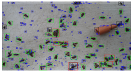
Analysis of visual contents

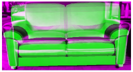
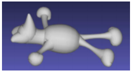
Complementary information for computer vision applications

Sylvie Chambon

October 24th, 2023

SAUTOS

Institut de Recherche en Informatique de Toulouse

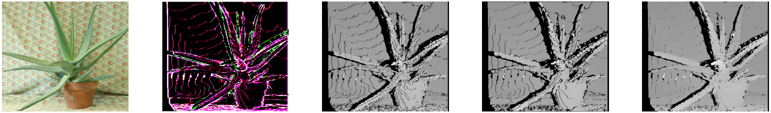
Université de Toulouse

IRIT
Institut de Recherche
en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J

TOULOUSE INP N7

Who am I?

Matching : Combine correlation measure to be robust to occlusions

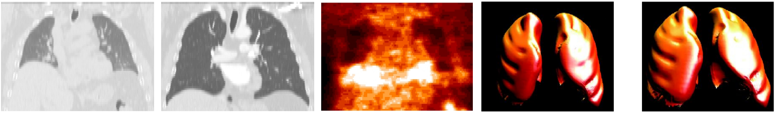


2005

Post-Doctorate
IRIT, Toulouse
2005-2006

Who am I?

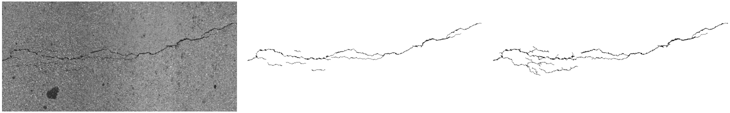
Registration : Feature detection in medical multi-modal images for precise organ tracking



2006

Post-doctorate
LTCI-UPM, Paris
2006-2007

Segmentation : Minimal path selection for thin object segmentation



2007 2011

Researcher
IFSTTAR, Nantes
2007-2011

Who am I?

Segmentation : Combine multiple features (line, plan, horizon) for urban scene comprehension





2011

2022

Associate professor
IRIT, ENSEEIHT
Toulouse

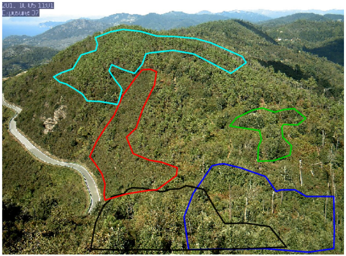
What am I interested in?

Analysis of complementary static and dynamic visual content

Understand visual scenes to provide tools
helping humans in their environment

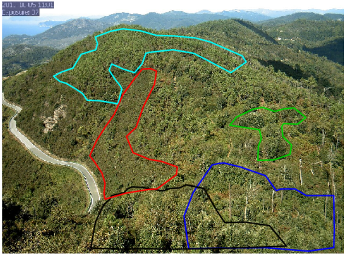
Analysis of complementary static and dynamic visual content

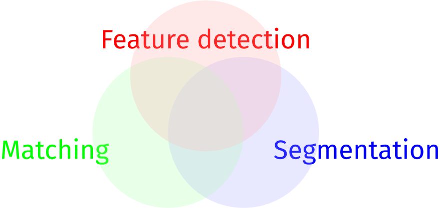
Understand visual scenes to provide tools
helping humans in their environment



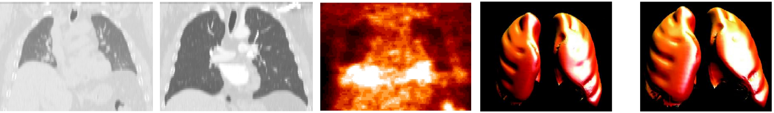
Analysis of complementary static and dynamic visual content

Understand visual scenes to provide tools
helping humans in their environment



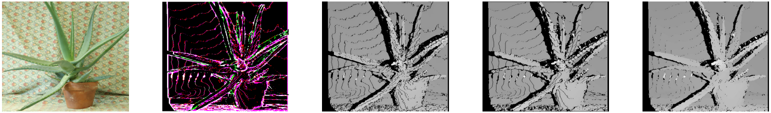


1. Combine all accessible information sources [Chambon 07]



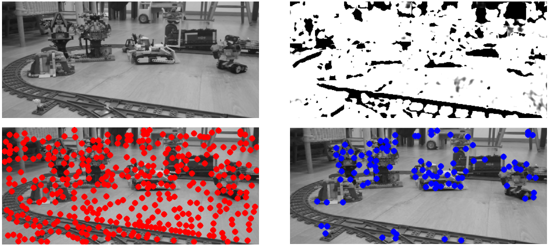
2. Combine all types of intermediate results [Bauda 15]

3. In order to match/to combine all the available information in the **best possible way** [Chambon 11a]

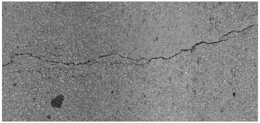
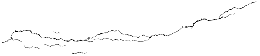
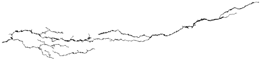


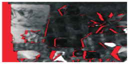
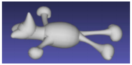
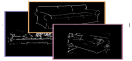
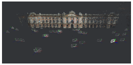
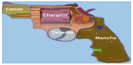
Important points
for academic and applied research

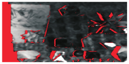
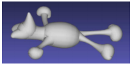
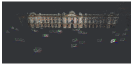
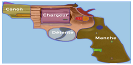
Comprehensive bibliography always linked with teaching



Intensive evaluations and comparisons
for real applications [Chambon 11b]

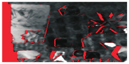
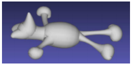
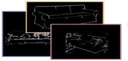
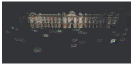
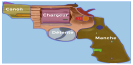
 Original image	 Reference segmentation
 Result of the state of the art	 Our result

Academic Research				
Macthing	Skeleton	View quality	3D Model	Segmentation
				

Academic Research				
Macthing	Skeleton	View quality	3D Model	Segmentation
				

Industrial research		
Planar superpixels	Multiple Videos	Faces
		

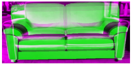
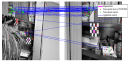
Academic Research

Macthing	Skeleton	View quality	3D Model	Segmentation
				

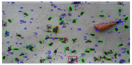
Industrial research

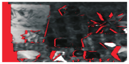
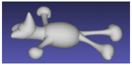
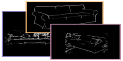
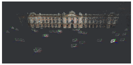
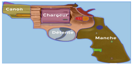
Planar superpixels	Multiple Videos	Faces
		

Regional projects

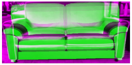
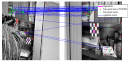
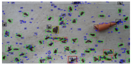
MobVille	INVISO
	

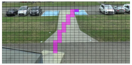
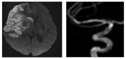
National projects

MARIO	VITI
	

Academic Research				
Macthing	Skeleton	View quality	3D Model	Segmentation
				

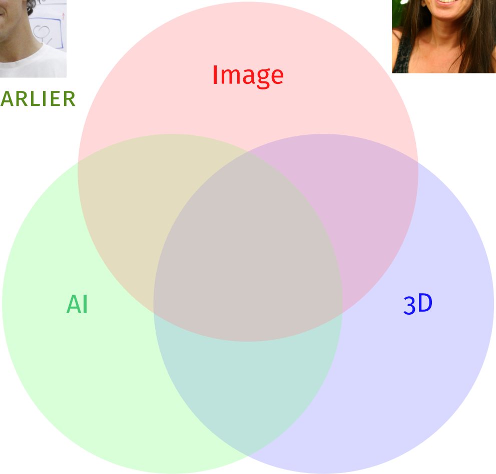
Industrial research		
Planar superpixels	Multiple Videos	Faces
		

Regional projects		National projects	
MobVille	INVISO	MARIO	VITI
			

European projects			
TRIMM	VICTORIA	Ceasefire	TARGET
			

Who is working with me?

