Thermal Video Processing (Object Detection)

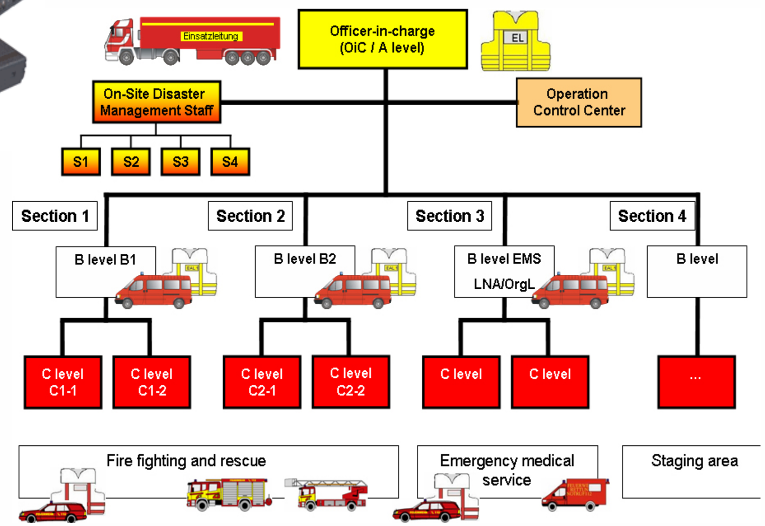
András Hajdu
University of Debrecen, Hungary

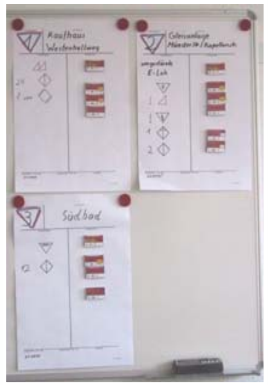
Summer School on Image Processing
10/07/2008

IMAGE PROCESSING GROUP OF DEBRECEN
http://ipgd.inf.unideb.hu

Tools for Rescue Operation Management – Today

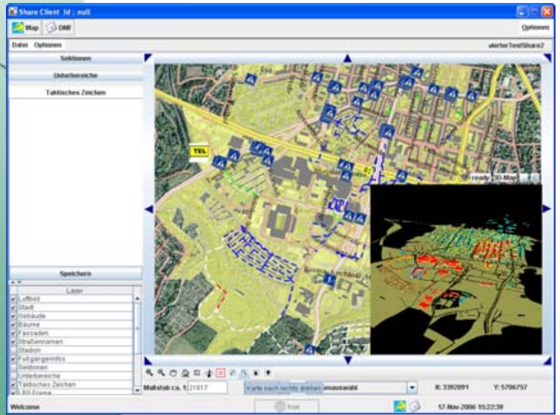
10/07/2008
SSIP 2009

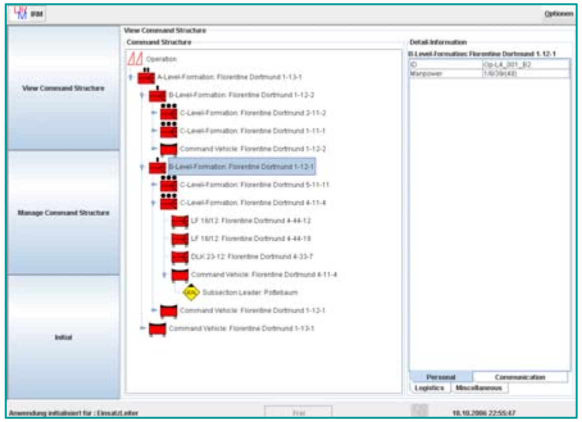


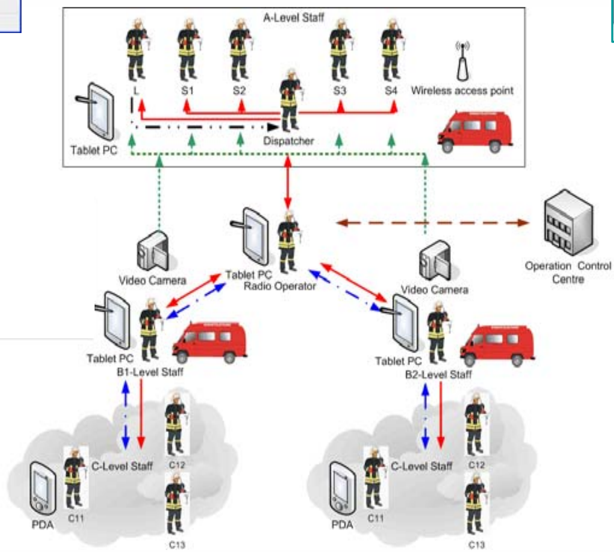


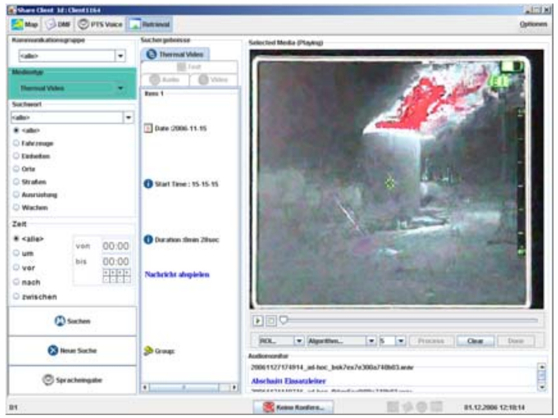
Tools for Rescue Operation Management – SHARE

10/07/2008
SSIP 2009



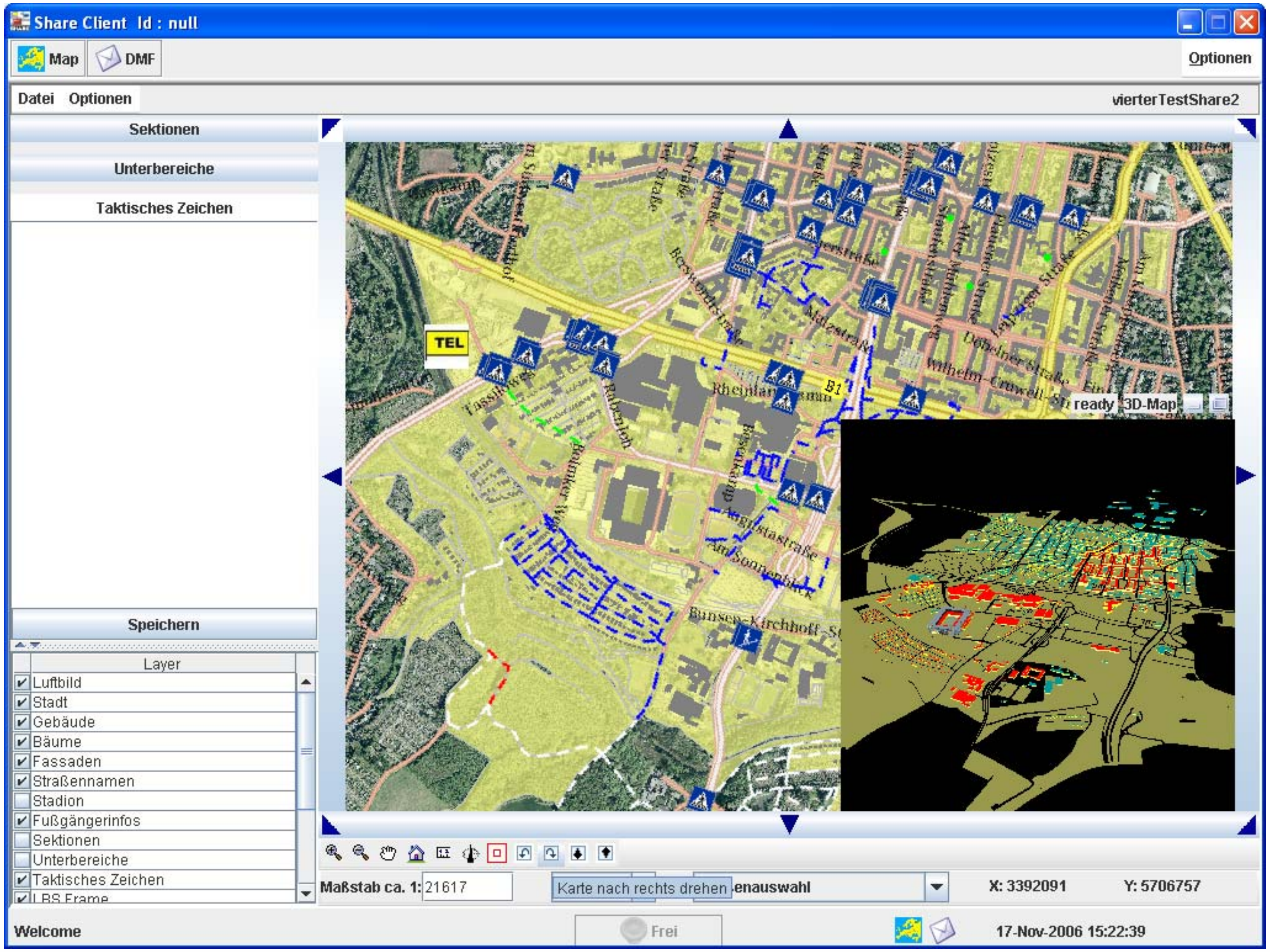






SHARE Overview: MAP3D

10/07/2008
SSIP 2009

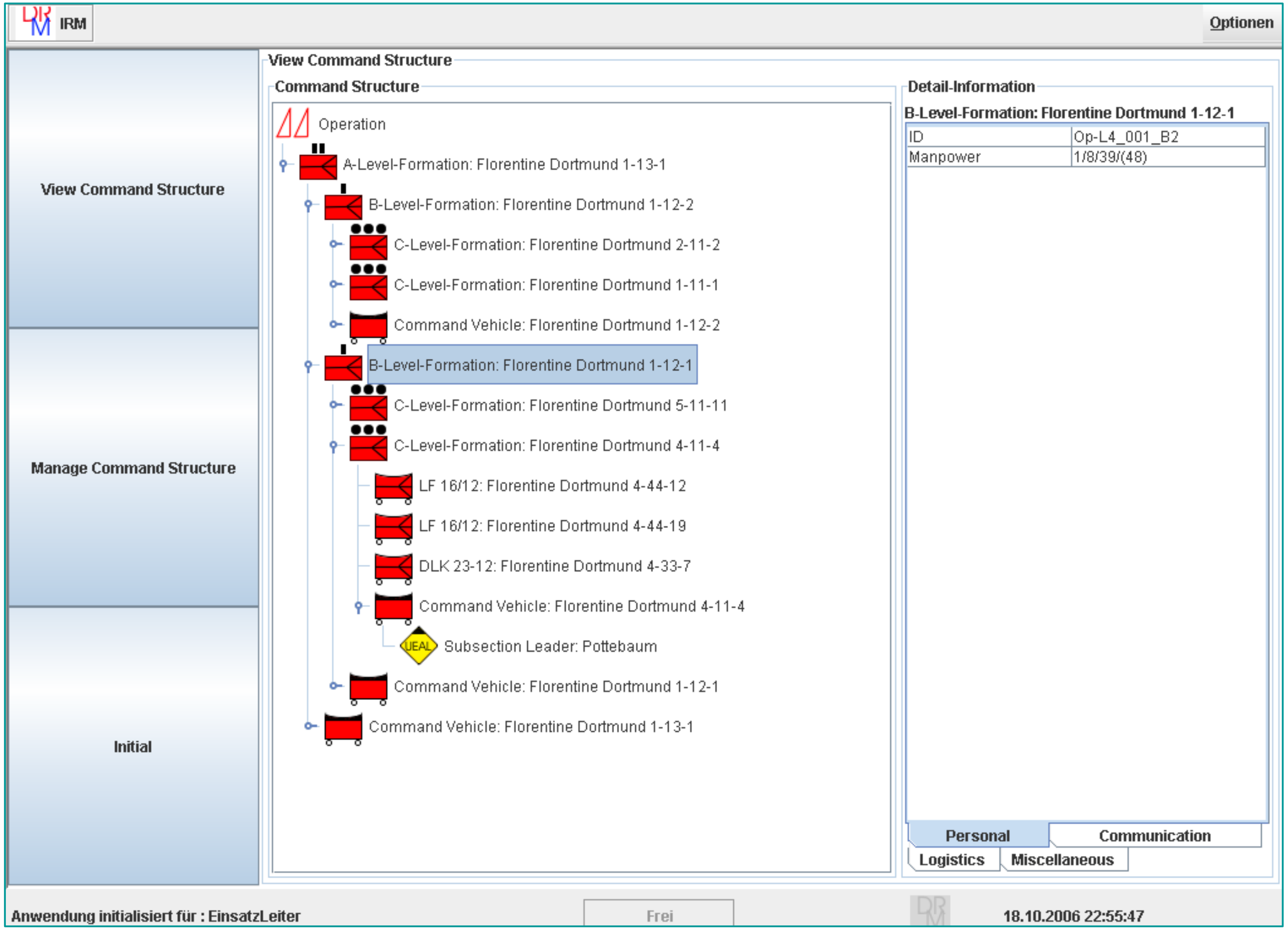


SHARE Overview:
PTS Video

10/07/2008
SSIP 2009

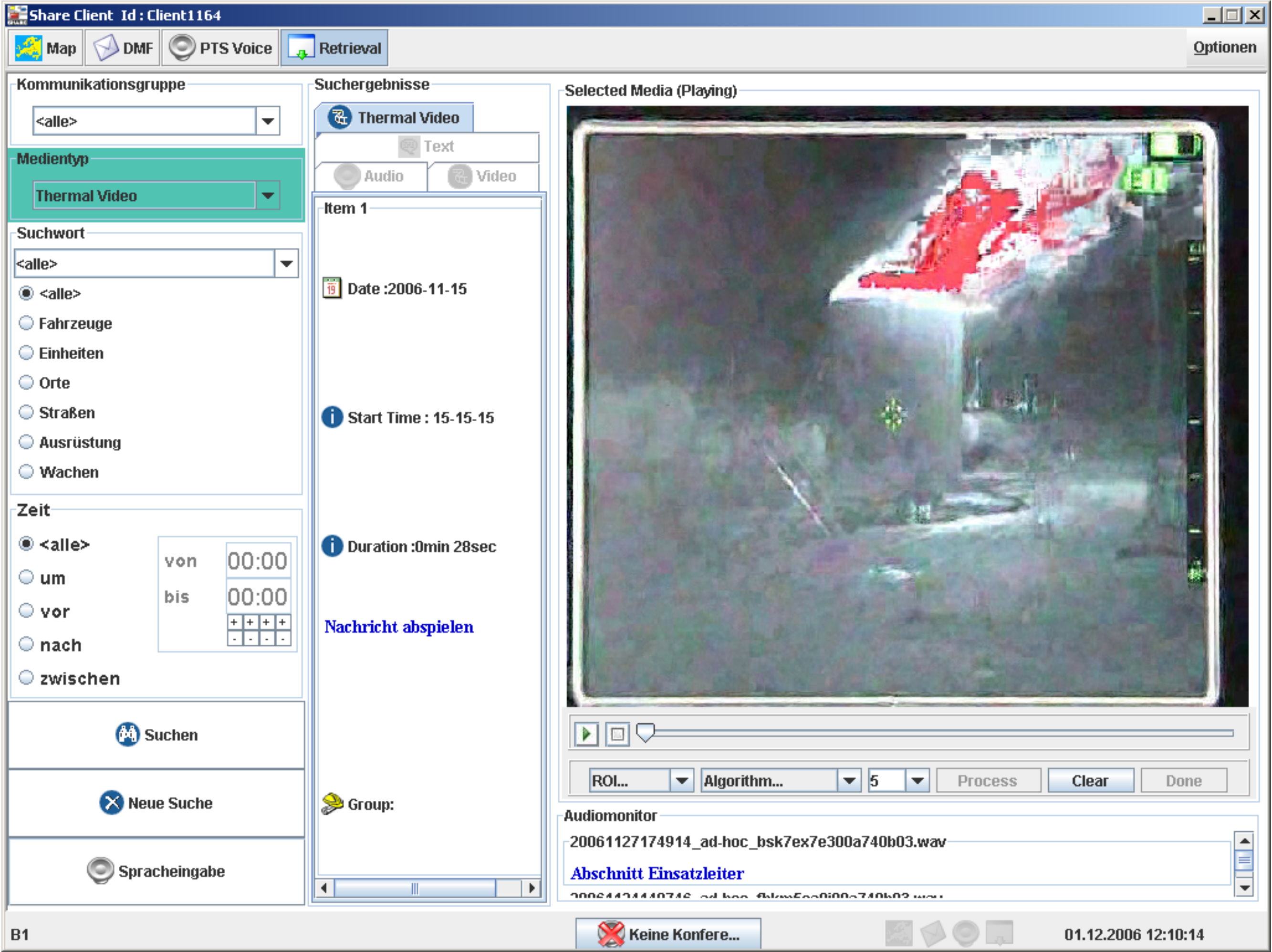
SHARE Overview: Interactive Resource Management

10/07/2008
SSIP 2009



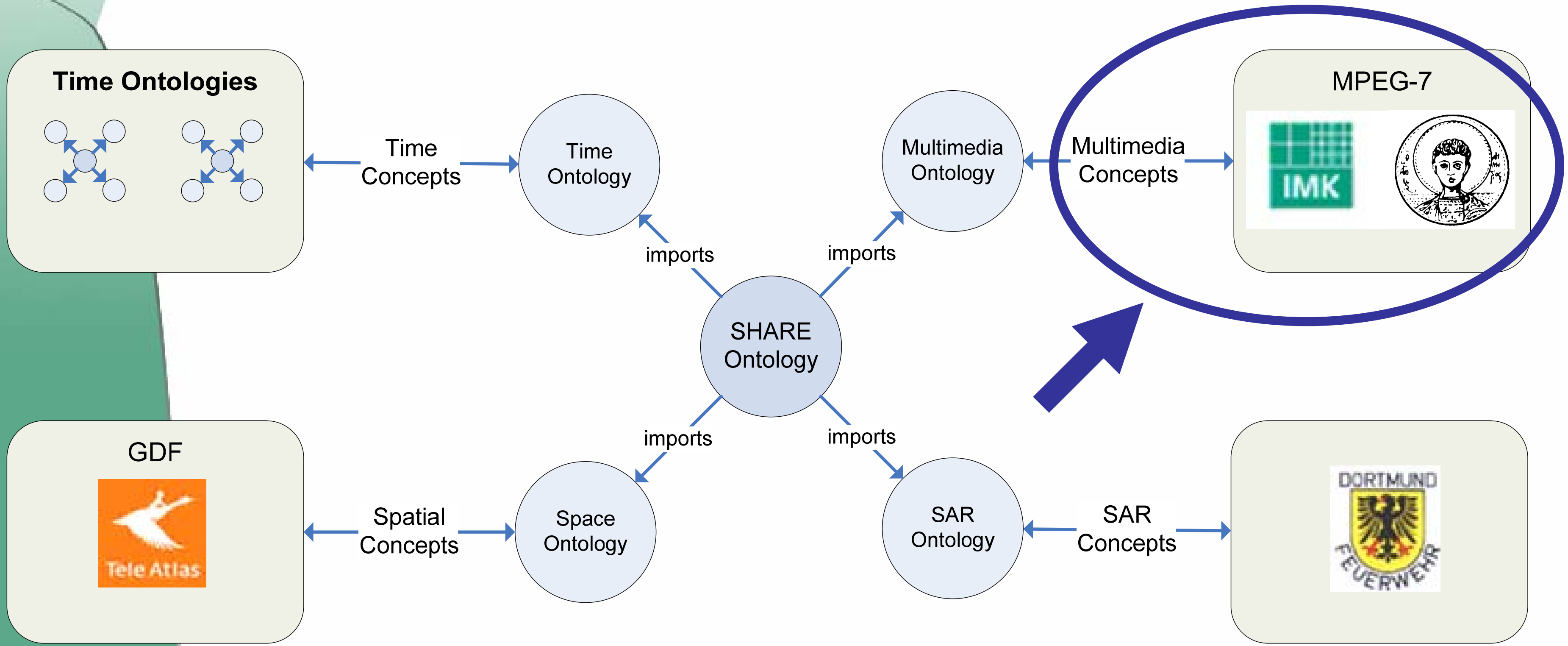
SHARE Overview: Thermal Video Processing

10/07/2008
SSIP 2009



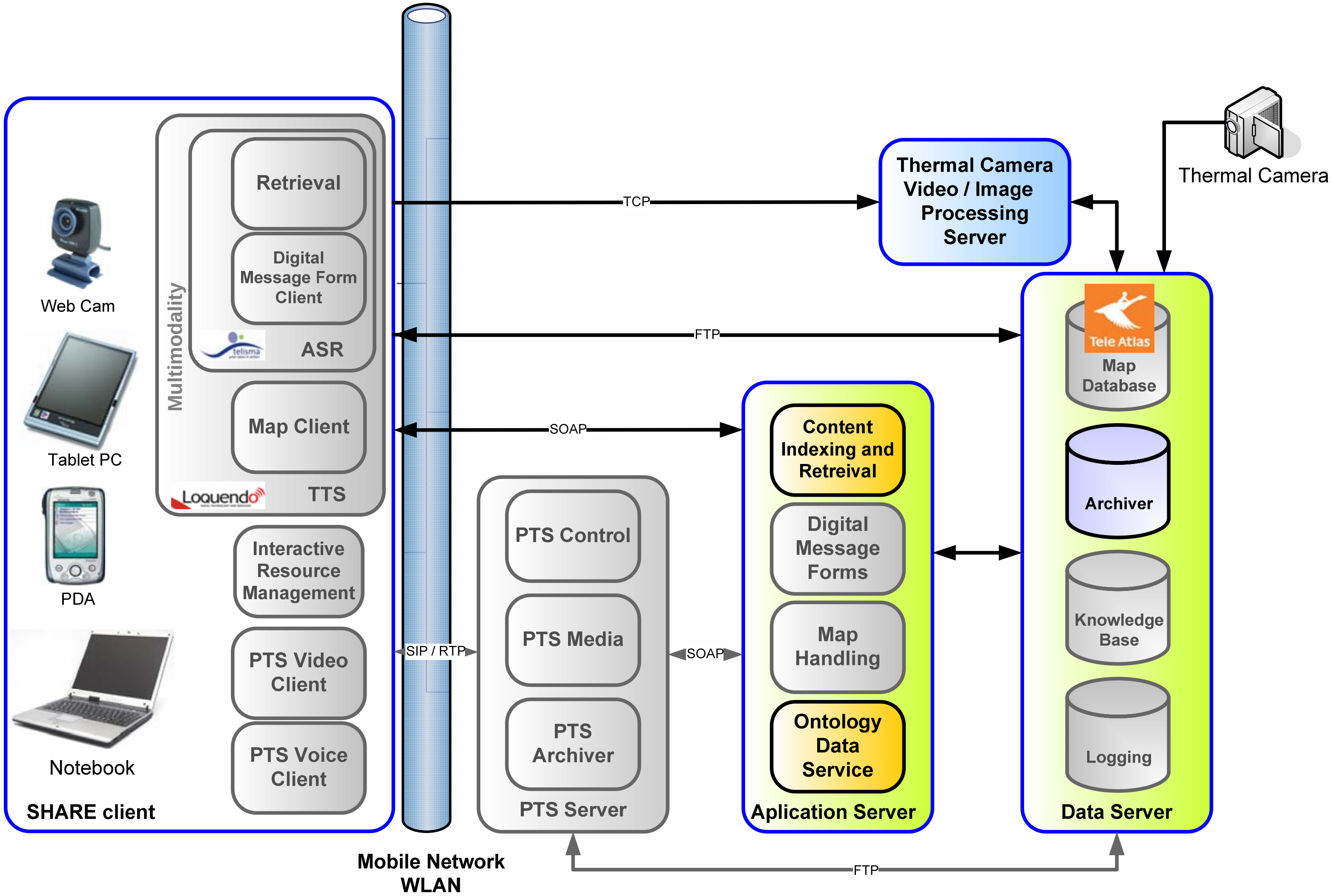
SHARE architecture

10/07/2008
SSIP 2009



SHARE architecture
Thermal image and video processing

10/07/2008
SSIP 2009



Objectives

- **To provide tools** for image and video analysis for assisting a fire-brigade or another rescue team during and after the rescue operation.

- **Integration** of state-of-the-art techniques for analyzing images/videos acquired using a thermal camera in a fire/smoke environment.




