |
| Near frontal view acquisition |
| Within a sequence the same user |
| Gestures performed mostly by arms and hands |
| Several available modalities: audio, skeletal model, user mask, depth, and RGB |
| Several instances of each gesture for training |

| Challenging |
| --- |
| *Within each sequence*: |
| Continuous gestures without a resting pose |
| Many gesture instances are present |
| Distracter gestures out of the vocabulary may be present in terms of both gesture and audio |
| *Between sequences*: |
| High inter and intra-class variabilities of gestures in terms of both gesture and audio |
| Variations in background, clothing, skin color, lighting, temperature, resolution |
| Some parts of the body may be occluded |

# Schedule

- **April 30th, 2013**: Beginning of the challenge competition, release of first data examples.
- **May 20th, 2013**: Full release of training and validation data. Training data with ground truth labels.
- **August 1st, 2013**: Encrypted Final evaluation data and ground truth labels for the validation data are made available.
- **August 15th, 2013**: End of the challenge competition. Deadline for code submission. The organizers start the code verification by running it on the final evaluation data and obtaining the team scores.
- **August 25th, 2013**: Deadline for fact sheets.
- **September 1st, 2013**: Release of the verification results to the participants for review.

| | # Sequences | # Gesture samples |
|---|---|---|
| Development | 393 | 3362 |
| Validation | 287 | 7754 |
| Test | 276 | 2742 |

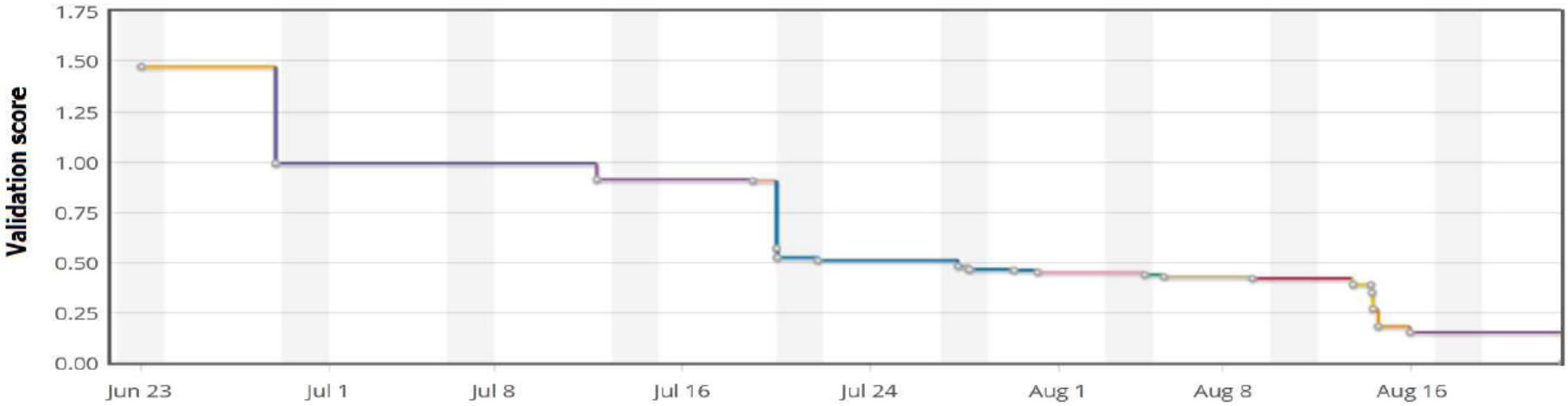
# Evaluation metric and participant entries

