HAND GESTURE RECOGNITION WITH LEAP MOTION AND KINECT DEVICES

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

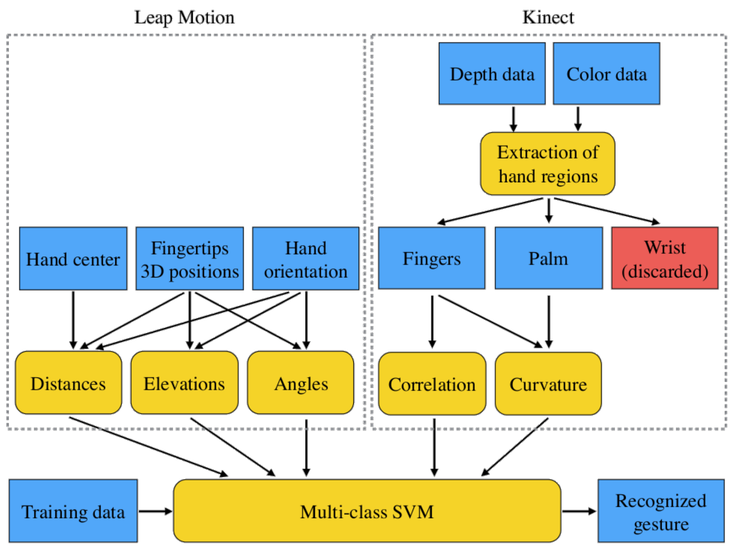
Giulio Marin, Fabio Dominio, Pietro Zanuttigh - University of Padova
{maringiu,dominiof,zanuttigh}@dei.unipd.it

LTTLM

Abstract

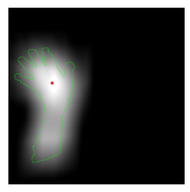
The recent introduction of novel acquisition devices like the Leap Motion and the Kinect allows to obtain a very informative description of the hand pose that can be exploited for accurate gesture recognition. This paper proposes a novel hand gesture recognition scheme explicitly targeted to Leap Motion data. An ad-hoc feature set based on the positions and orientation of the fingertips is computed and fed into a multi-class SVM classifier in order to recognize the performed gestures. A set of features is also extracted from the depth computed from the Kinect and combined with the Leap Motion ones in order to improve the recognition performance. Experimental results present a comparison between the accuracy that can be obtained from the two devices on a subset of the American Manual Alphabet and show how, by combining the two features sets, it is possible to achieve a very high accuracy in real-time.

Gesture recognition pipeline



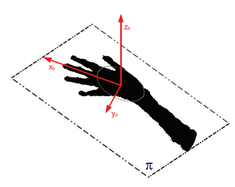
Extraction of the hand region

Extracted hand samples (Green dot: closest sample)



A Gaussian filter with a large standard deviation is applied to the mask and the maximum is selected as the starting point

A circle is fitted on the hand in order to extract the palm, finger and wrist regions



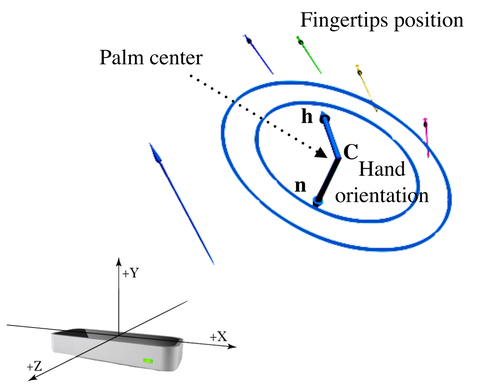
Reference system and palm plane estimated from the PCA output

Leap Motion Hand Data

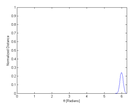
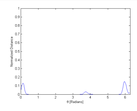
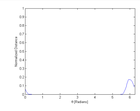
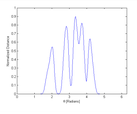
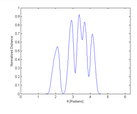
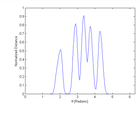
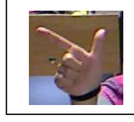
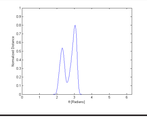
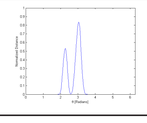
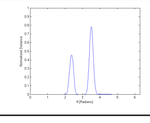
The Leap Motion does not return a complete depth map but only a set of relevant hand points and some hand pose features

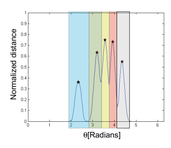
- **Position of the fingertips** ($\mathbf{F}_i$)
- **Palm center** ($\mathbf{C}$)
- **Hand orientation** ($\mathbf{h}$ and $\mathbf{n}$)

Note that the sensor is not always able to recognize all the fingers and can confuse objects with fingers.



Kinect: Distance Features

Gesture	Rep. 1	Rep. 2	Rep. 3
			
			
			

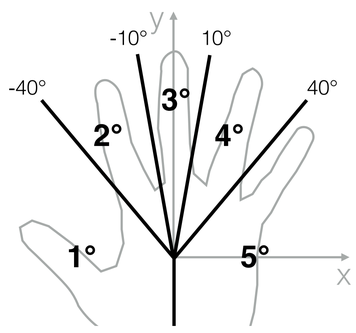


a) b)

1) An histogram of the finger edges distances from the centroid is built (a).
2) The regions corresponding to the different features (i.e., the different fingers) are computed (b).
3) The maximum for each region is selected (feature points are highlighted with red stars).
The approach is detailed in [1],[2].