Normal vs. thermal view: An example

General aims

Fire focus

Fire occurrences according to hotspot detection to localize the fire, estimate its spreading and determine/estimate its change in time.

Person detection

Human occurrences in thermal videos to localize human silhouettes to derive rescue and positioning information.

Person tracking

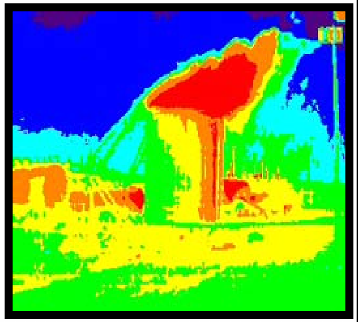
Human movement analysis in the fire videos according to the trajectory analysis of them. E.g. slower/faster movements, directions.

End-user requirements

10/07/2008
SSIP 2009

Colorization

The thermal camera view is generally B&W. The grey tones make it difficult to get a quick overview of thermal characteristics of different objects.



Interactive object detection and recognition

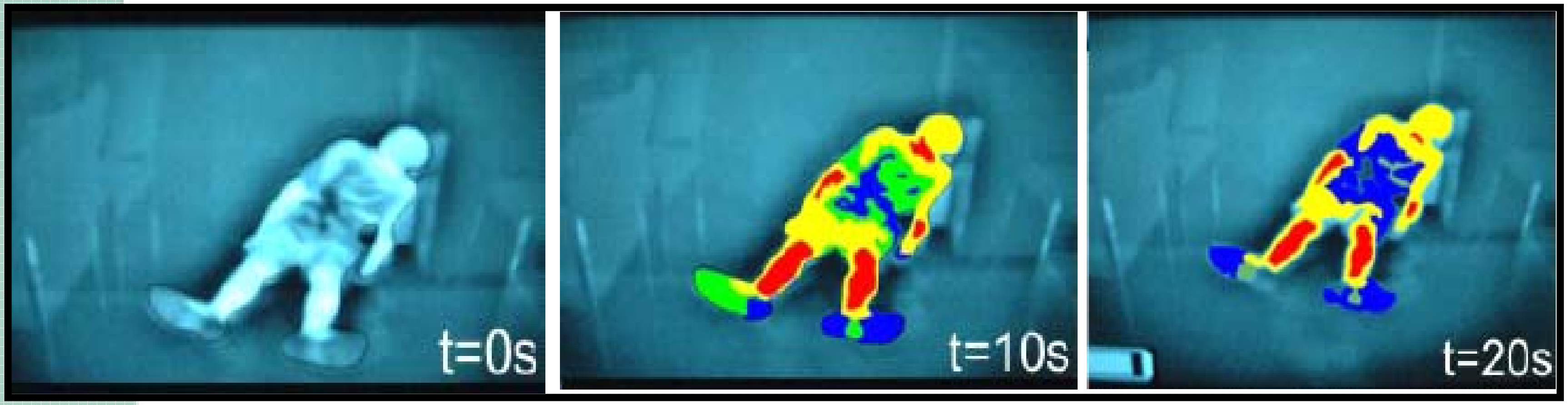
The recognition of objects is difficult. The user has to click on any area and the recognition module should mark the object. A further requirement is identification i.e. as a person.

End-user requirements

10/07/2008
SSIP 2009

Tactical effectiveness

A very important function is the time control of tactical effectiveness. This function is needed to see the thermal condition of an object within different periods (e.g. 10 seconds steps). The thermal status of the marked object is visualized by colored faces. This helps the officer in charge to verify the effectiveness of his tactic, and to correct the methods very fast.

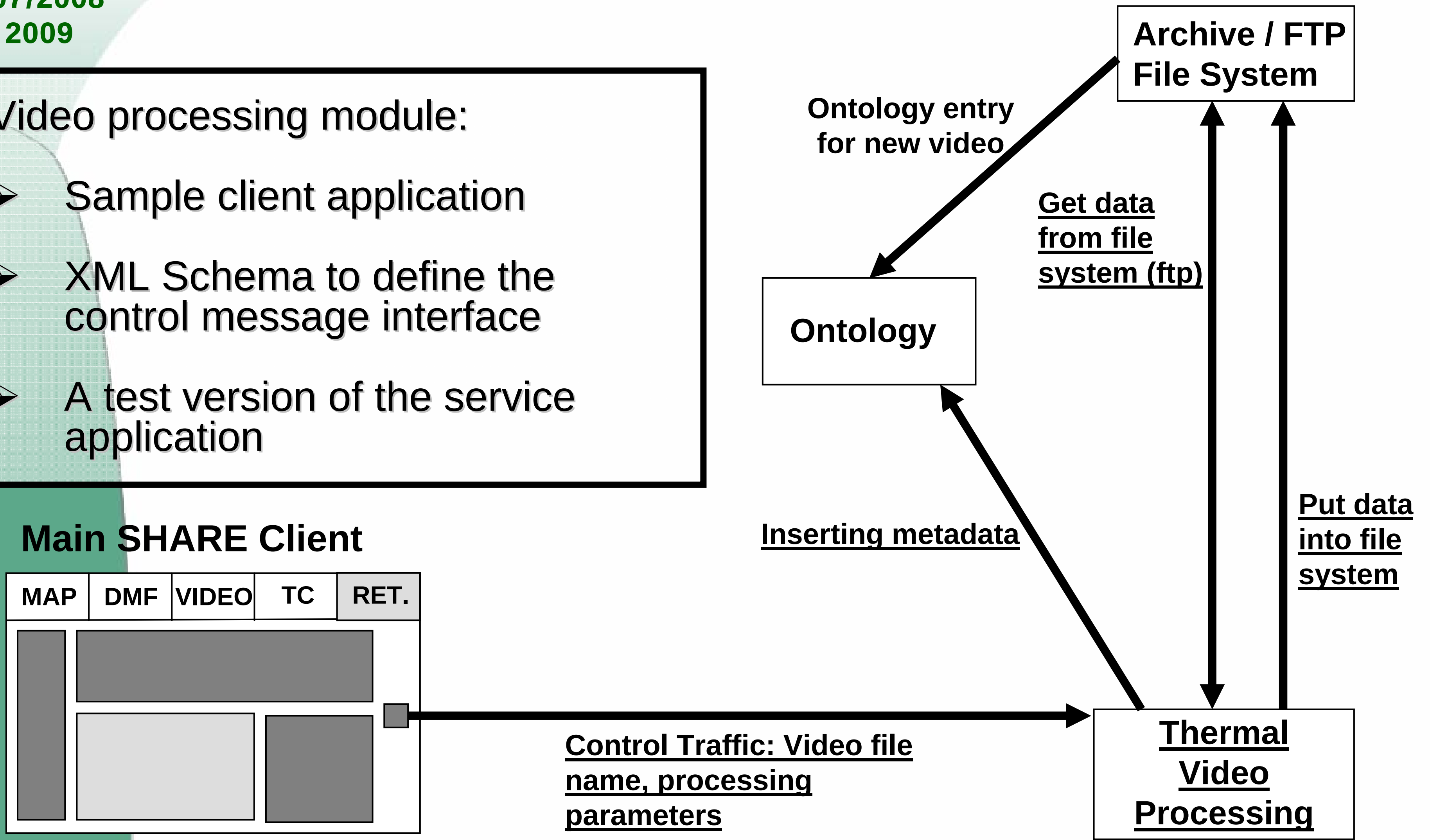


Integration of the thermal video processing component

10/07/2008
SSIP 2009

Video processing module:

- Sample client application
- XML Schema to define the control message interface
- A test version of the service application



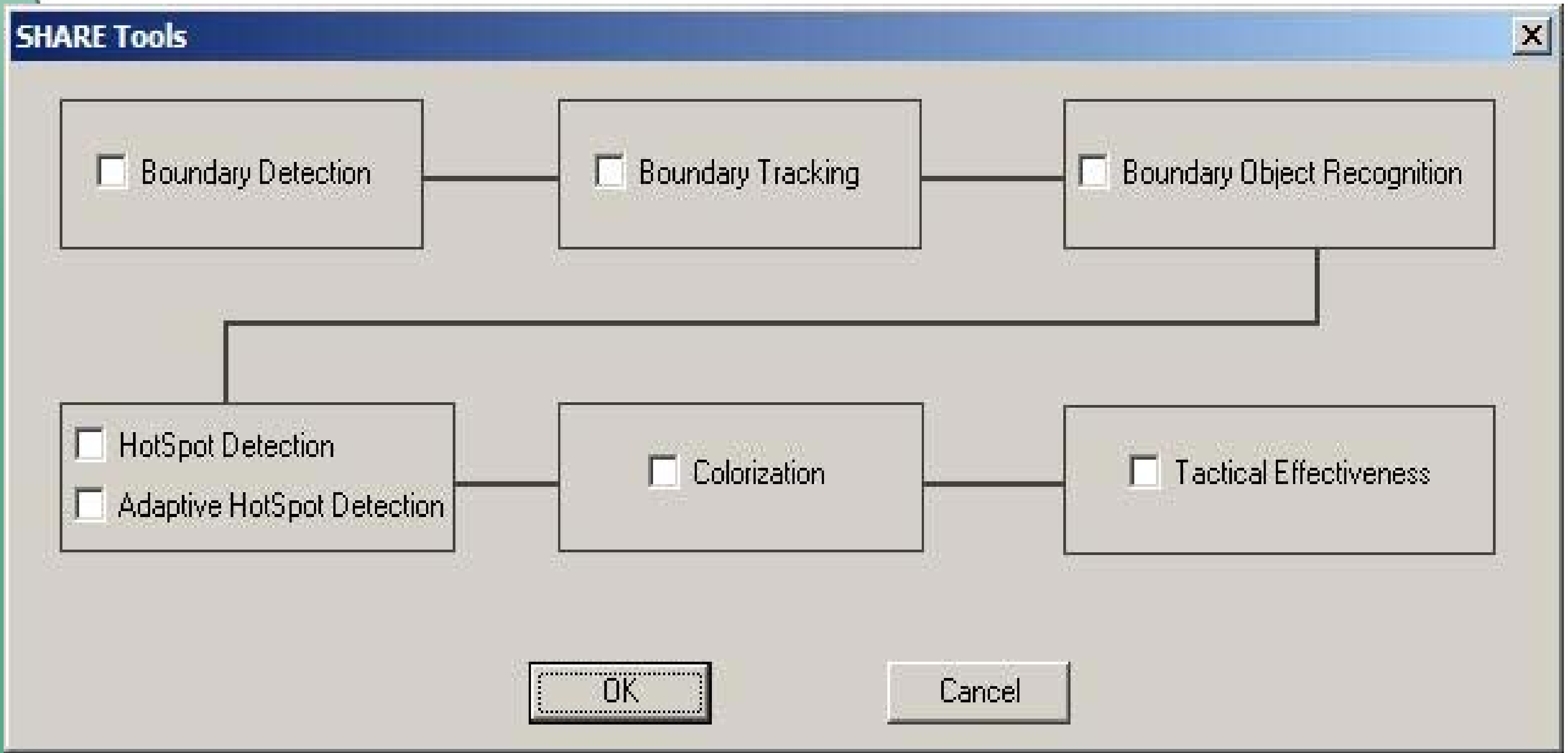
Integration of the thermal video processing component

- Defining GUI requirements for the SHARE client
- Creating a service module with supporting video processing tasks for the second prototype
- Creating a command message interface
- Releasing a sample client
- Adding XML parsing functionalities to interpret the control message and prepare Java source for ODS

Requirements for the client GUI

10/07/2008
SSIP 2009

- SHARE thermal video processing tools (1st prototype GUI):



Requirements for the client GUI

- Selection of regions of interest (1st prototype GUI):

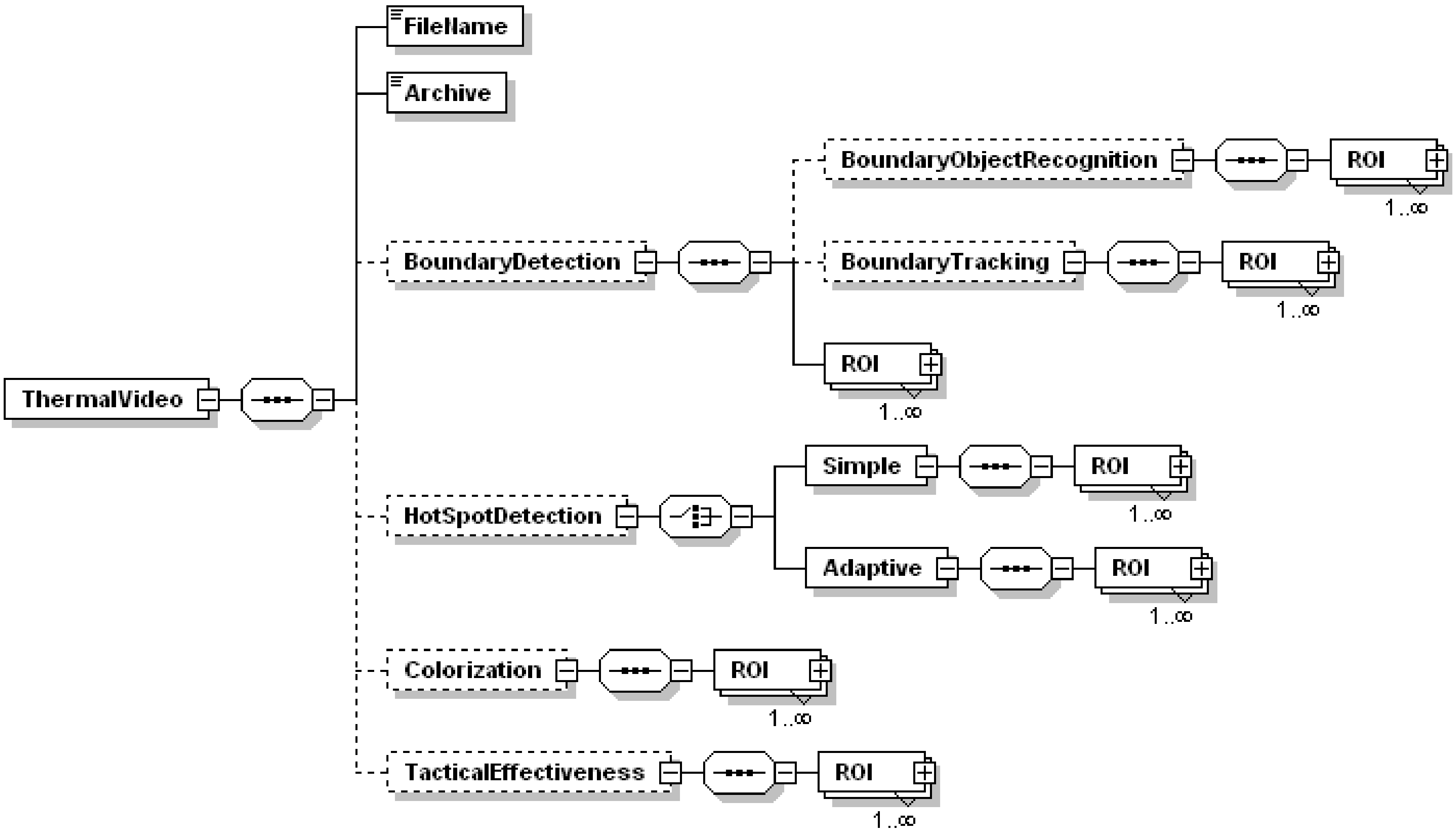
Bounding box	Snake
	

Definition of a control message interface

- Information sent from the client to the service:
  - Data server and directory information for the input video (video output is generated here, as well),
  - Name of the file to be processed,
  - Processing tasks together with their ROIs and scope (start frame/end frame information),
  - The control message is defined in terms of an XML Schema, and thus the message is sent as an XML file (IP package).

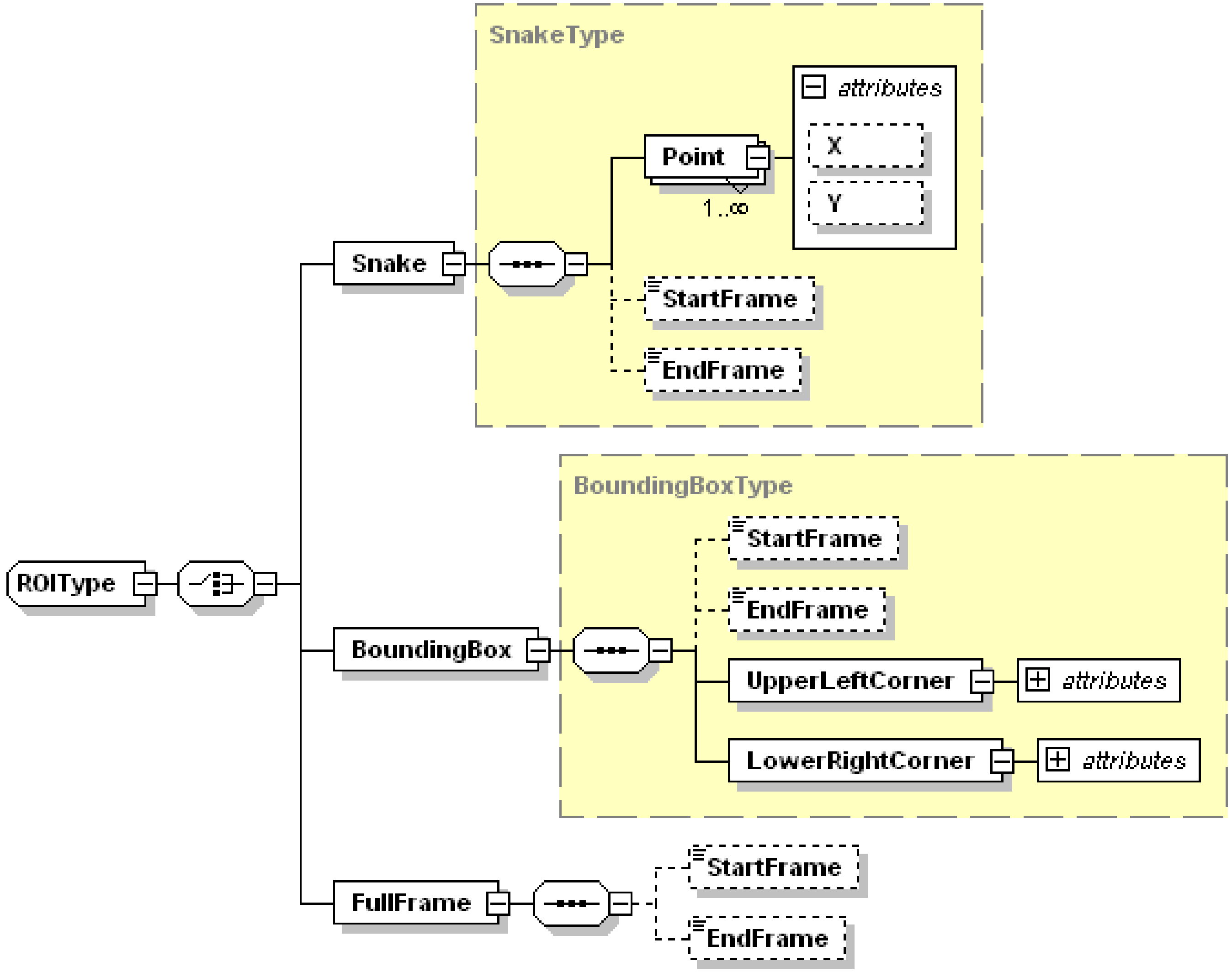
Control message interface, root class

10/07/2008
SSIP 2009



Control message interface, ROI class

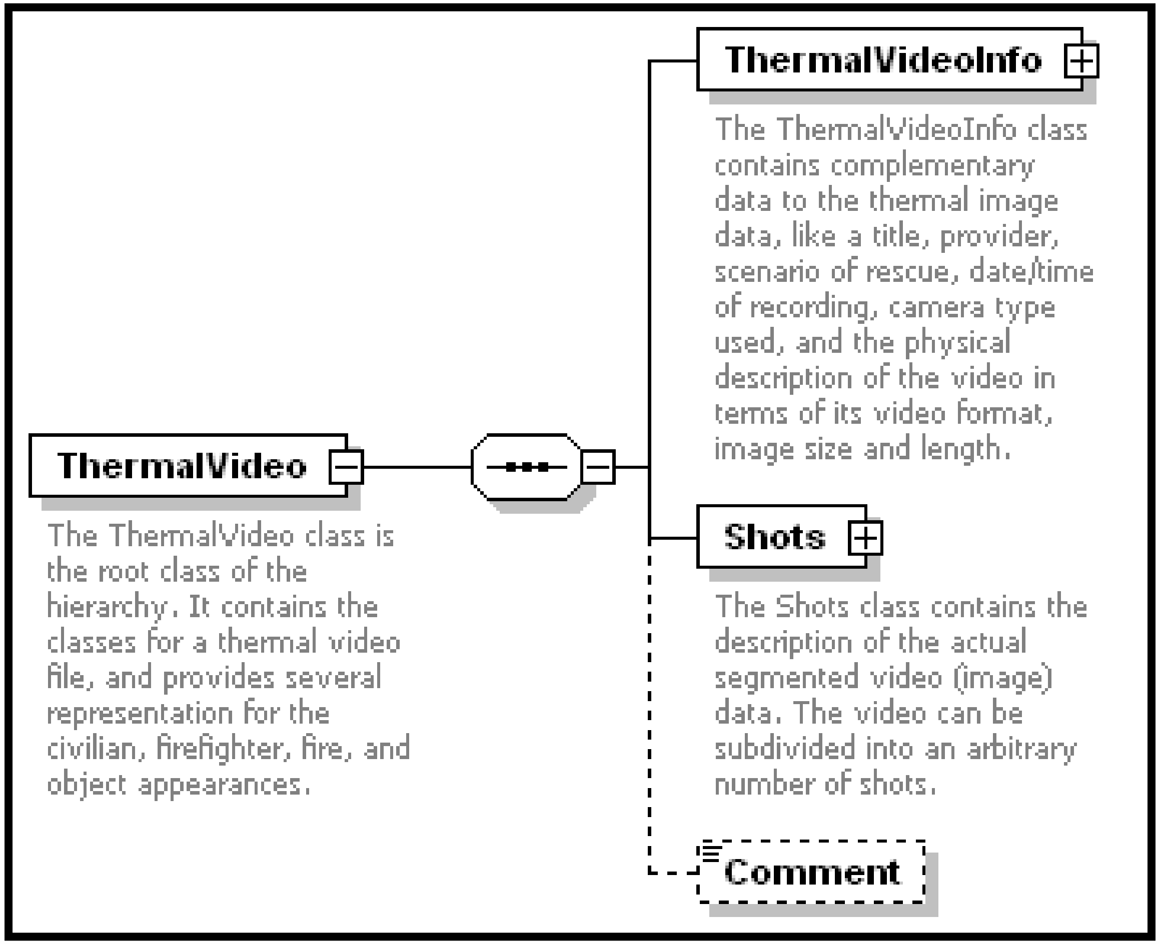
10/07/2008
SSIP 2009



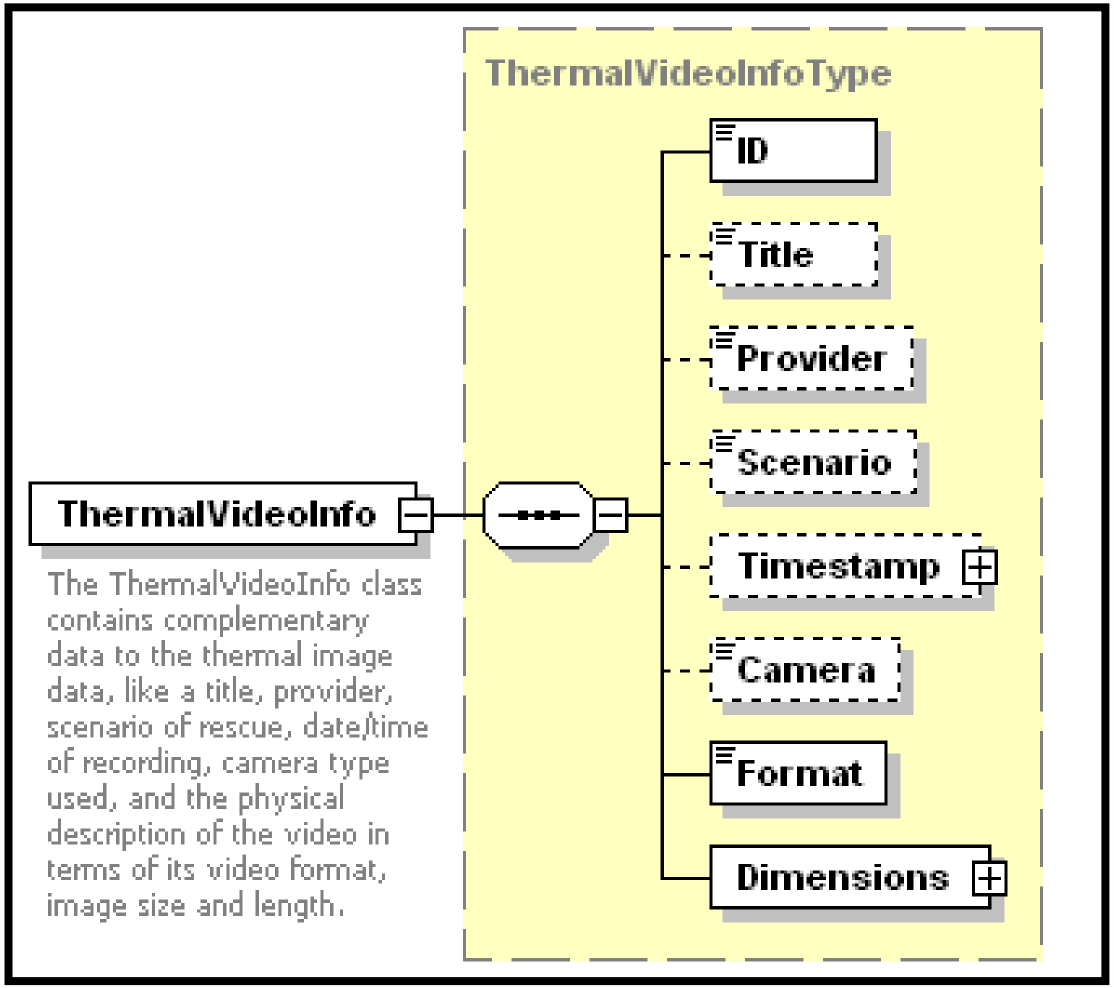
Video processing tasks

- Content description of thermal videos
- Fire detection with Fourier analysis
- Supervised object extraction by snakes
  - Tracing quadtrees for better concavity performance
  - Content adaptive heterogeneous snakes
- Automatic object detection and recognition
  - Creating human pose database
  - Hierarchical clustering for template databases
  - Textual description of database for faster matching
  - Divide and conquer strategy for affine distortions
  - Object simplification for faster matching