Axel Carlier

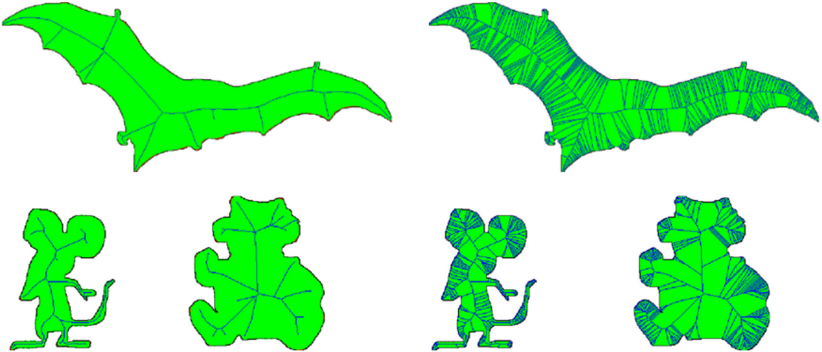


Simone Gasparini

Géraldine Morin

One step compact skeletonization 

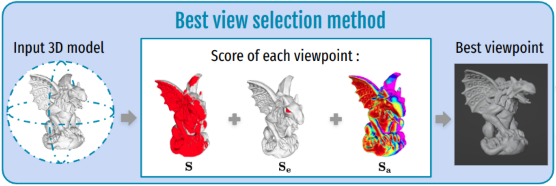
A robust, guaranteed approach, preserving details



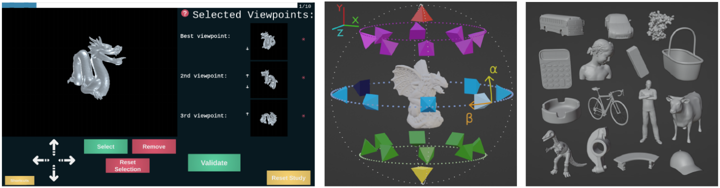
Our computation of the skeleton
And the ε-approximation (Hausdorff)

Voronoï computation of the skeleton

View-Dependent 3D Saliency Approach

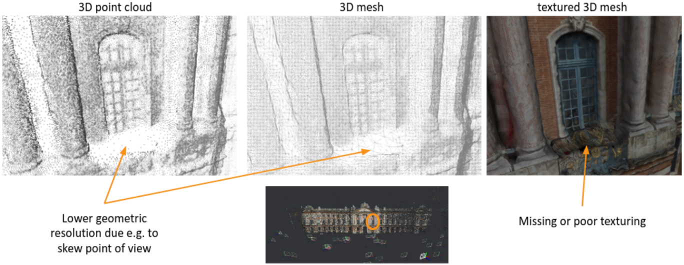


User study

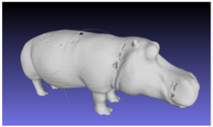


(a) Graphical interface of our user study (b) Cameras set-up (c) Extract of our 3D models database

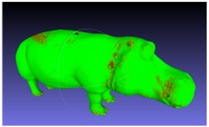
3D RECONSTRUCTION QUALITY ASSESSMENT



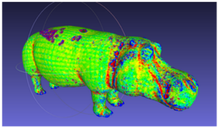
Leverage intrinsic metrics to detect the regions needing improvement



Reconstructed geometry

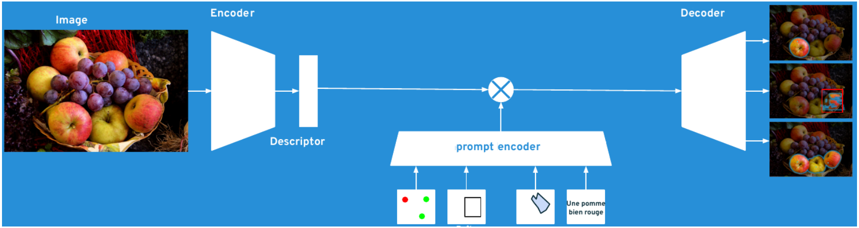


Roughness [Lavoué 2009]



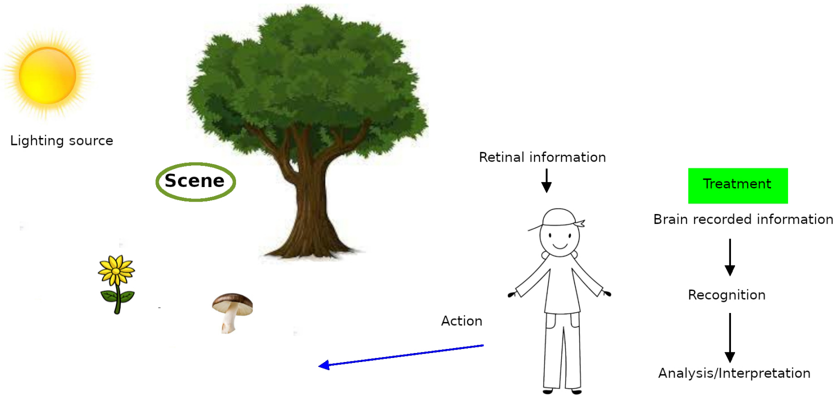
Curvature [Meyer 2003]

Semantic segmentation of object parts



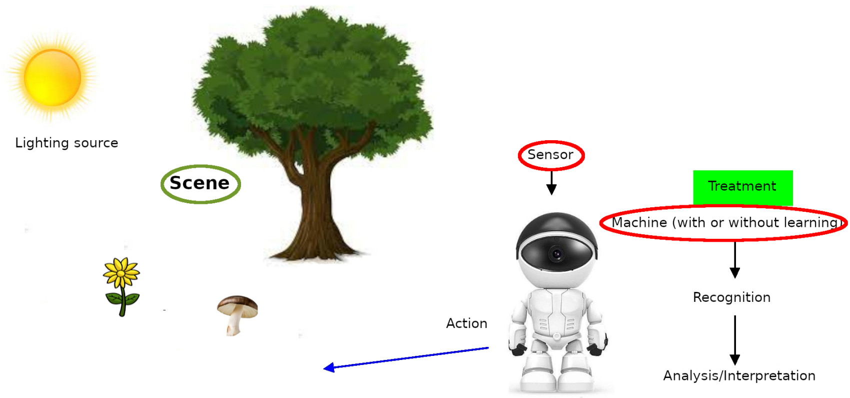
Presentation outline

Introduction



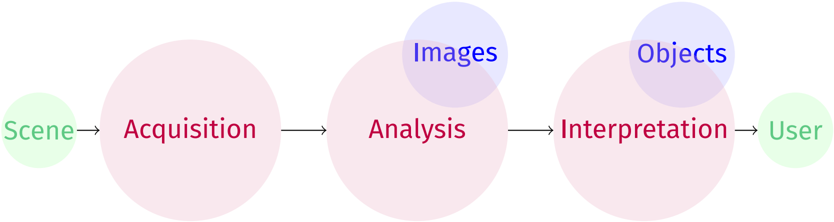
- **Performances**
  - Human physical system mature at 1 year old
  - $\neq$ Human visual system mature at 4 years old
  - $\approx$ 1 méga pixels
  - Identification time: $\approx$ 100ms
  - Number of memorised scenes: $\approx$ 100000

