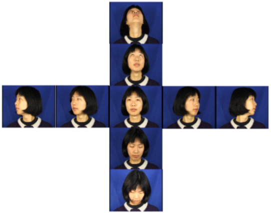
A manifold based clustering algorithm and application to object discovery in RGBD data

Rahul Erai

4 July 2012

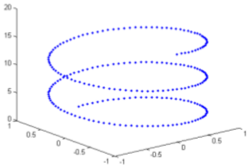
Introduction



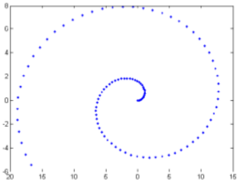
Outline

What are manifolds?

- Topological structures such that each point on them has a neighborhood that is homeomorphic to Euclidean space.
- Usually lies in a high dimensional space, but has a low intrinsic dimensionality.



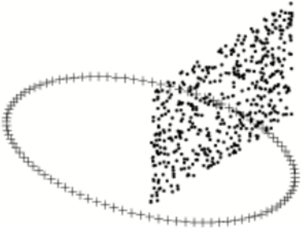
Helix

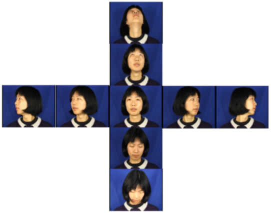


Spiral

Manifold clustering

- Traditional clustering algorithms like *k-means clustering* assumes data is distributed as a spherical blob.
- Manifold clustering is challenging because,
  - Data sampled from manifold may not be uniform.
  - Manifold could intersect with other manifolds and itself.





Kinect and pointclouds

- A part of Microsoft's XBOX gaming console.
- Uses structured light to perceive depth.
- Pointcloud: A set of 3D points with optional color information.

Kinect RGB-D dataset

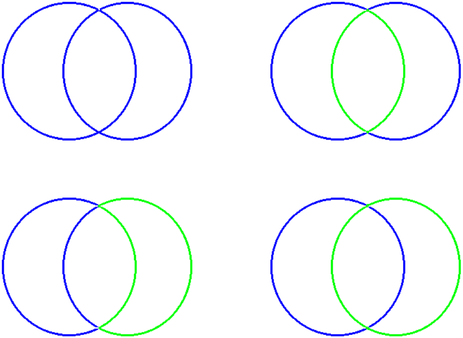
- Washington University's RGB-D dataset on household objects
- Contains pointclouds, RGB images, and depth images of 51 categories
- Multiple instances in each category
- Each instance has 3 image sequences shot at three different heights

Problem statement

- Given $X = \{x_1, x_2, .., x_n\}$, label each x_i as the manifold it could be belonging to.
- Assumptions:
  - Each point x_i is sampled from an underlying manifold.
  - Each point x_i belongs to one and only one manifold.
  - Manifolds involved are "pure", ie, they have a constant dimensionality throughout.
  - Number of manifolds involved or their dimensionality may not be known.
- We do NOT map the found clusters into their lower dimensional embedding.

Separating manifolds

- How do humans perceive shapes?



- This can be explained by gestalt perception.

Gestalt perception

- Our algorithm tries to model human perception.
- Gestalt laws of grouping

Law of proximity Law of continuity

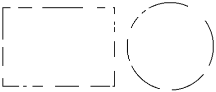
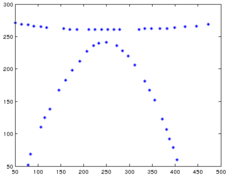


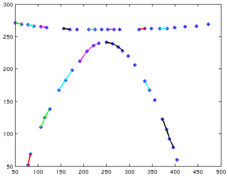
Figure: Low of closure

- Algorithm implements it as *dynamic branching factor* and *continuity scores*.