Content description of thermal videos

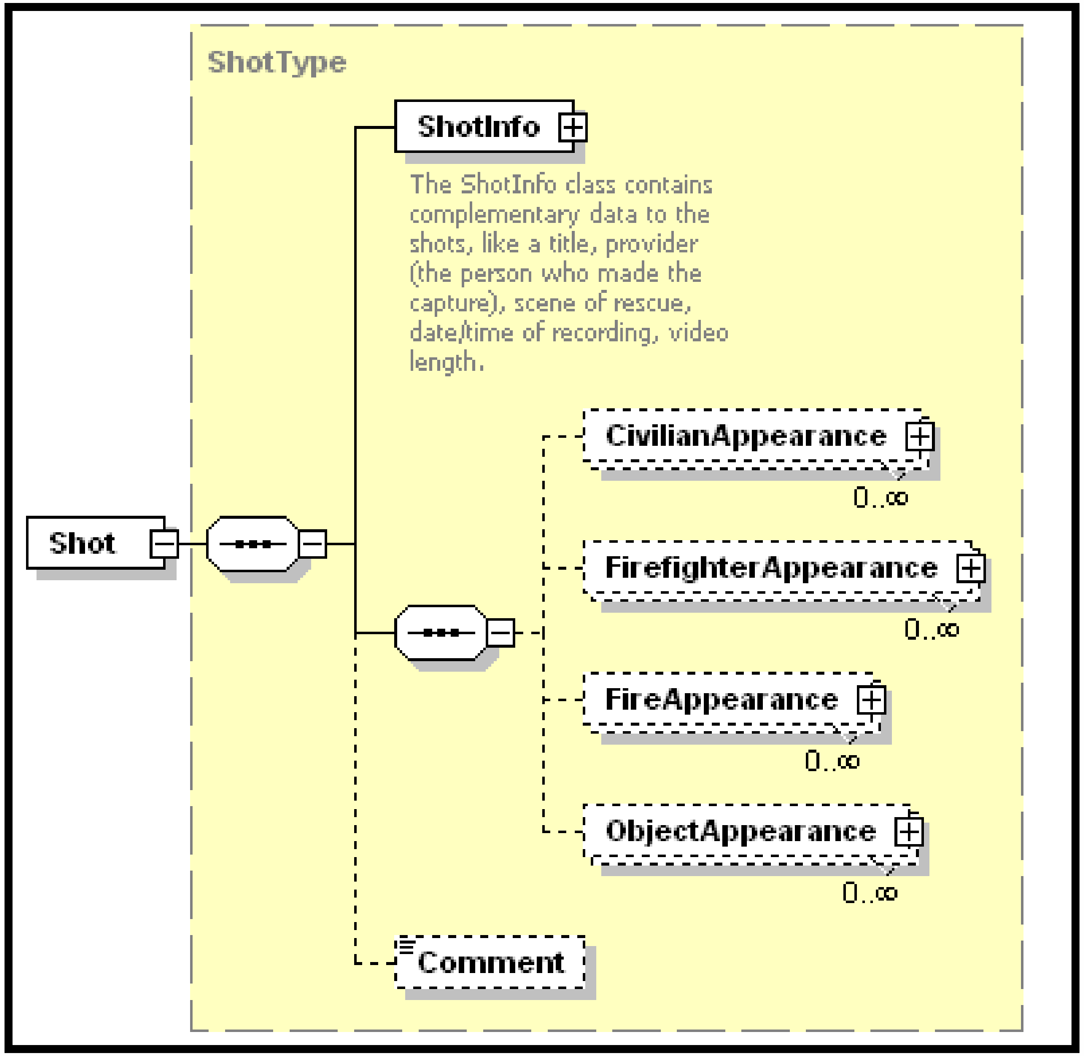


Content description of thermal videos



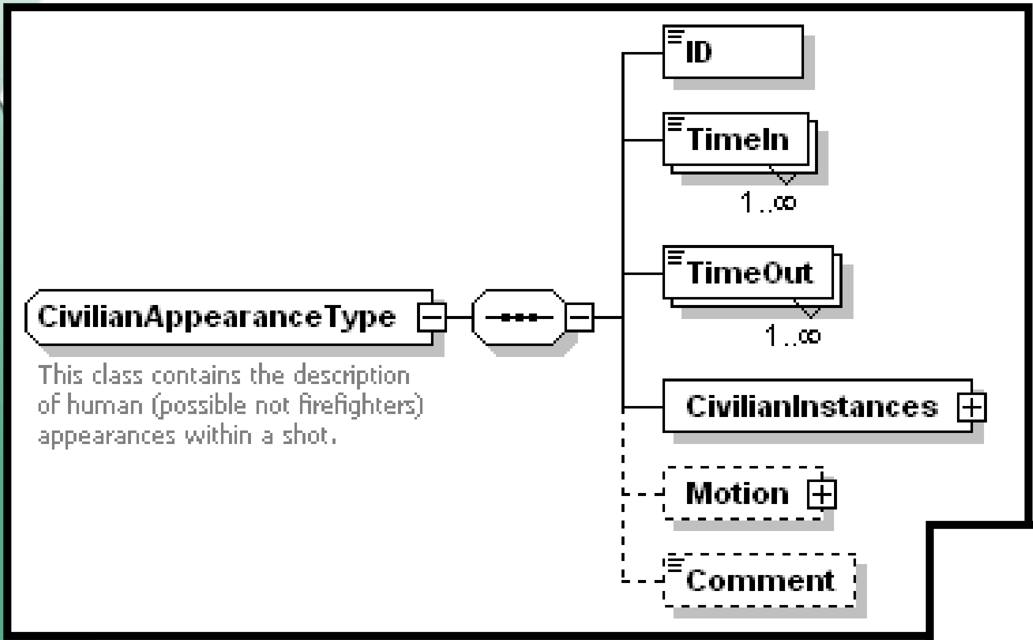
Content description of thermal videos

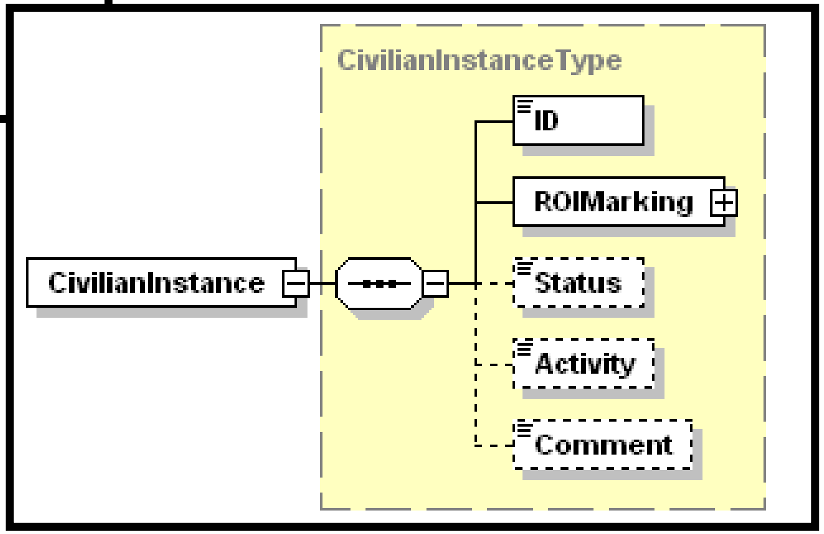
10/07/2008
SSIP 2009



Content description of thermal videos

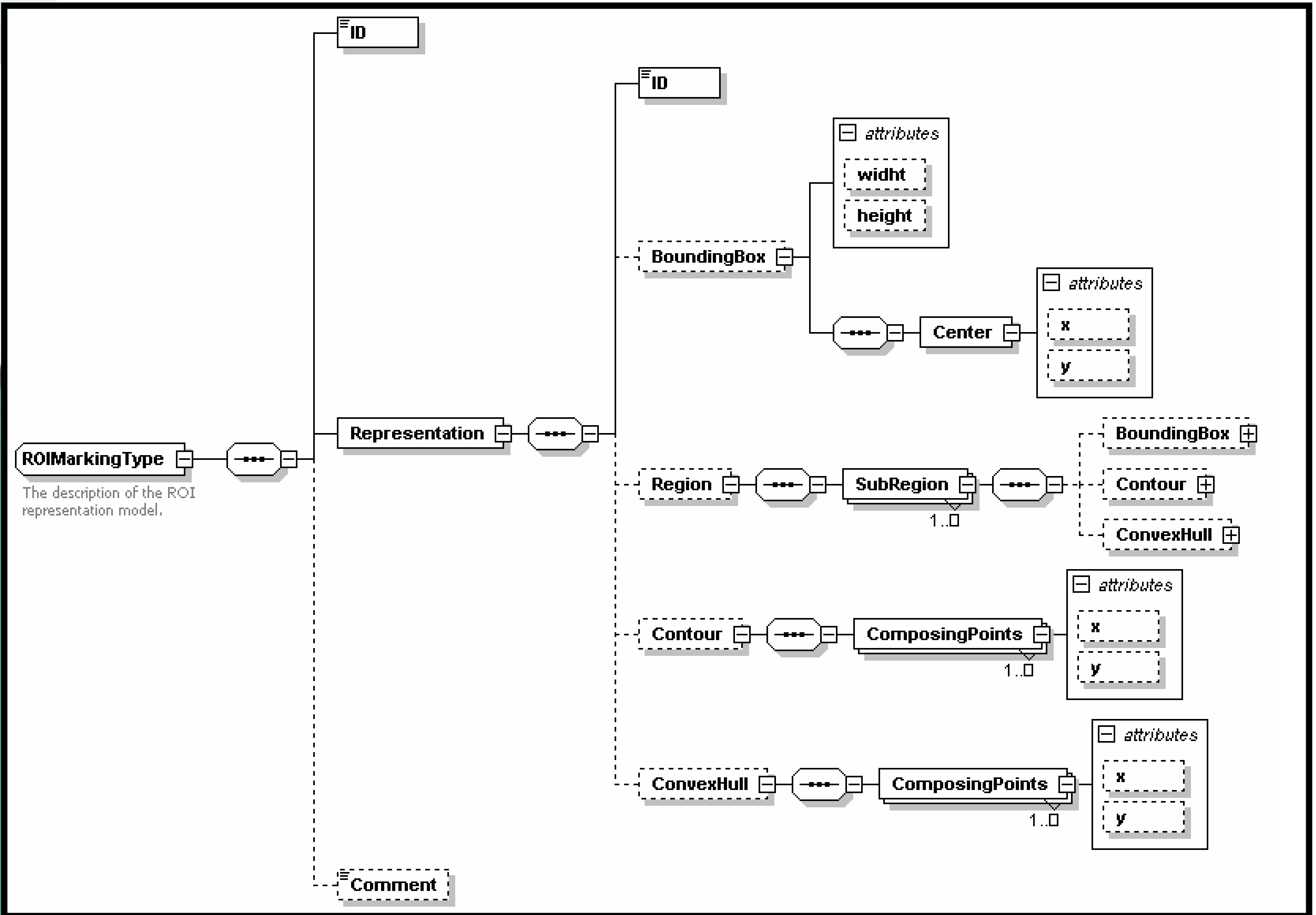
10/07/2008
SSIP 2009





Content description of thermal videos

10/07/2008
SSIP 2009



Querying based on the Schema

Questions on fire:

- Is there any fire in the XXX scenario?
- Is there any fire in the XXX scene?
- Is there any fire in the XXX scene having YYY status?
- Where is the fire in the XXX scene at the YYY time?
- Where is the fire in the XXX scene having YYY status?
- What is the status of the fire in the XXX scene?

Questions on human:

- Are there any humans in the XXX scenario?
- Are there any humans in the XXX scene?
- Are there any humans in the XXX scene having YYY status?
- Where are humans in the XXX scene at the YYY time?
- Are there any humans in the XXX scene having YYY status?
- Where are humans in the XXX scene having YYY status?
- What is the status of humans in the XXX scene?
- Are there any humans in the XXX scene having YYY activity?
- Where are humans in the XXX scene having YYY activity?
- What is the activity of humans in the XXX scene?

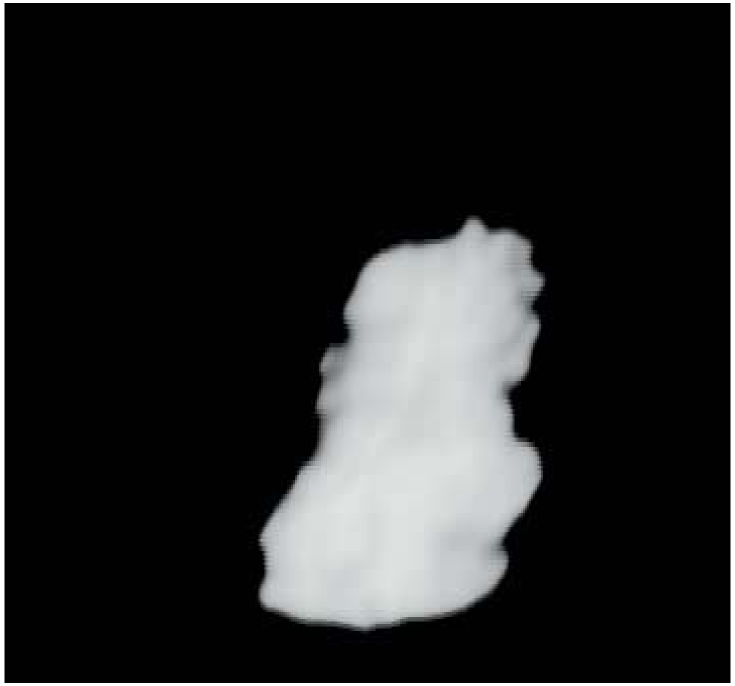
Querying based on the Schema

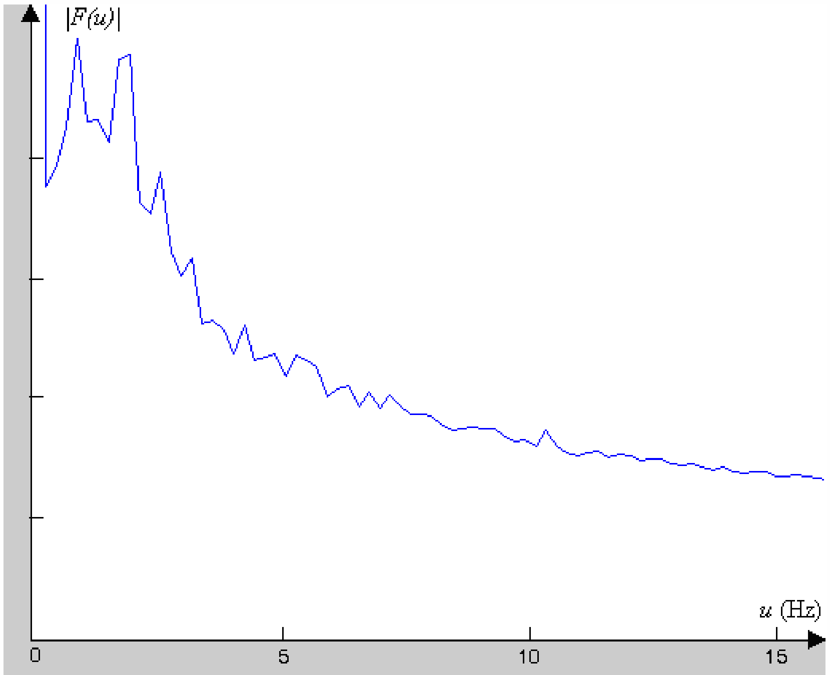
- Instead of using ontology to have the answers we can directly insert evaluation rules (schematrons) into the XML Schema. For example:

```
<xs:appinfo>
 <sch:pattern name="Health problem">
  <sch:rule context="Lesions">
   <sch:report test="((lesion1/detectable) or
      (lesion2/detectable)) and
      (lesion3/detectable)">
     Given disease is present.
   </sch:report>
  </sch:rule>
 </sch:pattern>
</xs:appinfo>
```

Fire detection with Fourier analysis

- Fire has a special behavior in the frequency domain, so temporal Fourier analysis can be executed to check it, as fire usually has 0.5-20Hz flickering. Artificial fire:

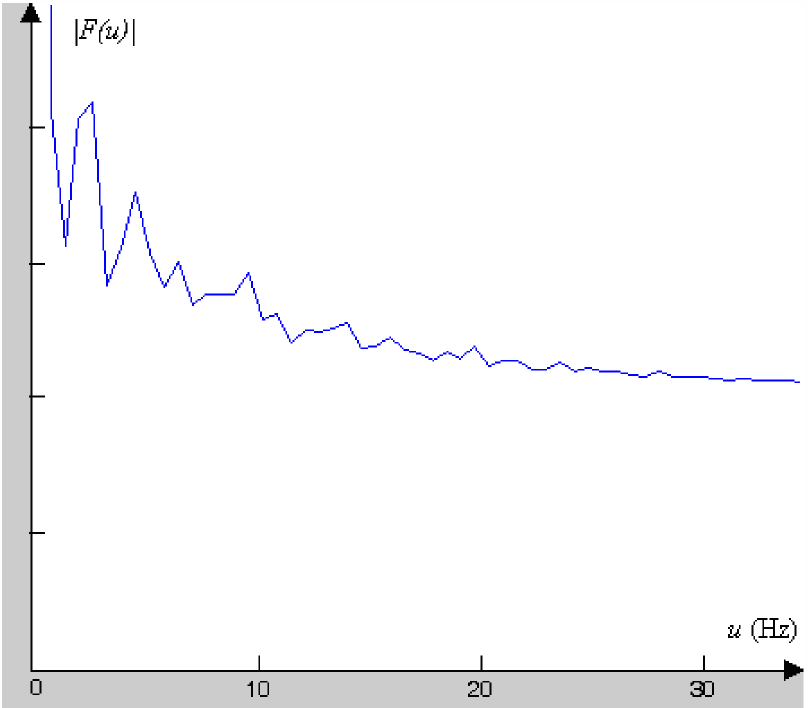




F(*u*) is the average temporal Fourier transform for the changing points.

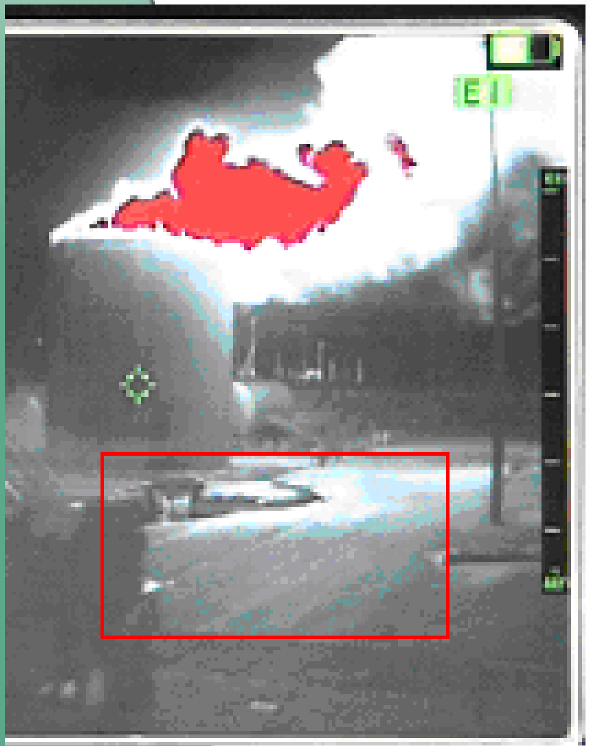
Fire detection with Fourier analysis

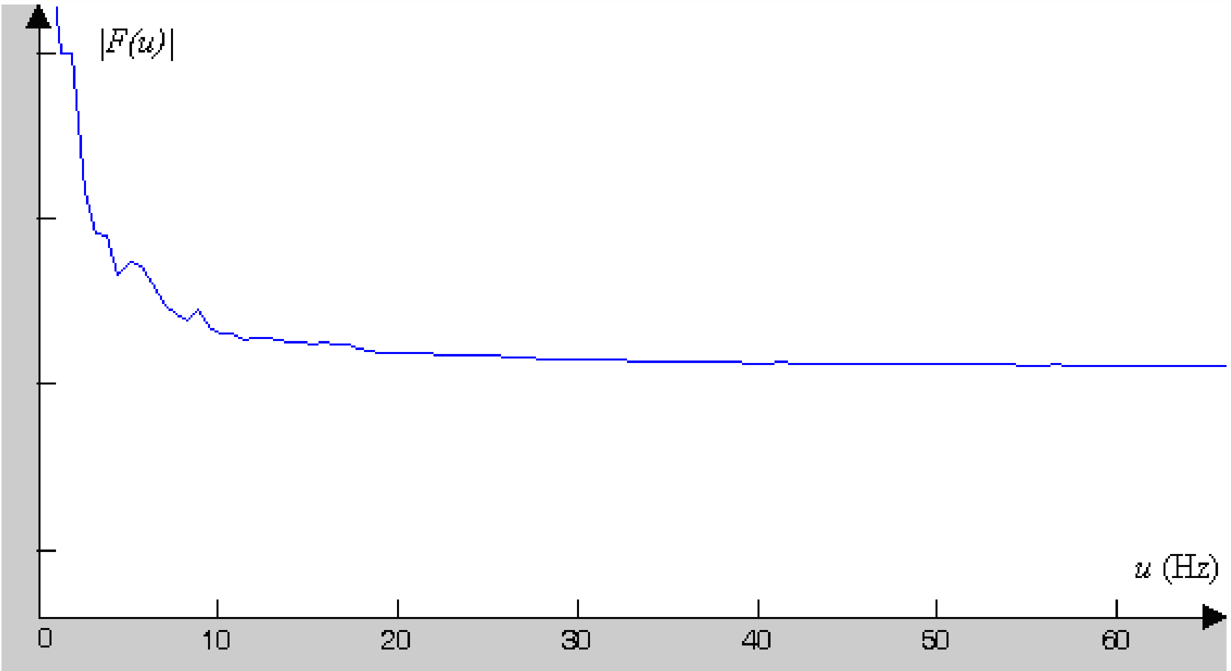
- Natural fire:



