Shape-Based Object Localization for Descriptive Classification

Geremy Heitz. Gal Elidan. Benjamin Packer. Daphne Koller.

International Journal in Computer Vision (2009) 84:40-62

Presented by : Apisit Aroonnual
ID : 107926 FoS: Mechatronics
Machine Vision for Robotics and HCI
Date : 09 July 2009

OUTLINE

- Introduction
- LOOPS Modeling
- Experimental Result
- Application of LOOPS

Introduction [1]

Discriminative task in computer vision

- Where is it ? (Detection)
- What is it ? (Categorization)

If we are interested in more refined descriptive questions

Example :

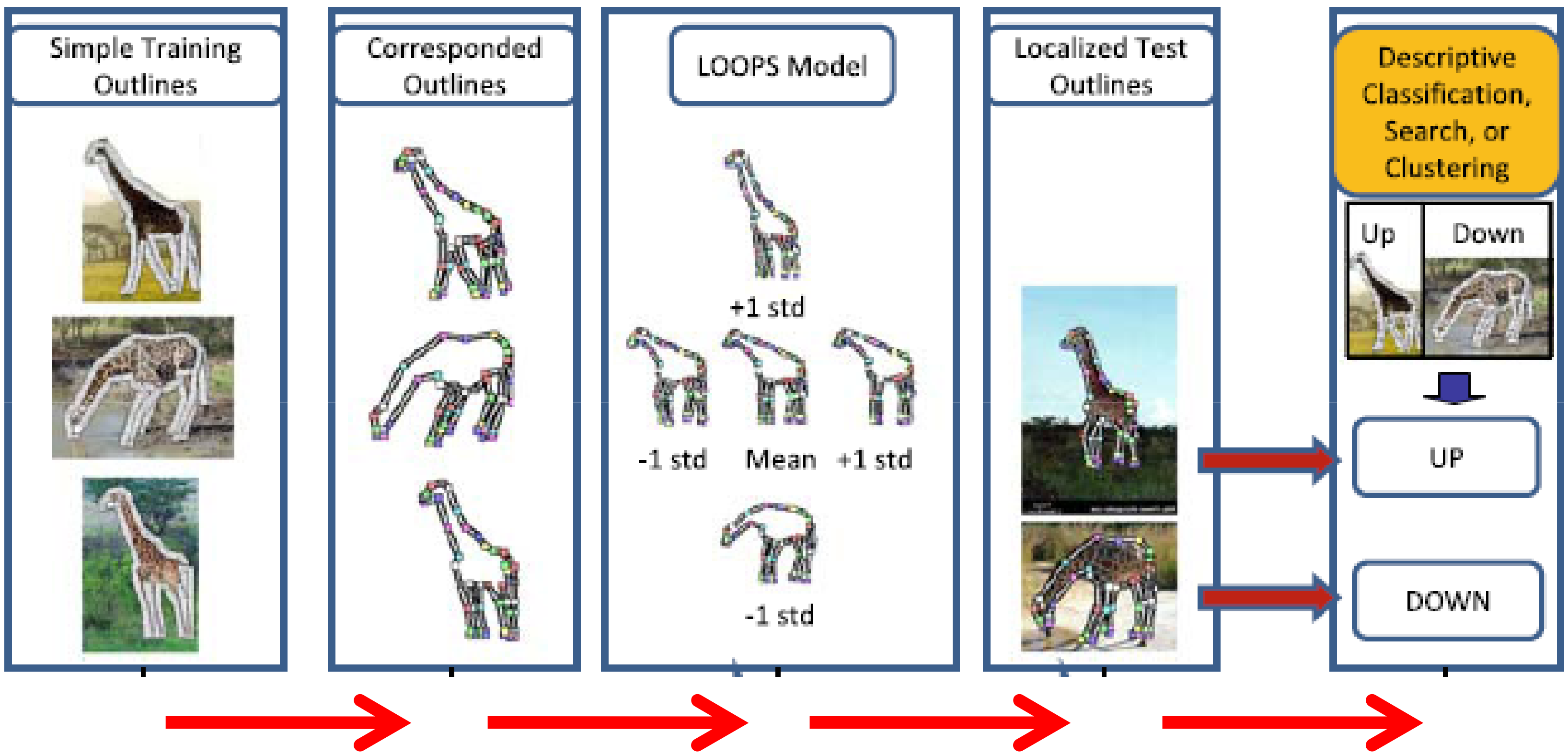
- Is the giraffe standing upright or bending over to drink ?
- Find me all lamps that have a beige rectangular lampshade