- For each unlabeled video, the participants were instructed to provide an ordered **list of labels R** corresponding to the recognized gestures.
- **We compared this list with the truth labels *T*** i.e. the prescribed list of gestures that the user had to play during data collection.
- **We computed the Levenshtein distance *L(R,T)***, that is the minimum number of edit operations (substitution, insertion, or deletion) that one has to perform to go from R to *T* (or vice versa).
- **The overall score** is the sum of the Levenshtein distances for all the lines of the result file compared to the corresponding lines in the truth value file, **divided by the total number of gestures in the truth value file.**

kaggle

$$L([124],[32]) = 2,$$
$$L([1],[2]) = 1,$$
$$L([222],[2]) = 2.$$

# Evaluation metric and participant entries




Best public score obtained in the validation set during the Challenge.

kaggle

$$L([124],[32]) = 2,$$

$$L([1],[2]) = 1,$$

$$L([222],[2]) = 2.$$

# Results

- Participation

• The challenge attracted high level of participation, with a total of **54 teams and near 300 total number of entries**.

• Finally, **17 teams successfully submitted their prediction in final test set, while providing also their code for verification and summarizing their method by means of a fact sheet questionnaire**.

• After verifying the codes and results of the participants, the final scores of the top rank participants on both validation and test sets were made public.

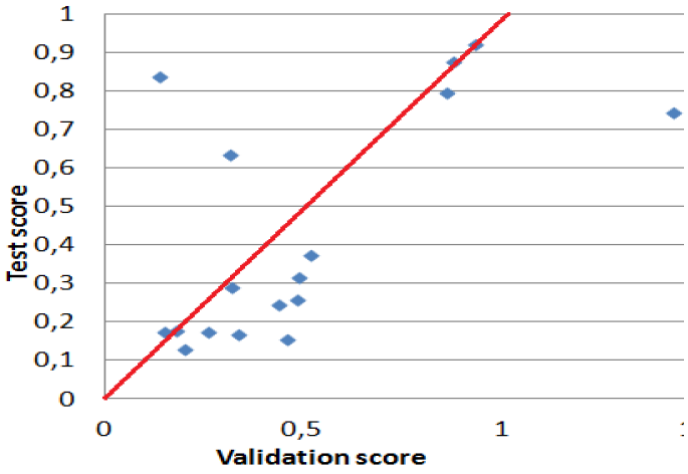
• In the end, **the final error rate on the test data set was around 12%.**

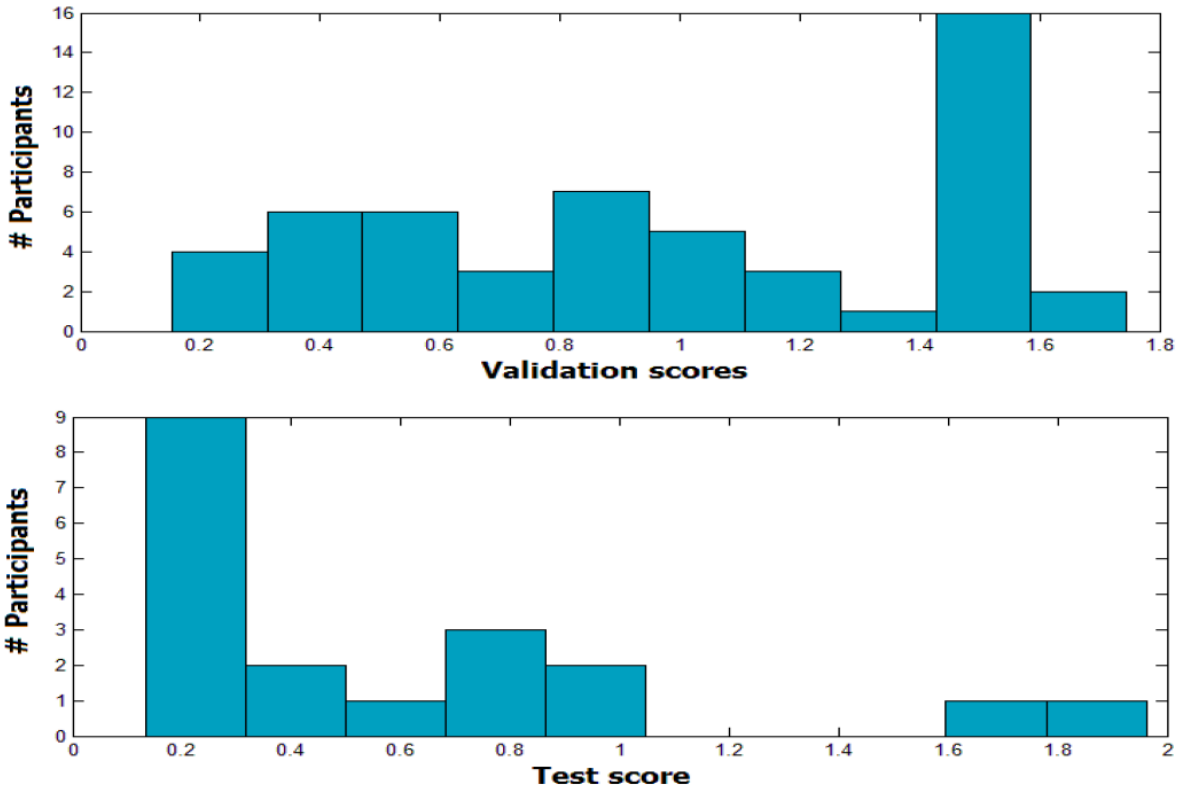
Top rank results on validation and test sets.

