5. Object recognition in videos

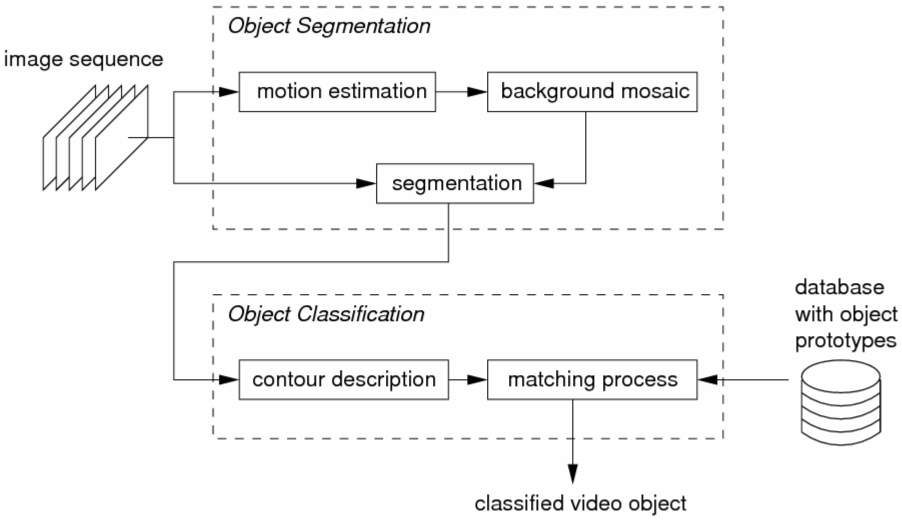
Semantic video analysis

Stephan Kopf

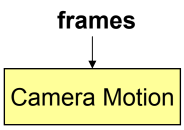
Outline

- General idea
- Object segmentation and parameterization
- Classification of objects
- Experimental results

General approach

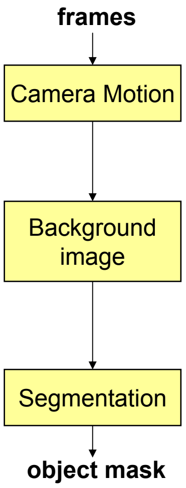


Segmentation (I)



- Assumption: At least half of the visible area in each frame is background.
- Estimate the camera motion between consecutive frames.
- Calculate parameters of the camera model.

Segmentation (II)



- Assumption: At least half of the visible area in each frame is background.
- Estimate the camera motion between consecutive frames.
- Calculate parameters of the camera model.
- Apply a median filter on the transformed frames to construct the background image.
- Compare the background image with the transformed frame to get the object mask.

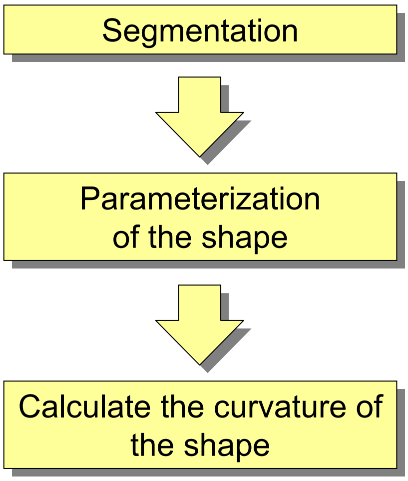
Object classification approach

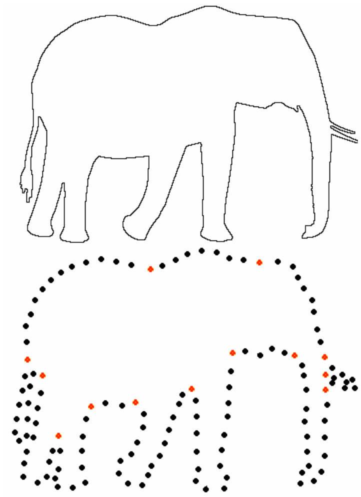
1. Sample the shape of an segmented object with a fixed number of sample points.
2. Identify feature points to describe the curvature of the shape.