Introduction [2]

To deal with refined descriptive questions

- In principle, it is possible to do, but we have to pay significant human effort.
- The new approach called "Localized Object Outlines using Probabilistic Shape (LOOPS)" was proposed.

Overview of the LOOPS Method



Automatic Landmark Selection

LOOPS Model

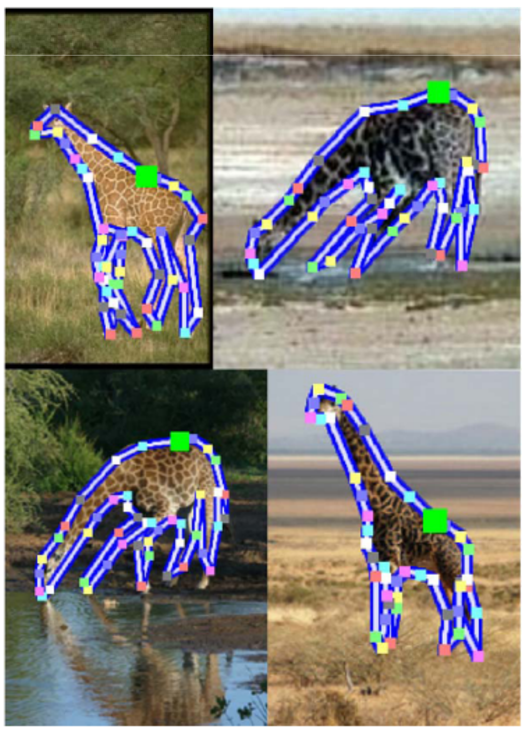
Localization of Object Class

Descriptive Task

Automatic Landmark Selection [1]

- Transform simple outlines into corresponded outlines over a relatively small and consistent set of landmark

- Use a simple two-step process



Find a correspondence between high resolution outlines (Arc-Length Correspondence)

Prune down to a small number of salient landmark (Landmark Pruning)

Automatic Landmark Selection [2]

Arc-Length Correspondence :

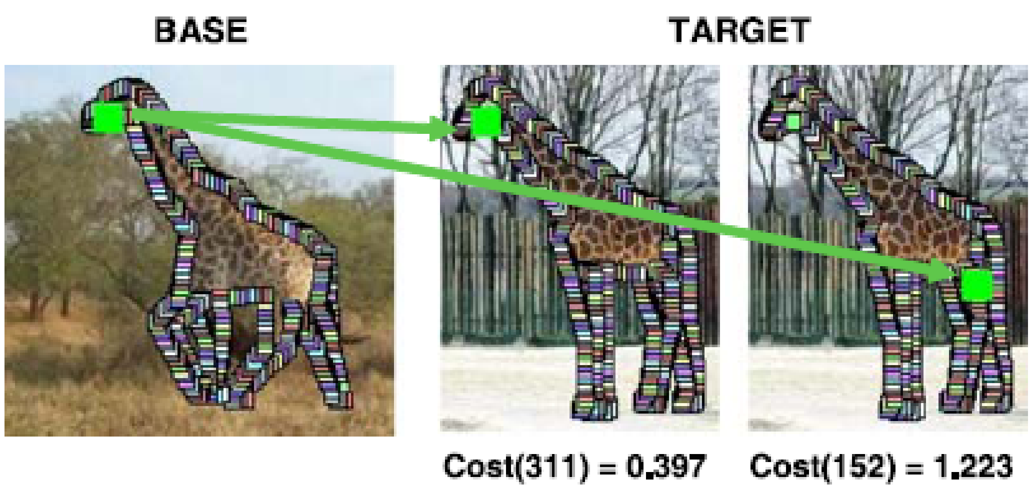
1) Select a base image, make high resolution outline (eg.500 points) , and landmark numbering

