HAND GESTURE RECOGNITION WITH DEPTH DATA

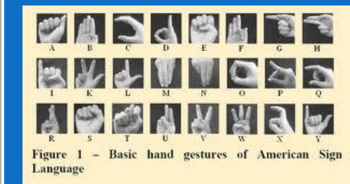
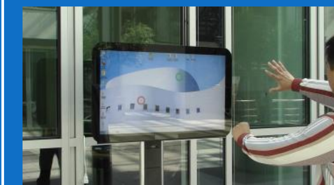
Fabio Dominio, Mauro Donadeo, Giulio Marin,
Pietro Zanuttigh and Guido M. Cortelazzo

LTTM

Università degli Studi di Padova

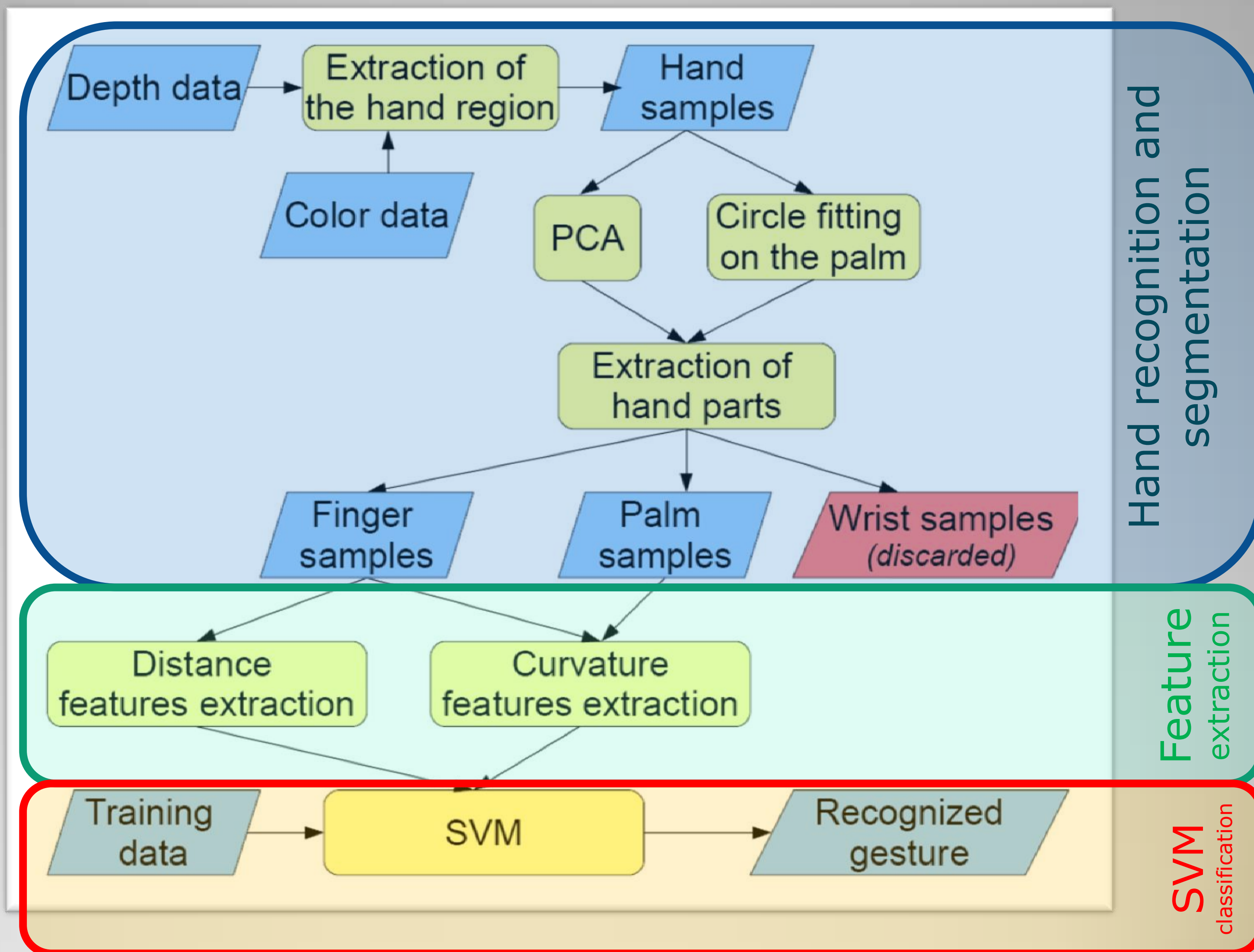
Department of Information Engineering
University of Padova

Hand gesture recognition with depth data

				
Sign language recognition	Computer interfaces	Gaming	Healthcare applications	Robotics