Features are calculated based on the ***Curvature Scale Space Image***

3. Compare the features of the unknown shape with features of objects that are stored in a database

Parameterization



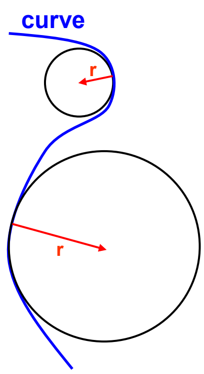


Definition of the curvature (I)

- The curvature at a given point has a magnitude equal to the reciprocal of the radius of an osculating circle (the circle touches the curve):

$$K = \frac{1}{r}$$

- The curvature is a vector pointing to the direction of the circle's center.
- A small circle corresponds with a high curvature's magnitude, a straight line has a curvature of zero.



Definition of the curvature (II)

- Consider a plane curve ***u(t)*** that lies completely within a 2D plane. ***u(t)*** is parameterized by the arc length ***t***.
- The curve ***u(t)*** is parametrically defined by two functions ***x(t)*** and ***y(t)***:

  u(t) = (x(t), y(t)).

- Curvature ***K*** for a plane curve ***u(t):***

$$K = \frac{\dot{x} \cdot \ddot{y} - \dot{y} \cdot \ddot{x}}{\left(\dot{x}^2 + \dot{y}^2\right)^{3/2}}$$

$\dot{x}$ and $\dot{y}$ define the first derivative (gradient),
$\ddot{x}$ and $\ddot{y}$ the second derivative (change of the gradient).

Definition of the curvature (III)

- We can derive a less general definition of the curvature if we explicitly use plane curves defined by ***y = f(x)***. We get the following definition of the curvature for each point ***(x, f(x))***:

$$K = \frac{f''(x)}{\left(1 + \left(f'(x)\right)^2\right)^{3/2}}$$

- This form is widely used in engineering:
  - to approximate the fluid flow around surfaces, e.g. in aerodynamics (gases) or hydrodynamics (liquids),
  - to derive the characteristic behavior when bending structural elements, e.g., put weight on a beam and analyze the flexure.

Definition of the curvature (IV)

Example

- Consider the parameterized curve ***u(t) = (x(t), y(t)) = (t, t^2)***.
  The explicit definition of this curve is: ***y = f(x) = x^2***.
- Calculate curvature based on the **parameterized curve**:
  First and second derivatives: $\dot{x}$= 1, $\ddot{x}$= 0, $\dot{y}$= 2t, $\ddot{y}$= 2.

$$K(t) = \frac{\dot{x}\cdot\ddot{y} - \dot{y}\cdot\ddot{x}}{\left(\dot{x}^2 + \dot{y}^2\right)^{3/2}} = \frac{1\cdot 2 - 2t\cdot 0}{\left(1^2 + (2t)^2\right)^{3/2}} = \frac{2}{\left(1+4t^2\right)^{3/2}}$$

- Calculate curvature based on the **explicit definition**:

$$f'(x) = 2x, \quad f''(x) = 2 \quad K(x) = \frac{f''(x)}{\left(1+\left(f'(x)\right)^2\right)^{3/2}} = \frac{2}{\left(1+4x^2\right)^{3/2}}$$

Calculation of the curvature (V)

- Approximation of derivatives for discrete values (parameterized shapes):

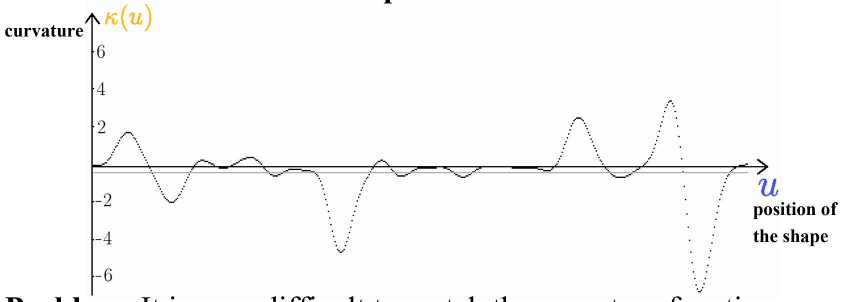
$$\dot{x}(t) = \frac{x(t+1) - x(t-1)}{2\cdot hx}$$

$$\dot{y}(t) = \frac{y(t+1) - y(t-1)}{2\cdot hy}$$

- Parameter ***t*** is defined for whole numbers $(t \in \mathrm{N})$.
- ***hx*** and ***hy*** normalize the derivatives depending on the distance between sample points.