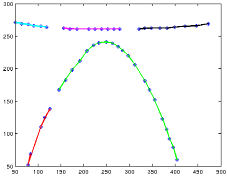
Dynamic branching factor I

- Static *k* can not handle gaps in the manifolds

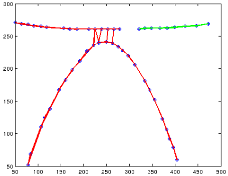




k=2



k=3



k=4

Dynamic branching factor II

- Solution: Change k dynamically
- Algorithm tries to find k_{min} *unvisited* by varying k from k_{min} to k_{max}

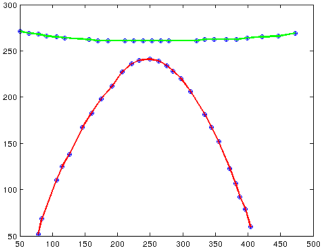
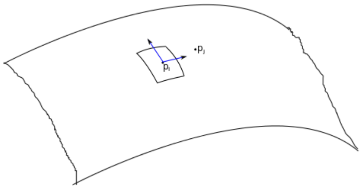


Figure: $k_{min} = 1$, $k_{max} = 10$

- Captures gestalt principle of closure.

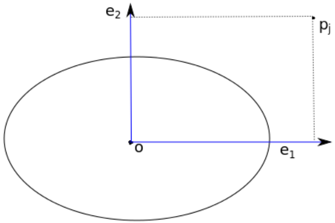
Continuity score I



$$continuity\_score(S_i, x_j) = \alpha \cdot SPD(S_i, x_j) + (1 - \alpha) \cdot embedding\_error(S_i, x_j)$$

- SPD: how far a point is from a local *patch* around another point.
- Embedding error: How well the point fits into the local *patch* around another point.

Scaled Projected Distance(SPD)



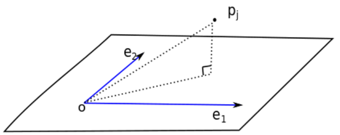
$$Scaled\_projected\_distance(S_i, x_j) = \sum_{k=1}^{d} \frac{\Lambda}{\sqrt{\lambda_k}} \cdot < e_k, (x_j - \mu_i) >$$

where,

$S_i = \{(e_1, \lambda_1), (e_2, \lambda_2), .., (e_d, \lambda_d)\}$ be the eigen decomposition at x_i

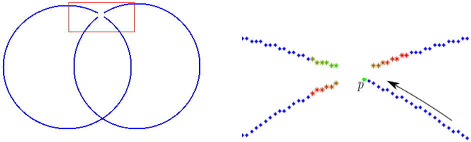
μ_i is the origin of the tangent space at x_i.

Embedding Error



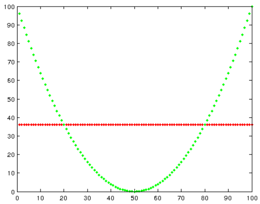
$$Embedding\_error(S_i, x_j) = ||x_j - \sum_{k=1}^{d}(< e_k, (x_j - \mu_i) > \cdot e_k)||$$

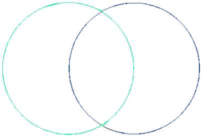
Continuity score in action

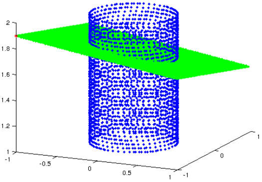


- Continuity score captures gestalt principle of continuity.

Some results





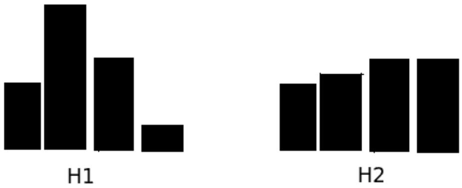


Kinect RGB-D dataset I

- RGBD data of 51 household objects

Kinect RGB-D dataset II

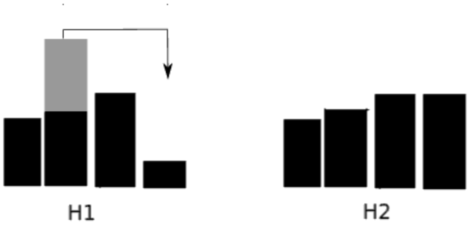
CEMD



- EMD is the minimum work needed to be done to convert a histogram to another.
- CEMD: ground distance g_{ij} as the correlation distance, $1 - corr(H_i, H_j)$, instead of the traditional $|i - j|$

$$
\begin{aligned}
g_{ij} &= 1 - corr(S^i, S^j) \\
&= 1 - \frac{Cov(S^i, S^j)}{\sigma_i \sigma_j}
\end{aligned}
\tag{1}
$$