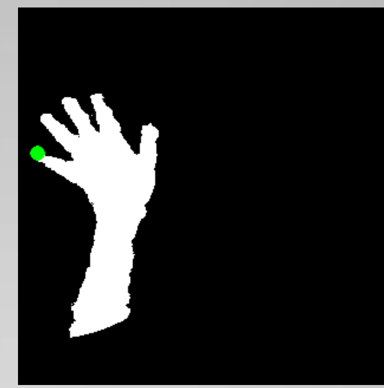
- *Hand gesture recognition* is an intriguing problem with many applications
- Large amount of research on hand gesture recognition from image and videos but it remains a challenging task due to the complex geometry of the hand and to the inter-occlusions
- Now *Depth data* is easily available from low cost devices
- *Depth data* offers a very accurate representation of the hand shape and allows to improve the accuracy of gesture recognition schemes

Overview of the Approach



Extraction of the Hand (1)

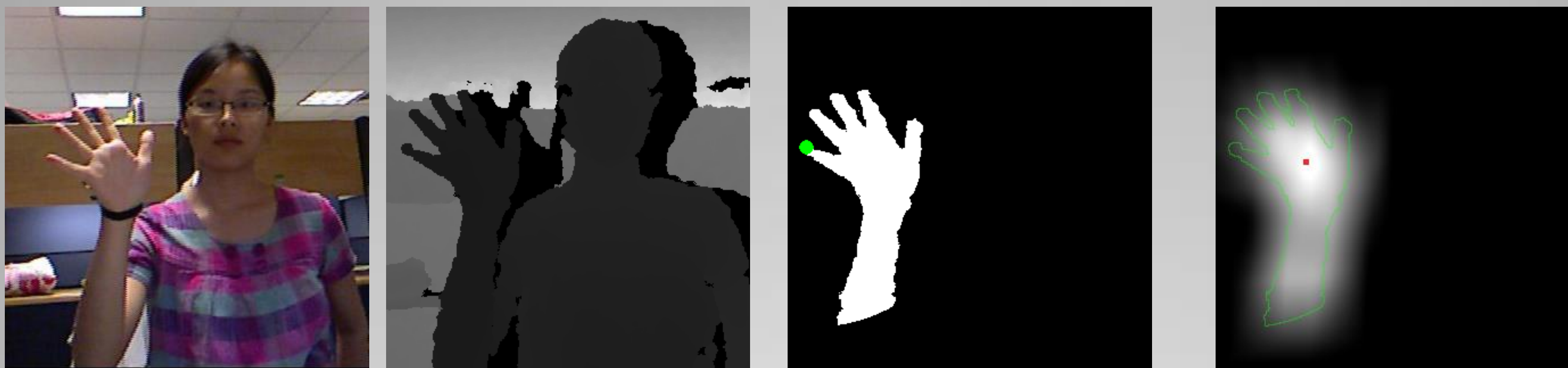




- Both color and depth data are used for hand recognition
- Start from closest point X_{min} (if it is an isolated point a new point is selected)
- Thresholding for initial hand estimation
  - On the **depth** value
  - On the **distance** from the closest point in 3D space

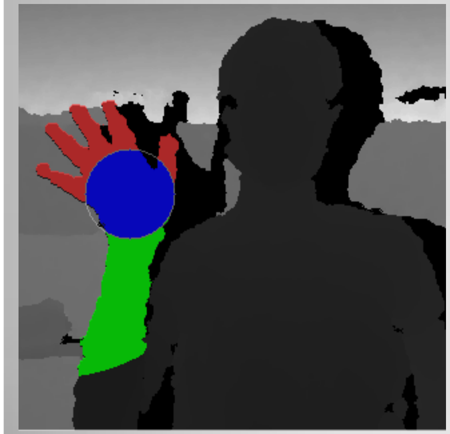
$$\mathcal{H} = \{\mathbf{X_i} \mid D(X_i) < D(X_{min}) + T_{h1} \wedge \|X_i - X_{min}\| < T_{h2}\}$$

Extraction of the Hand (2)



- Hand compatibility check:
  - Detected object size must be compatible with the hand
  - Detected object color in the CIELAB space must be compatible with the skin color
  - *Wrist and part of the forearm could be included*
- Initial palm center position detection
  - Low pass Gaussian filtering of the mask to find highest density region
  - σ depends on the hand distance
  - Center of the palm: point of highest density closest to X_{min}