- **Objectifs :** Achieve human system performance
- **Goals :** Use machines to perform difficult or laborious human tasks

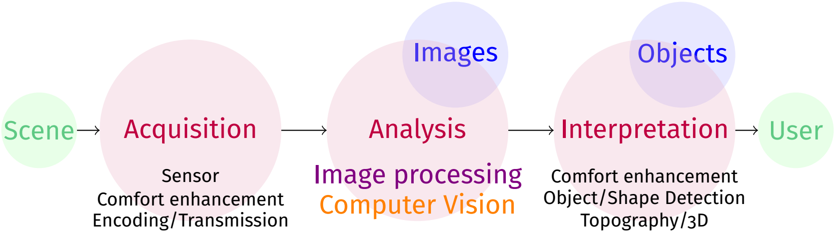


- **System :** scene + lighting source + sensor
- **Pipeline :** acquisition, treatment (recognition + interpretation), decision


Scene
Acquisition
Analysis
Interpretation
User


Scene
Acquisition
Images
Analysis
Objects
Interpretation
User

Introduction



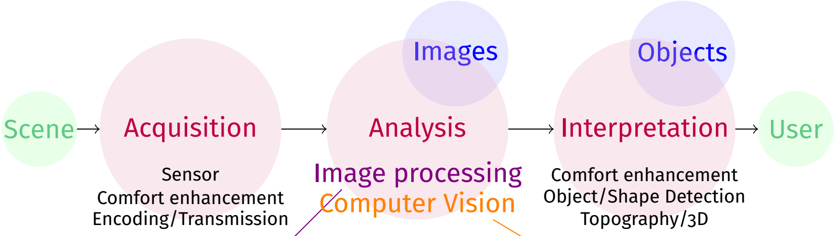


Image Transformation

Low Level (Near acquisition)

Compression / Quantization

Restoration / Enhancement

Transformation / Filtering

High Level (Near analysis)

Edge detection

Points of interest detection

Segmentation

Image comprehension

3D reconstruction

Calibration

Matching

Reconstruction

Shape recognition

Scene understanding

Introduction

Scene → Acquisition → Analysis → Interpretation → User

Images

Objects

Acquisition

Sensor
Comfort enhancement
Encoding/Transmission

Analysis

Image processing
Computer Vision

Interpretation

Comfort enhancement
Object/Shape Detection
Topography/3D

Image Transformation

Low Level (Near acquisition)

Compression / Quantization

Restoration / Enhancement

Transformation / Filtering

High Level (Near analysis)

Edge detection

Points of interest detection

Segmentation

Image comprehension

3D reconstruction

Calibration

Matching

Reconstruction

Shape recognition

Scene understanding

- Supervised learning

Def. Produce programs capable of performing a task without explicitly coding it

- Program learns from its experience to perform the task

only if Performance measure (a cost) increasing with experience [Mitchell 97]

- Requires two elements:
  1. Learning set (annotated data)
  2. Construction of a predictor for minimizing the difference

between the actual labels/values and the predicted labels

- **Most famous algorithms**
  - Decision trees [Quinlan 86]
  - Random forests [Breiman 01]
  - Neural network [McCulloch 43]

Why Neural Networks is popular since 2010?

- **Neuron**: closely related to the concept of neurons in biology
- **Perceptron**: several inputs, stored in the vector $\mathbf{x}$
  weigthed, with weights stored in $\mathbf{w}$
  allowing one output response y [Rosenblatt 58]:

$$y = f(\mathbf{w}^T \mathbf{x} + b), \tag{1}$$

where b a bias
and f the activation function

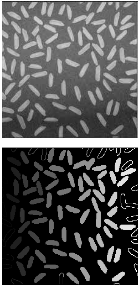
Rk: Perceptron model does not allow the resolution of non-linear problems [Marvin 69]

- [Rumelhart 86]: taking into account several layers:
  - Input layer (the data)
  - Output layer (the result)
  - Intermediate layers (hidden layers)
- **Deep learning**: Networks with at least 2 hidden or intermediate layers
- LeNet network [Lecun 98]: 5 layers
- AlexNet network [Krizhevsky 12]: several dozen of layers

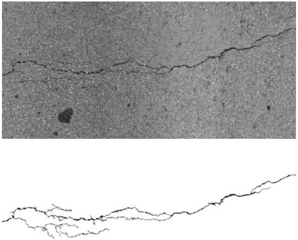
Why Neural Networks is popular since 2010?

- **Neuron-based approaches requirements/drawbacks**
  1. Important calculations
  2. Annotated database of sufficient size to carry out the learning phase correctly
- **Key elements in overcoming these obstacles**
  1. Increase of access to data acquisition systems
  2. Increase of access data sets (data sharing and transfer)
  3. Data anotation easier (crowdsourcing platforms)
  4. Computing power greater
  5. A better understanding and use of the learning activation function
     sigmoid, partly responsible for the initial poor results
     reLU function, rectified Linear Unit [Nair 10]
- Famous works: [LeCun 15, Goodfellow 16]

Is the problem well defined?
Can I give explicit caracteristics of the objects to detect/to segment?



Edge detection

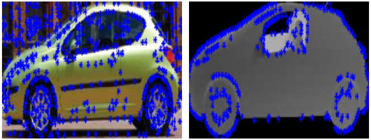


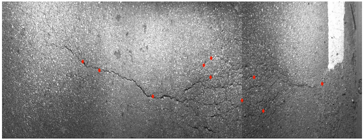
Minimal Path Selection [Amhaz 16]

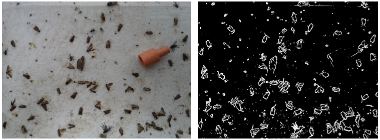
Points of interest detection and description

What is 2D/3D points of interest?