Shape matching based on curvatures

Curvature function of a shape

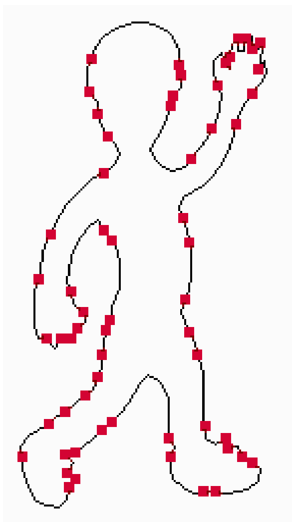


Problem: It is very difficult to match the curvature functions of two shapes.

→ Identify significant curvature features.
We use the curvature scale space technique.

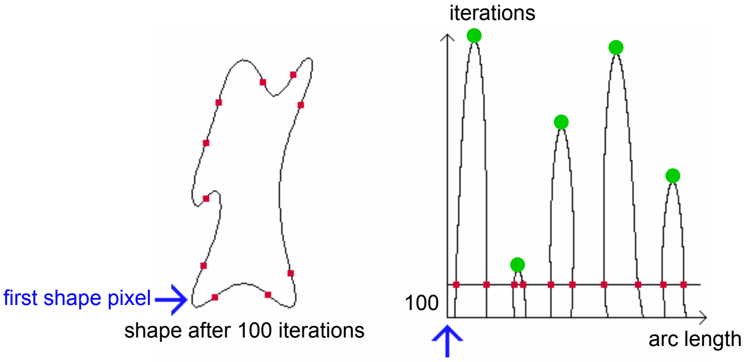
Curvature scale space (I)

- Analyze the outer shape of an object.
- Smooth the shape with a Gaussian kernel.
- The **inflection points** in each iteration are used as features to describe the object.



Curvature scale space (II)

- A **curvature scale space image** is a visual representation of the inflection points during the smoothing process.



- The **peaks** are used as features to describe the object.

Curvature scale space (III)

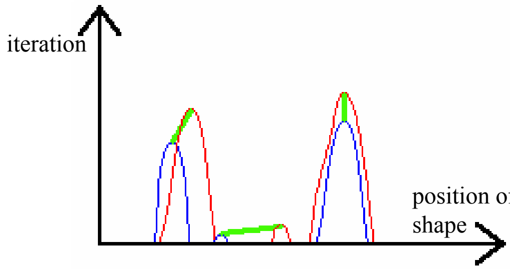
Characteristics of CSS images

- The peaks in the curvature scale space image describe **concave segments** of a shape.
- The peaks are used as features to describe the shape.
- Each peak characterizes
  - a **position**
    relative position compared to other peaks,
  - a **value**
    strength of the concave segment.

Comparison of two shapes

1. Shift one curvature scale space image until the largest peaks match (makes the approach invariant to rotation)
2. Calculate the Euclidian distance between two peaks

$$D = \sqrt{d_u \cdot d_u + d_t \cdot d_t}$$



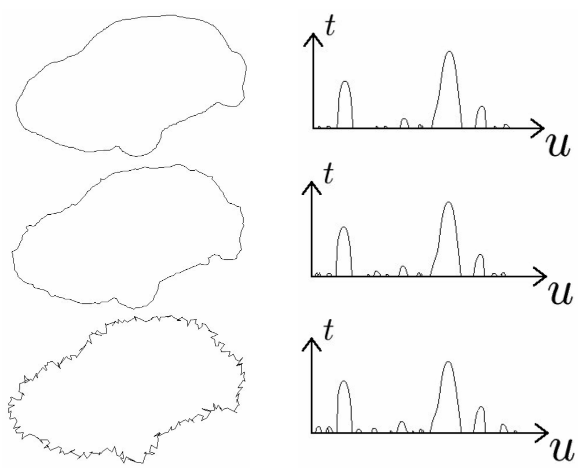
3. Summarize the distances

Semantic video analysis – Stephan Kopf – HS2006 – Chapter 5: Object recognition

17

Features of CSS images (I)

- Original shape of a car
- Shape with noise
- Shape with severe noise



→ CSS images are very similar

Semantic video analysis – Stephan Kopf – HS2006 – Chapter 5: Object recognition

18

Features of CSS images (II)