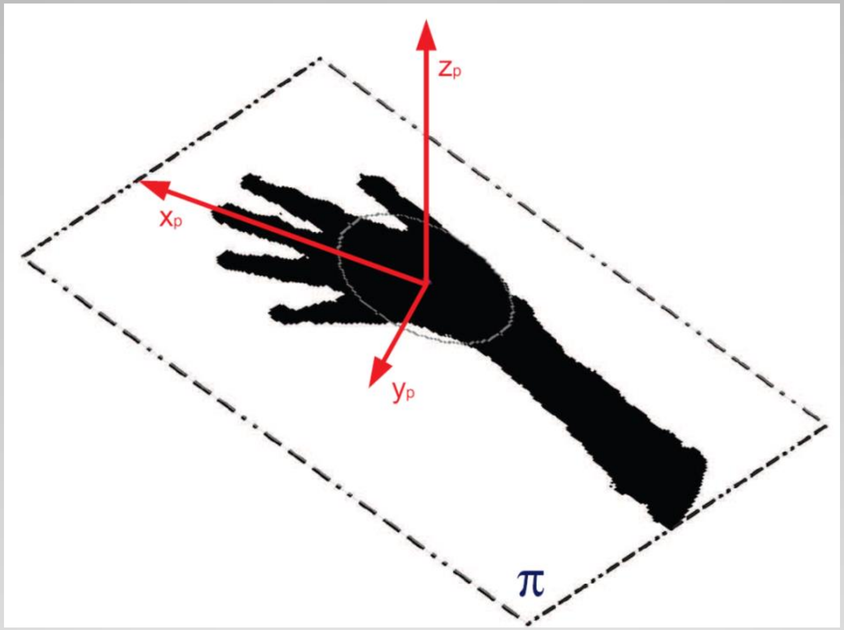
Extraction of the Palm and Fingers



- A circle is fitted on the palm starting from the estimated palm center
- Search for the maximum size circle that can be fitted on the palm area
  - 95% of the circle must be inside the detected region
- Refinement of the palm center position
  - 2 iterated phases: move/enlarge
- Subdivision into ***palm***, ***fingers*** and ***wrist*** regions
- Improvement of the palm recognition with ellipse fitting [1]

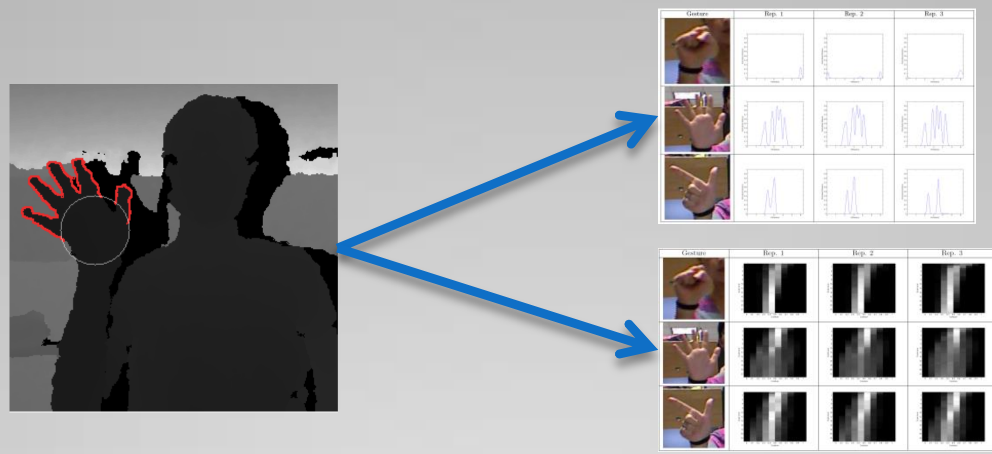
[1] G. Marin, M. Fraccaro, M. Donadeo, F. Dominio, P. Zanuttigh, "Palm area detection for reliable hand gesture recognition", Proc. of MMSP 2013

Hand orientation



- The rough orientation of the hand is detected using Principal Component Analysis (PCA)
- A plane is fitted on the hand's palm using SVD and RANSAC
  - RANSAC ensures robustness to Kinect's artifacts
- A new reference system is built

Feature Extraction



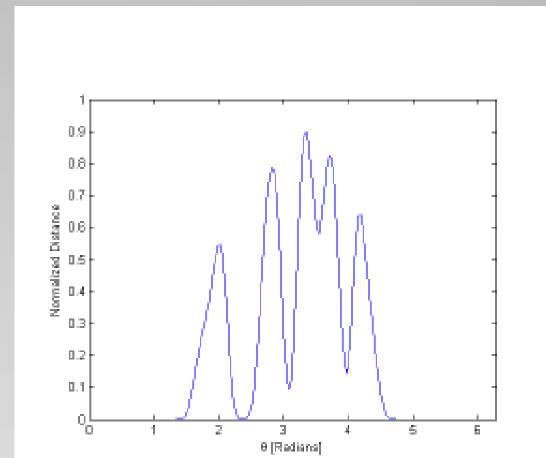
Two different types of features are extracted:

1. Distance features computed on the finger samples
2. Curvature features computed on the hand contour

Two additional types of features (palm area features and elevations from the hand's plane) have been added in journal extension on Pattern Recognition Letters [2]

[2] F. Dominio, M. Donadeo, P. Zanuttigh, "Combining multiple depth-based descriptors for hand gesture recognition", accepted for publication on Pattern Recognition Letters

Distance Features (1)