- **Primitive** = an element that can be extracted from an image in order to obtain information characteristic of the scene such as the presence of an object, a distance, a relief (i.e. an altitude or a depth)
- primitives = pixels, regions, contours, polygons, any set of points to obtain the desired information
- **Primitive of interest** - **points of interest**

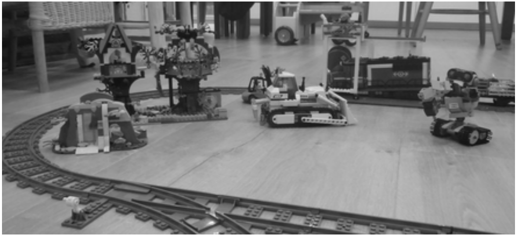
(a)


(b)


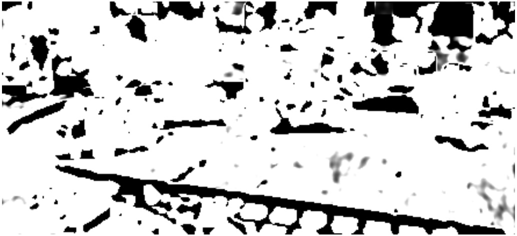
(c)


(d)

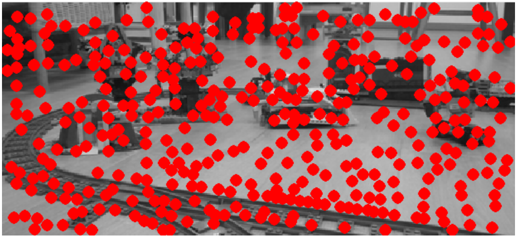

Original Image



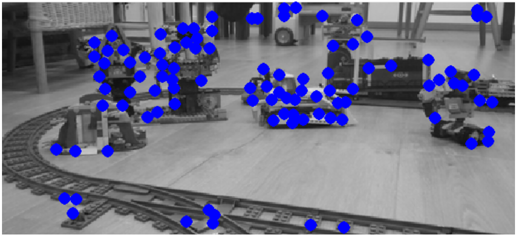
(1) Probability map calculation



(2) Non-local maxima suppression



(3) Post-processing (Selection)



How points of interest are detected?

- Without Non-local maxima suppression

  Badly distributed points
- Without Selection
  - Points well distributed across the entire image

  But Some features are not sufficiently salient to be tracked correctly (high risk of tracking errors)

  And Some parts of the image have no points of interest to follow

> *The **repeatability** of a primitive corresponds to its ability to be detected regardless of the image or data in which it appears. More precisely, if a primitive is detected in a given representation, then it must also be detected in another representation.*

If we note:

- $\mathbf{p}$, a point of interest in image I_1
- $\mathbf{p}^{'}$ his theoritical homologous in an other image, I_2 then the primitive is repeatable if it exists a point $\mathbf{q}$, detected in I_2 such as:

$$\|\mathbf{p}^{'} - \mathbf{q}\| \leq \epsilon.$$

Quality of the detection?

- Accuracy: excat location
- Robust to occlusions
- Invariant to image transformations
- Robust againt noise, blur or compression
- Dense point distribution guranteed
- Fast

- **Scale:**

= Level of detail used to observe/analyze the image

Related to convolution/filtering of the image

The greater the smoothing, the more fine details are lost, the smoother the image

- **Résolution :** Different resolutions of the images or objects manipulated

= Size of the image studied

Gradual downsizing of the image

The smaller the image, the more fine details are lost

Multi-resolution or multi-scale?

Detection *versus* description

What is the purpose of the descriptor relative to the detector?

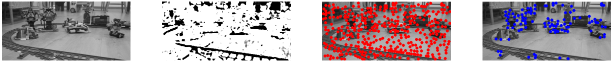
Detection *versus* description

What is the purpose of the descriptor relative to the detector?

- Detector
- = Determines if it is a point of interest or not
- Descriptor
- = All the caracteristics used for tracking or matching
- All the features used for tracking/matching
- Using the detector response to track is is not efficient

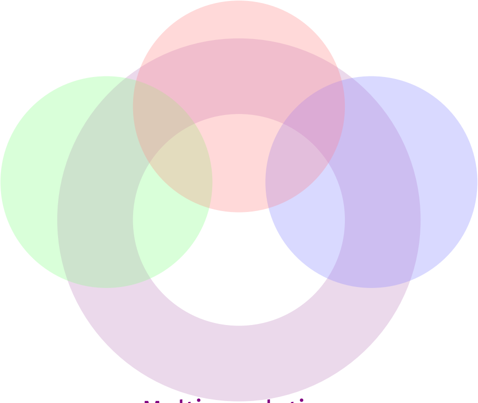
  Two solutions for tracking

  1. Simple correlation measure of intensities/colors
  2. Descriptors like SIFT, Scale-invariant feature transform



Region
[Tuytelaars 04, Deng 07]

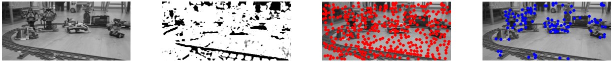
First-Order
[Harris 88]
[Mikolajczyk 04]



Second-Order
[Kitchen 82]
[Lowe 04]
[Deng 07]

Multi-resolution

2D/3D MATCHING: STATE OF THE ART



Region
[Tuytelaars 04, Deng 07]

First-Order
[Harris 88]
[Mikolajczyk 04]

Second-Order
[Kitchen 82]
[Lowe 04]
[Deng 07]

Multi-resolution

Detector families

<table>
<tr><th></th><th>First order</th><th>Region</th><th>Second order</th></tr>
<tr><td>Single scale</td><td>Moravec 1980, Harris 1988</td><td>SUSAN 1997, FAST 2006, MSER Matas2002, IBR 2004</td><td>Beaudet 1978, Kitchen 1982</td></tr>
<tr><td rowspan="2">Multi-scale</td><td rowspan="2">Harris-laplace 2004</td><td>Kadir 2004, EBR 2004</td><td>Hessien-Laplace 2004, SIFT 2004, SURF 2008, CSS 1998, MFC 2017</td></tr>
<tr><td colspan="2">PCBR 2007</td></tr>
</table>

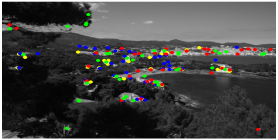
Invariant detectors

Detector	Photometrical Affine	Geometrical Non-affine	Geometrical affine	Scale transformation
Moravec,Harris				
Harris-laplace				×
SUSAN				
FAST				
MSER	×	×		
IBR	×	×	×	
Kadir	×	×	×	×
EBR	×	×	×	×
PCBR				×
Beaudet				
Kitchen1982				
Hessien-Laplace				×
SIFT		×		×
SURF				×
CSS	×	×	×	×
MFC				×

Detector complementarity

Image

Multiples points of interest

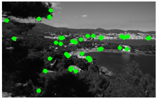


Harris

Beaudet

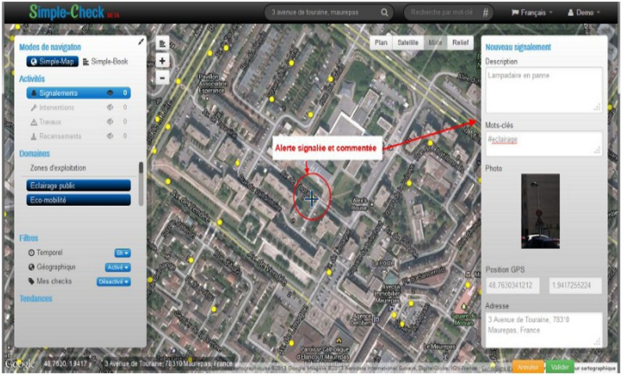
KR

SURF



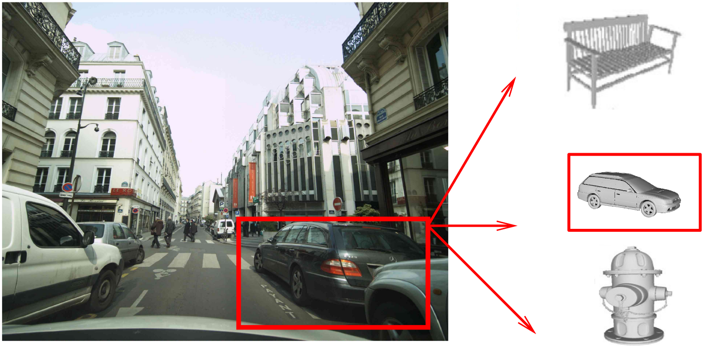
2D/3D POINT MATCHING

Collaborative application for taking into account geolocalised visual alert of citizens



