Master Computer Vision & Artificial Intelligence

Real-Time Hand Pose Recognition using Depth Sensors combined with Spherical Shape Model Descriptor

Oscar Lopes

Advisors: Sergio Escalera

Jordi Gonzàlez

September 27, 2012

CVC Computer Vision Center

UAB Universitat Autònoma de Barcelona

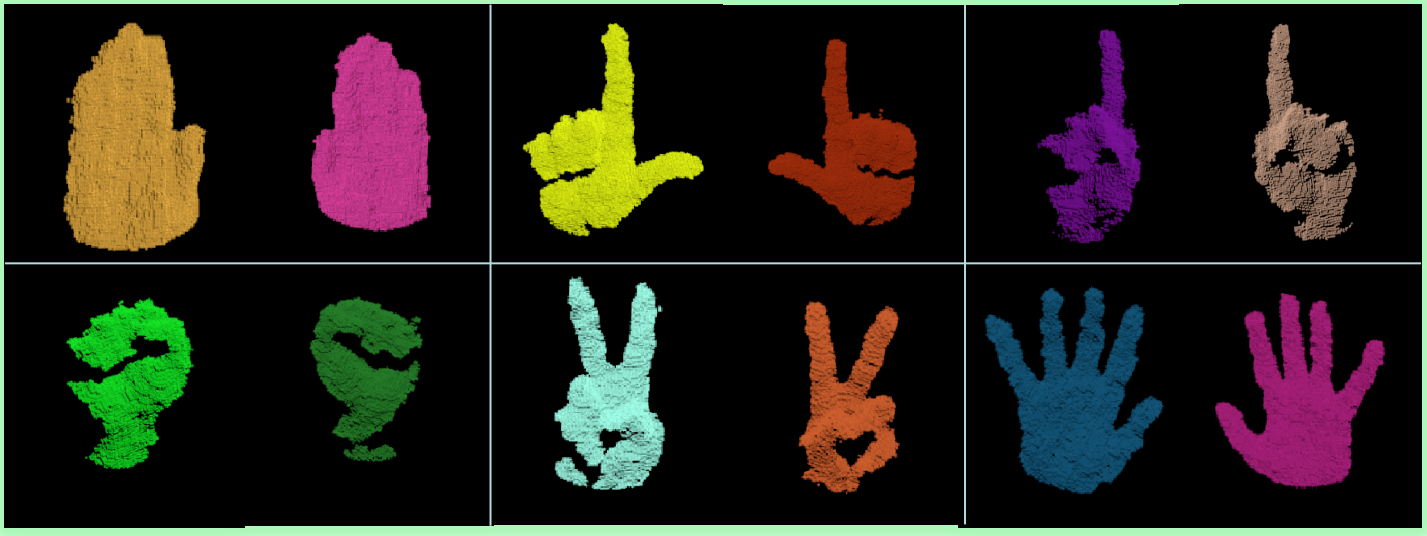

overview: g-speak
oblong industries


XBOX 360
KINECT
KINECT
for XBOX 360

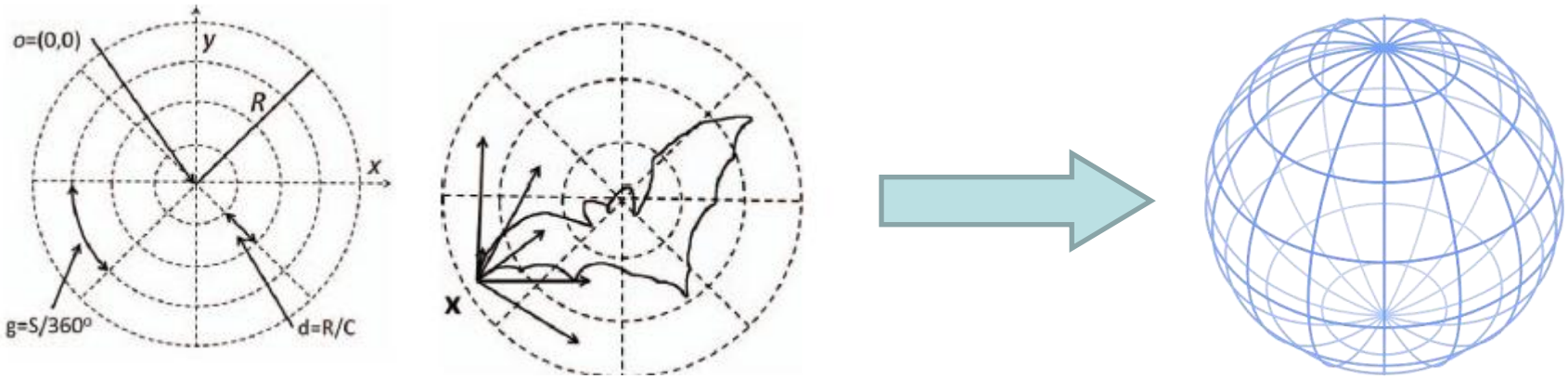
Model
Hand Detector
Spherical Shape Model Descriptor
Classifier
Hand Pose Label
Train
Test

- First the closest point is searched
  - Perform **a radius search** of 10cm, until the depth starts to rapidly increase.
  - If less **than 100 points** are obtained, then the sample is discard.
- Repeat the same procedure to find the second hand: the distance to camera must be within 30 cm of the first hand's closest distance and must more than 20 cm from the center of the first hand.