Pros

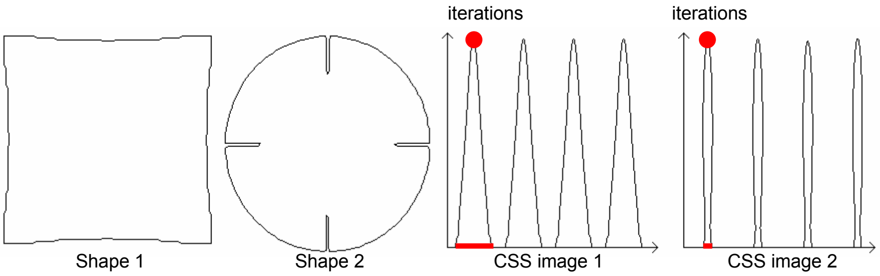
- Only a few values are required to describe complex objects.
- The approach is invariant to rotation or scaling.
- Low computation time.

Cons

- Bad classification results with some shapes:
  - shallow and deep concavities
  - convex regions

Ambiguities of CSS Images (I)

Shallow vs. deep concavities

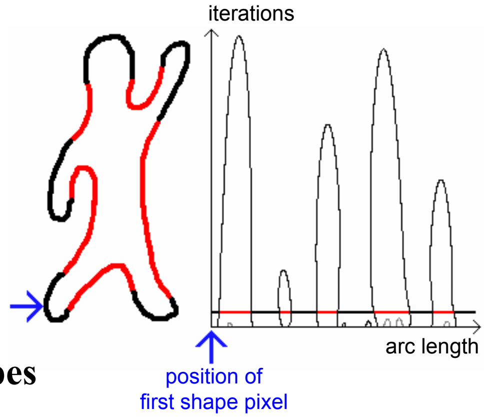


→ Solution: Use **position**, **height** and **width** of each peak as feature.

Ambiguities of CSS Images (II)

Convex regions

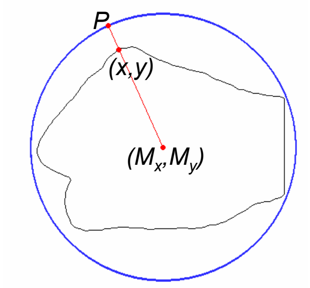
- Poor representation of convex regions of a shape.
- Convex objects cannot be represented at all.

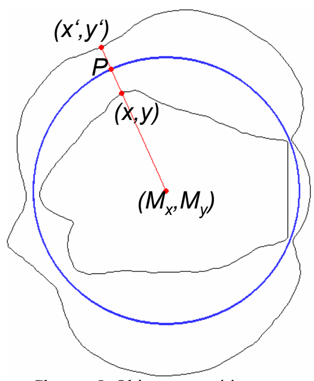


→ **Solution: Mapped shapes**

Mapped Shapes (I)

- **Idea**: Reflect each shape pixel and create a new shape.
- Strong convex segments of the original shape become concave segments of the mapped shape.

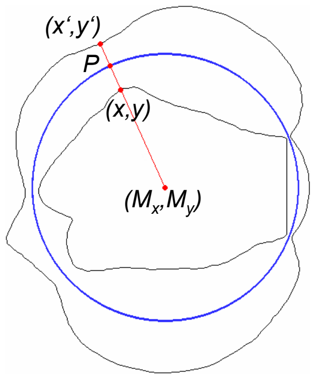




Mapped Shapes (II)

Calculation of mapped shapes

$$
\begin{aligned}
D_{x(u),y(u)} &= \sqrt{(M_x - x(u))^2 + (M_y - y(u))^2} \\
x'(u) &= \frac{2 \cdot R - D_{x(u),y(u)}}{D_{x(u),y(u)}} \cdot (x(u) - M_x) + M_x \\
y'(u) &= \frac{2 \cdot R - D_{x(u),y(u)}}{D_{x(u),y(u)}} \cdot (y(u) - M_y) + M_y
\end{aligned}
$$



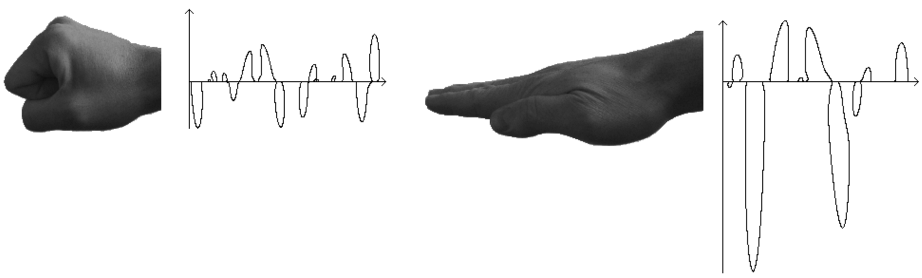
R: radius of the circle

D: distance between (M_x, M_y) and (x,y)

Mapped Shapes (III)

CSS images with mapped shapes

- Get standard curvature scale space features.
- Calculate features for the mapped shape.