Fire detection with Fourier analysis

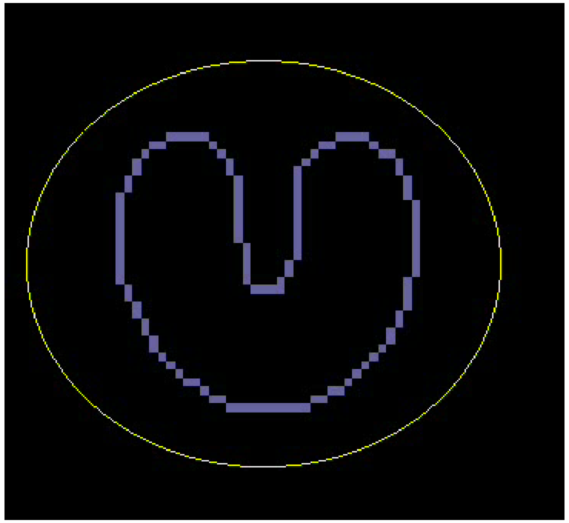
- Non-fire videos does not contain strong peaks in the investigated frequency domain





Supervised object extraction by snakes

- GVF snakes are used to extract objects starting from some user defined initial points
- Insufficient number of iterations can miss concavities
- Improvement of concavity performance and reduction of iteration steps are useful



Snake (active contour) model: $v(s) = (x(s), y(s)) \quad (0 \le s \le 1)$

Energy functional to minimize:

$$E_{snake} = \int_0^1 E_{int}(v(s)) + E_{image}(v(s))ds$$

Internal snake forces:

$$E_{int} = (\alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2)/2$$

External force field (GVF):

$$G(x;y) = (q(x;y); r(x;y))$$

which minimizes:

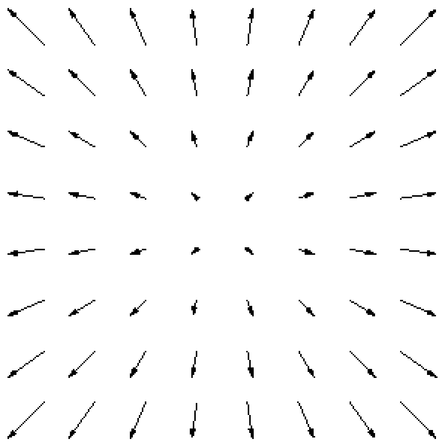
$$\int\int \mu(q_x^2 + q_y^2 + r_x^2 + r_y^2) + |\nabla E|^2 |G - \nabla E|^2 dxdy \qquad q = \frac{\partial E_{image}}{\partial x} \quad r = \frac{\partial E_{image}}{\partial y}$$

Divergence of GVF field

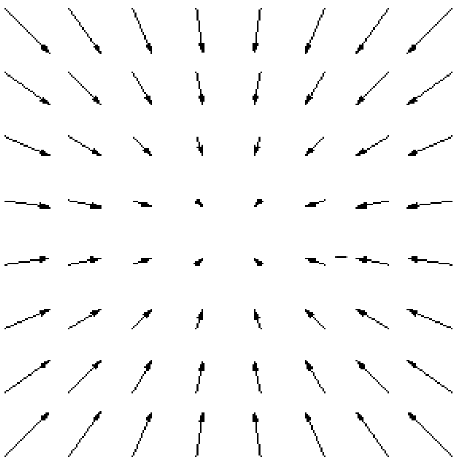
For a vector field: $$F(x,y) = P(x,y)\mathbf{i} + Q(x,y)\mathbf{j}$$

Its divergence is defined as: $$divF = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}$$

High divergence:



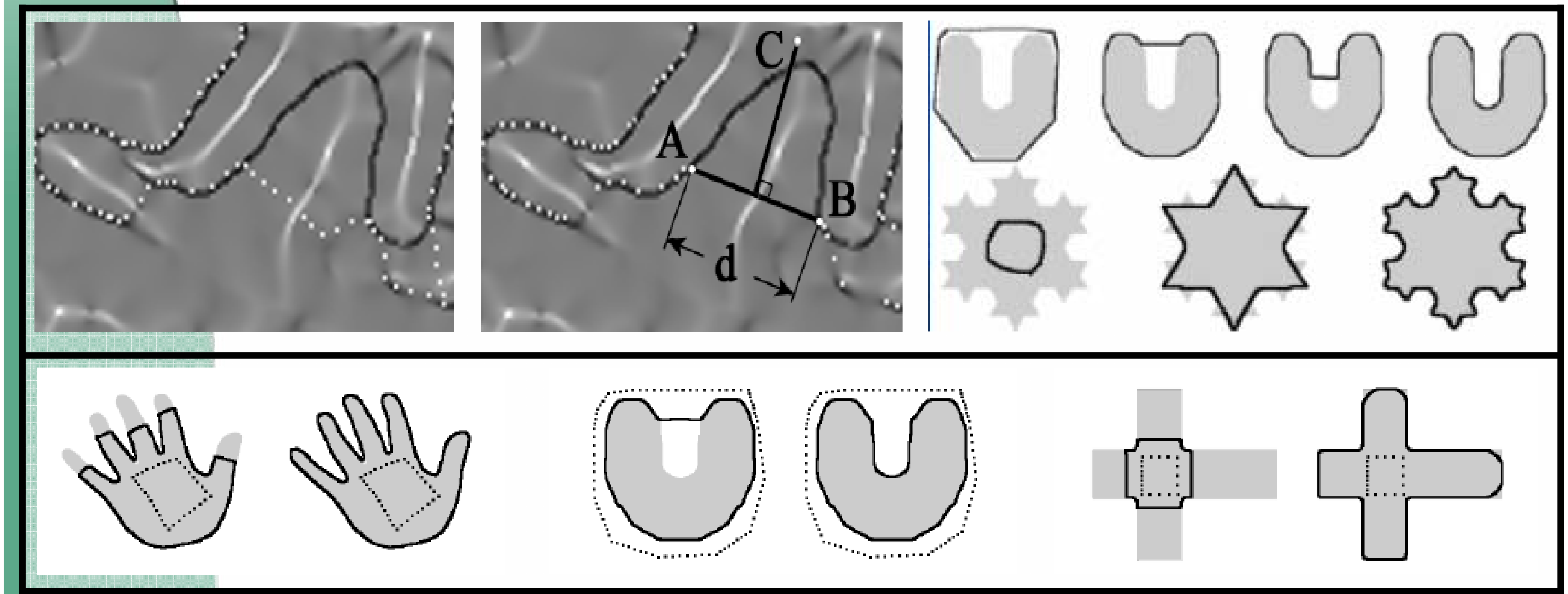
Low divergence:



Thresholding to decide: $divF < \theta$

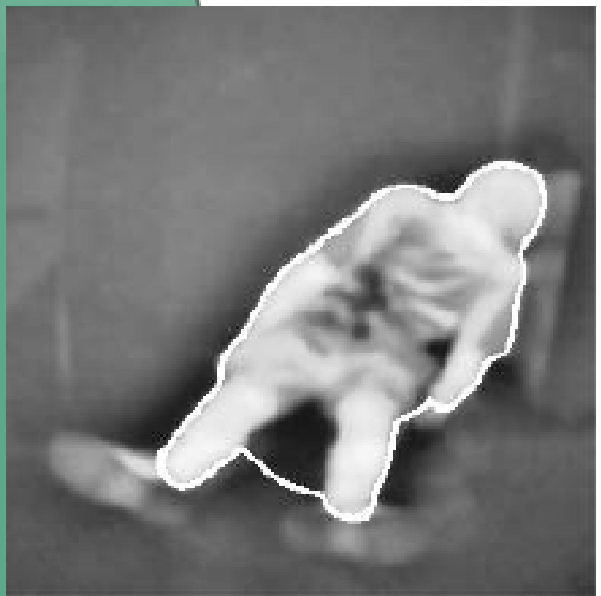
Improving GVF snake
Triangle step

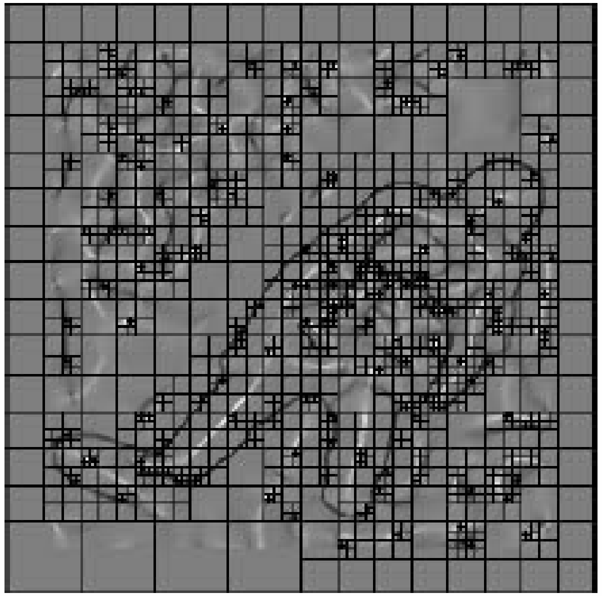
- We detect bad snake segments based on divergence, estimate snake directions and select a new snake point inside the concavity

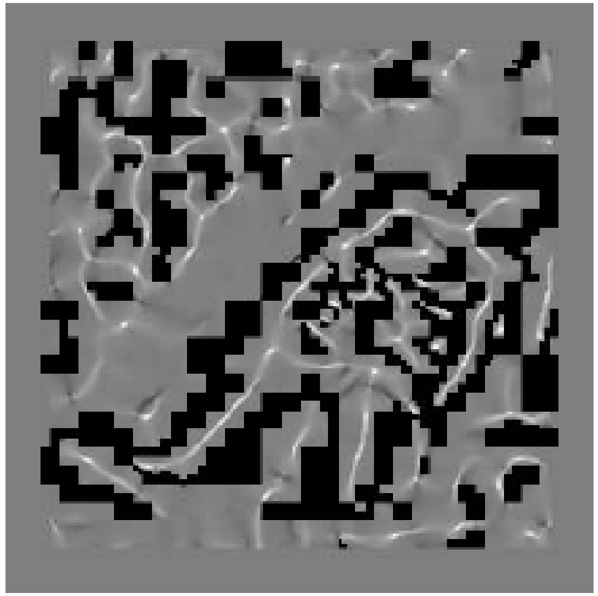


Supervised object extraction by snakes

- Tracing quadtrees for better concavity performance
  - We cover the expected object boundary with a grid with larger scale (quadtree decomposition)

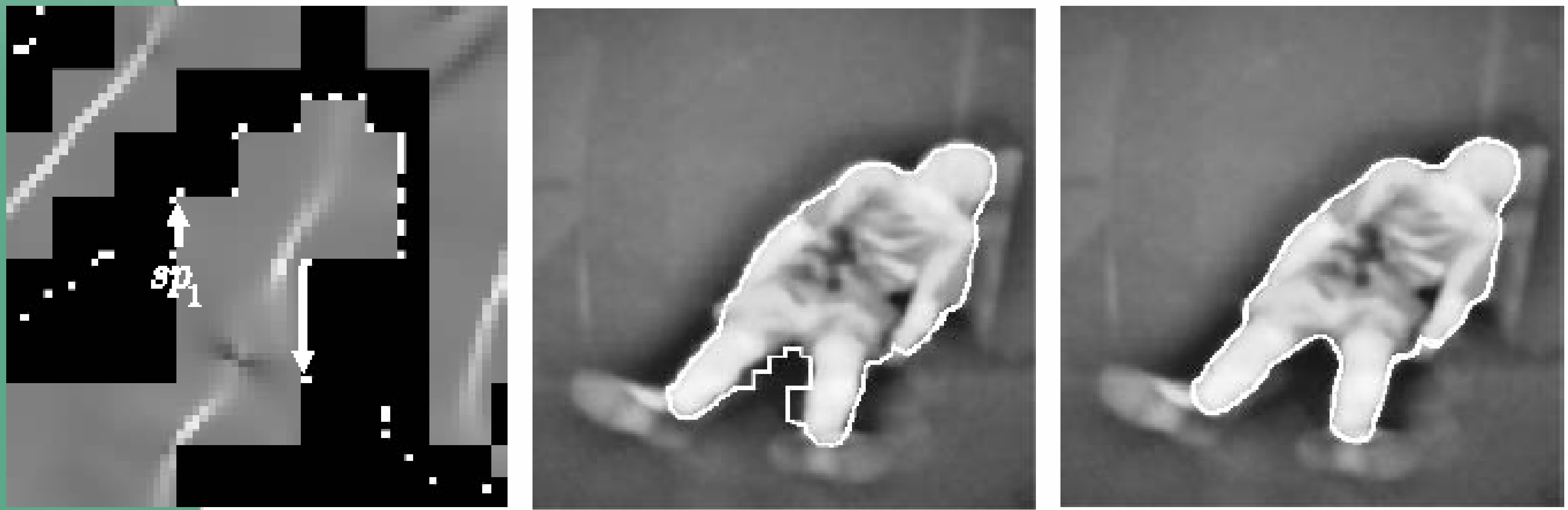






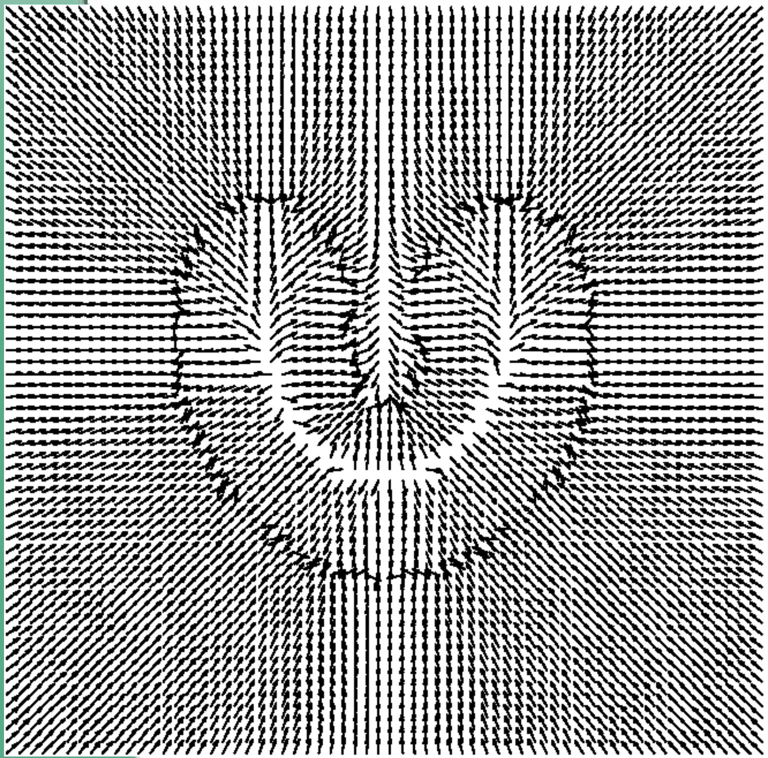
Supervised object extraction by snakes

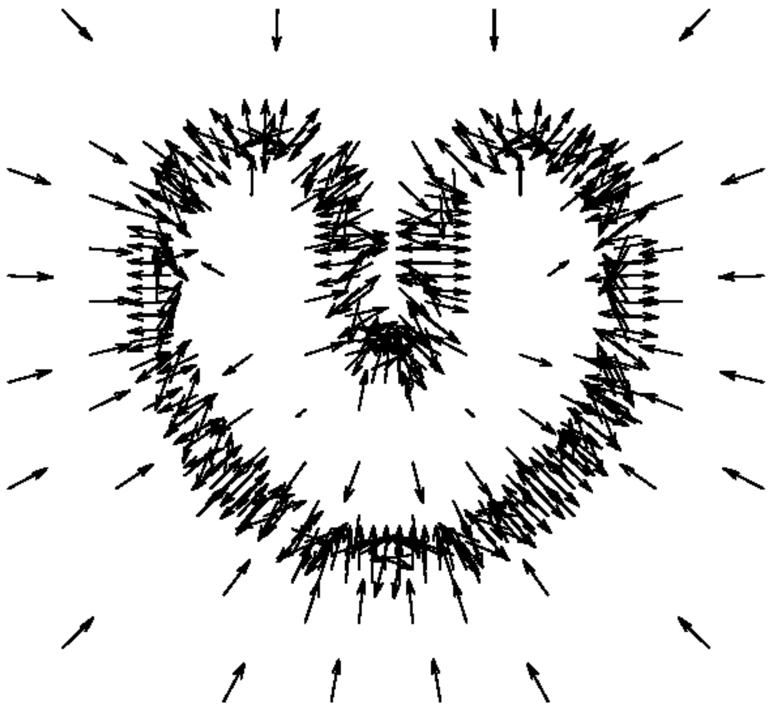
- Tracing quadtrees for better concavity performance
  - A boundary tracer is generalized for the „thick arc" representation, followed by pixelwise iterations



Supervised object extraction by snakes

- Content adaptive heterogeneous snakes
  - We can make faster the snake iteration within homogeneous subregions (regions can be defined e.g. by quadtrees)





Supervised object extraction by snakes

- Content adaptive heterogeneous snakes
  - The snake captures the rough boundary in a few iteration steps on the larger grid (e.g. four steps is needed for U-shape)

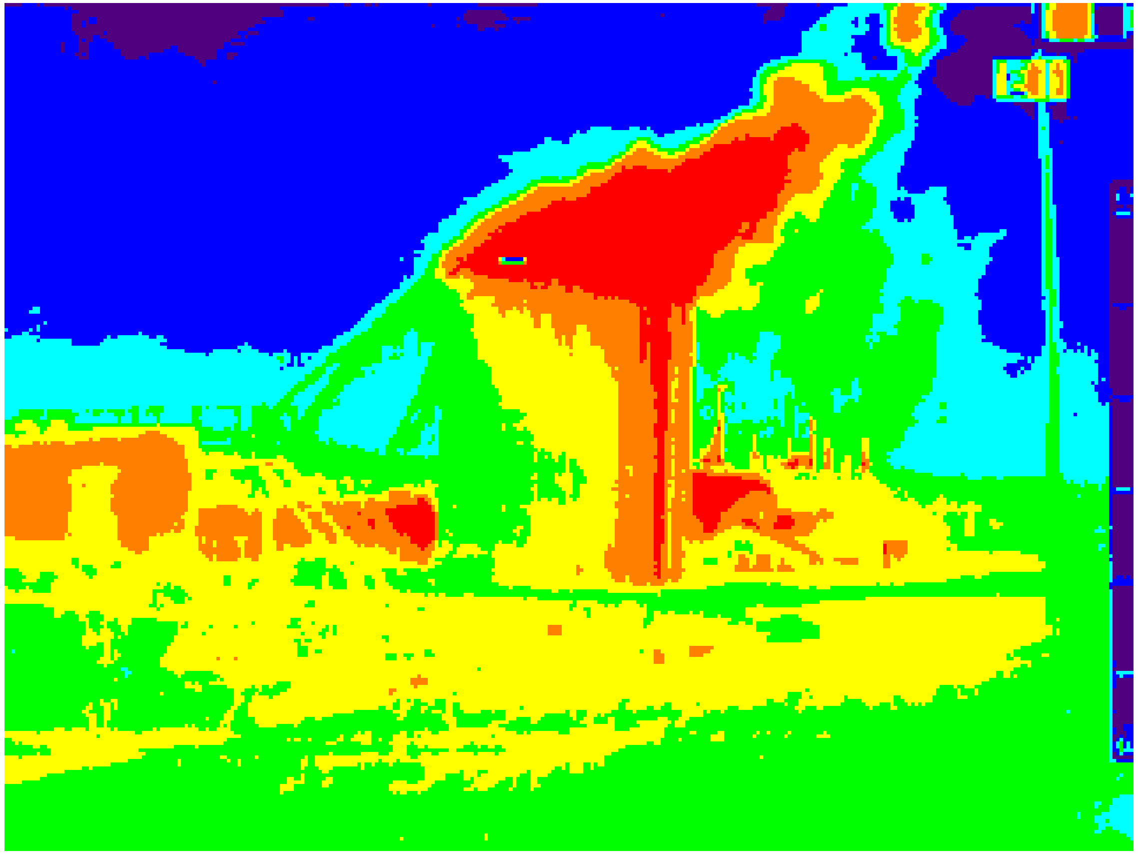
Software module – 2nd prototype

- Tasks in the 2nd prototype
  - Colorization
  - Hotspot detection
  - Fire detection
  - Boundary detection
  - Boundary tracking
  - Boundary object recognition
  - Automatic object detection

Software module – 2nd prototype

10/07/2008
SSIP 2009

- Colorization tool



Software module – 2nd prototype

10/07/2008
SSIP 2009

- Hotspot detection tool

Software module – 2nd prototype

- Fire detection tool


```
- <FireAppearance>
    <ID>1</ID>
    <TimeIn>10:00:23</TimeIn>
    <TimeOut>10:00:30</TimeOut>
  - <FireInstances>
    - <FireInstance>
        <ID>1</ID>
      - <ROIMarking>
          <ID>1</ID>
        - <Representation>
            <ID>1</ID>
          - <BoundingBox>
              <Width>70</Width>
              <Height>19</Height>
            - <Center>
                <X>67</X>
                <Y>170</Y>
              </Center>
```

- Boundary detection tool

Software module – 2nd prototype

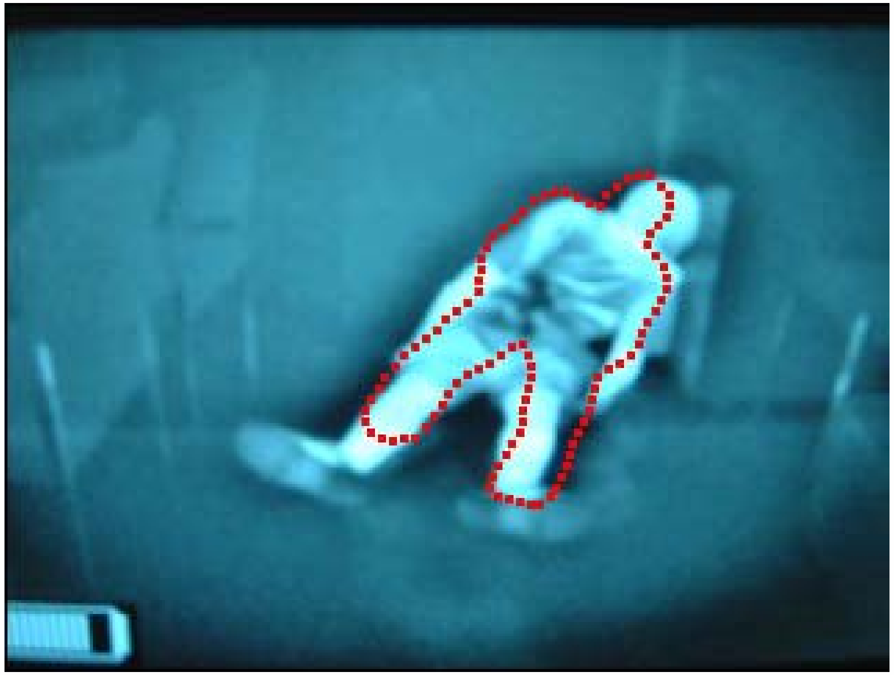
- Boundary object recognition tool

– <CivilianAppearance>
<ID>1</ID>
<TimeIn>10:00:23</TimeIn>
<TimeOut>10:00:23</TimeOut>
– <CivilianInstances>
– <CivilianInstance>
<ID>1</ID>
– <ROIMarking>
<ID>1</ID>
– <Representation>
<ID>1</ID>
– <Contour>
<ComposingPoints>(147,25)
</ComposingPoints>

Software module – 2nd prototype

- Boundary tracking tool

- Automatic object detection



Object detection and recognition

- Creating human pose database
  - For detection, based on whole object (human) silhouettes a database is created with different activities and views
  - Activities:
    - Running
    - Walking
    - Falling/lying
    - Sitting (testing reasons)
    - Punching (testing reasons)
  - Views:
    - Front
    - Side
    - Top