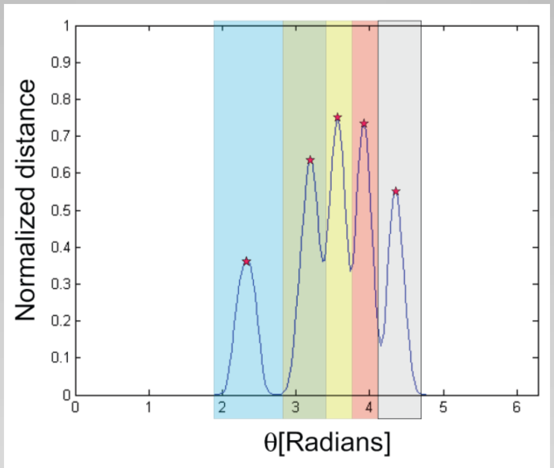
- We consider the distances of the finger samples from the hand centroid *in the 3D space* for each angular direction
- An histogram with the maximum value for each direction is built
- Alignment with reference templates for precise hand orientation
- Histograms flipping to handle left/right hand and palm/dorsum facing the camera

$$L(\theta_q) = \max_{\theta_q - \frac{\Delta}{2} < \theta_{X_i} \le \theta_q + \frac{\Delta}{2}} d_{\mathbf{X}_i}$$

$$\begin{aligned} \Delta_g &= \arg\max_{\Delta}(\rho(L(\theta), L_g^r(\theta + \Delta)) \\ \Delta_g^{rev} &= \arg\max_{\Delta}(\rho(L(-\theta), L_g^r(\theta + \Delta)) \end{aligned}$$

LTTM
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
UNIVERSITÀ DEGLI STUDI DI PADOVA

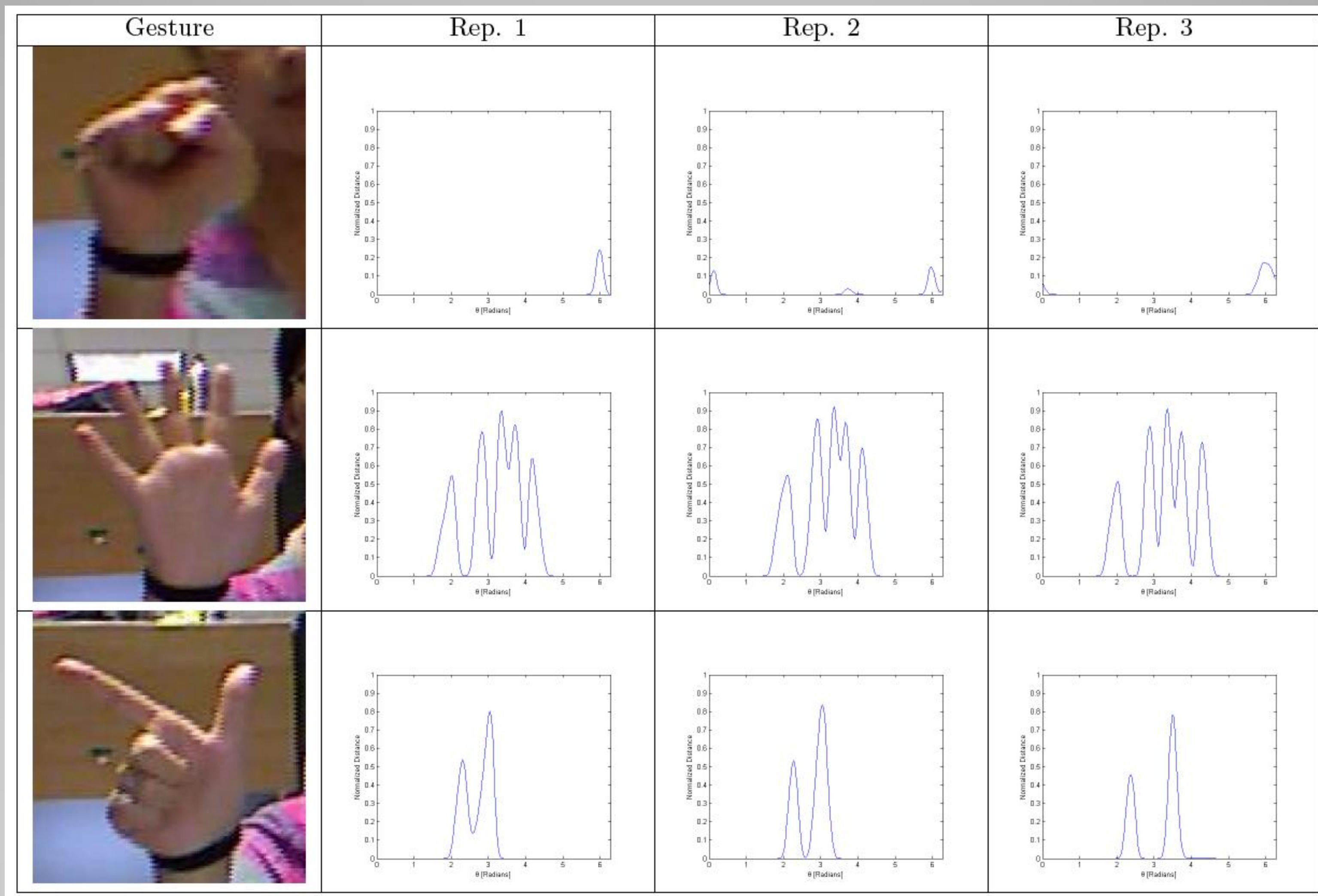
Distance Features (2)



$$f^{l}_{g,j} = \frac{\max\limits_{\theta^{min}_{g,j} < \theta < \theta^{max}_{g,j}} L^{g}_{a}(\theta) - r_f}{L_{max}}$$

