Computer Science
Image and Interaction
Laboratory

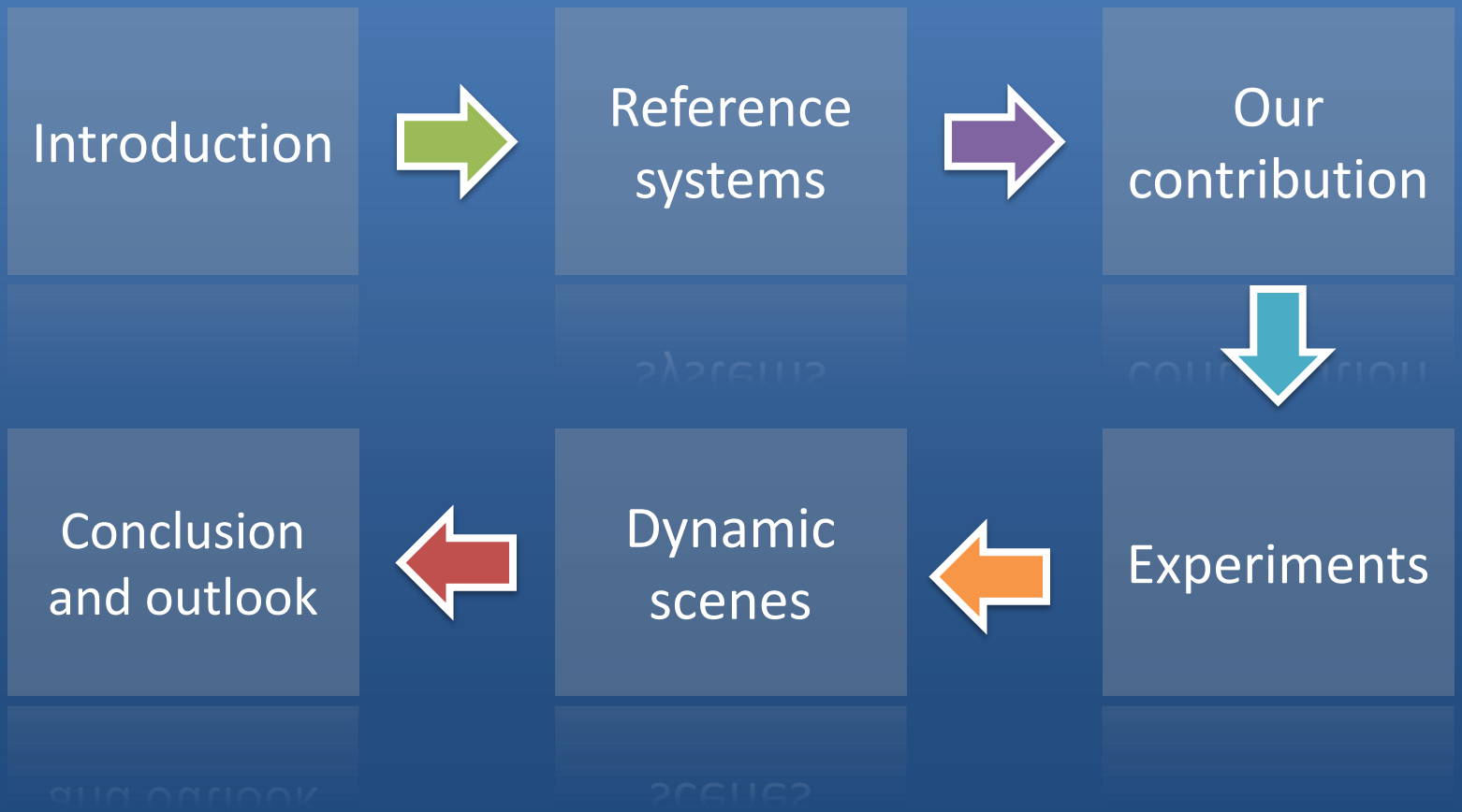
Real-time computational attention model for dynamic scenes analysis

Matthieu Perreira Da Silva – **Vincent Courboulay**

19/04/2012

Photonics Europe 2012 Symposium,
Brussels, 16-19 April

Overview


Introduction
Reference systems
Our contribution
Experiments
Dynamic scenes
Conclusion and outlook

Introduction

Reference systems

Our contribution

Conclusion and outlook

Dynamic scenes

Experiments

INTRODUCTION

What are we talking about ?