2) For the other image, make high resolution outline

3) Because each outline can be landmark 1, compute the cost for all possible choices, and select the minimun one as a correspondance set

$Cost(O^1,O^2)=\sum |d^1_{ij} - d^2_{ij}|^2$

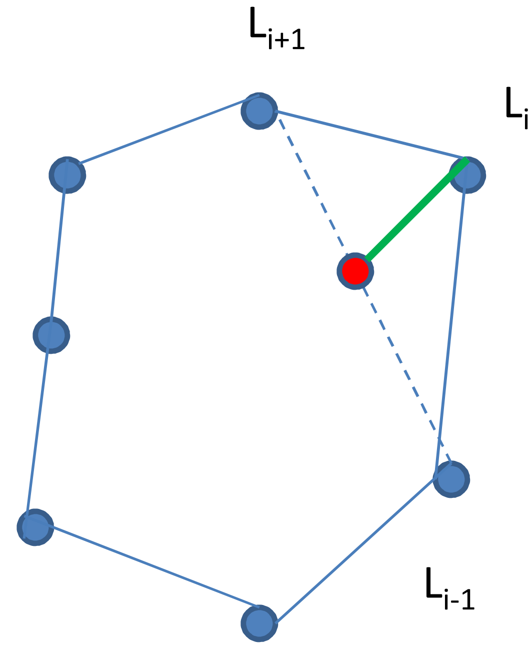
d^m_{ij} is the vector offset between landmark $_i$ and landmark $_j$ on contour m.



Automatic Landmark Selection [3]

Landmark Pruning :

Reduce the number of landmarks used to represent each outline. Points that contribute little information to the outlines is removed.



1) Suppose landmark i of every image is candidate for removal
2) Make a point at the middle of the line between L_{i-1} and L_{i+1}
3) Calculate the mean segment-to-outline squared distance (green line), sum this value for every image, find the cost fucntion

$$C_i = (1/M)*\sum distO_m(l^m_{i-1},l^m_{i+1})$$

4) Remove the landmark whose cost is lowest
5) Teminal the process when the cost of the next removal is more than threshold

$distO_m(l^m_{i-1},l^m_{i+1})$ is mean segment-to-outline squared distance. M is number of instance

LOOPS Model [1]

Correspondance outlines are learned shape and appearance model by LOOPS

LOOPS model combines two components

- Explicit representation of the object's shape (2D silhouette)
- Set of image-based features

  By using Markov Random Field,MRF, the LOOPS model defines a conditional probability distribution over 2N vector coordinates,L

Landmark detector feature

$$
\begin{aligned}
&P(\mathbf{L} \mid \mathcal{I}, \mathrm{w}, \mu, \boldsymbol{\Sigma}) \\
&= \frac{1}{Z(\mathcal{I})} \boxed{P_{\text{Shape}}(\mathbf{L}; \mu, \boldsymbol{\Sigma})} \prod_i \exp\left(w_i \boxed{F_i^{\text{det}}(l_i; \mathcal{I})}\right) \\
&\times \prod_{i,j} \exp\left(w_{ij} \boxed{F_{ij}^{\text{grad}}(l_i, l_j; \mathcal{I})}\right), \qquad (1)
\end{aligned}
$$

Object Shape Model

Gradient Feature

L : 2N vector of image coordinate
w : training weight
u : mean of multivariate Gaussian distribution over landmark locations
∑ : covariance
I : image
Z : normalization factor for the Gaussian density

LOOPS Model [2]