- No hand pose dataset is publicly available.
- New point cloud hand pose dataset must be creatred!
- The dataset was created using and adaptation of the hand detector.
- Includes **6 classes** with **2000 samples** (1000 per hand):
  - Each class includes both hands.
  - High hand orientation variability.
- Plus a *No-Pose* class.



- Current state-of-art point cloud descriptors (e.g PFH) have great computation overhead: ***O(N²)***.
- The design of a novel descriptor is necessary!
- Circular Blurred Shape Model Descriptor [Escalera et al]
  - Very good discriminative power
  - Low computational requirements



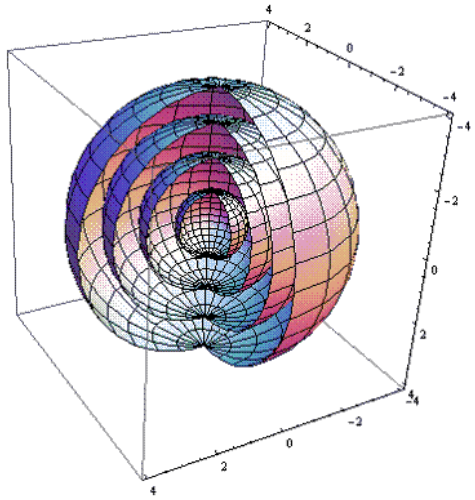
S. Escalera, A. Fornes, O. Pujol, A. Escudero, and P. Radeva, "Circular blurred shape model for symbol spotting in documents," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, nov. 2009, pp. 2005 –2008.

- Novel *Spherical Blurred Shape Model* Descriptor (SBSM)

$$P = \{p_i \mid p_i \in \mathbb{R}^3\}$$

N_L number radial layers
N_θ number of θ angular divisions
N_ϕ number of ϕ angular divisions
S_{Radius} sphere radius

$$S_R = S_{Radius}/N_L$$
$$S_\theta = 2\pi/N_\theta$$
$$S_\phi = 2\pi/N_\phi$$

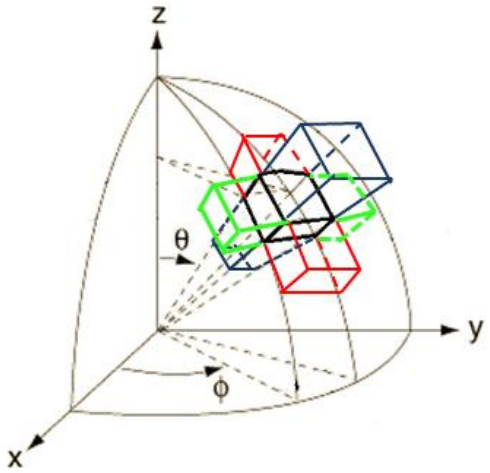


B the ordered set of bins for the spherical description of P^*

$b^*_{\{i,j,k\}}$ the centroid of the section $b_{\{i,j,k\}} \in B$.

$W_n = 0, n \in \{1, \ldots, N_L N_\phi N_\phi\}$

```
foreach p_n ∈ P* do
    b_x : b_x ∈ B, p_n ⊂ b_x
    W(b_x) = 1
    foreach b_{i,j,k} ∈ N(b_x) do
        d_{i,j,k} = d(b_{i,j,k}, p_n) = ||p_n − b*_x||
        W(b_{i,j,k}) = W(b_{i,j,k}) + 1/d_{i,j,k}
    end
end
```