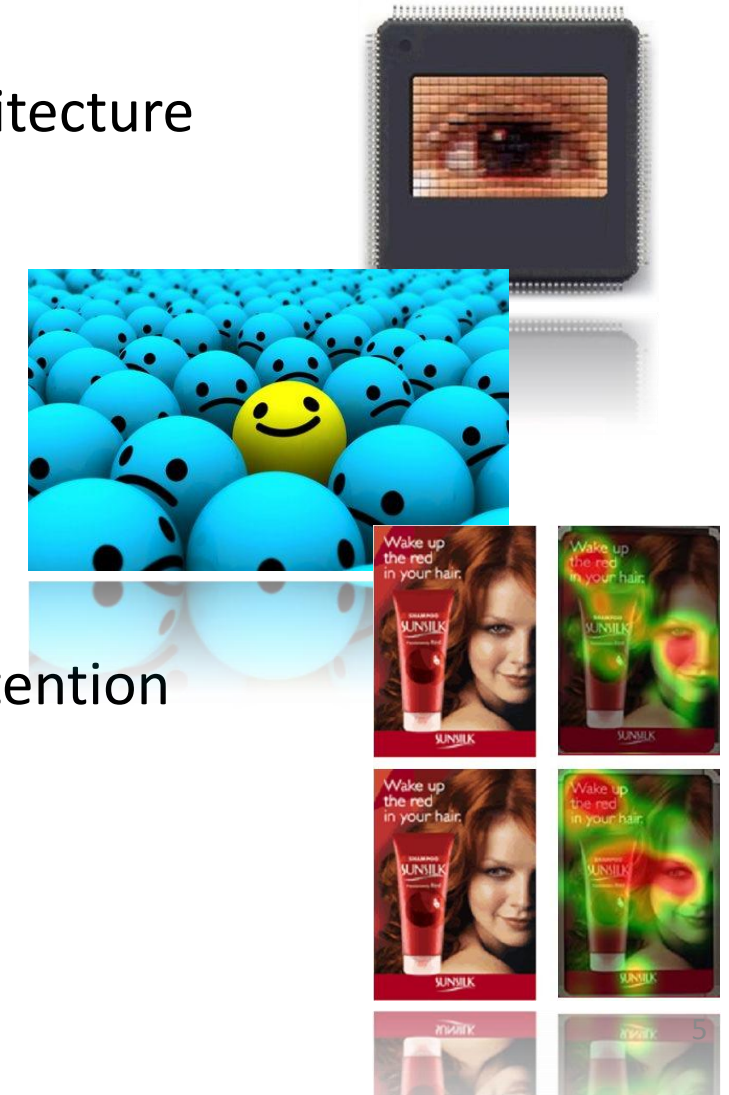
- What is *visual* attention ? A tool for
  - Selectively concentrating on one aspect of the *visual* environment while ignoring other ones
  - Allocating processing resources

- Links with saliency *maps*
  - Describes how important a part of the visual signal is
  - Some theory claim the existence of such a *map(s)* in our brain

- 2 types of visual attention
  - **Overt** : eye movement
  - Covert : mental focus
- Saccades vs fixations (cf previous pres)
- 2 types of attention *driving*
  - ***Bottom-up***
    - *stimulus based (involuntary)*
    - *Rarity / surprise / novelty*
  - *Top-down*
    - *goal directed (voluntary)*