Features Extraction from Leap Motion Data

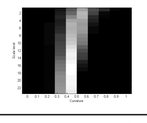
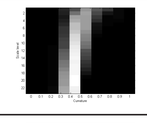
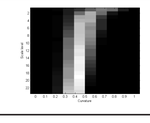
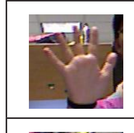
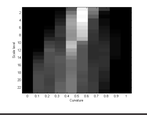
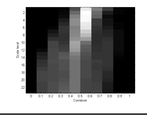
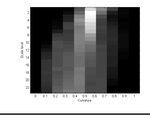
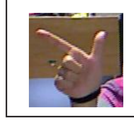
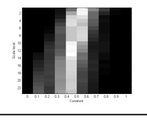
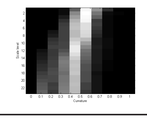
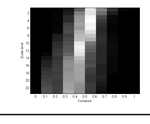
We define 5 angular regions and assign each captured finger to a specific region. Note that there is not always a one-to-one matching between sectors and fingers.



The following features are extracted from Leap Motion data :

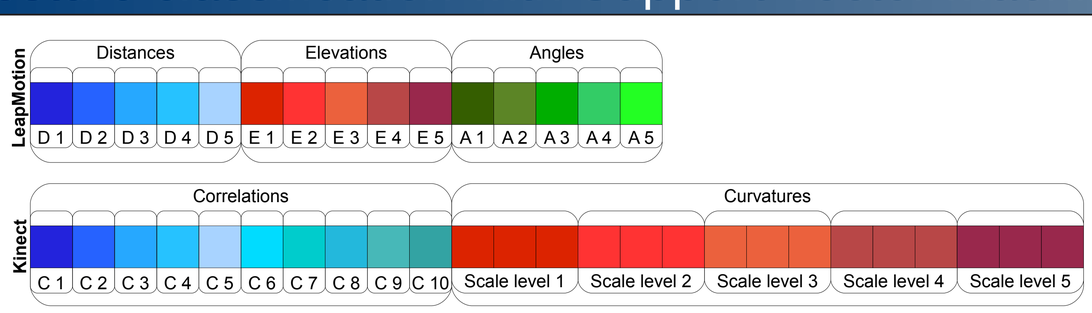
- **Fingertips angle**: $A_i = \angle(\mathbf{F}_i^{\pi} - \mathbf{C}, \mathbf{h}), i = 1, \ldots, 5$, this feature set contains the angles corresponding to the orientation of the projected fingertips with respect to the hand orientation.
- **Fingertips distance**: $D_i = ||\mathbf{F}_i - \mathbf{C}||/S, i = 1, \ldots, 5$, this feature set contains the 3D distances of the fingertips from the hand center.
- **Fingertips elevation**: $E_i = \text{sgn}((\mathbf{F}_i - \mathbf{F}_i^{\pi}) \cdot \mathbf{n})||\mathbf{F}_i - \mathbf{F}_i^{\pi}||/S, i = 1, \ldots, 5$, this feature set contains the distances of the fingertips from the plane corresponding to the palm region.

Kinect: Curvature Features

Gesture	Rep. 1	Rep. 2	Rep. 3
			
			
			

This descriptor is based on the curvature of the hand shape edges. A set of circular masks with varying radius centered on each edge sample is built and the number of hand samples inside each mask is used to compute the curvature values. Features are given by the amount of samples with a certain curvature value at a certain scale level. As expected, the amount of curvature depends on the number of raised fingers and on their arrangement, thus giving an accurate description of the gesture.

Gesture classification with Support Vector Machines



In order to recognize the performed gestures, the extracted sets of features are classified with a Support Vector Machine (SVM). Each gesture is described by a set of features from the Leap Motion (fingertips distances, angles and elevations), and a set of features from the Kinect (distance and curvature). Those sets are used both separated and concatenated to classify each gesture into one of the G classes using a multi-class SVM classifier based on one-against-one approach.
The parameters selection for the non-linear Gaussian Radial Basis Function kernel has been carried out by means of grid search and cross-validation on the training set.

Experimental results

Acquisition Setup

Acquired dataset

Leap Motion		Kinect	
Feature set	Accuracy	Feature set	Accuracy
Fingertips distances (D)	**76.07%**	**Curvature (C)**	**87.28%**
Fingertips angles (A)	74.21%	Correlation (R)	65.00%
Fingertips elevations (E)	73.07%		
D + A	78.78%	**C + R**	**89.71%**
E + A	77.28%		
D + E	**80.20%**		
D + A + E	**80.86%**		

Feature set (Kinect and Leap Motion)	Accuracy
D + A+ E + C + R	**91.28%**

Accuracy of the approach with different combinations of features

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
G1	**0.99**		0.01							
G2		**0.96**	0.03		0.01	0.01				
G3		0.02	**0.96**		0.01		0.01			
G4		0.01	0.01	**0.91**	0.01		0.01	0.03		0.01
G5		0.03		0.01	**0.94**	0.01		0.01		
G6		0.01	0.01		0.02	**0.86**			0.04	0.07
G7			0.01	0.02	0.01	0.01	**0.90**	0.05		
G8				0.03			0.07	**0.86**		0.04
G9						0.01		0.01	**0.97**	0.01
G10					0.01	0.19		0.03		**0.78**

Confusion matrix for the complete features set

[1] F. Dominio, M. Donadeo, G. Marin, P. Zanuttigh, G.M. Cortelazzo, "Hand Gesture Recognition with Depth Data", ACM Multimedia ARTEMIS Workshop, 2013
[2] F. Dominio, M. Donadeo, P. Zanuttigh, "Combining multiple depth-based descriptors for hand gesture recognition", Pattern Recognition Letters, 2014

http://lttm.dei.unipd.it