Object detection and recognition

10/07/2008
SSIP 2009

- Creating human pose database
  - Realistic 3D modeling software was used

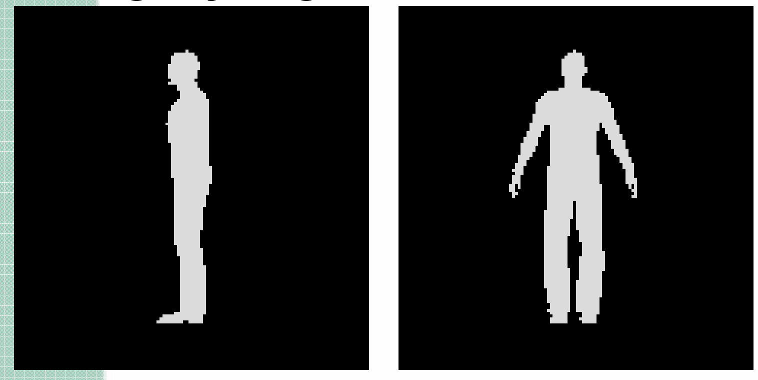
Object detection and recognition

- Creating human pose database

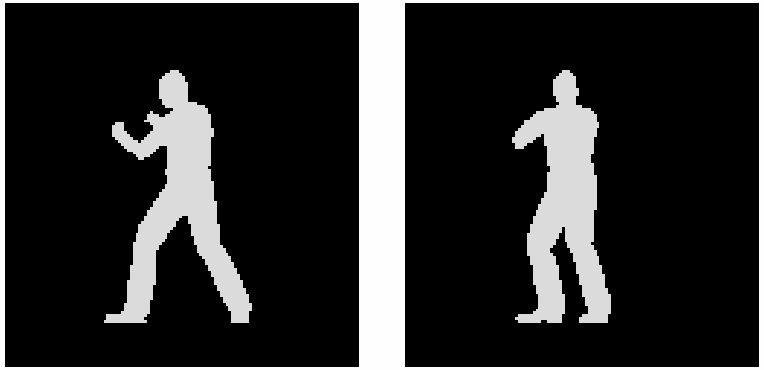
Walking

Running

Falling/lying

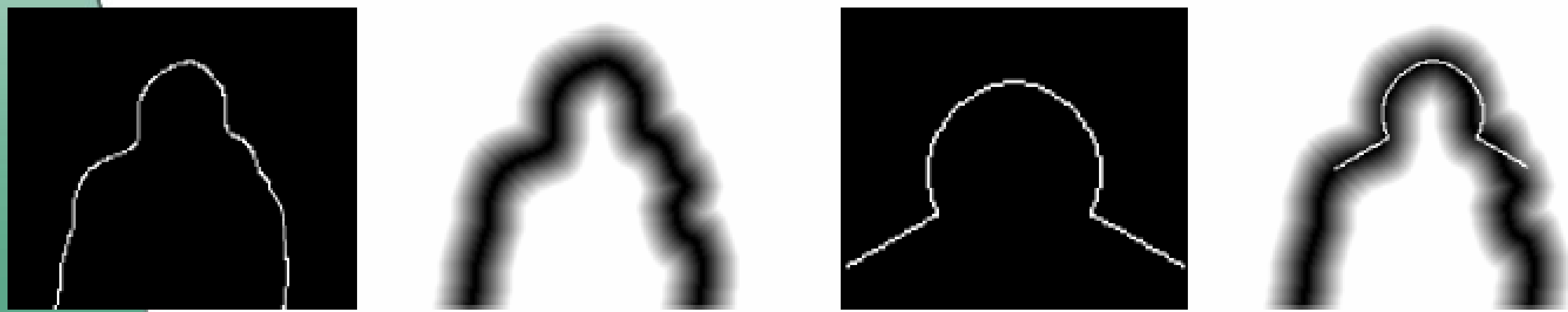


Punching (demo)



Object detection and recognition

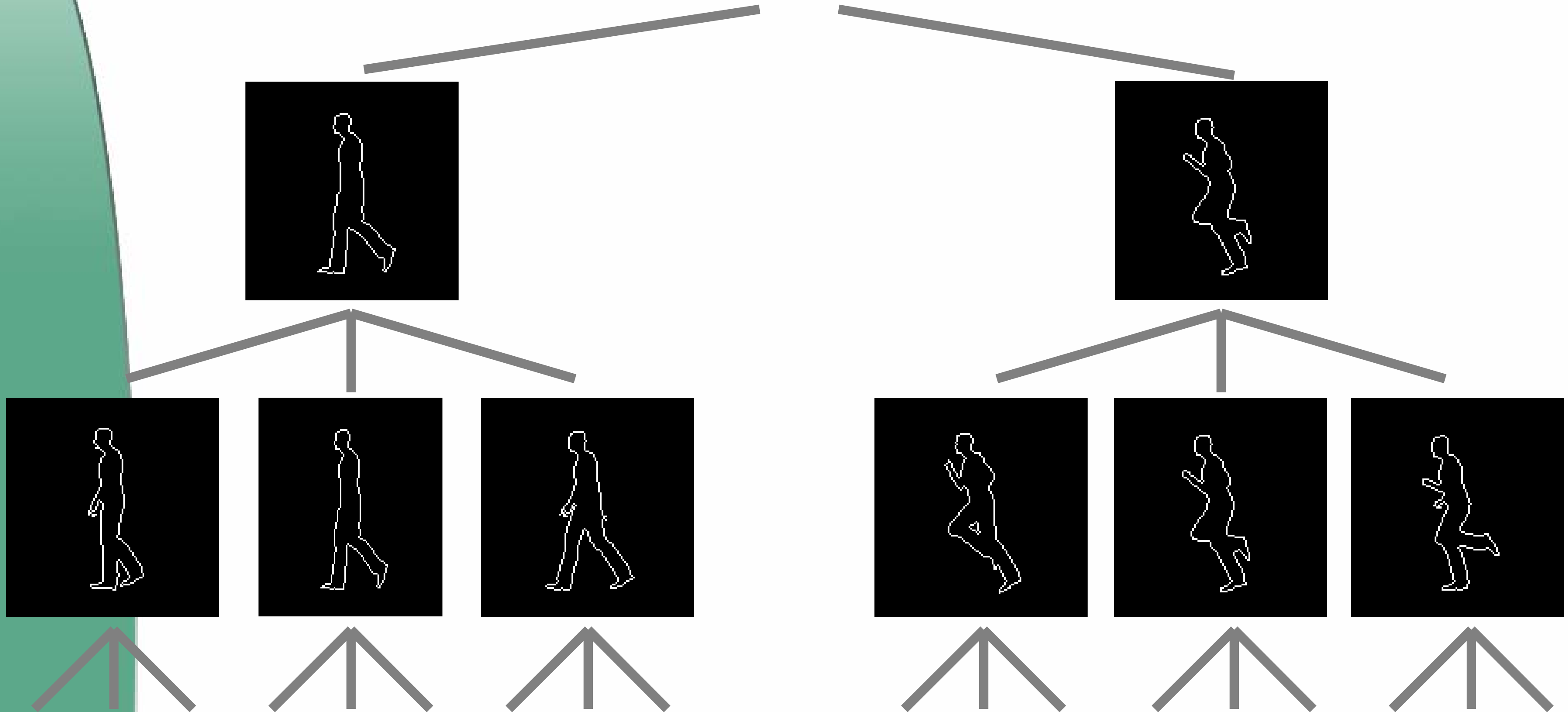
- Creating human pose database
  - Since chamfer matching is used to locate database templates…



  - …boundary tracing is applied to extract database elements

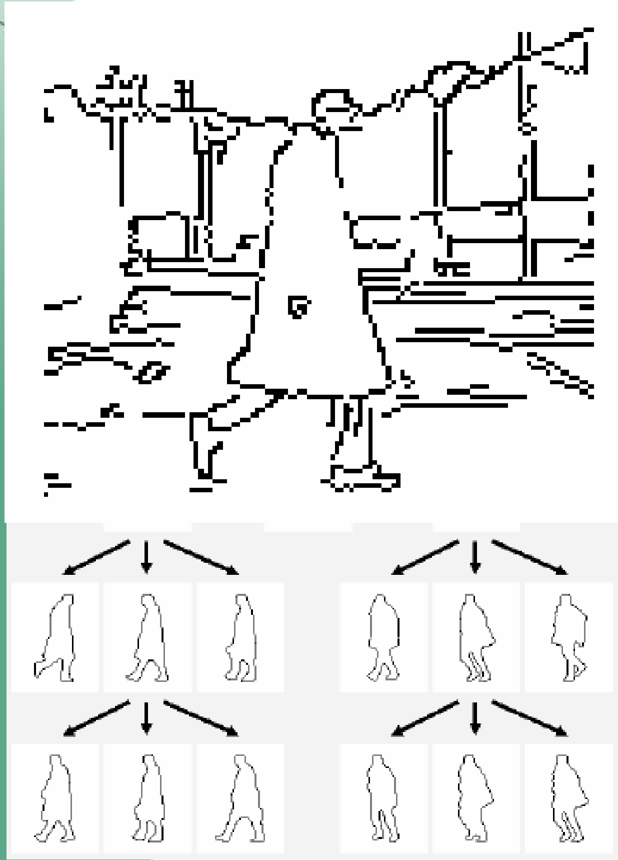
Object detection and recognition

- Hierarchical clustering for template databases
  - Templates are organized into hierarchy for faster matching



Object detection and recognition

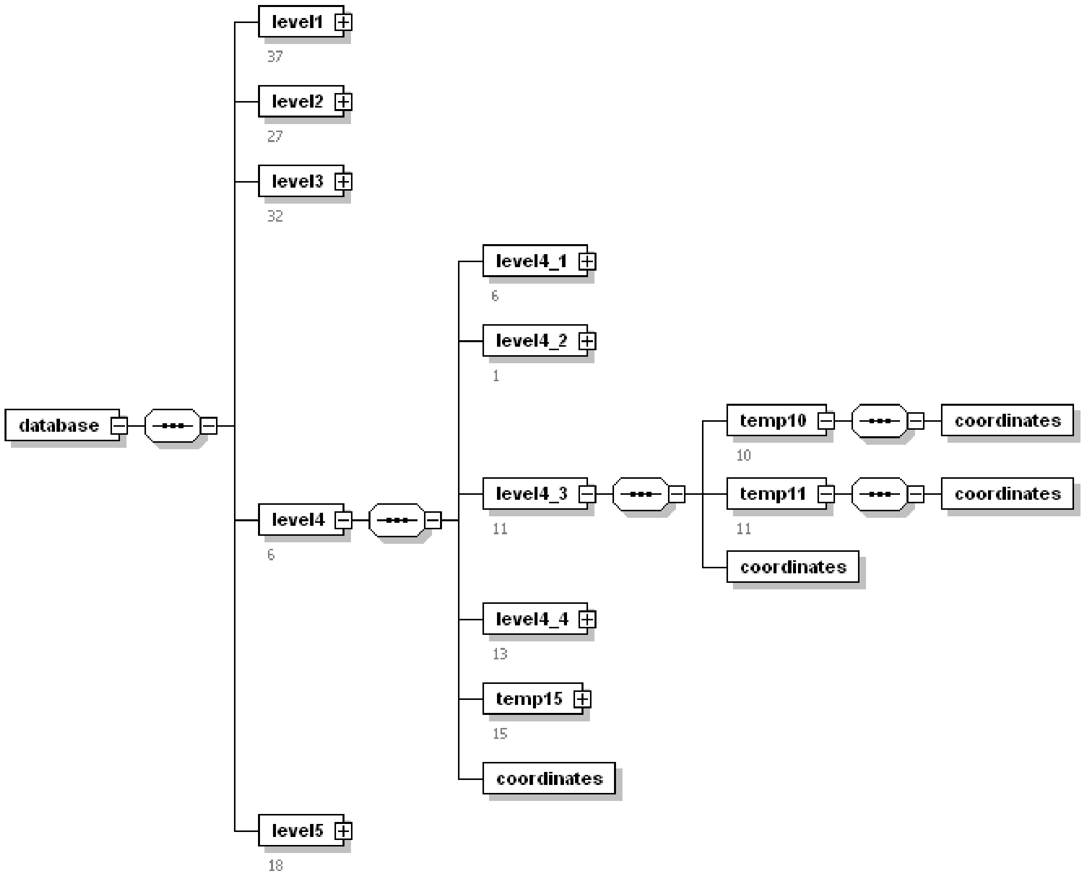
10/07/2008
SSIP 2009



Object detection and recognition

- Textual description of database for faster matching
  - Database is stored as an XML file containing silhouette vectors

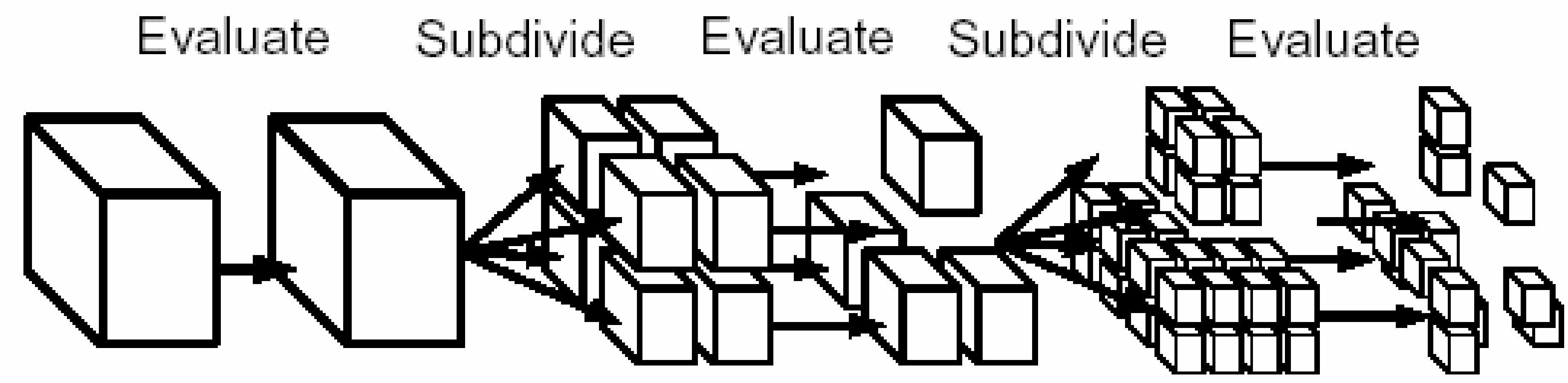
```
- <database>
  - <level1>
    + <level1_1>
    + <level1_2>
    + <temp39>
    + <temp36>
      <coordinates>(-54,-3)
      </coordinates>
    </level1>
  + <level2>
  + <level3>
  + <level4>
  + <level5>
  </database>
```



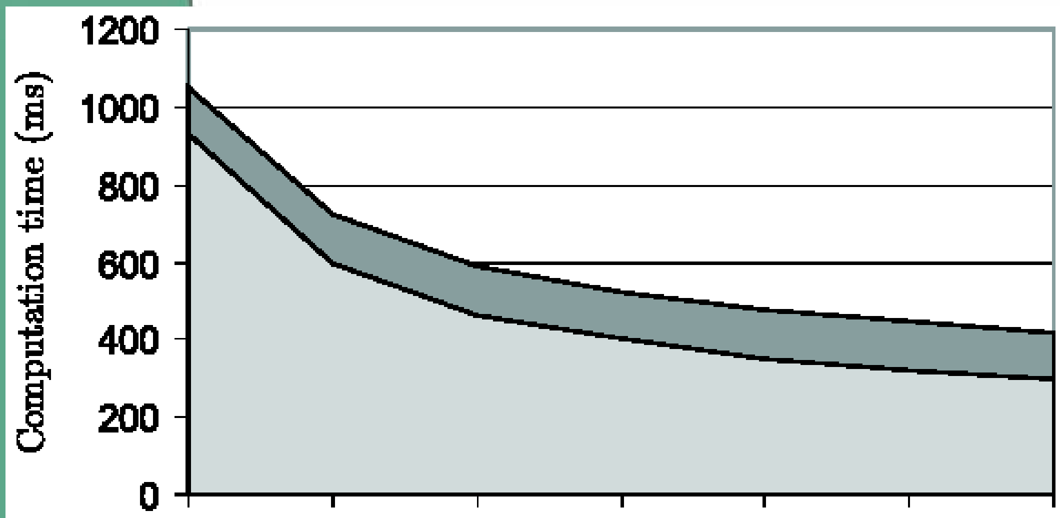
Object detection and recognition

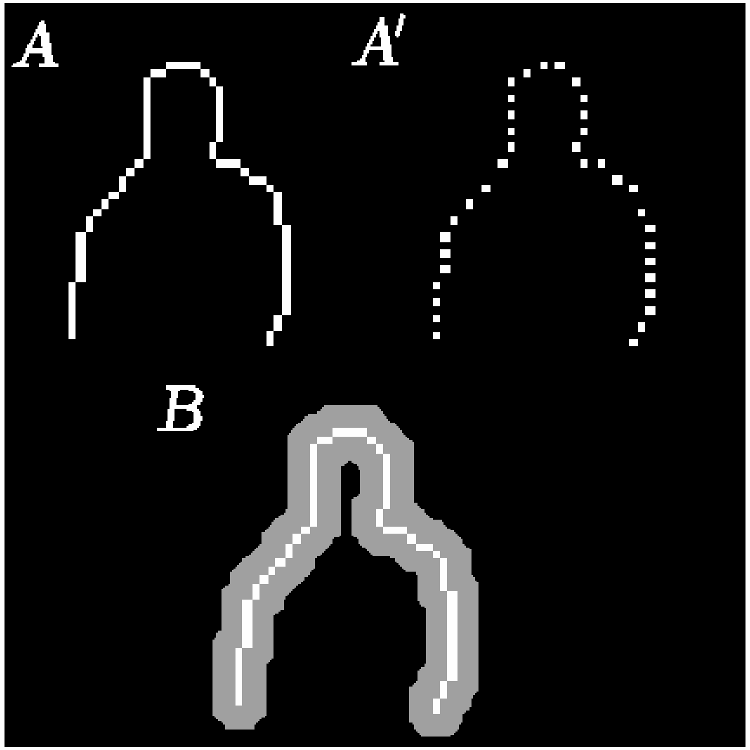
10/07/2008
SSIP 2009

- Divide and conquer strategy for affine distortions
  - The templates may occur in a geometrically distorted way
  - Usually, affine transformations are sufficient to be considered
  - A „divide and conquer” strategy is used to restrict the search of the affine (6D) parameter space



- Object simplification for faster matching
  - For further speed up a point reduction method was developed which preserves the most optimal subset for chamfer matching
  - It can be applied to any sets in arbitrary dimension. Contours:





Centroidal Voronoi Tessellation framework

Given

- $\Omega \in \mathbb{R}^d$
- a density function ρ
- k points $\{\mathbf{x}_i\}_{i=1}^k$ in Ω

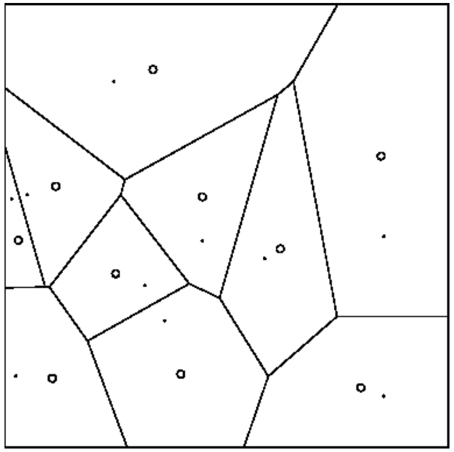
the set $\{V_i\}_{i=1}^k$ is a Voronoi tesselation (**VT**) of Ω corresponding to the points $\{\mathbf{x}_i\}_{i=1}^k$ if, for each i,

$$V_i = \{\mathbf{x} \in \Omega \ : \ |\mathbf{x}-\mathbf{x}_i| \leq |\mathbf{x}-\mathbf{x}_j| \text{ for } i \neq j\}$$

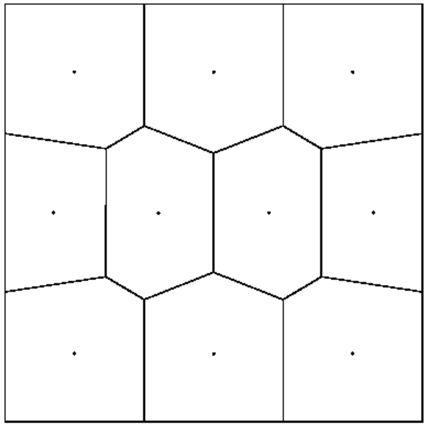
the Center of Mass of each Voronoi set V_i, $i = 1, \ldots, k$, is given by

$$\mathbf{z}_i = \frac{\int_{V_i} \mathbf{x}\rho(\mathbf{x})\, d\mathbf{x}}{\int_{V_i} \rho(\mathbf{x})\, d\mathbf{x}}$$

*** VT = CVT if $\mathbf{z}_i = \mathbf{x}_i$ for all i ***



generic VT



CVT

Centroidal Voronoi Tessellation framework

Given

- any set $\{\mathbf{y}_i\}_{i=1}^k$ of k points in Ω
- any tessellation $\{W_i\}_{i=1}^k$ of Ω

let

$$F(\{\mathbf{y}_i, W_i\}) \equiv \sum_{i=1}^{k} \int_{W_i} \rho(\mathbf{x}) |\mathbf{x} - \mathbf{y}_i|^2 \, d\mathbf{x}$$

---

electrical engineer: distortion value

statistician: variance

planner, manager: cost

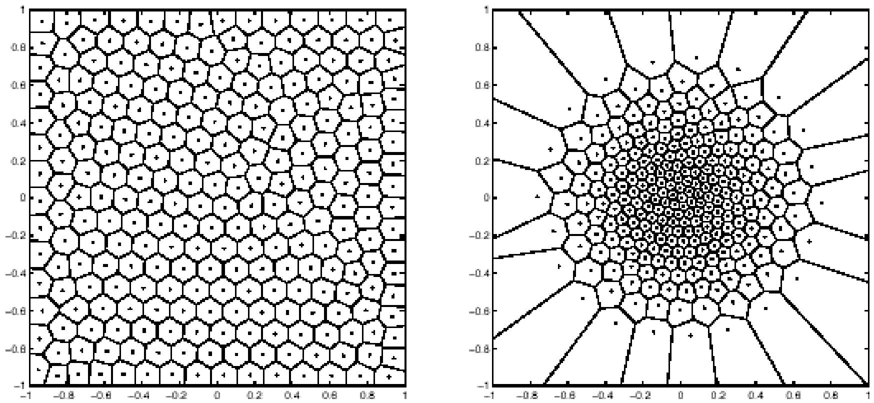
---

if $F(\{\mathbf{z}_i, V_i\}) = \min_{\{\mathbf{y}_i, W_i\}} F(\{\mathbf{y}_i, W_i\})$

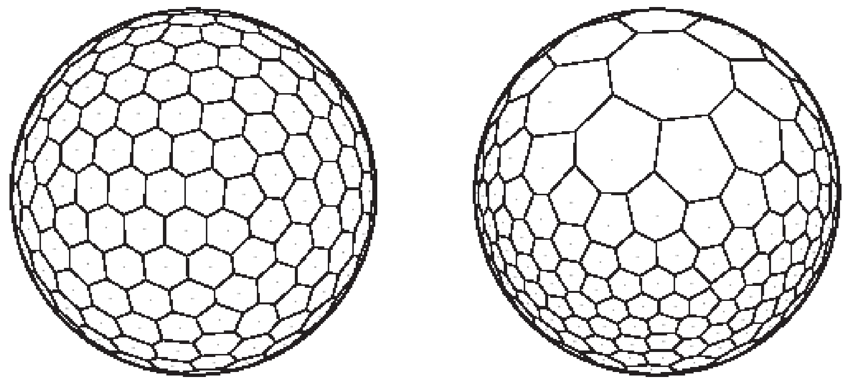
then

$\{V_i\}_{i=1}^k$ is a CVT with generators $\{\mathbf{z}_i\}_{i=1}^k$

tessellations of a square



tessellations on a sphere



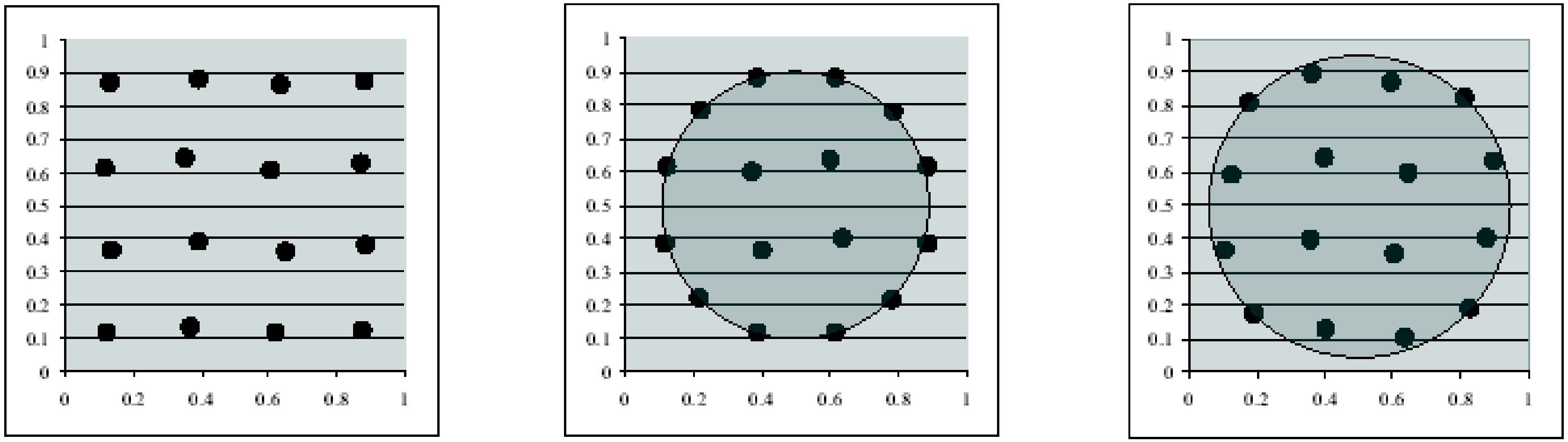
Centroidal Voronoi Tessellation framework

We should remove points from the original template in such a way that is optimal with respect to the matching function.