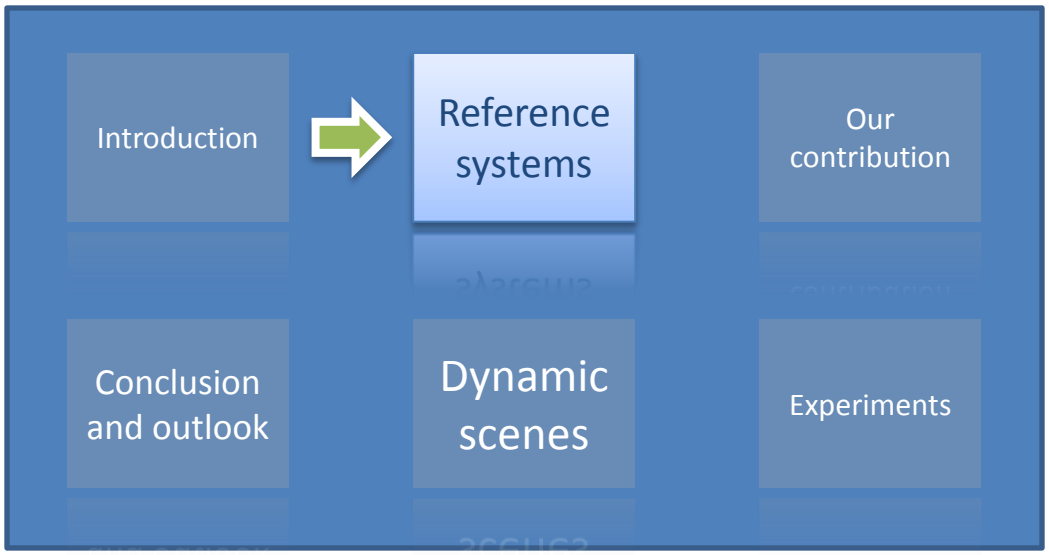
What for ?

- A building block for
  - Computer vision / robotics architecture
    - Smarter
      - Scene exploration
      - Resource allocation
  - Artificial intelligence
    - Detect novel / important data
    - Learn from it…
  - Understanding human visual attention
    - Replace eye-tracking
  - MM Applications (smart TV, …)



WHY ANOTHER MODEL ?

- Many visual attention models
  - **[Itti1998]**, [Ouerhani2003], [Tsotsos2005], [LeMeur2005], [Hamker2005], **[Frintrop2006]**, [Mancas2007],[Bruce2009] and others…
  - Cf. presentation of Mr Stentiford
- Usually
  - 1 model = 1 set of constraints / hypothesis
- In our case
  - Real time
  - Image and video
  - Focus points (no saliency map)
  - **Dynamical results**