Problem

- Object recognition based on 2D images and 3D models

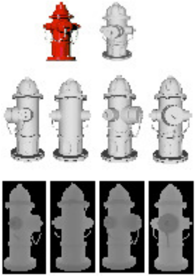


- **Objectives**
  1. Transform 2D and 3D data in order to obtain comparable data
  2. Match 2D and 3D data

Solution for the comparable representation

To be robust to texture and lighting

Depth map



focus/defocus principle

Solution for the matching

Existing approaches

- In 2D: Harris, SIFT, Scale Invariant Feature Transform
- In 3D: generalization of Harris, SIFT or adapted to 3D modelling
- between 2D and 3D: use of depth images or rendering images, based on learning
- **Proposition : detector based on geometric aspects to be adapted both to 2D and 3D**

Multi-scale curviness saliency

- **Motivation**
  - To be robust to scale transform: multiple scale, like SIFT
  - To use geometry: curvature
- **Notation**

  Image as the regular surface given by the graph of its intensity function $\mathcal{I}$ in x and y:

$$f(x,y) = (x, y, \mathcal{I}(x,y))$$

- **Assumption** $\mathcal{I}$ twice differentiable
- **Definition** Principal curvatures $\kappa_1(x_p, y_p)$ and $\kappa_2(x_p, y_p)$ of a point $p = f(x_p, y_p)$ are the eigenvalues of the *Hessian matrix* of $\mathcal{I}$ at (x_p, y_p):

$$\mathrm{H} = \begin{pmatrix} \mathcal{I}_{xx} & \mathcal{I}_{xy} \\ \mathcal{I}_{xy} & \mathcal{I}_{yy} \end{pmatrix} \tag{2}$$

  where $\mathcal{I}_{xx}$, $\mathcal{I}_{xy}$, $\mathcal{I}_{yy}$, the second-order partial derivatives
- **Analysis of this matrix**
  - Positive determinant: local mimimum, corner, point of interest
  - Negative determinant: saddle point
  - Near zero: flat regions

- **An other way to express the same idea: curvature tensor**
- **Notations** Two eigenvalues λ_1 and λ_2 and corresponding eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$ of H, assuming $\lambda_1 > \lambda_2$
- **Assumption** H is not singular
- **Definition** a conic centered at the origin with principal axis directions $\mathbf{e}_1$ and $\mathbf{e}_2$:

$$(x, y)\ \mathrm{H}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = 1$$

- **Analysis of this conic**
  - if $\lambda_1\lambda_2 > 0$: ellipse
  - if $\lambda_1\lambda_2 < 0$: hyperbola
- The shape of the conic indicates the structure

- **An other way to express the same idea: curvature tensor**
- **Notations** Two eigenvalues λ_1 and λ_2 and corresponding eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$ of H, assuming $\lambda_1 > \lambda_2$
- **Assumption** H is not singular
- **Definition** a conic centered at the origin with principal axis directions $\mathbf{e}_1$ and $\mathbf{e}_2$:

$$(x, y)\ \mathrm{H}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = 1$$

- **Analysis of this conic**
  - if $\lambda_1\lambda_2 > 0$: ellipse
  - if $\lambda_1\lambda_2 < 0$: hyperbola
- The shape of the conic indicates the structure
- Curviness Saliency is defined by:

$$CS = \lambda_1 - \lambda_2 \tag{3}$$

2D/3D matching: Curvilinear Saliency (CS) [Rashwan 19]

Definition

$$CS(x, y) = \kappa_1(x, y) - \kappa_2(x, y)$$

- κ_1 and κ_2 are principal curvatures
- Points of interest belong to elongated surface elements
- The higher CS, the higher the probability to be a point of interest

Multiscale and focus

- Mutliscale analysis
- Focus curve concept: estimation of the scale of blur

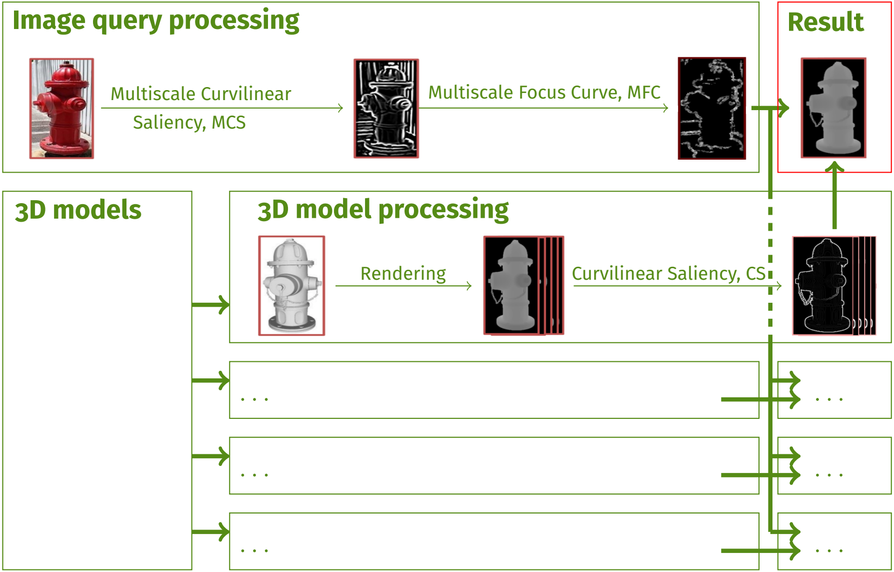
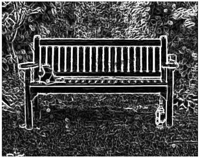


Image query processing
Multiscale Curvilinear Saliency, MCS
Multiscale Focus Curve, MFC
Result
3D models
3D model processing
Rendering
Curvilinear Saliency, CS
. . .
. . .
. . .
. . .
. . .
. . .

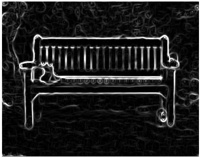
Multi-scale curviness saliency: illustration