| TEAM | Validation score | Test score |
|---|---|---|
| IVA MM | 0.20137 | **0.12756** |
| WWEIGHT | 0.46163 | **0.15387** |
| ET | 0.33611 | **0.16813** |
| MmM | 0.25996 | 0.17215 |
| PPTK | 0.15199 | 0.17325 |
| LRS | 0.18114 | 0.17727 |
| MMDL | 0.43992 | 0.24452 |
| TELEPOINTS | 0.48543 | 0.25841 |
| CSI MM | 0.32124 | 0.28911 |
| SUMO | 0.49137 | 0.31652 |
| GURU | 0.51844 | 0.37281 |
| AURINKO | 0.31529 | 0.63304 |
| STEVENWUDI | 1.43427 | 0.74415 |
| JACKSPARROW | 0.86050 | 0.79313 |
| JOEWAN | 0.13653 | 0.83772 |
| MILAN KOVAC | 0.87835 | 0.87463 |
| IAMKHADER | 0.93397 | 0.92069 |

# MMGR Challenge

## Results



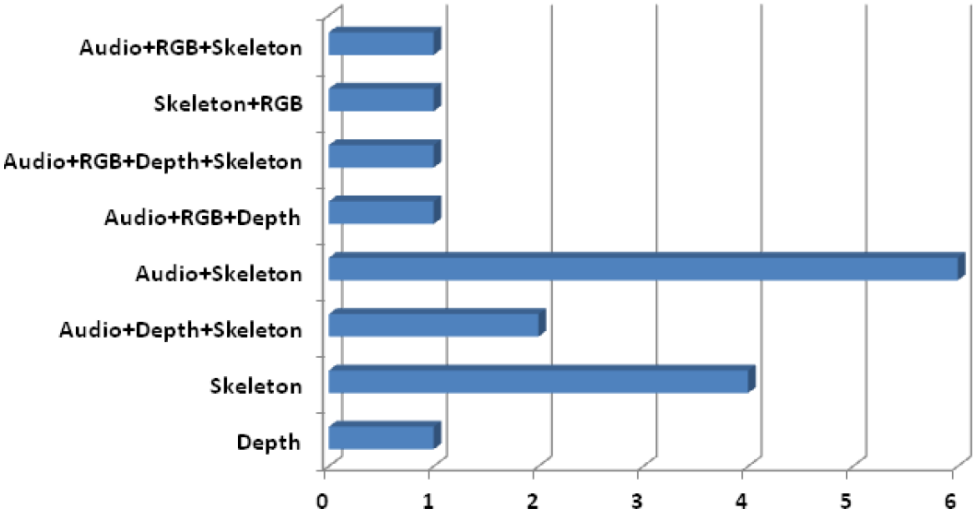




Validation and test scores histograms.