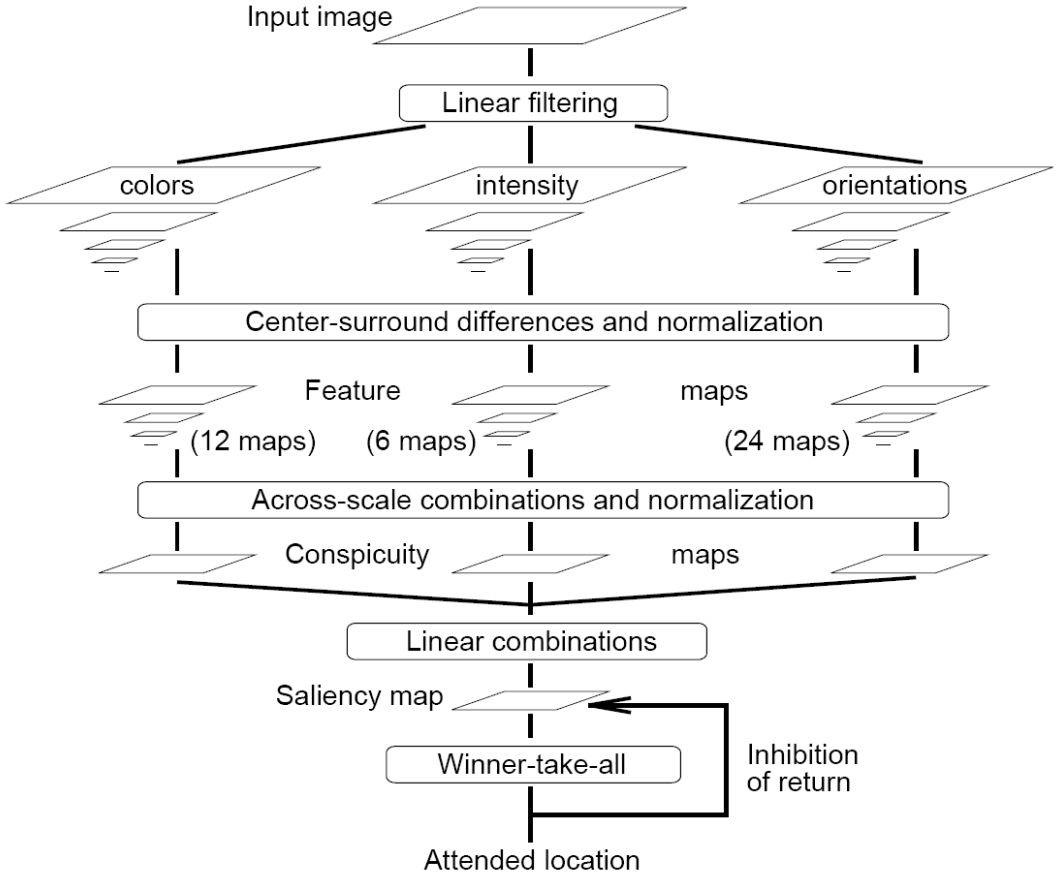
REFERENCE SYSTEMS

L.Itti's original architecture

- Well known attention model (1998)
- Open source implementation
- Biologically inspired
- Quite fast

But
- normalization,
- fusion
- no dynamic in simulation

One more time the famous Itti Architecture

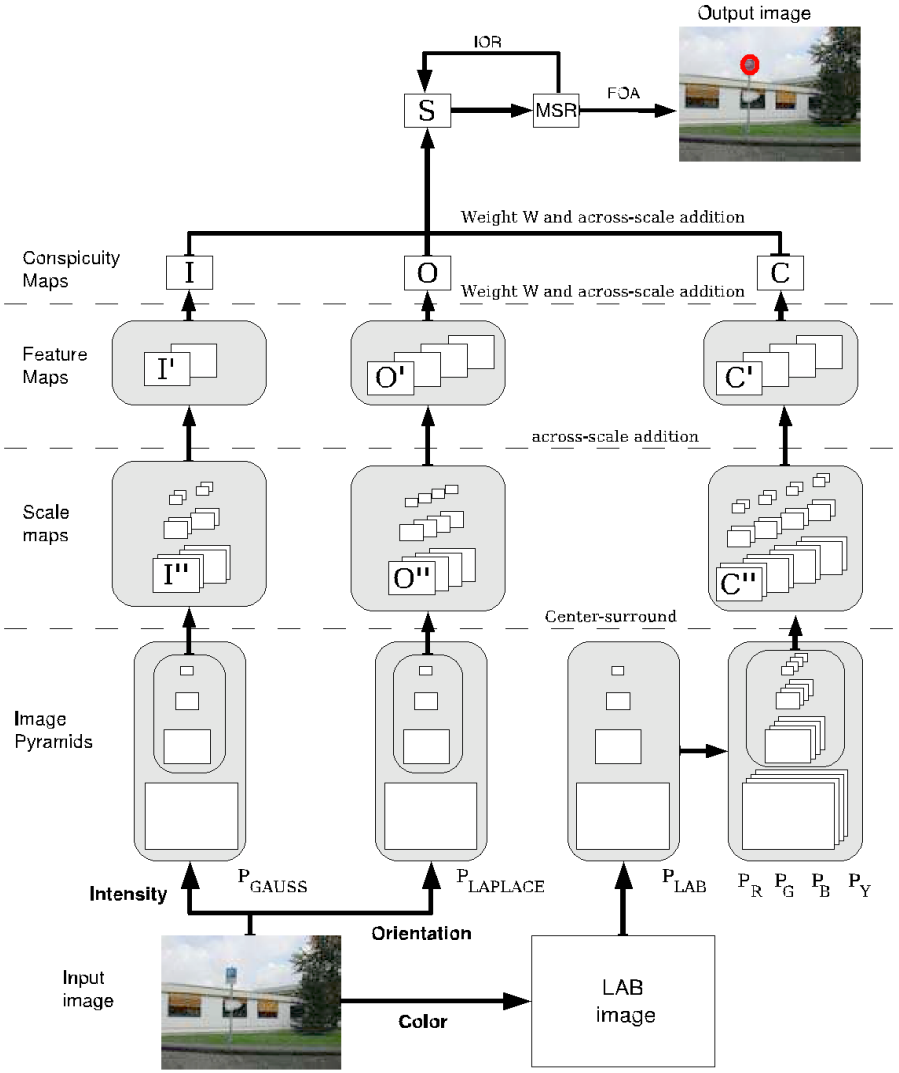


S. Frintrop improvements (VOCUS)

- Almost the same architecture
- Better normalization operator
- Better center-surround “filtering”