- Angular directions are divided into regions corresponding to the fingers of interest in the considered gesture
- Feature values are the normalized maxima in the region corresponding to each finger
- There is one feature for each finger in each gesture hypotheses (i.e., there can be up to G*5 features)

Distance Features: Examples



Curvature Features (1)

- Computed on the edges of the hand region (palm and fingers)
- Multi-scale descriptor of the curvature of the edges
- Curvature $V(x_i, s)$: ratio between hand area inside a circular mask centered on x_i and the mask size
  - $V < 0{,}5$: convex region
  - $V = 0{,}5$: straight edge
  - $V > 0{,}5$: concave region

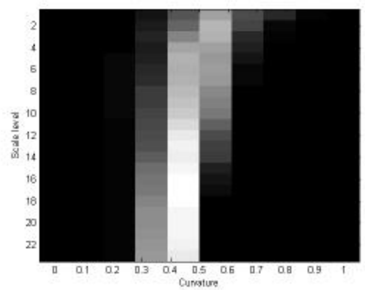
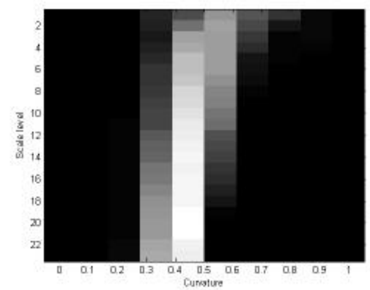
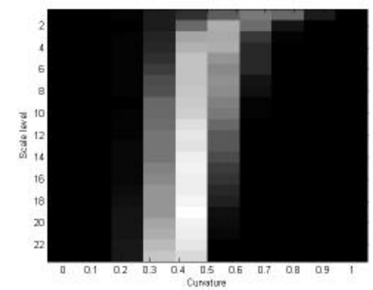
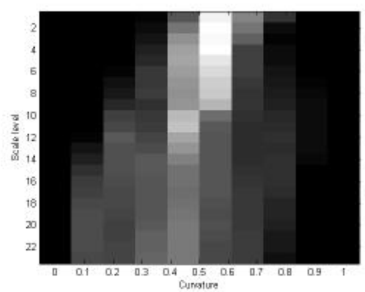
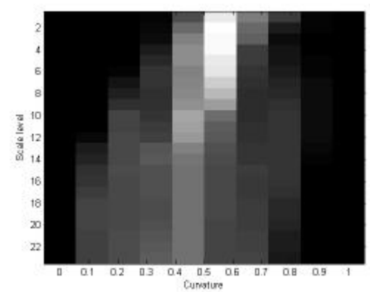
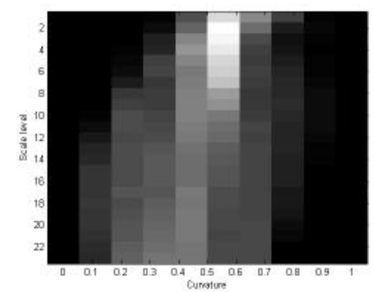
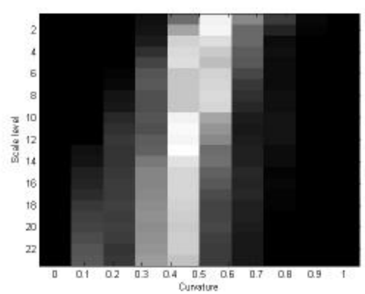
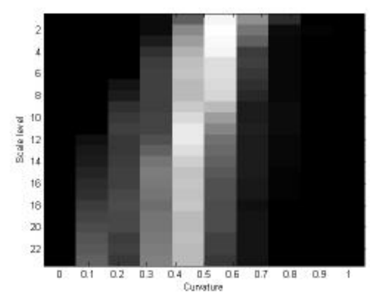
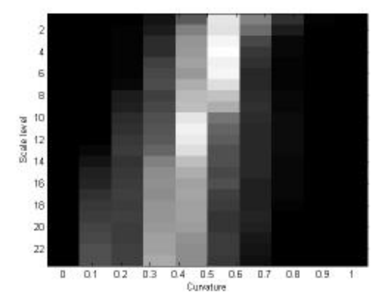
$$V(\mathbf{X_i}, s) = \frac{\sum_{X \in M_s(\mathbf{X_i})} h_m(\mathbf{X})}{|M_s(\mathbf{X_i})|}$$