Object recognition in videos (I)

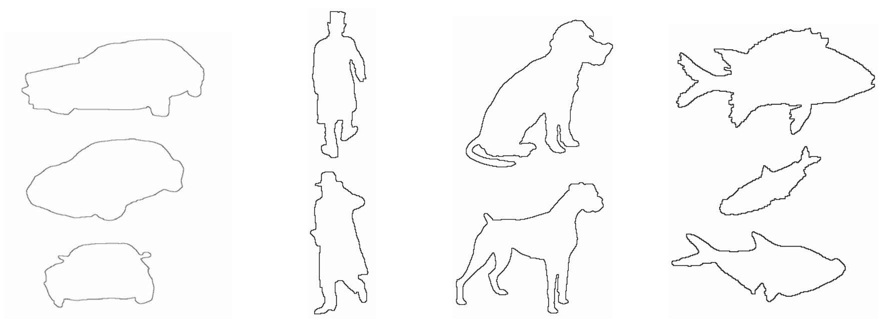
Approach

- Compare each object in the sequence with the objects in the database.
- Calculate average difference of the objects in the database to each object class.
- Display most similar objects.

Object recognition in videos (II)

Objects in the database

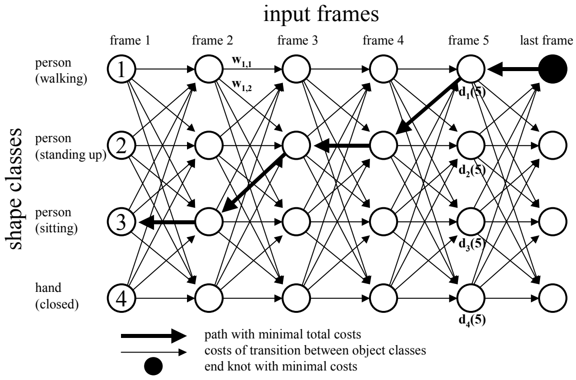
- 300 objects are stored in the database.
- 13 object classes group similar objects.



car person dog fish

Object recognition in videos (III)

Aggregation of object classification results



- Calculate distance $d_c(i)$ between input object i and object class c.
- Transition costs $w_{n,m}$ occur for each change of the object class.

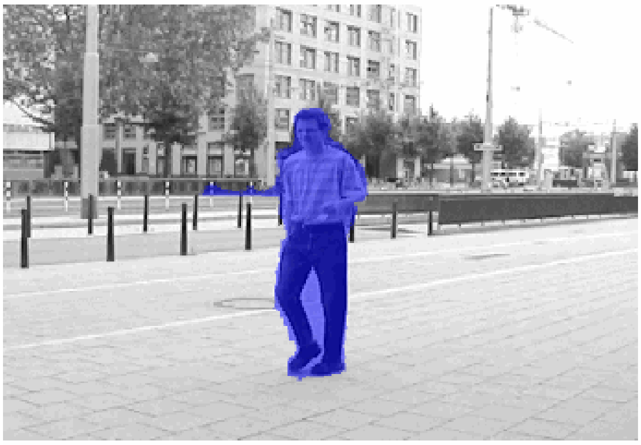
- Solve minimization problem: $\min_c \sum_{i=1}^{N} (d_{c(i)}(i) + w_{c(i),c(i-1)})$

Object recognition in videos (IV)

Recognition rates

- Recognition rates between 25-95 % (depends on the complexity of the object and the number of objects in the database).
- The recognition rates of rigid objects (e.g., a car) is much higher compared to deformable objects.
- The curvature scale space approach is invariant to scaling and rotation, and it is very robust to noise
- The comparison is very fast (smooth shape once and calculate Euclidean distances)

Object recognition in videos (V)



standing
walking
turning around
sitting down
sitting
open hand
closed hand
fist
thumb

Questions ?