(a) Intensity map of NVT (b) Intensity map of VOCUS

- Faster



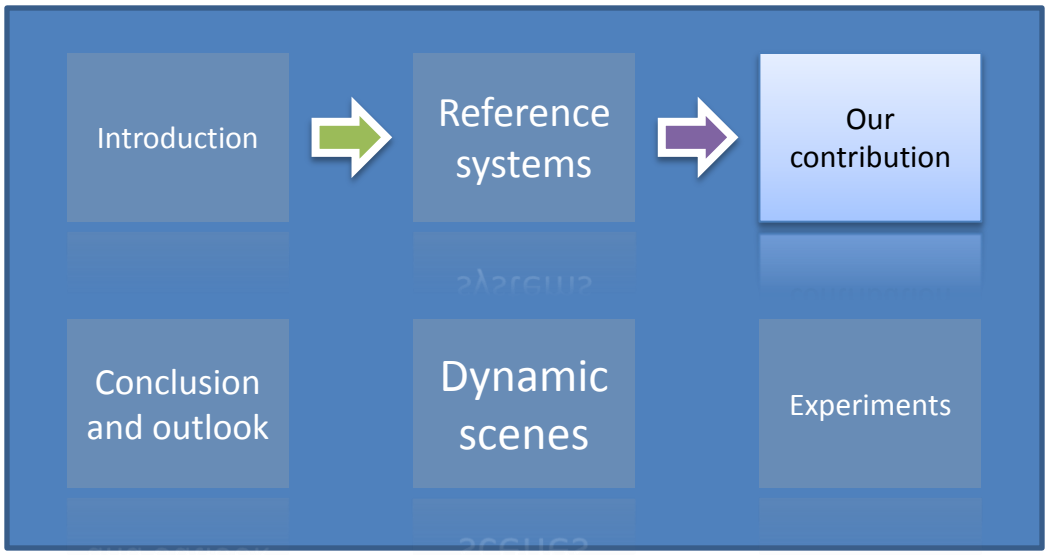
What can we do next ?

- Better conspicuity maps *fusion*

  - Normalization + linear combination are difficult to adjust in the absence of prior knowledge