Optimization problem

Problem: Reduce the set $A \subseteq B$ with $|A| = N$ to $|A'| = K$, such that the distance map of A' remains as close as possible to the distance map of A within B.

Centroidal Voronoi Tessellation framework



Iteration:

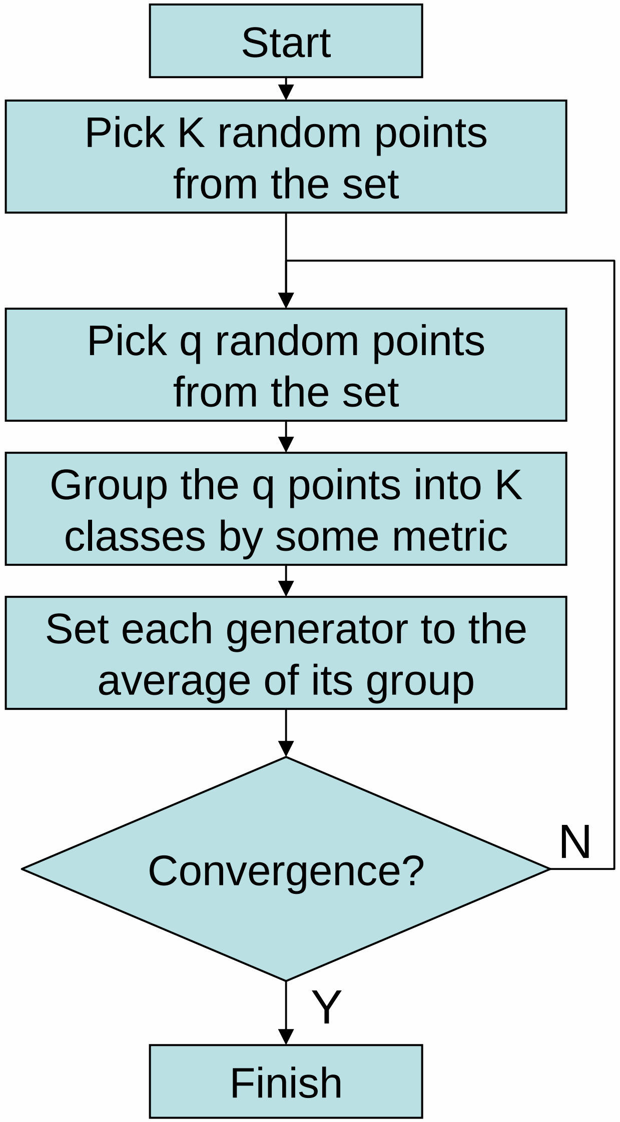
1. Construct the Voronoi tessellation $\{V_i\}_{i=1}^{K}$ of B with generators $\{\underline{\mathbf{z}}_i \in A \mid i = 1, \ldots, K\}$;
2. Define the new set of generators as the points of A closest to the centroids of $\{V_i\}_{i=1}^{K}$;
3. Repeat steps 1 and 2 until some stopping criterion is met.

Other algorithms (e.g. random sampling) are also possible.

Subsampling approach
CVT implemantation (McQueen 1967)

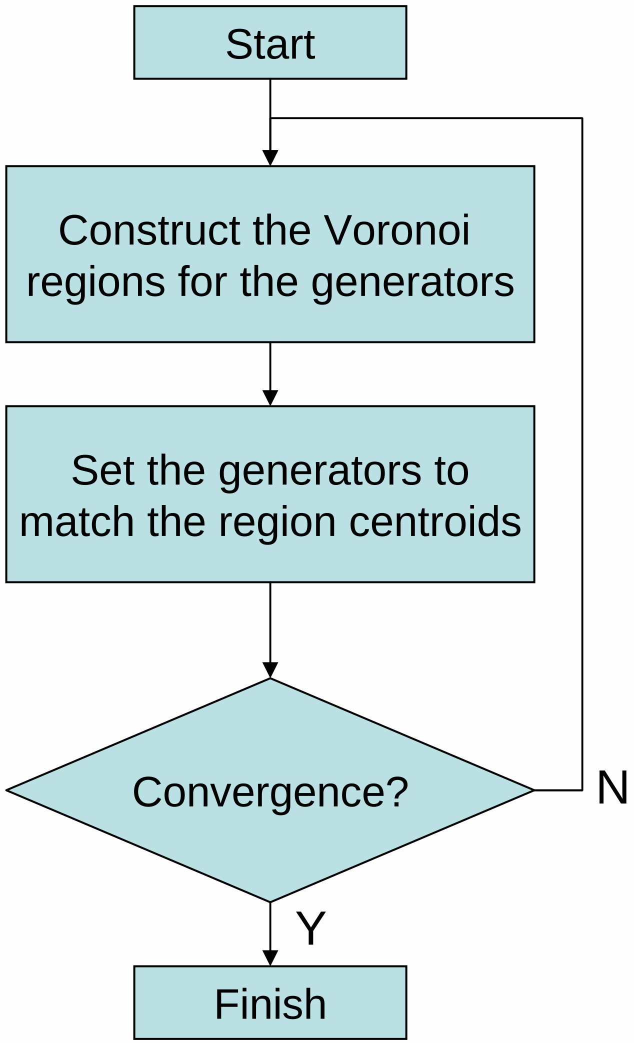
10/07/2008
SSIP 2009

- Input
  - Original set
  - Number of generators
  - Density function
- Output
  - CVT generators
- Properties
  - Many samples might be necessary



Subsampling approach
CVT implemantation (Lloyd 1982)

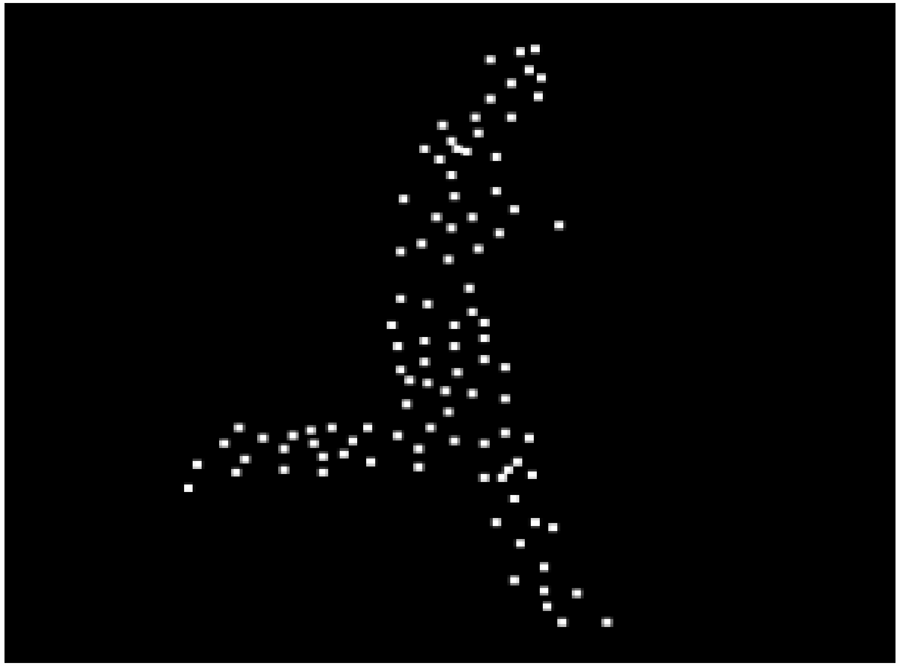
- **Input**
  - Original set
  - Initial generators
  - Density function
- **Output**
  - CVT generators
- **Properties**
  - One has to compute the Voronoi regions



Subsampling approach
Uniform subsampling

- We considered constant density function (uniform subsampling), and implemented the Llyod algorithm.
- A 2D example (starting from random points):

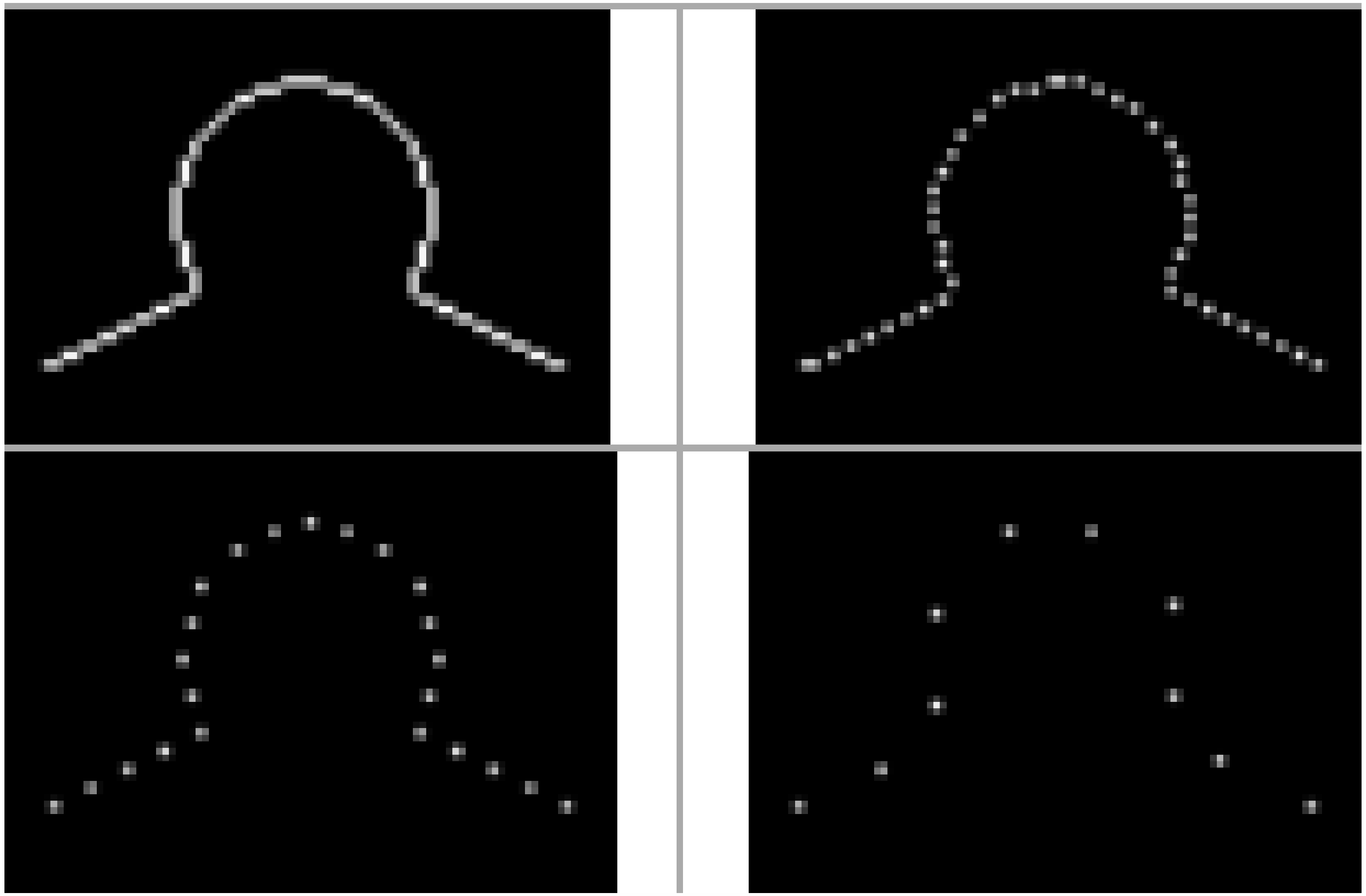
Original image



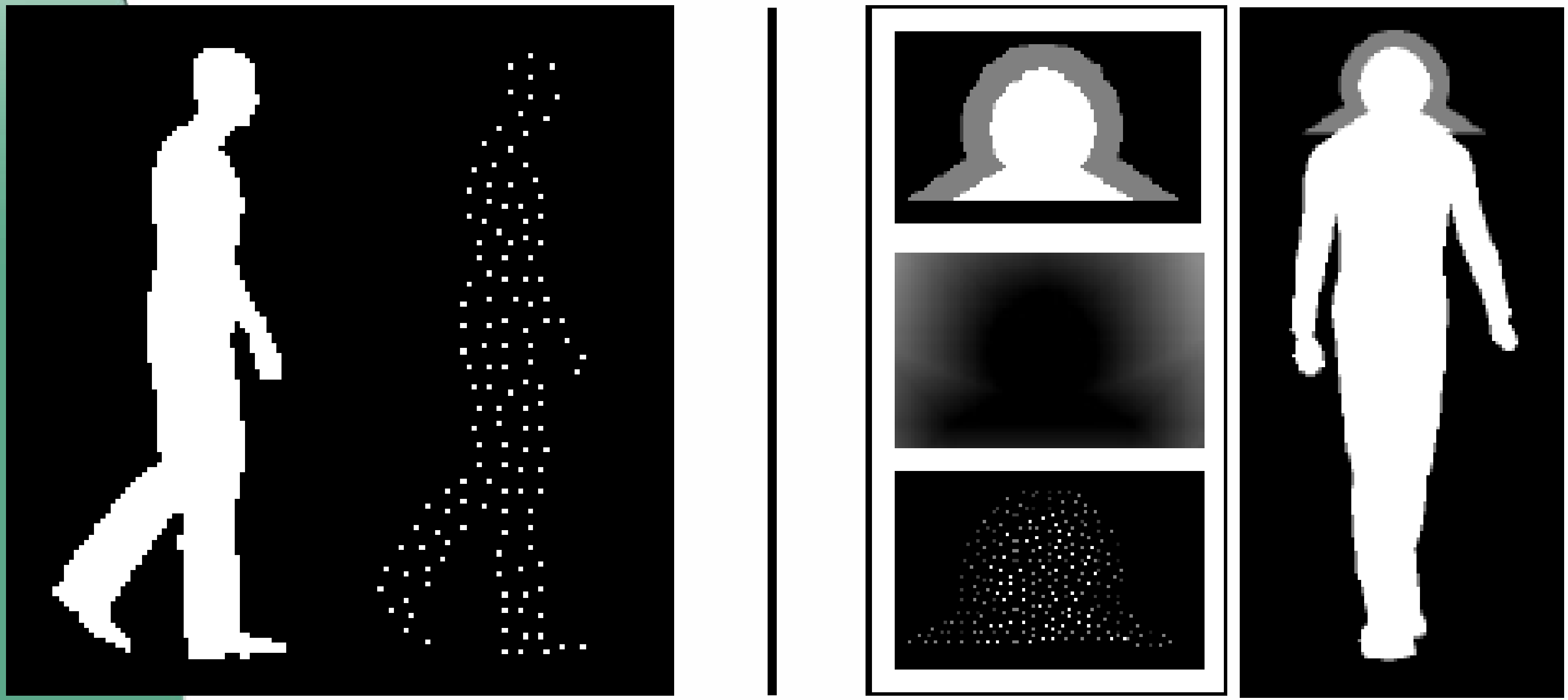
Subsampling

Object detection and recognition

10/07/2008
SSIP 2009



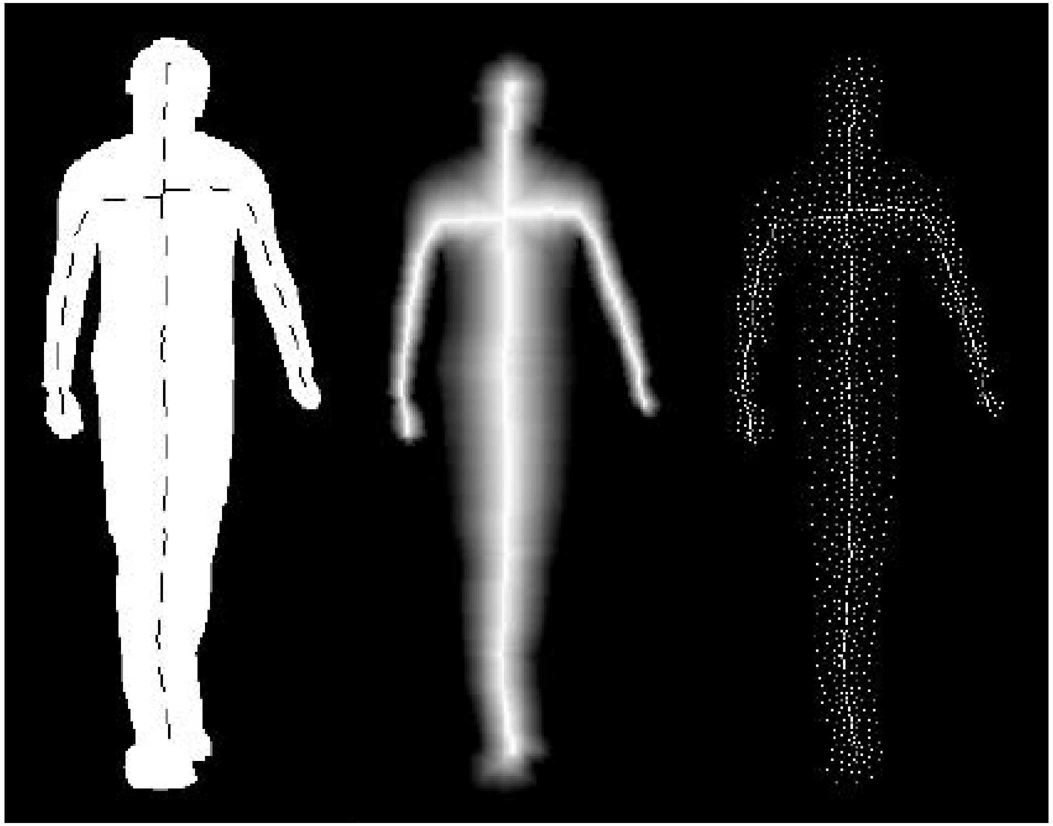
- Object simplification for faster matching
  - … and regions:



- Object simplification for faster matching
  - Simplification can be done with weighting the set differently

$$\widetilde{\varrho}(\mathbf{x}) = \frac{d(\mathbf{x}, \overline{B^c})}{d(\mathbf{x}, \mathrm{sk}(A)) + d(\mathbf{x}, \overline{B^c})}$$

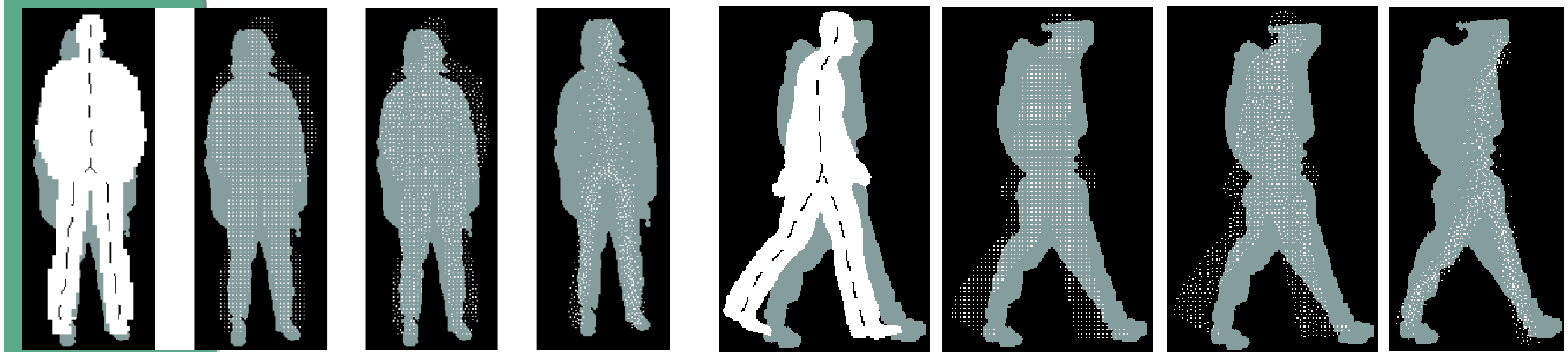
$$\varrho(\mathbf{x}) = \frac{\widetilde{\varrho}(\mathbf{x})}{\int_B \widetilde{\varrho}(\mathbf{y})\,d\mathbf{y}}, \quad \text{for } \mathbf{x} \in B$$



Object detection and recognition

- Simplification using a skeleton-based weight function

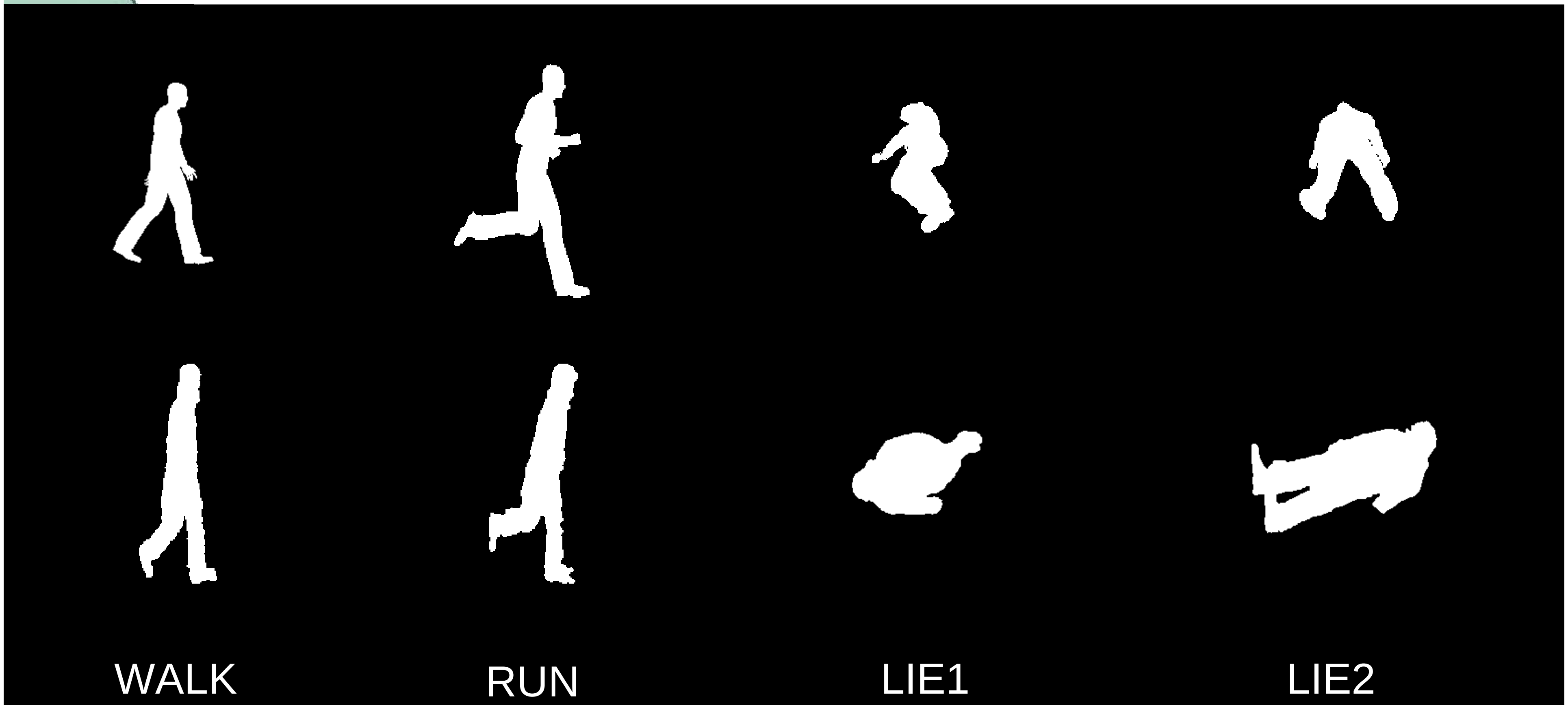

	Simplification			
	No (original)	Unifom (trivial)	Unifom (CVT)	Weighted (skeleton)
Standing	80,9%	81,2%	81,1%	94,0%
Walking	85,8%	86,2%	86,2%	94,6%