Normalize the vector W

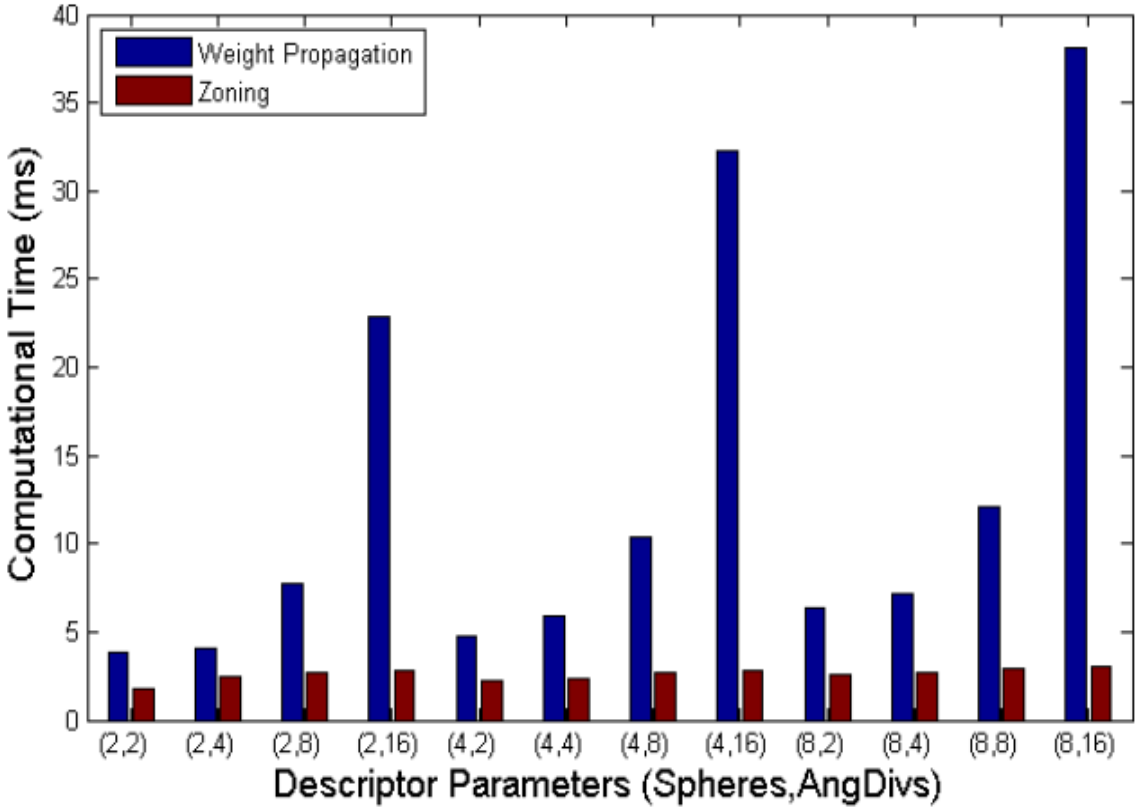
$\frac{W_i}{\#P}, i \in \{1, \ldots, N_L N_\phi N_\phi\}$

- Test settings:
  - Descriptor testing considered

$$\begin{cases} N_L = \{2, 4, 8\} \\ N_\theta = N\phi = \{2, 4, 8, 16\} \end{cases}$$

  - The classification was performed as multiclass one-versus-one, using the libSVM framework.
  - Each combination pair was executed 10 times for cross-validation test.
    - Each execution considered 70% train data of each dataset class samples (randomly picked).
    - Every test run comprises a cross-validation of the train data for fine tune C-SVM (RBF Kernel) parameters: C and γ .
  - Previous settings were considered in two descriptor modalities:
    - Weight Propagation.
    - Zoning.

Descriptor Configuration			Average Accuracy	
Layers	Angles	Features	Weight Propagation	Zoning
2	2	9	78.425	**89.278**
2	4	33	**94.027**	90.797
2	8	129	98.046	**98.709**
2	16	513	98.680	**99.498**
4	2	17	**98.599**	96.941
4	4	65	98.637	**99.539**
4	8	257	**99.847**	99.479
4	16	1025	**99.818**	99.796
8	2	33	**99.657**	99.627
8	4	129	**99.772**	99.713
8	8	513	**99.839**	99.527
8	16	2049	**99.785**	99.598



KINECT

OpenNI™

pointcloudlibrary

Eigen

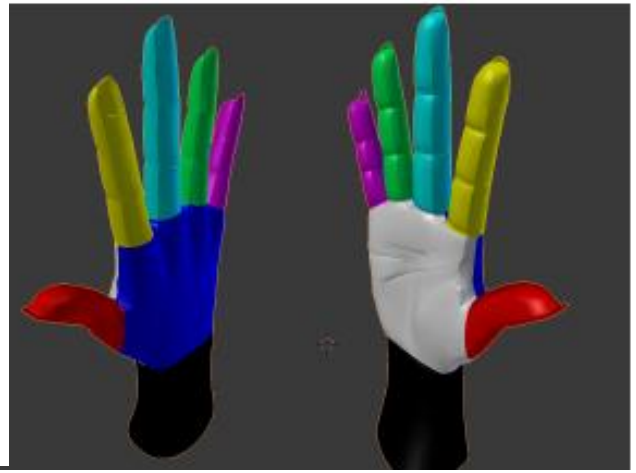
Processor:	Intel(R) Core(TM) i5-2430M CPU @ 2.40GHz
Installed memory (RAM):	4.00 GB
System type:	64-bit Operating System

CVC Computer Vision Center

UAB Universitat Autònoma de Barcelona

- The proposed system achieves the proposed design goal of a real-time hand pose recognition, with an end-to-end performance of 14fps.
- The SBSM descriptor is crucial for the results obtained:
  - High discriminative power for hand pose point cloud.
  - Small computational overhead.
  - Slight advantage of blurring aspect versus Zoning, encourage further studies.

- As future work…
  - Creation of a more difficult **multi-user dataset**.
  - Implement the descriptor algorithm using the **GPU** (Grapics processing unit) for a performance boost.
  - Include a pre-description phase, for a per-pixel classification using **Random Forest** classifier, in order to perform **Label Blurring**, to increase the robustness of the overall pipeline.



Thank You!

Oscar Lopes
(oscar.pino.lopes@gmail.com)