1) Landmark Detector Features

$$F_i^{\text{det}}(l_i;\mathcal{I}) = H_i(l_i).$$

$$H_i(p) = \sum_{t=1}^{T} \alpha_t h_i^t(p),$$

$h_i^t(p)$: a weak feature detector for landmark i

T : number of boosting rounds

$H_i(p)$: strong clasifier

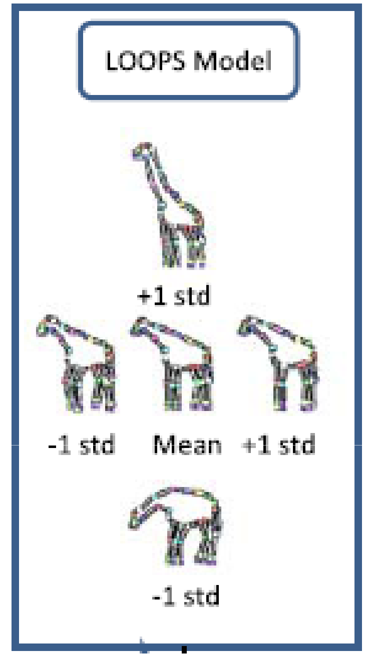
Each landmark detector is trained by constructing a strong boosted classifer from a set of the weak single feature detector such as, SIFT, Shape template, Boundary frangment.

LOOPS Model [3]

2) Object Shape Model

$$P_{\text{Shape}}(\mathbf{L} \mid \mu, \mathbf{\Sigma}) \quad = \frac{1}{Z} \prod_{i,j} \phi_{i,j}(x_i, x_j; \mu, \mathbf{\Sigma}),$$

$\phi_{i,j}$ is potential of shape



3) Gradient Feature

$$F_{ij}^{\text{grad}}(l_i, l_j; \mathcal{I}) = \sum_{r \in \overline{l_i l_j}} |\mathbf{g}(r)^T \mathbf{n}(l_i, l_j)|,$$

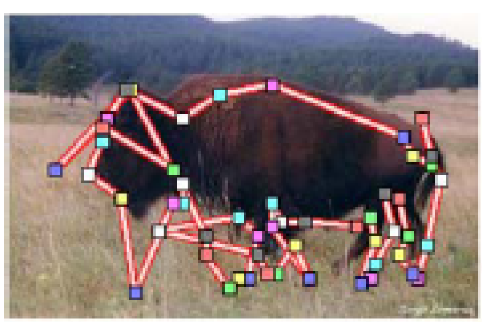
r : point between l_i and l_j
g(r) : gradient at point r
$n(l_i,l_j)$: normal to the edge between l_i and l_j

This term estimates the likelihood of the the edge connecting the landmark
High value of this term encorages the boundary lie along segment of high gradient

Localization of Object Classes [1]

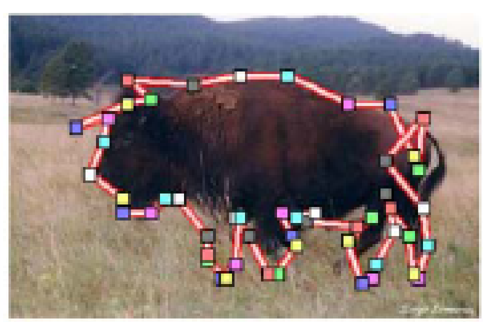
- Outline objects by using probabilistic inference to find the most probable assignment $\mathbf{L}^* = \arg\max_{\mathbf{L}} P(\mathbf{L} \mid \mathcal{I}, \mathbf{w}).$

- Use a simple two-step process
  - First, approximate the a problem and find coarse solution using discrete inference
  - Refine the solution using continuous optimization



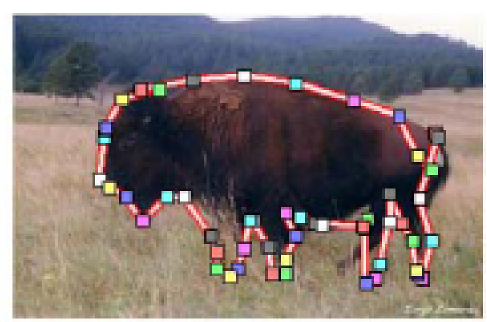
Candidate ⇒

Figure : a



Discrete ⇒

Figure : b



Refinement

Figure : c

Localization of Object Classes [2]

Discrete Inference :