Use of the detection at different scales

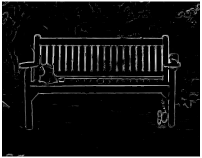
Image



Scale 1

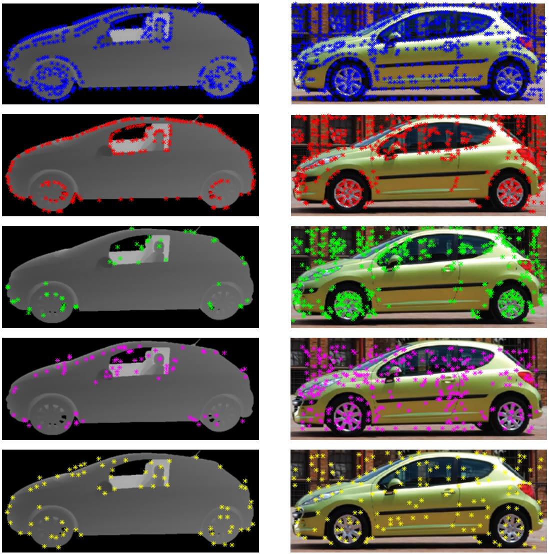


Scale2



MCS

Visual results



MCS

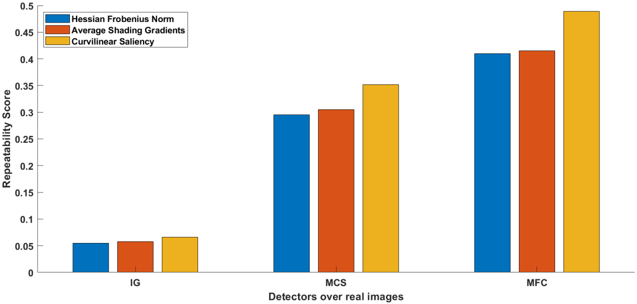
Curviness

SIFT

SURF

Harris

PASCAL3D+ dataset
Repeatability between 2D/3D



PASCAL3D+ dataset

THrre visual results for pose estimation

PASCAL3D+ dataset

Comparison with a CNN model

Models	mean *Acc*	mean *MedErr*
Render [Su 15]	0.82	13.6
ONet [Tulsiani 15]	0.81	11.7
Our Model with MFC	0.80	09.5

- *Acc*: pose estimation accuracy (the higher the better)
- *MedErr*: viewpoint error (the smaller the better)

Contributions

- Proposition of an approach for 2D/3D matching
- Methodology for validating the repeatability between 2D and 3D data

Perspectives

- Mobville project: detection of changes/problems
- Application of this concept in other works:
  - [Abdulwahab 19]
  - [Bakkay 18]
  - [Pelissier-Combescure 23]
- Experiment other 2D/3D common representation [Grabner 19b]
- Experiment other tools for focal length estimation [Grabner 19a]

MULTI-VIDEO ANALYSIS

VICTORIA

VIDEO ANALYSIS FOR INVESTIGATION OF CRIMINAL AND TERRORIST ACTIVITIES

- 15 collaborators with complementary competences: audio, vidéos de synthèse, métadata...

Context

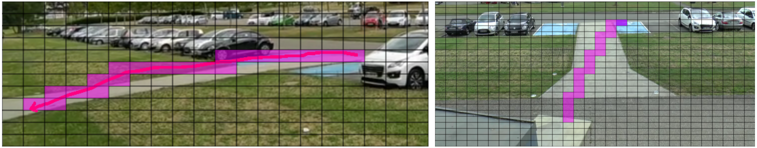
- Various and different image data: cameras, smartphone, videos, images
- How can we use this redundant data to extract significant information?
  - Watching all the data is too long and expensive
  - Methods needed to analyse automatically or help analysis

Context

- Various and different image data: cameras, smartphone, videos, images
- How can we use this redundant data to extract significant information?
  - Watching all the data is too long and expensive
  - Methods needed to analyse automatically or help analysis
- **Needs for investigators**
  - Naviguate efficiently inside a video collection
  - From one current video:
    1. Which videos allow us to view the same elements but from different angles?
    2. Which video offers the best view of an element of interest?

Videos from multiple devices

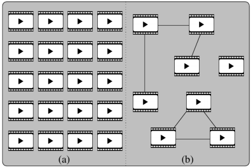
Which videos in a collection allow to better visualise a query trajectory?



- a list of videos ranked in descending order of relevance
- the reformulated trajectory for each video

Videos from multiple devices

How to help a user to navigate through a video collection in order to extract relevant information?



Existing works

- Re-identification [Cho 19]
- Camera network analysis based on activity profile [Loy 09]
- Homographies between ground planes

Existing works

- Re-identification [Cho 19]
- Camera network analysis based on activity profile [Loy 09]
- Homographies between ground planes

In this complex video configurations, tools unuseful!

Existing works

- Re-identification [Cho 19]
- Camera network analysis based on activity profile [Loy 09]
- Homographies between ground planes

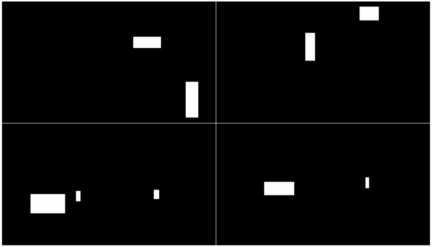
In this complex video configurations, tools unuseful!

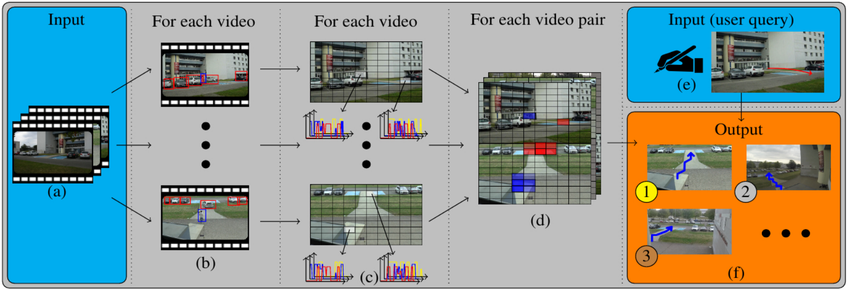
Assumptions

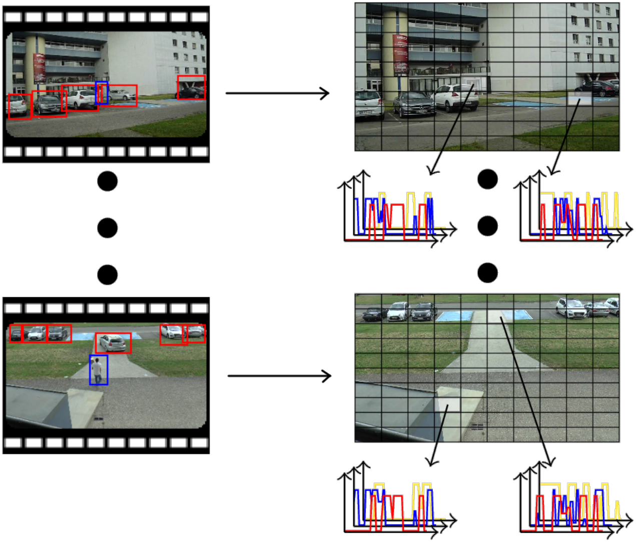
- Videos are static and synchronised
- Previous detections based on [Redmon 18] or [He 20]

Motivation of the proposed approach

Two regions extracted from two different cameras systematically and simultaneously occupied or unoccupied are matching regions




Input
(a)
For each video
(b)
For each video
(c)
For each video pair
(d)
Input (user query)
(e)
Output
1
2
3
(f)



Correspondance maps

For each cell pair $(c_{\mathbf{i}}^{V_1}, c_{\mathbf{i}'}^{V_2})$, correlation score by category ω:

$$\mathcal{C}_\omega(c_{\mathbf{i}}^{V_1}, c_{\mathbf{i}'}^{V_2}) = \mathrm{corr}(a_{\mathbf{i}}^{V_1,\omega}, a_{\mathbf{i}'}^{V_2,\omega})$$