Curvature Features (2)

- Feature value: number of samples with a certain curvature at the selected scale level
- Curvature values interval divided into B bins of equal size
- Feature vector: 2D array with count of samples with a certain curvature at a certain scale level

$$\mathcal{V}_{b,s} = \{\mathbf{X_i} | \frac{(b-1)}{B} < V(\mathbf{X_i}, s) \le \frac{b}{B}\}$$

Curvature Features: Examples

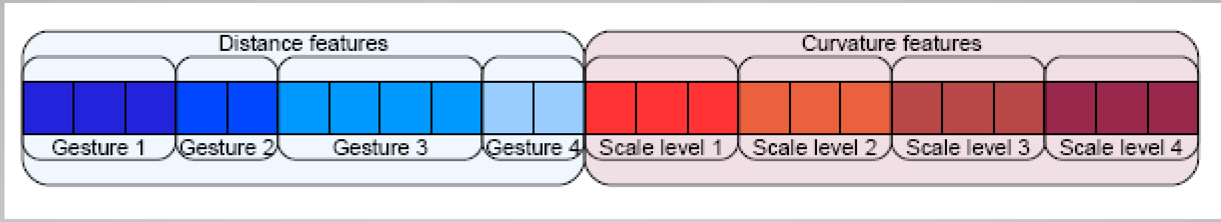
Gesture	Rep. 1	Rep. 2	Rep. 3
			
			
			

LTTM

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

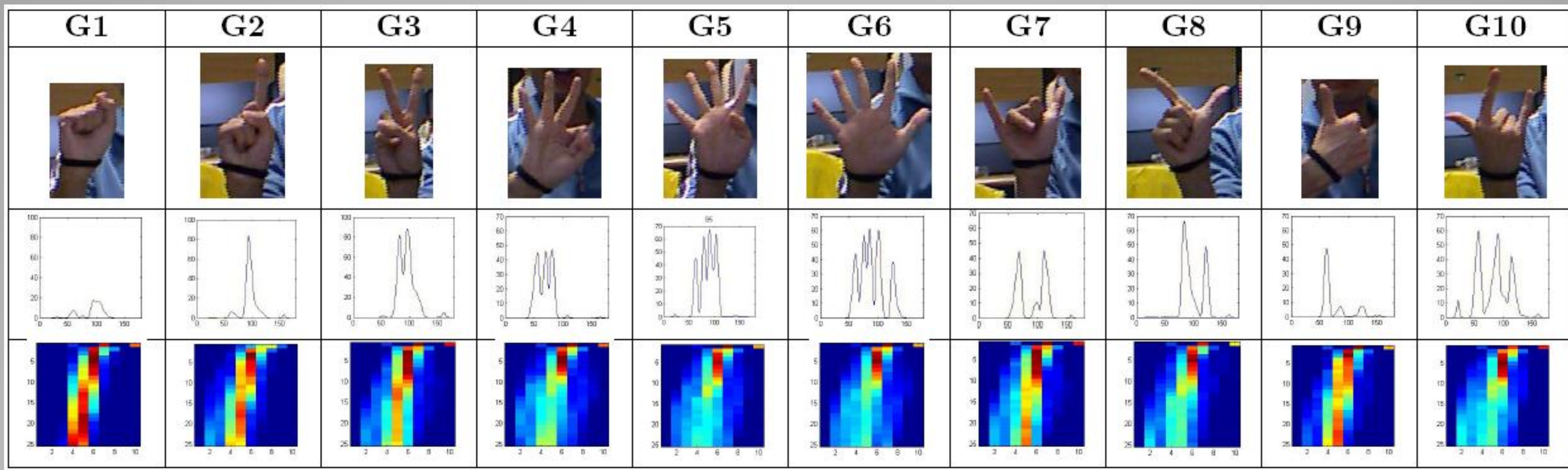
UNIVERSITÀ DEGLI STUDI DI PADOVA

Classification



- Feature vectors are built by concatenating the different distance and curvature features
- Distance features: one for each relevant finger in each gesture hypothesis
- Curvature features: one for each curvature bin and scale level
- Classification with a multi-class support vector machine
  - *One-against-one* approach
  - Kernel: Gaussian Radial Basis Function (RBF)
  - Grid-search with cross validation for parameters tuning