CEMD



- EMD is the minimum work needed to be done to convert a histogram to another.
- CEMD: ground distance g_{ij} as the correlation distance, $1 - corr(H_i, H_j)$, instead of the traditional $|i - j|$

$$
\begin{aligned}
g_{ij} &= 1 - corr(S^i, S^j) \\
&= 1 - \frac{Cov(S^i, S^j)}{\sigma_i \sigma_j}
\end{aligned} \tag{1}
$$

CEMD

- EMD is the minimum work needed to be done to convert a histogram to another.
- CEMD: ground distance g_{ij} as the correlation distance, $1 - corr(H_i, H_j)$, instead of the traditional $|i - j|$

$$\begin{aligned} g_{ij} &= 1 - corr(S^i, S^j) \\ &= 1 - \frac{Cov(S^i, S^j)}{\sigma_i \sigma_j} \end{aligned} \quad (1)$$

CEMD

- EMD is the minimum work needed to be done to convert a histogram to another.
- CEMD: ground distance g_{ij} as the correlation distance, $1 - corr(H_i, H_j)$, instead of the traditional $|i - j|$

$$
\begin{aligned}
g_{ij} &= 1 - corr(S^i, S^j) \\
&= 1 - \frac{Cov(S^i, S^j)}{\sigma_i \sigma_j}
\end{aligned}
\tag{1}
$$

CEMD

- CEMD is calculated by optimizing following expression

$$CEMD(H_1, H_2) = \frac{\sum_{i=1}^{d} \sum_{j=1}^{d} f_{ij} \cdot g_{ij}}{\sum_{i=1}^{d} \sum_{j=1}^{d} f_{ij}} \quad (2)$$

subjected to following constrains.

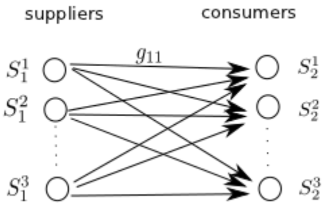
$$f_{ij} \geq 0 \text{ ; such that } 1 \leq i, j \leq d \quad (3)$$

$$\sum_{j=1}^{d} f_{ij} \leq H_1^i \text{ ; such that } 1 \leq i \leq d \quad (4)$$

$$\sum_{i=1}^{d} f_{ij} \leq H_2^j \text{ ; such that } 1 \leq j \leq d \quad (5)$$

$$\sum_{i=1}^{d} \sum_{j=1}^{d} f_{ij} = min(\sum_{i=1}^{d} H_1^i, \sum_{j=1}^{d} H_1^j) \quad (6)$$

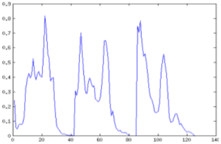
Optimizing CEMD for manifold growing

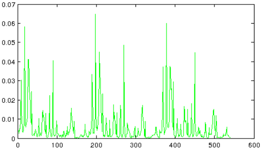


- Original complexity: $O(d^3 logd)$
- Optimization was done in two steps
  - We used an approximated $O(d^2)$ greedy algorithm.
  - To find k neighbors of a point in CEMD distance, we first found K neighbors in $L1$ distance($K > k$),with in which we calculated k CEMD neighbors.

Color features

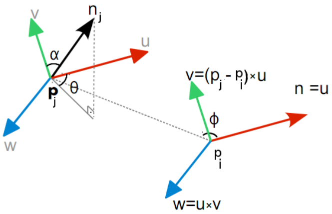
- RGB histogram: Captures global color distribution of the image
- PHOG: Gradient histogram over image pyramid. Captures color variation.





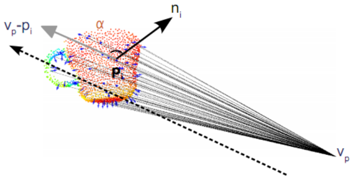
Viewpoint Feature Histograms(VFH) I

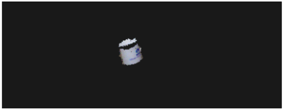
- Given a pointcloud, define normals at every point.
- For every pair of points(p_i, p_j), define a local coordinate system.
- Calculate pan, tilt, yaw, and distance for every pair of points and bin them.