- Assume that landmarks will fall on interesting points in the image, and consider only points found by SIFT operator
- Adjust the setting of SIFT operator to produce 1000-2000 descriptors
- Make use of MRF appearance based feature function F_i^{det}

  to indentify the most promising candiate pixel assignment for each landmark
- Use the corresponding F_i^{det} to score all the SIFT keypoint pixels and choose the top 25 local optima as candidate assignments (Figure a)
- Performing approximate max product inference by using Residual Belief Propagation (Elidan et al. 2006b).(Figure b)

Refinement :

- Use a greedy hill climbing algorithm in which we iterate across each landmark, moving it to the best candidate location, while holding the other landmards fixed (Figure c)

Experimental Evaluation of LOOPS Outlineing [1]

1) **Compare accuracy of LOOPS outline with two state-of-the-art methods**

   - OBJCUT method (Prasad and Fitzgibbon, 2006)

     - uses an exemplar-based shape model of the object class together with a texture model to find the best match

   - kAS Detector (Ferrari et al., 2008)

     - uses adjacent contour segments as features for detector.

Experimental Evaluation of LOOPS Outlineing [2]

Results from LOOPS, OBJCUT, and kAS

Experimental Evaluation of LOOPS Outlineing [3]

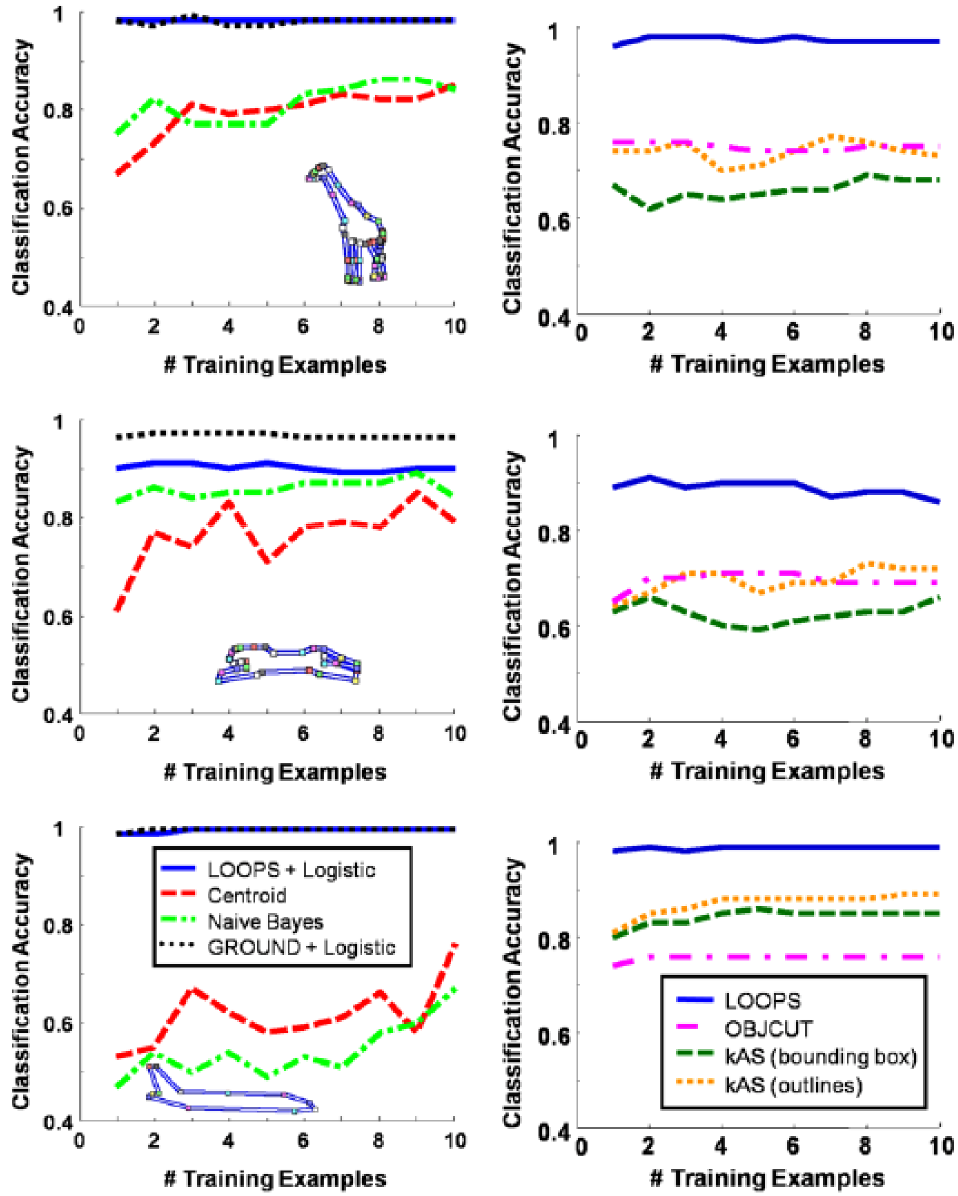
- Measure the symmetric root mean squared (rms) distance between the produced outlines and the hand-labeled groundtruth