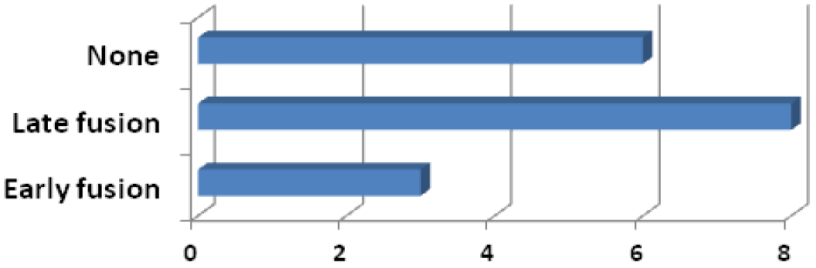
# MMGR Challenge

- Fact sheets statistics

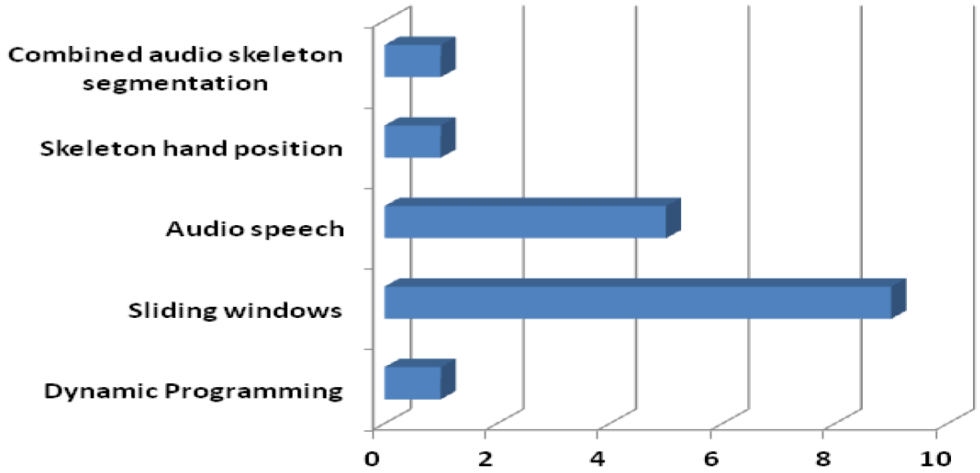
## Results



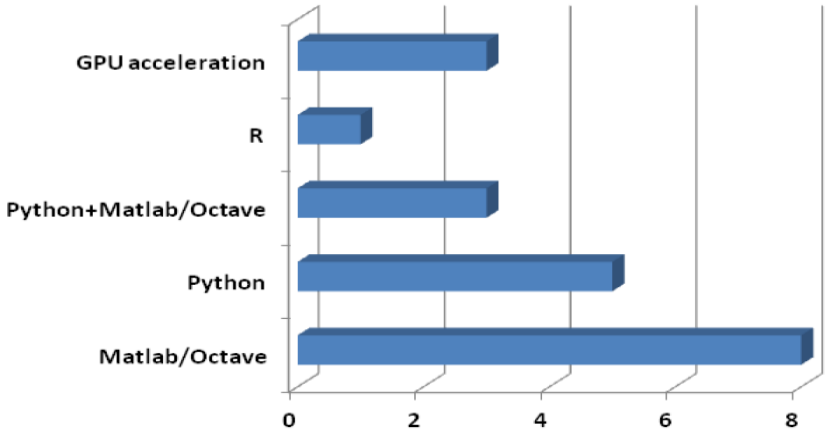

Modalities considered.




Fusion strategy.