Object detection and recognition

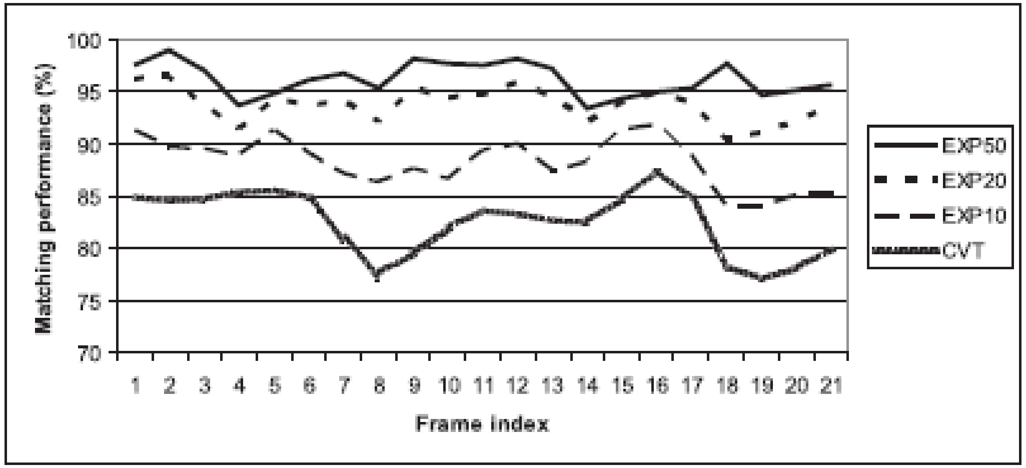
10/07/2008
SSIP 2009

- Simplification using a skeleton-based weight function

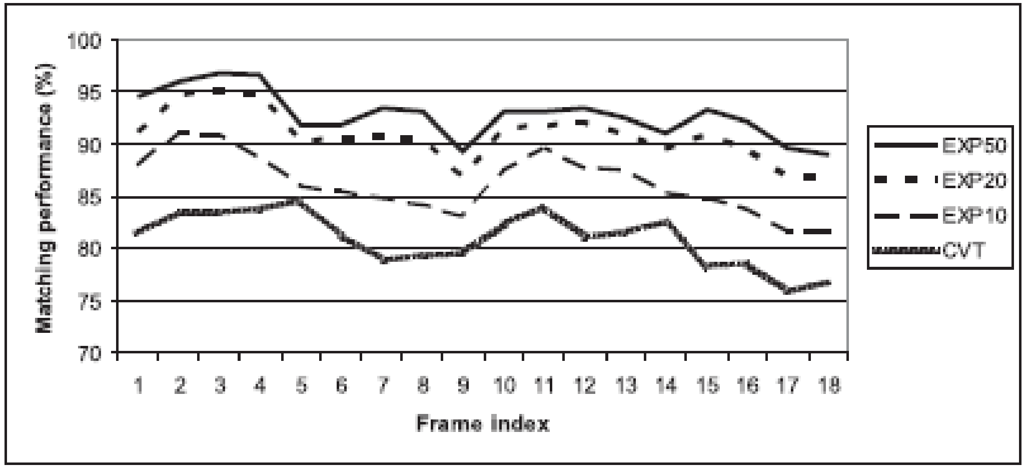


Object detection and recognition

- More focused simplification leads to better matching performance



WALK

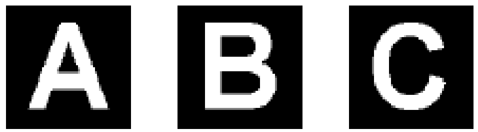


RUN

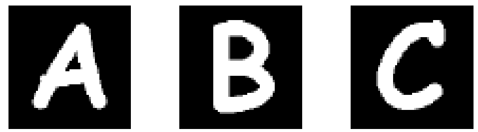
Object detection and recognition

- Simplification using a skeleton-based weight function
  - Different font alphabets can be matched better using this subsampling

Times New Roman



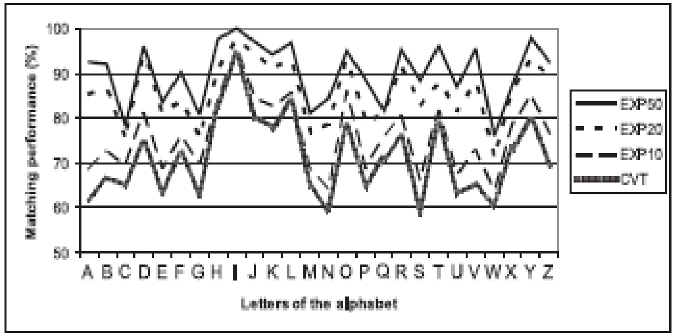
Arial

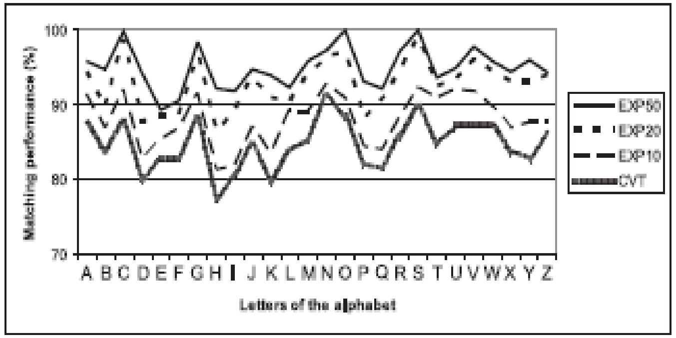


MS Comic Sans

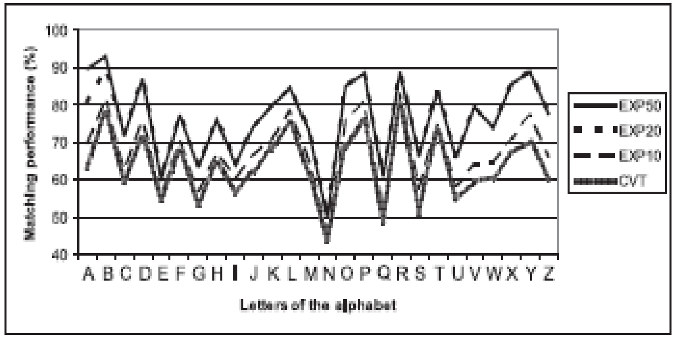
Object detection and recognition

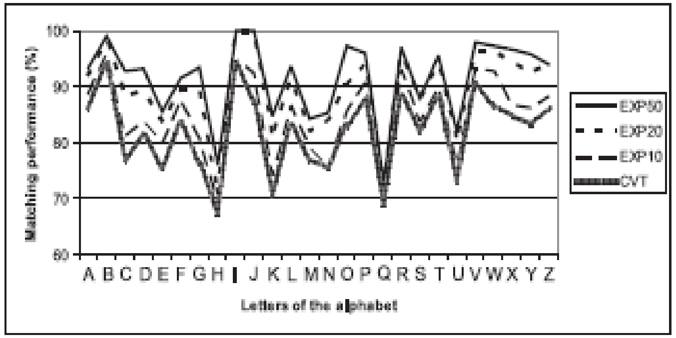
- More focused simplification leads to better matching



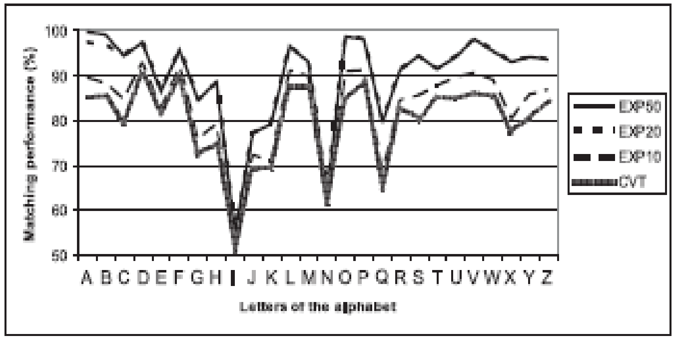


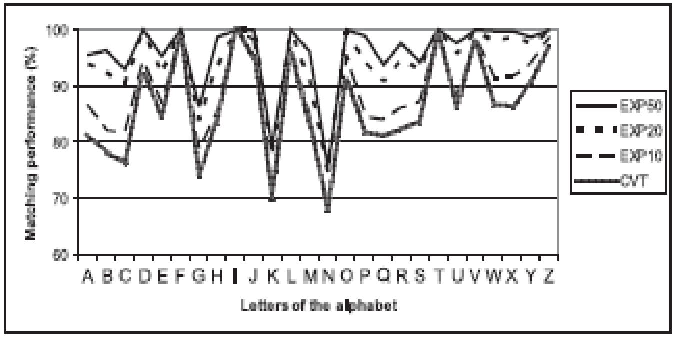
TNR/AR





TNR/COMIC





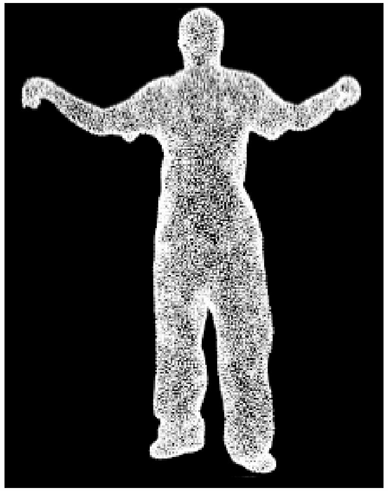
AR/COMIC

Registration of objects

- To investigate the impact of an object simplification (subsampling) approach in registration using the Iterative Closest Point (ICP) algorithm
- Fields of application:

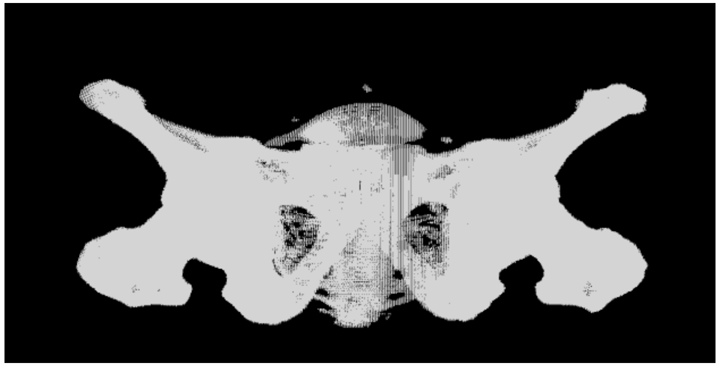
Body posture estimation

Registering 3D human models



Clinical registration

Registering 3D medical data

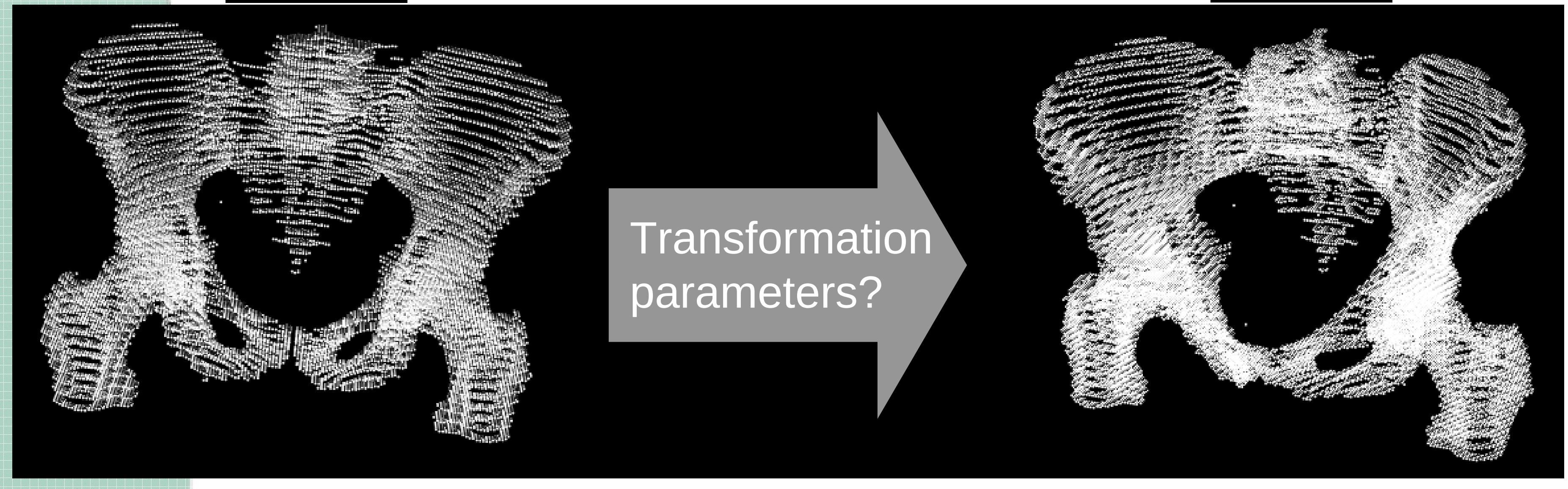


ICP registration

10/07/2008
SSIP 2009

- The registration task is to find the geometric transformation parameters (rotation and translation) between two objects represented by point clouds.
- The paired point-based ICP algorithm determines the geometric parameters through an iterative process.

Object 1 **Object 2**



Benefit and questions

- Benefit of subsampling:
  - Since the ICP algorithm considers matching point pairs the computational time drops with using less points.
- Arising questions:
  - Is the subsampled object capable to find the geometric transformation (does it substitute the original object)?
  - What level of subsampling leads to poor performance?
  - Does subsampling take less time than registration (is it possible to perform subsampling online)?

Offline approach

Only one of the objects are subsampled (the one taken earlier).

Online approach

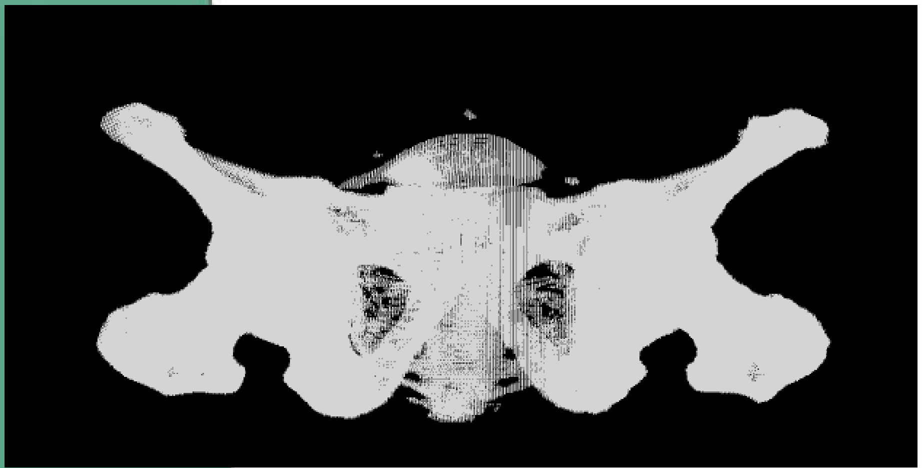
Both of the objects are subsampled.

Subsampling levels

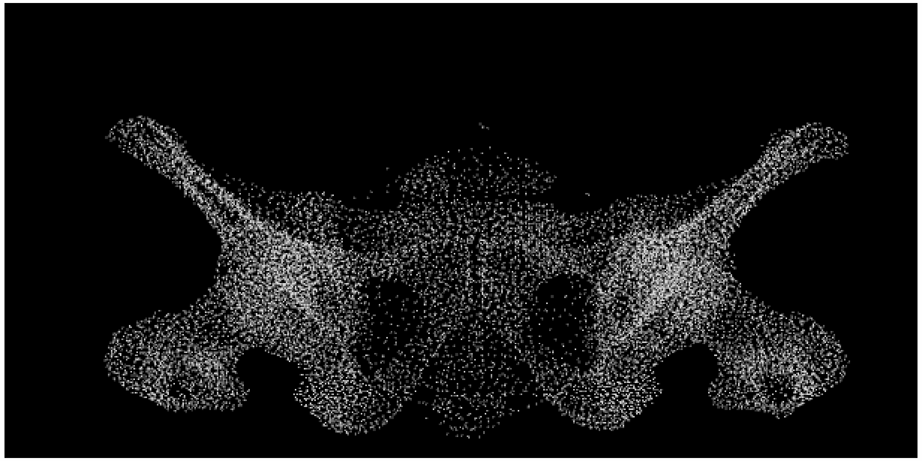
- The following 20 subsampling levels were considered:

100%	95%	90%	85%	80%	75%	70%	65%	60%	55%
50%	45%	40%	35%	30%	25%	20%	15%	10%	5%

(% of points retained)



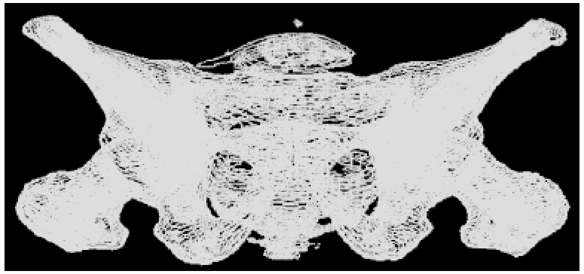
Original set



Subsampled at 5% level

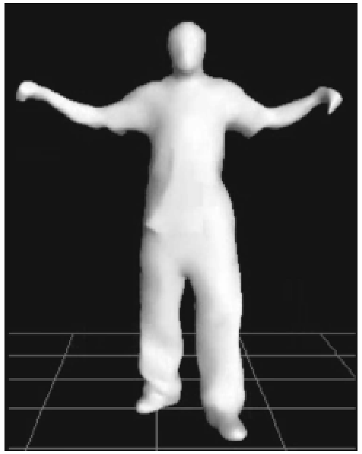
Experimental results
Test bed

- **Both simulated an real data were considered.**
- **Simulations applied for 3D medical data (hip):**



– SIM1:	Rotation	(0.017, 0.017, 0.017)
	Translation	(2, 2, 2)
– SIM2:	Rotation	(0.035, 0.035, 0.035)
	Translation	(4, 4, 4)
– SIM3:	Rotation	(0.07, 0.07, 0.07)
	Translation	(8, 8, 8)

- **Simulations applied for 3D human:**



• SIM4:	Rotation	(0.017, 0.017, 0.017)
	Translation	(0.1, 0.1, 0.1)
• SIM5:	Rotation	(0.14, 0.14, 0.14)
	Translation	(0.8, 0.8, 0.8)

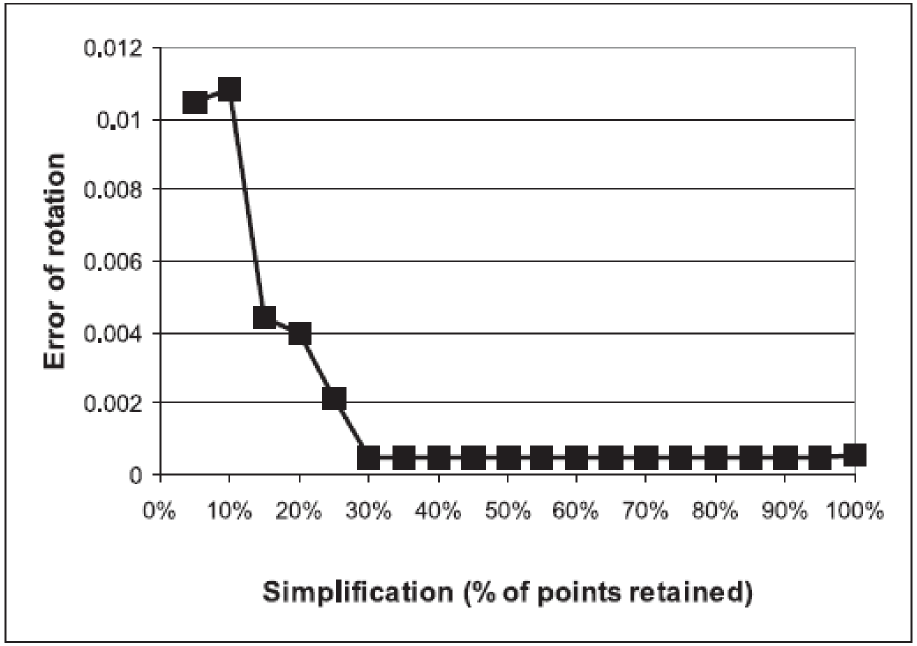
Experimental results
Error measurement

- In the simulated cases we know the actual transformation parameters:
  - Rotation: (rd_1, rd_2, rd_3) Translation: (td_1, td_2, td_3)
- If the transformation parameters found by the ICP algorithm are:
  - Rotation: (rf_1, rf_2, rf_3) Translation: (tf_1, tf_2, tf_3)
- Then the error terms are defined as:
  - Rotation: Translation:

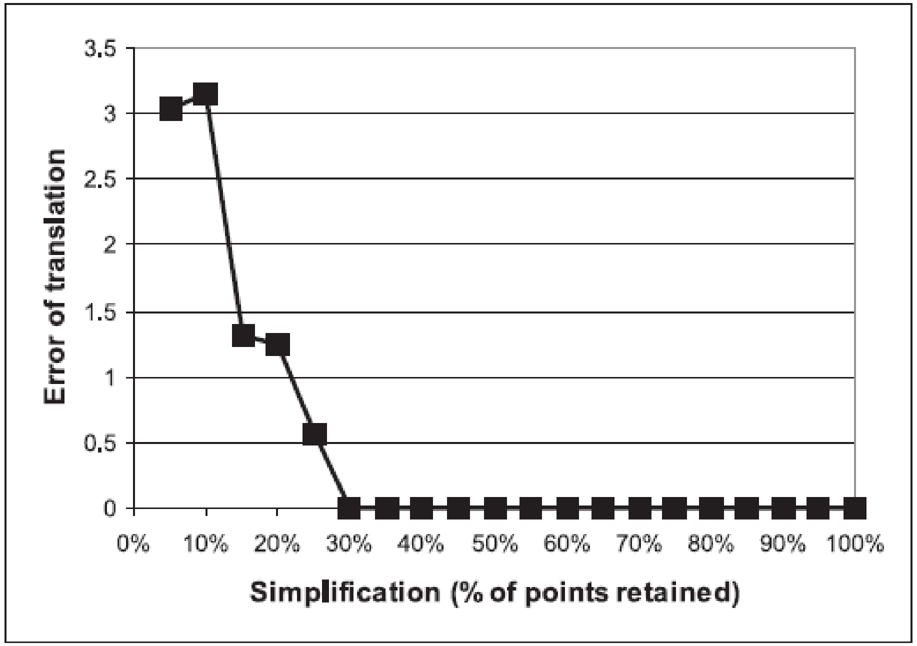
$$E_R = \sqrt{\sum_{i=1}^{3}(rd_i - rf_i)^2}$$

$$E_T = \sqrt{\sum_{i=1}^{3}(td_i - tf_i)^2}$$

Experimental results
SIM1 (geometric degradation for hip)



Error of rotation

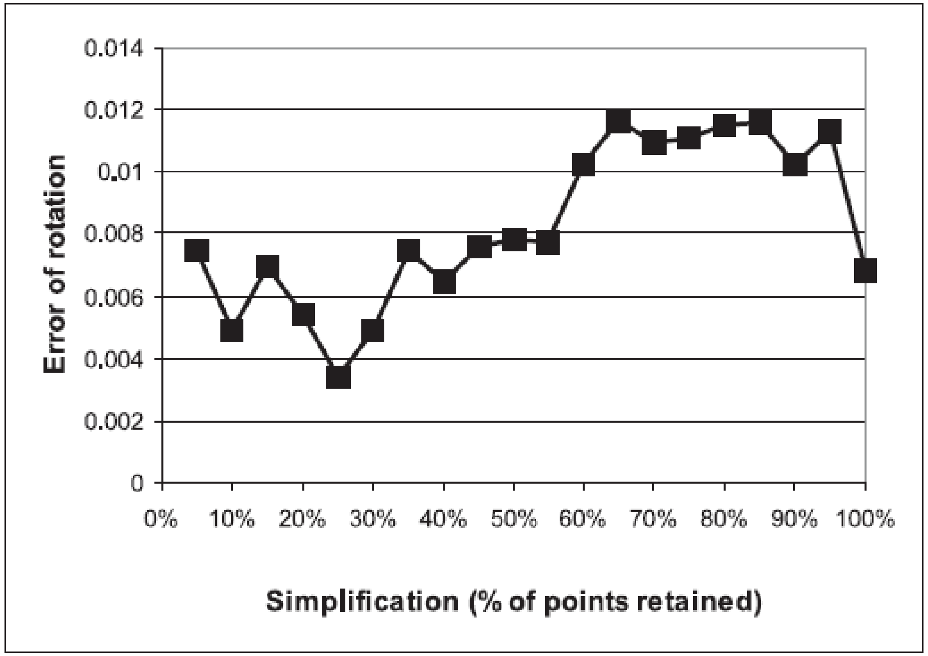


Error of translation

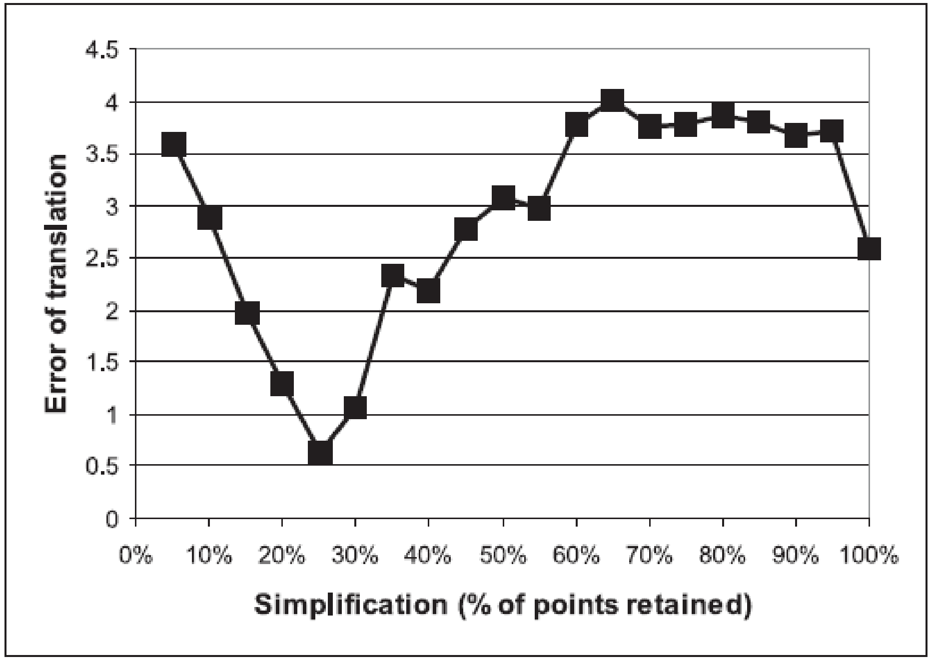
- ICP accurately found the true parameters down to some subsampling level.
- Since the geometric degradation was minor, even the 5% level performed quite well.