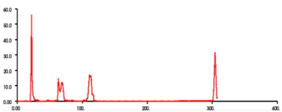


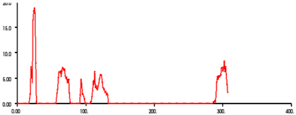
Viewpoint Feature Histograms(VFH) II

Viewpoint Component









Viewpoint Feature Histograms(VFH) III

Table: Object class detection results using K nearest neighbor classifier

Features used	Accuracy
VFH	54.4%
RGB Histogram	34.93%
PHOG	57.90%
VHF+RGB+PHOG	70.68%

- Combining RGB histogram, PHOG, and VFH results in a strong signature for 3D images.
- Detection rate was comparable with the current state of the art(84.1%)[L. Bo et al., with hierarchical kernel descriptors])

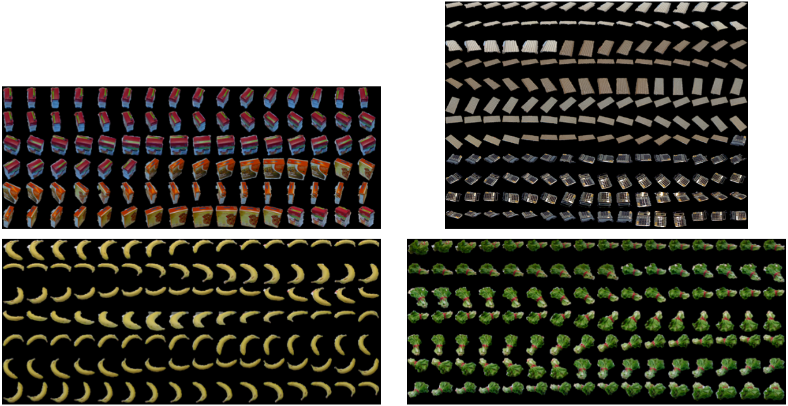
Object discovery from RGBD dataset I

- One instance from all the 51 classes were selected
- All its views were selected
- One frame per 5 consecutive frames was chosen.

	K means clustering	Manifold growing
No of clusters	100	100
NMI	78.42%	90.30%
Purity	69.36%	82.60%

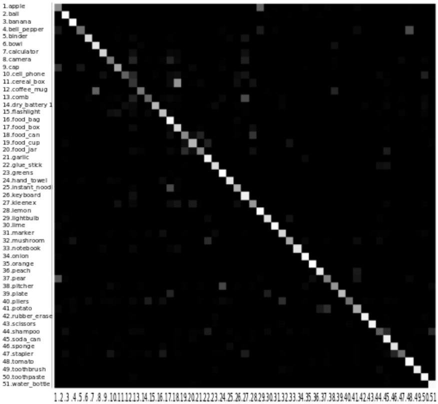
Object discovery from RGBD dataset II

- Some discovered clusters:



Object discovery from RGBD dataset III

Confusion matrix



Object discovery from RGBD dataset IV

Most confused pairs

Object discovery from RGBD dataset V

Least confused objects

Banana (100%)

Lemon (100%)

Orange(100%)

Greens(99.52%)

Onion (98.75%)

Effect of CEMD and Dynamic branching

- One instance from every class was chosen.
- One view from the three available views was chosen

Table: Results with dynamic branching turned off

Distance metric	k	Purity
Euclidean	2	72.26%
Cityblocks	2	86.87%
CEMD	2	92.27%

Table: Dynamic branching turned on($k_{min} = 1$)

Distance metric	k_{max}	Purity
Euclidean	27	91.32%
Cityblocks	24	98.47%
CEMD	21	98.70%

Supervised vs Unsupervised learning I

- First 10 categories were considered
- One instance per category was reserved for testing, all other were used in training
- Supervised *KNN* accuracy: **74.04%**

Unsupervised learning by manifold growing	
Training	
No of clusters	55
NMI	0.952
Purity	91.65%
Testing	
KNN accuracy(Unsupervised)	67.24%

Conclusion

- A manifold based clustering algorithm inspired from gestalt perception.
- Discovering object classes from an RGBD dataset.
- CEMD and dynamic branching factor inproved the results.

Possible future works

- Extention to manifold mapping.
- k_{min} could be calculated based on local dimensionality.

Thanks

Questions?