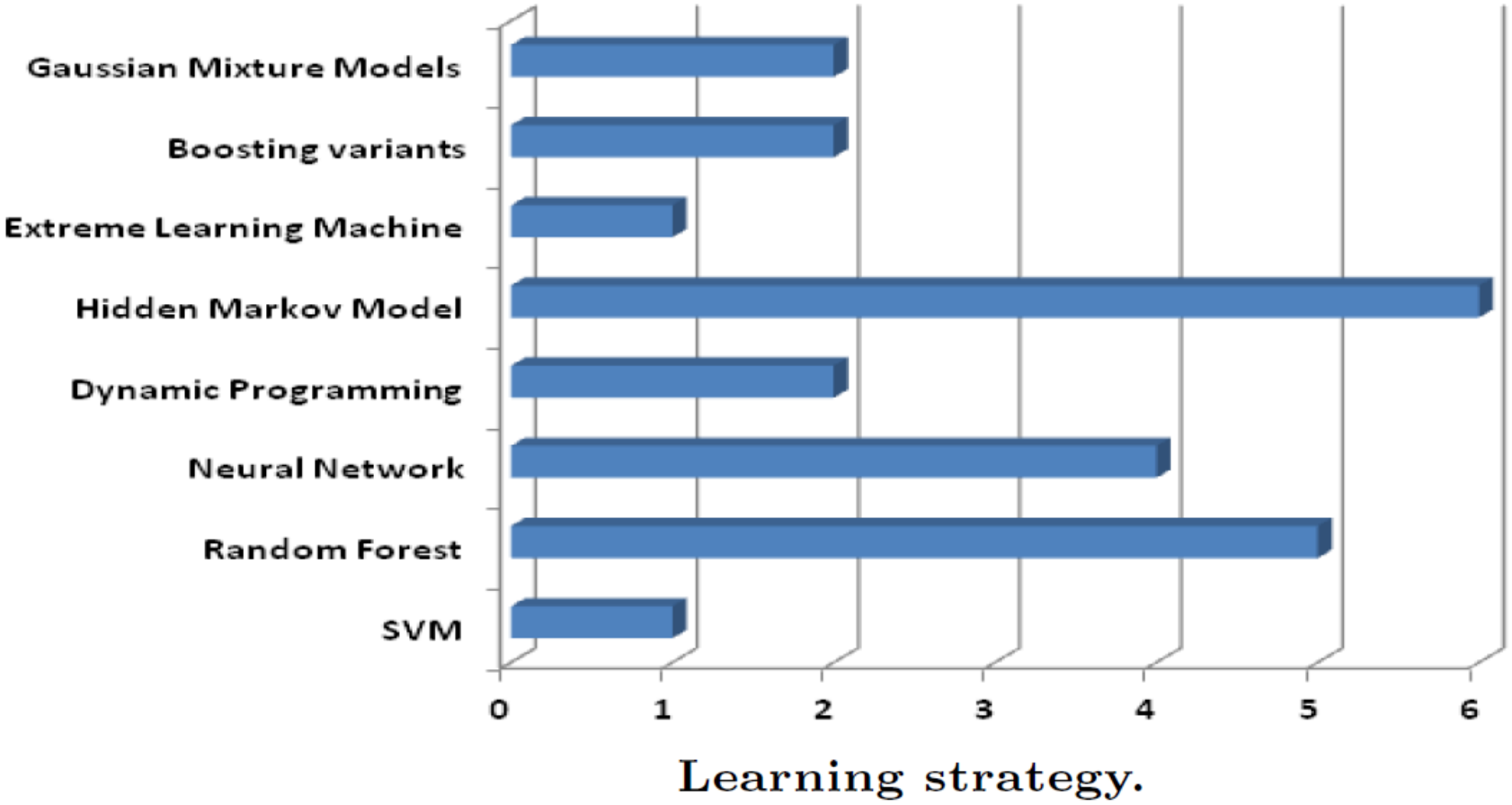

Segmentation strategy.




Programming language.