Experimental results
SIM2 (geometric degradation for hip)

10/07/2008



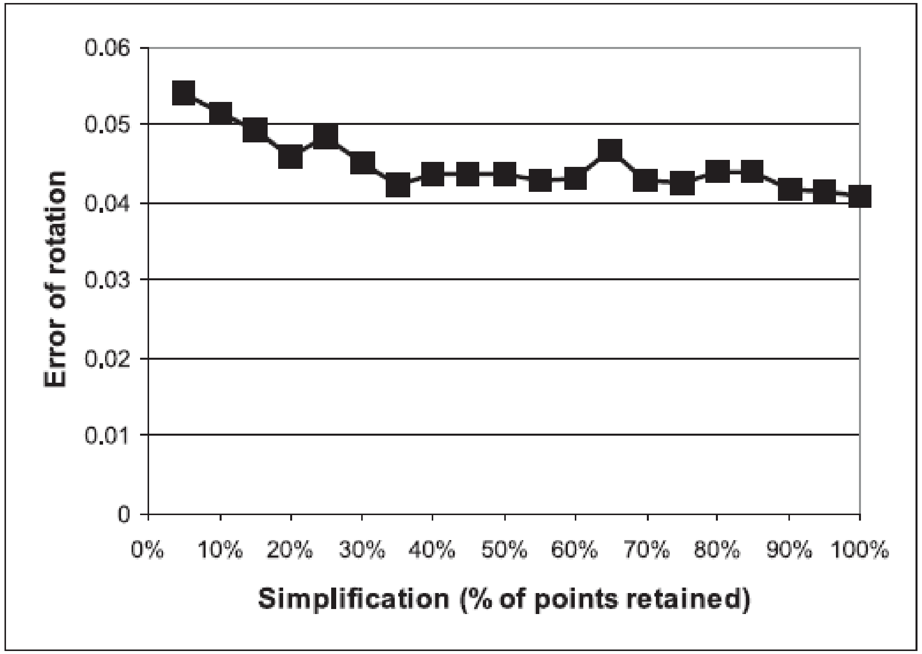
Error of rotation



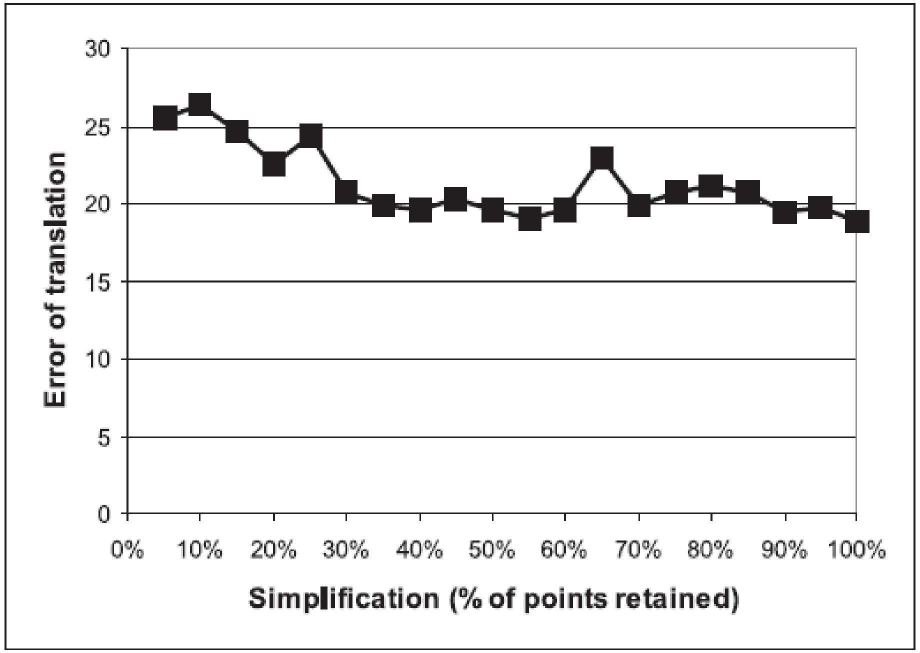
Error of translation

- For the larger geometric distortion ICP was less successful.
- Even a simpler version can perform better than the original one.

Experimental results
SIM3 (geometric degradation for hip)



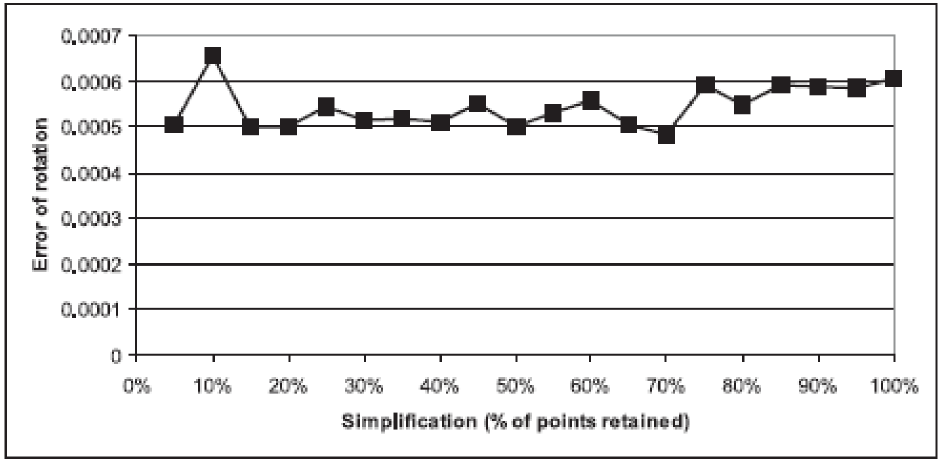
Error of rotation

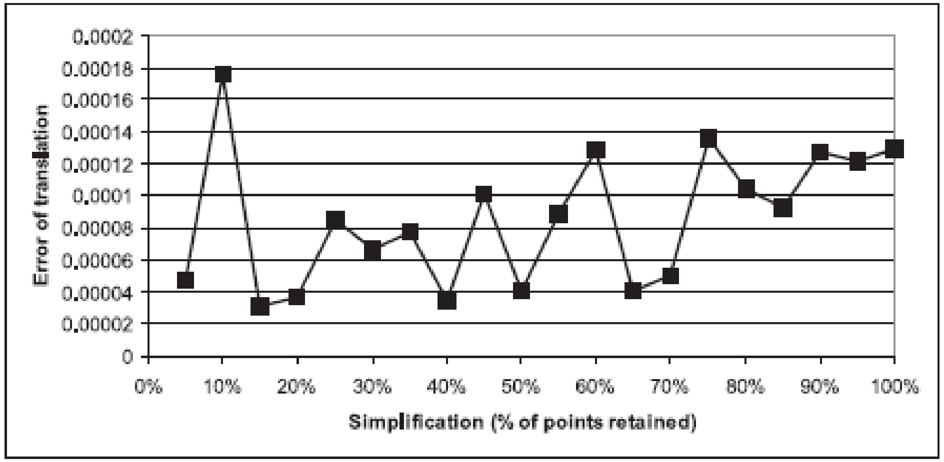


Error of translation

- For the even larger geometric distortion ICP was even less successful.
- The subsampled variants behaved similarly to the original one.

Experimental results
SIM4 (geometric degradation for human)

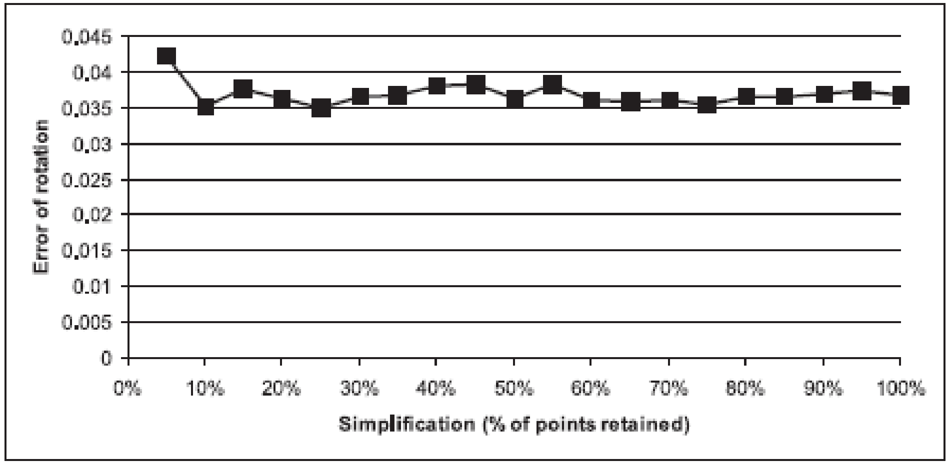


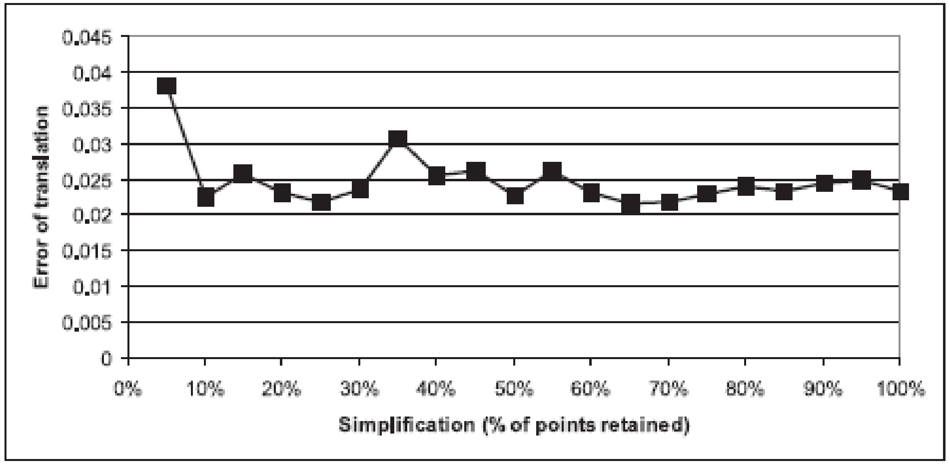


Error of rotation **Error of translation**

- Similar results as for the medical data:
  - For the minor transformation, the subsampled versions behaved well.
  - It is possible that a subsampled version behaves better than the original one.

Experimental results
SIM5 (geometric degradation for human)



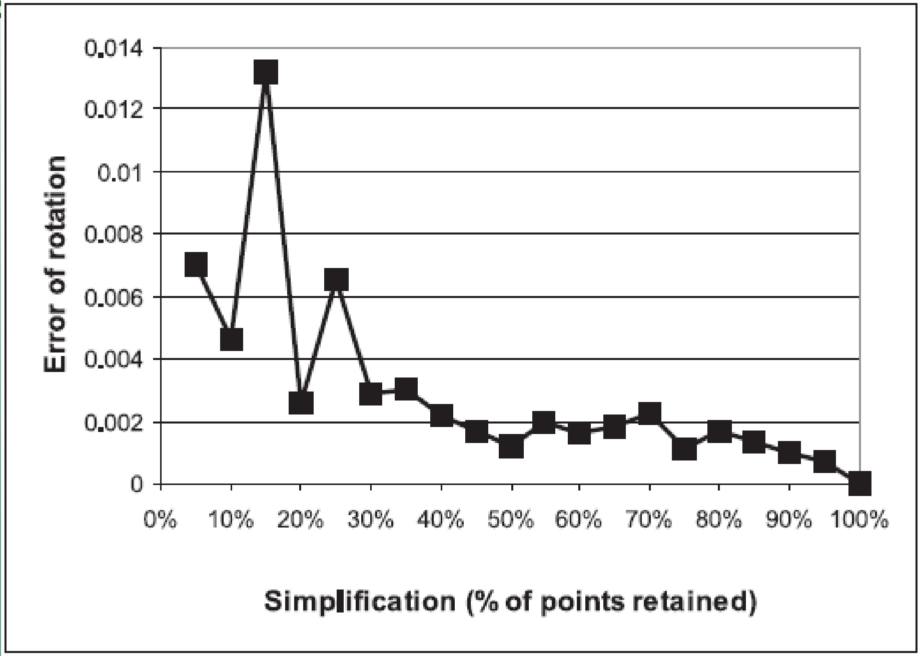


Error of rotation **Error of translation**

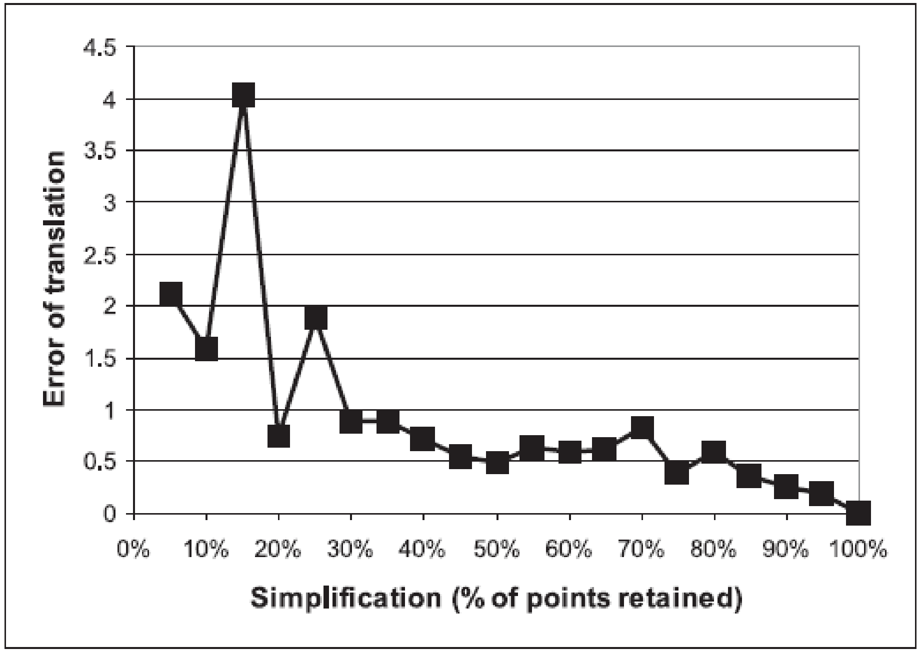
- **Similar results as for the medical data:**
  - More severe geometric transformation results in worse registration performance.
  - The subsampled versions substituted the original one well.

Experimental results
REAL (geometric degradation for hip)

10/07/2008



Error of rotation

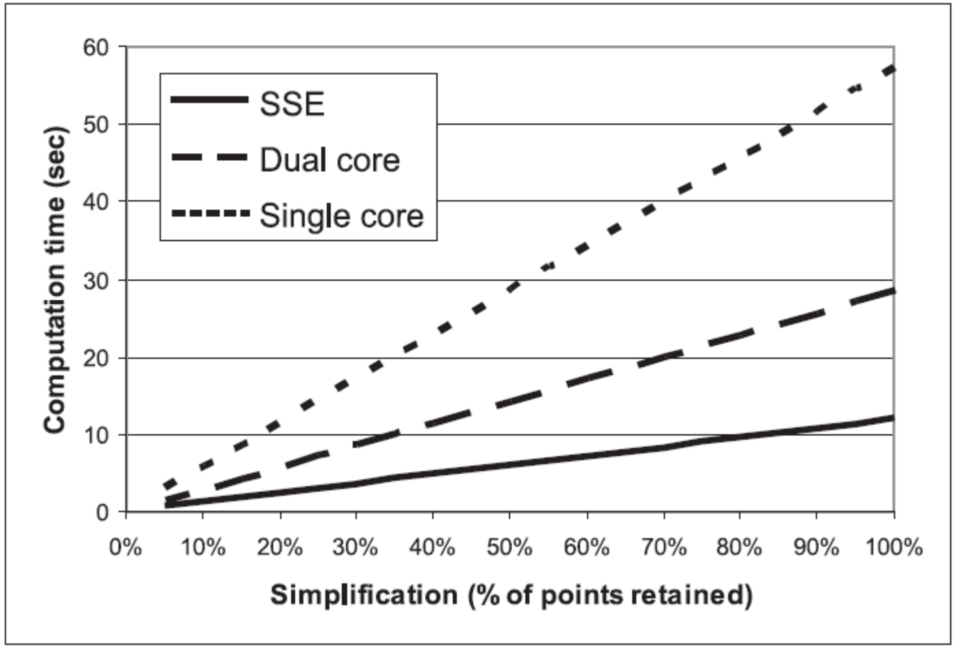


Error of translation

- In the real test the two sets were captured by CT scans at different times.
- The transformation parameters found for the original version were considered as having zero error.
- Down to a level, subsampling performed well.

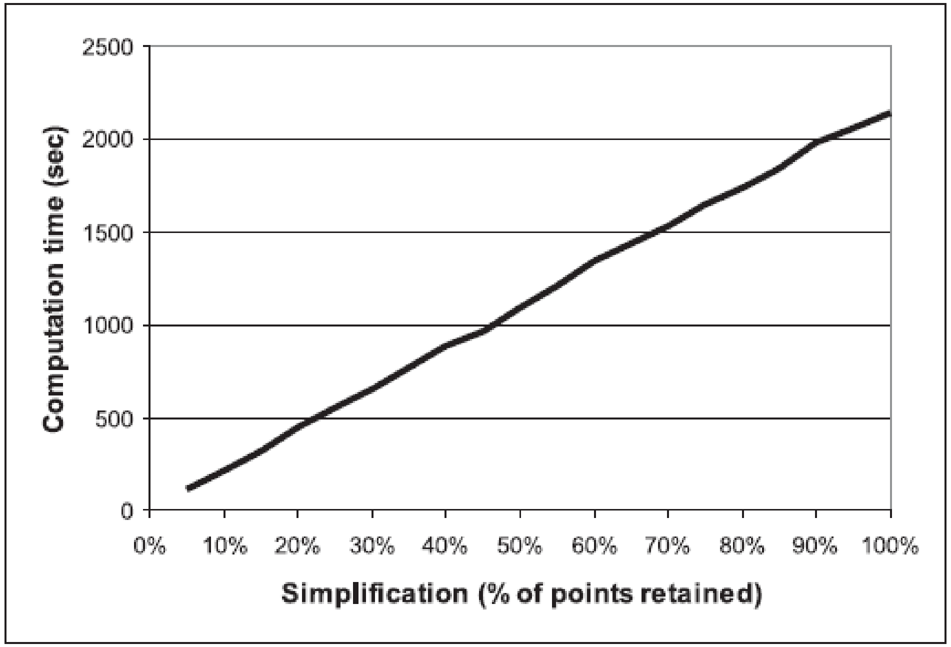
Experimental results
Computational times for subsampling

- **Implementations of the subsampling method:**
  - *Single core*: **single core CPU environment,**
  - *Dual core*: **dual core CPU environment,**
  - *SSE3*: **dual core CPU environment using SSE3 instructions.**
    (Intel Core 2 Duo CPU @ 2.00GHz (T7200), 1.5 GB RAM)
- **Computational times (object of 20565 voxels):**



Experimental results
Computational times for ICP

- **Implementations of the ICP:**
  - **The Insight Toolkit (ITK) library was utilized (Levenberg-Marquardt optimizer of the VNL class library),**
  - **The maximal number of iterations were fixed to be 15,**
  - *Single core*: **single core CPU environment, (Intel Core 2 Duo CPU @ 2.00GHz (T7200), 1.5 GB RAM)**
- **Computational times (object of 20565 voxels):**



Experimental results
Comparing computational times

- **The** *single core* **implementation of subsampling was 35 times faster than ICP.**
- **The** *SSE3* **implementation of subsampling was 160 times faster than ICP.**
- **Thus, the subsampling of the model captured later (online approach) looks to be valid.**

Analysing thermal photographs

- Thermal cameras are considered in more fields, like:
  - Medical applications,
  - Architecture,
  - (Millitary applications,)
  - (Mechanical/electrical engineering,)
  - (Biometrics,)
  - (Dangerous materials,)
  - ...

Analysing thermal photographs

- Medical thermal cameras:
  - are calibrated for the domain [30ºC--40ºC],
  - have (state-of-the-art) thermal resolution of 0.001ºC,
  - have moderate resolution (384x288).
- Usual tasks:
  - detecting breast cancer,
  - face recognition (blood vessels),
  - monitoring of disease processes,
  - evaluation of complex regional pain syndrome,
  - storage and retrieval of medical thermograms,
  - applications in dentistry,
  - applications in surgery,
  - monitoring during exercise,
  - fever monitoring (e.g. against bird flue),
  - ...

Approaches for thermal analysis

- Statistical:
  - Histogram based first order statistics:
    - mean, variance, skewness, kurtosis, energy, entropy, …
  - Second order statistics defined on co-occurrence matrix (joint probability of two pixels):
    - energy, variance, difference variance, correlation, inverse difference, entropy, …
  - Classification:
    - ANN, *k*NN, PCA, LDA, …
- Anything else…
  - (snakes, edge detaction, masked filters, …)

Thank you!

Thank you for your attention.