$\Longrightarrow$ Correspondance maps

Reformulation score

$$\underset{(\mathbf{i}'_1,\ldots,\mathbf{i}'_M)}{\operatorname{argmax}} \frac{\frac{1}{M}\sum_{k=1}^{M} \mathcal{C}_\omega(c_{\mathbf{i}_k}^{V_1}, c_{\mathbf{i}'_k}^{V_2})}{1 + \sum_{k=1}^{M-1} \max(0, ||\mathbf{i}'_k - \mathbf{i}'_{k+1}|| - 1)}$$

- **($\mathbf{i}_1$, ..., $\mathbf{i}_M$)** request cell sequence
- **($\mathbf{i}'_1$, ..., $\mathbf{i}'_M$)** reformulated cell sequence
- $\mathcal{C}_\omega(c_{\mathbf{i}_k}^{V_1}, c_{\mathbf{i}'_k}^{V_2})$ correspondance score

- **Hypothesis**

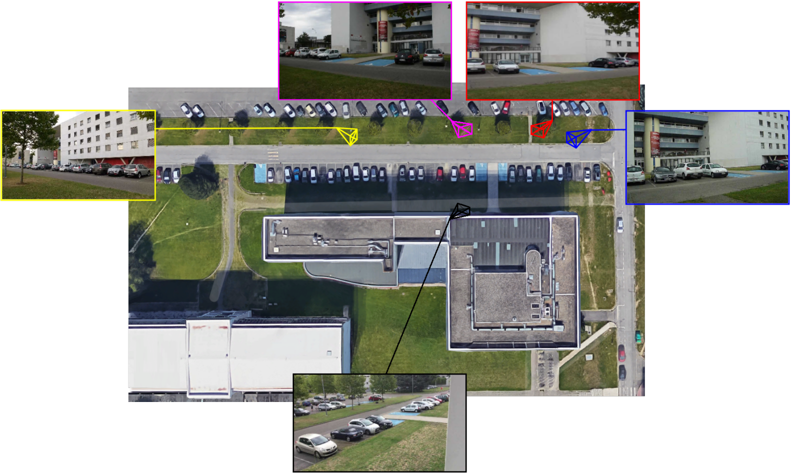
  Correlation between reformulated trajectory length and video interest
- **Visibility score**

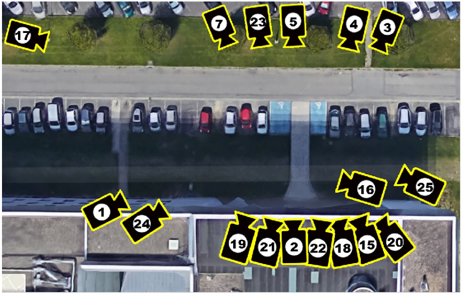
  Reformulation score

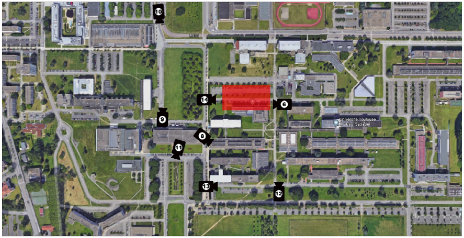
$\times$ length of the reformulated trajectory

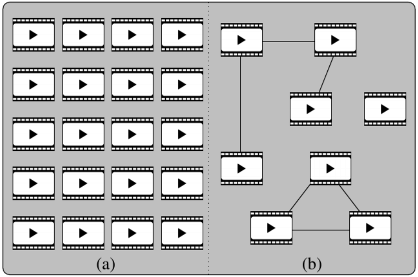
= Ranking of the video based on decreasing visibility score

- 25 cameras synchronised + a scenario
- Manual annotations with bounding boxes

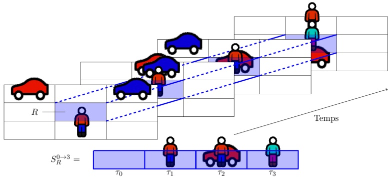



17
7
23
5
4
3
16
25
1
24
19
21
2
22
18
15
20

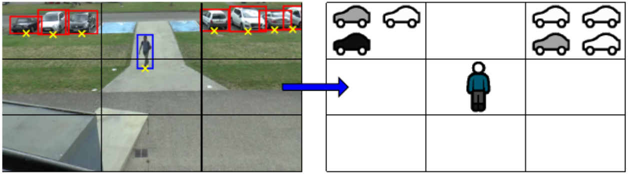

10
14
6
9
8
11
13
12

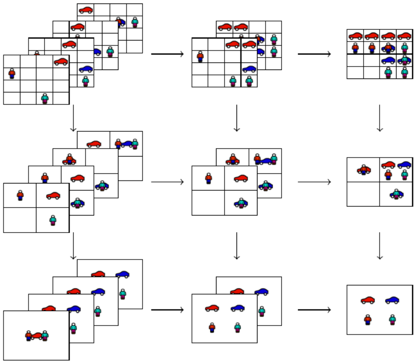


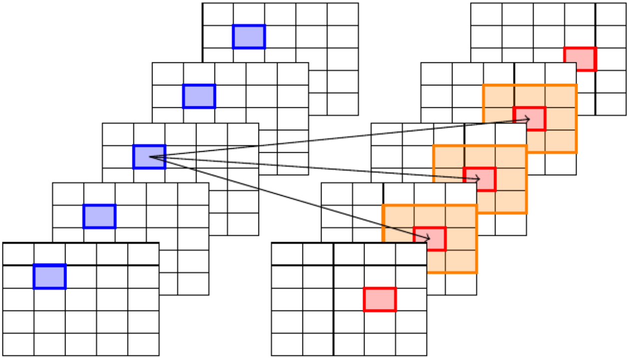
(a) Video Collection (b) Links between the videos



- R: a region
- T: Time step number
- $S_R^{0\to i}$: stories at each studied time step $\tau_0, \ldots, \tau_i$
- S_R^j: story at time step τ_j
- S_R: Stories of video







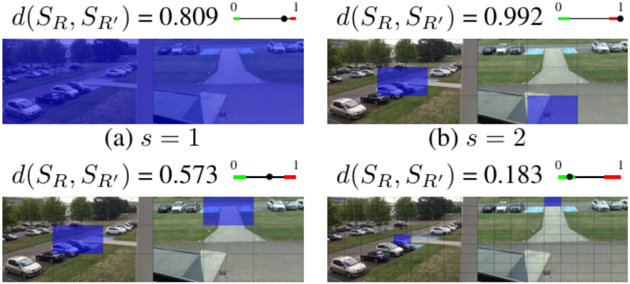
Distance between two stories

- S_R et $S_{R'}$: two stories
- O, O': two objects in two stories
- p_i: one attribute (person pose, car model)
- d_{p_i}: distance relative to p_i
- ω_{p_i}: weight of p_i

$$\delta(O, O') = 1 - \frac{\sum_{p_i} \omega_{p_i} d_{p_i}(O, O')}{\sum_{p_i} \omega_{p_i}} \tag{4}$$

- $C(S_R, S_{R'})$: Object number such as it exists at least one object O' in its spatial neighbourhoud such as $\delta(O, O') \leqslant \sigma_{dissimilarity}$

$$d(S_R, S_{R'}) = 1 - \frac{C(S_R, S_{R'}) + C(S_{R'}, S_R)}{|S_R| + |S_{R'}|} \tag{5}$$