# Results

- Fact sheets statistics

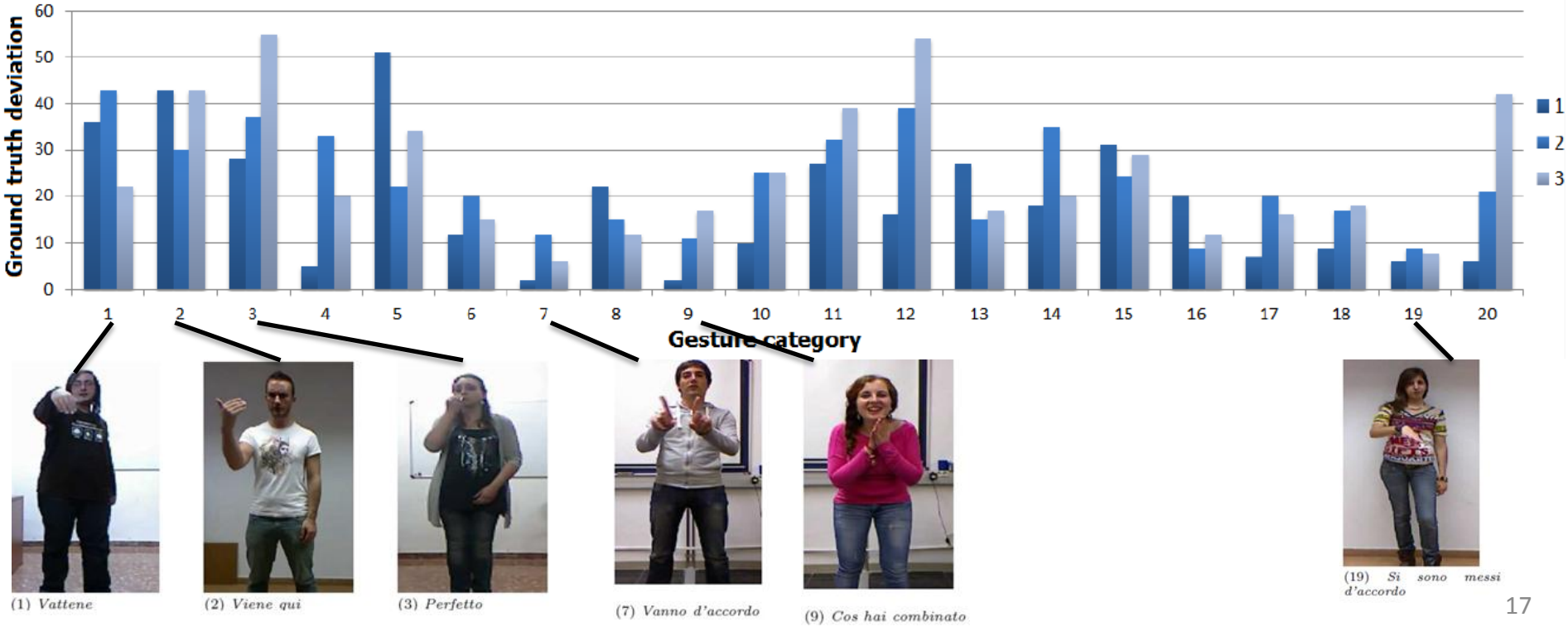


# Results

- Winner methods

| TEAM | Test score | Rank position | Modalities | Segmentation | Fusion | Classifier |
|---|---|---|---|---|---|---|
| IVA MM | 0.12756 | 1 | Audio,Skeleton | Audio | None | HMM,DP,KNN |
| WWEIGHT | 0.15387 | 2 | Audio,Skeleton | Audio | Late | RF,KNN |
| ET | 0.16813 | 3 | Audio,Skeleton | Audio | Late | Tree,RF,ADA |

Correlation of the number of recognized gestures per category and GT gestures (averaged among all sequences)



# Thank you

## Organizers









## Sponsors