Experimental Results

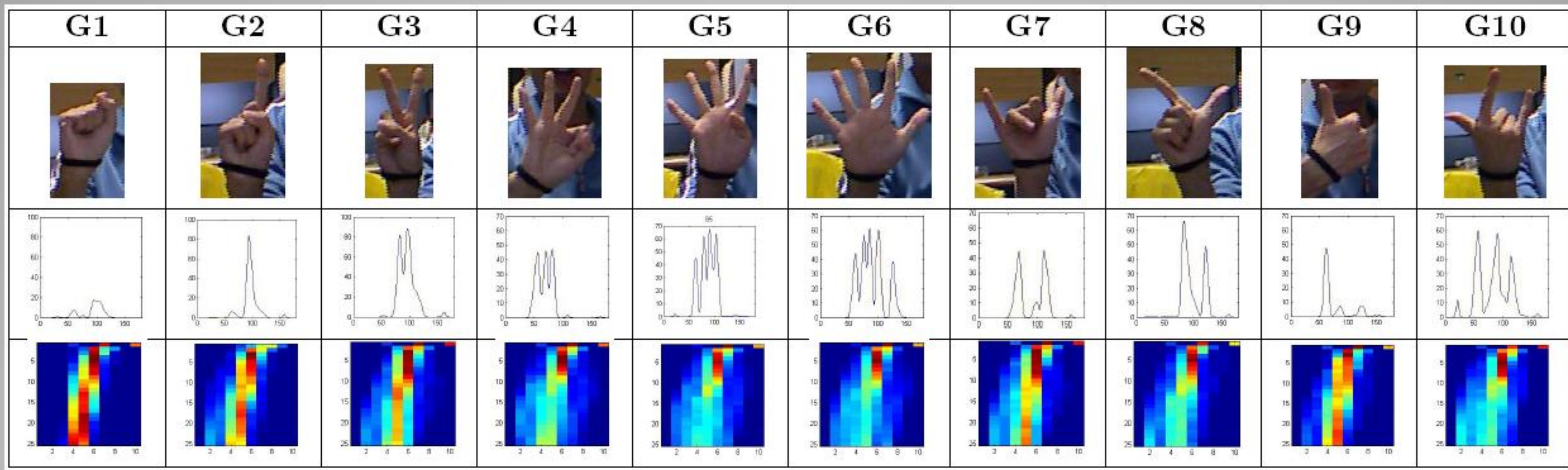


- Gesture database from the work of Ren et Al [3]
- 1000 samples : 10 gestures, 10 people and 10 repetitions for each gesture
- Results on a second more challenging dataset are included in the journal extension [2]

[2] F. Dominio, M. Donadeo, P. Zanuttigh, "Combining multiple depth-based descriptors for hand gesture recognition", accepted for publication on Pattern Recognition Letters

[3] Z. Ren, J. Yuan, and Z. Zhang. “Robust hand gesture recognition based on Finger-earth mover's distance with a commodity depth camera”, Proc. of ACM Multimedia 2011

SVM Training



- The database has been divided into 800 samples for training and 200 for testing
- 2 Subdivision modalities
  1. Random subdivision (*user training*)
  2. 8 people for testing and 2 for training (*generic training*)
- Grid search with cross validation for optimal parameters extraction

Experimental Results

Method	*Mean Accuracy*	
	Training with users	*Generic Training*
Distance features	96%	92,5%
Curvature features	97,5%	92%
Distance + Curvature	**99,5%**	**98,5%**
Shape context [1]	83,2%	
Near-convex Dec.+FEMD [3]	90,6%	
Thresholding Dec.+FEMD [3]	93,9%	

- Combined use of the two features: better performances
- 99,5% accuracy with user training
- Generic training more challenging, but the combined use of the two features leads to very good results (98,5%)
- Large performance improvement w.r.t. [3]

[3] Z. Ren et Al. "Robust hand gesture recognition based on Finger-earth mover's distance with a commodity depth camera", ACM Multimedia 2011

Confusion Matrices

(generic training)

Distance

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
G1	1	0	0	0	0	0	0	0	0	0
G2	0	1	0	0	0	0	0	0	0	0
G3	0	0,05	0,95	0	0	0	0	0	0	0
G4	0	0	0	0,8	0,05	0,15	0	0	0	0
G5	0	0	0	0	1	0	0	0	0	0
G6	0	0	0	0	0,05	0,95	0	0	0	0
G7	0	0	0,05	0	0	0	0,85	0	0,1	0
G8	0	0	0	0	0	0	0	1	0	0
G9	0	0	0,05	0	0	0	0	0,1	0,85	0
G10	0	0	0	0,05	0,1	0	0	0	0	0,85