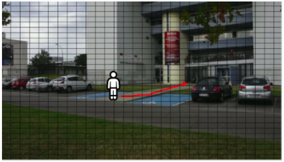
(a) $s = 1$ (b) $s = 2$

(c) $s = 2$ (d) $s = 3$

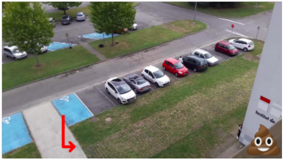
Function relationship (S^{V_1}, S^{V_2})
Input : Two stories at different scales S^{V_1} and S^{V_2}
Output: Relationship between V_1 and V_2
match = [ ]
candidate = $\{(V_1, V_2, 1)\}$
while *candidate* $\neq \emptyset$ **do**
    c = candidate[0] = $(R_1, R_2$, scale)
    if $d(S_{R_1}, S_{R_2}) \leqslant \sigma^s_{accept}$ **then**
        match = match $\cup\{(\mathrm{R}_1, R_2)\}$
    else
        if $d(S_{R_1}, S_{R_2}) \leqslant \sigma^s_{reject}$ **then**
            add all tuples $(r_1, r_2$, scale+1) where $r_1 \in R_1$ and $r_2 \in R_2$ to candidate
        else
            null
        end
    end
    delete candidates[0]
end

- V_i: region relative to the whole image
- *match*: set of matched videos
- *candidate*: set of triplets (two matched regions, one scale)
- *scale*: current scale
- *c*: first candidate studied

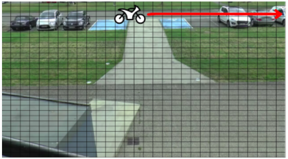
Results

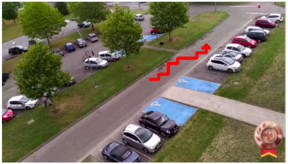




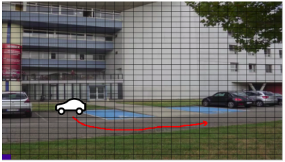


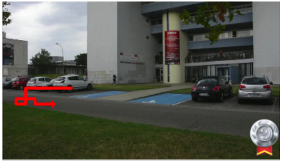
Results

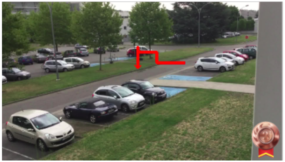


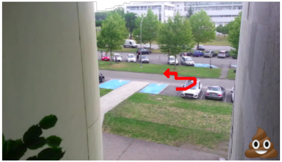


Results



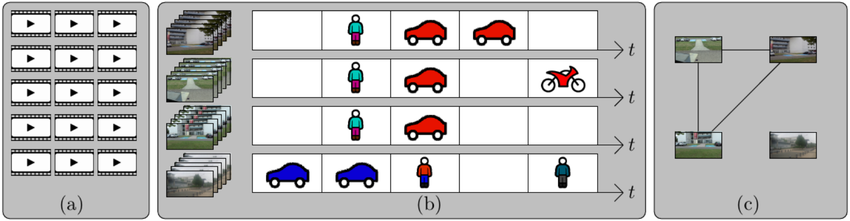




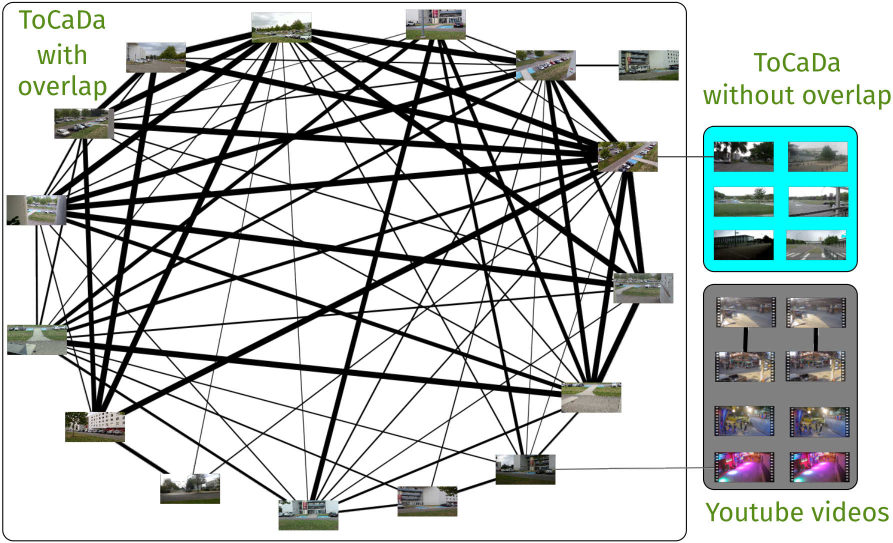


Story concept

Simultaneous occurrences of elements in the same category indicate an overlap between videos



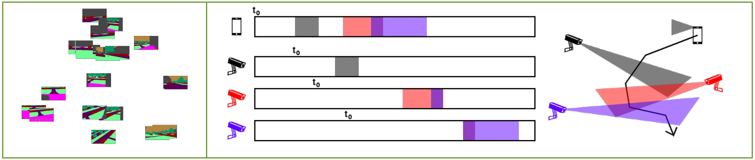
(a) (b) (c)


ToCaDa
with
overlap
ToCaDa
without overlap
Youtube videos

Contributions

- Reformulation of trajectory
- Ranking of videos based on interest
- Proposition of the story concept
- Graph of links between video
- **Without 3D reconstruction and re-identification**

Multi-video analysis: perspectives

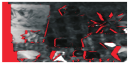


Multiple unsynchronised dynamic and static videos with or without overlap

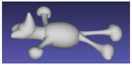
1. Dynamic video processing
2. Unsynchronised video processing
3. Event detection processing
4. Links between videos without overlap but with a dependency link
5. Taking into account the background scene

THANK YOU FOR YOUR ATTENTION ! QUESTIONS ?

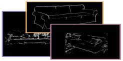
Macthing



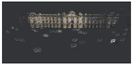
Skeleton



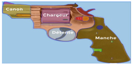
View quality



3D Model



Segmentation

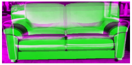


Planar superpixels

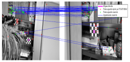
Multiple Videos

Faces

MobVille

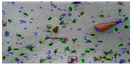


INVISO



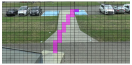
MARIO

VITI



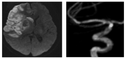
TRIMM

VICTORIA



Ceasefire

TARGET



References

S. Abdulwahab, H. Rashwan, J. Cristiano, S. Chambon and D. Puig.
Effective 2D/3D Registration using Curvilinear Saliency Features and Multi-Class SVM.
In International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISAPP, pages 354–361, 2019.

R. Amhaz, S. Chambon, J. Idier and V. Baltazart.
Automatic Crack Detection on Two-Dimensional Pavement Images: An Algorithm Based on Minimal Path Selection.
IEEE Transactions on Intelligent Transportation Systems, TITS, 17(10):2718–2729, 2016.

M. C. Bakkay, S. Chambon, C. Lubat and S. N. Barsotti.
Automatic detection of individual and touching moths from trap images by combining contour-based and region-based segmentation.
IET Computer Vision, 12(2):138–145, 2018.

M.-A. Bauda, S. Chambon, P. Gurdjos and V. Charvillat.
Geometry-based Superpixel Segmentation - Introduction of Planar Hypothesis for Superpixel Construction.
In International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISAPP, pages 227–232, 2015.

L. Breiman