Class	LOOPS	OBJCUT	kAS Detector (bounding box)	kAS Detector (outline)
Airplane	1.9	5.5	3.8	3.6
Cheetah	5.0	12.3	11.7	10.5
Giraffe	2.9	10.5	8.7	8.1
Lamp	2.9	7.3	5.8	5.3

Comparing rms distance from each technique, LOOPS clearly produces more accurate outline than the others

Experimental Evaluation of LOOPS Outlineing [4]

2) Compare descriptive task classification accuracy



Three descriptive classification tasks

1) Giraffes standing VS bending down
2) Cheetahs running VS standing
3) Airplanes taking off VS flying horizontally

From the experiment, LOOPS could performe the best result

Application of LOOPS Outlines [1]

1) User-defined Landmark Queries

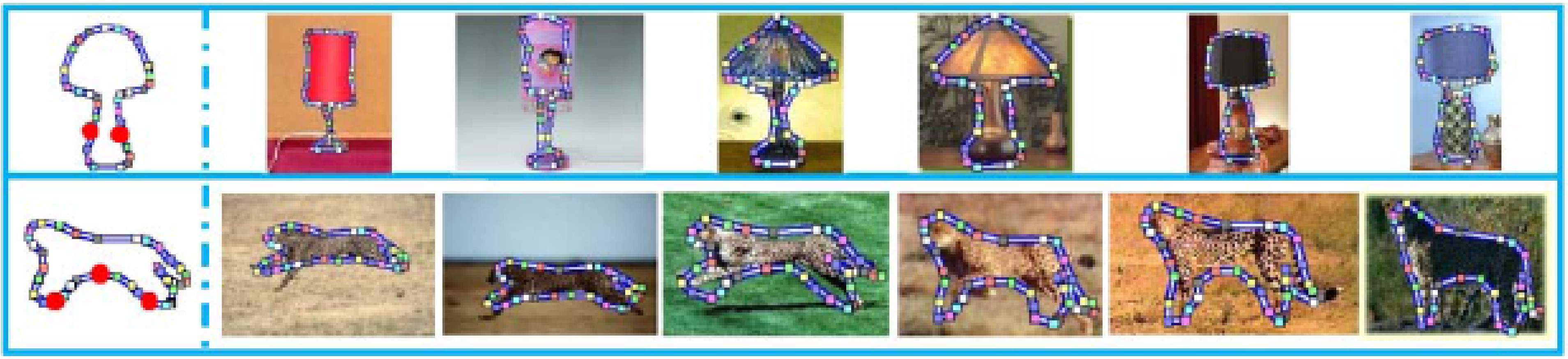
Allow the user to query the objects in our image using refined shape-based queries.

Example of question

Where is the giraffe’s head ?

Show the distance between 2 points on the lamp

Find the angle from 3 points at cheetah’s stromach



Application of LOOPS Outlines [2]

2) Shape Similarity Search

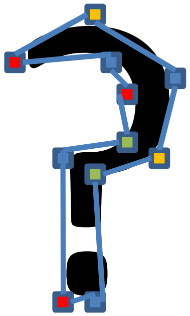
Application of LOOPS Outlines [3]

3) Descriptive Clustering

To make groups of similar looking from a large database

(c) (d)

Question



THANK YOU