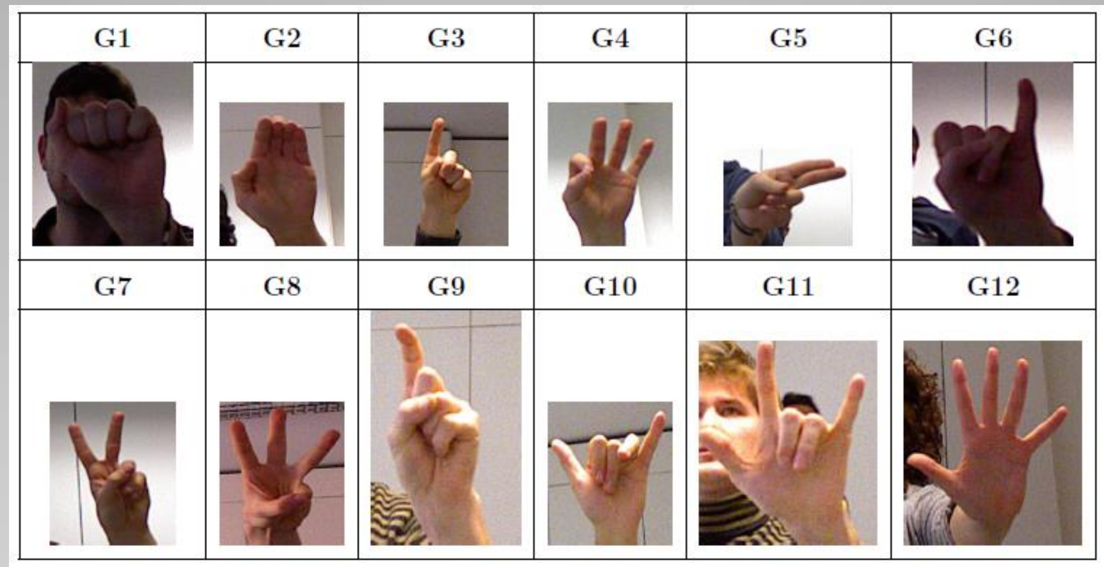
Curvature

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
G1	1	0	0	0	0	0	0	0	0	0
G2	0	1	0	0	0	0	0	0	0	0
G3	0	0	0,8	0	0	0	0,15	0,05	0	0
G4	0	0	0	0,95	0,05	0	0	0	0	0
G5	0	0	0	0,05	0,95	0	0	0	0	0
G6	0	0	0	0	0	1	0	0	0	0
G7	0	0	0	0	0	0	0,95	0	0,05	0
G8	0	0	0,25	0	0	0	0,05	0,7	0	0
G9	0	0,1	0	0	0	0	0	0	0,9	0
G10	0	0	0,05	0	0	0	0	0	0	0,95

Combined

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
G1	1	0	0	0	0	0	0	0	0	0
G2	0	1	0	0	0	0	0	0	0	0
G3	0	0	1	0	0	0	0	0	0	0
G4	0	0	0	0,9	0,1	0	0	0	0	0
G5	0	0	0	0	1	0	0	0	0	0
G6	0	0	0	0	0,05	0,95	0	0	0	0
G7	0	0	0	0	0	0	1	0	0	0
G8	0	0	5	0	0	0	0	1	0	0
G9	0	0	0	0	0	0	0	0	1	0
G10	0	0	0	0	0	0	0	0	0	1

New Dataset



- More challenging dataset (from extended version of the work [2])
- 1680 samples : 12 gestures, 14 people and 10 repetitions for each gesture
- Accuracy of 95% *with user training* and 89,6% *with generic training*
- *Up to 97,6% and 93,8% with 4 features*

LTTM

[2] F. Dominio, M. Donadeo, P. Zanuttigh, "Combining multiple depth-based descriptors for hand gesture recognition", accepted for publication on Pattern Recognition Letters

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

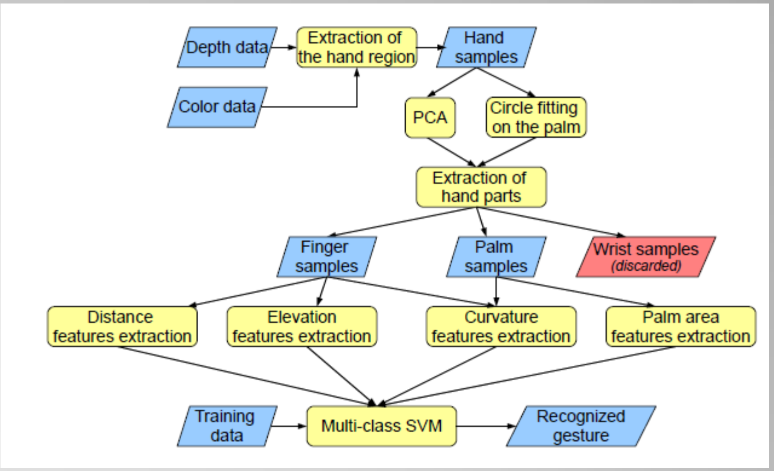
UNIVERSITÀ DEGLI STUDI DI PADOVA

Conclusions

- The hand is reliably extracted from the color and depth data
- Recognition of the palm, fingers and hand orientation
- Reliable feature descriptors based on 3D measures
- Distance and curvature features capture different clues: they are complementary
- Real-time computation (10 fps)
- Very high accuracy on datasets from the literature

Future Research

- Additional feature descriptors from depth and color data
  - Elevation features
  - Palm area features
  - Color-based features



- Better palm area identification
- Recognition of multiple interacting hands
- Advanced machine learning strategies
- Extension to dynamic gestures recognition

Thanks for your attention

For datasets and further infomation on our research: visit our website http://lttm.dei.unipd.it