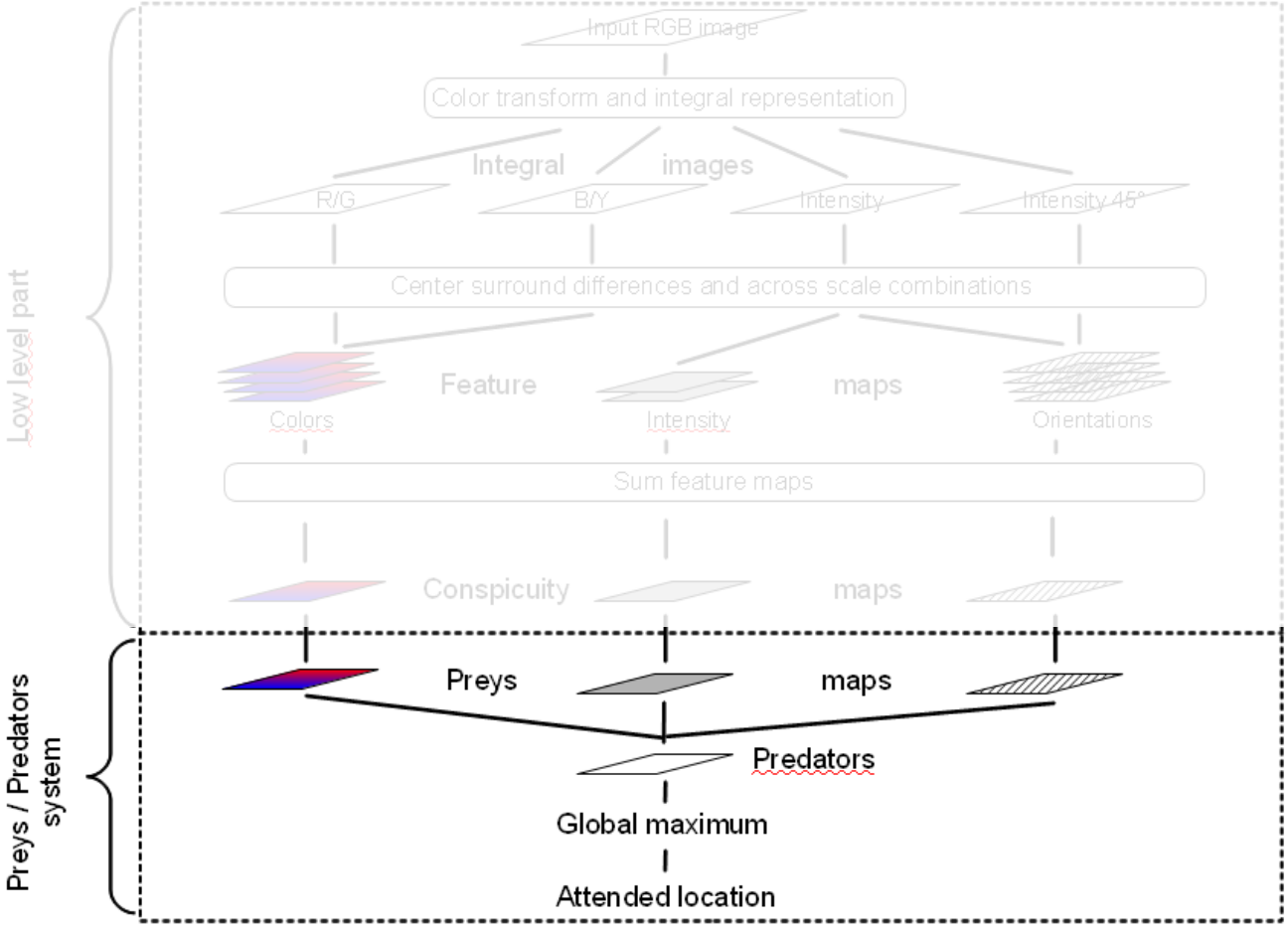
  - Maps fusion is a *competition* between different information to gain attention… why not using *"existing"* preys / predators models ?

- Dynamical scene analysis



OUR CONTRIBUTION

MODIFIED ARCHITECTURE



4 integral images

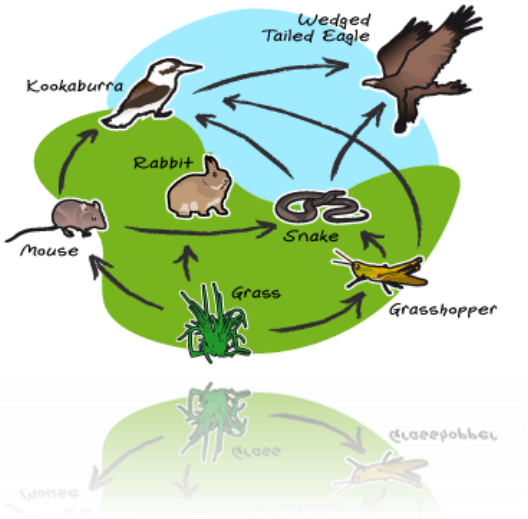
10 feature maps

3 conspicuity maps

Why preys / predators system for conspicuity maps fusion ?

- Dynamical system
  - Time evolution is intrinsically handled
  - Visual attention focus (max of predators population) can evolve dynamically

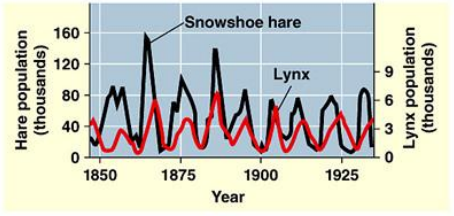
- Competition as a “default” fusion strategy
  - Different types of information to mix
  - Hard to find a good default fusion strategy
    - No top down information or pregnancy
  - Natural equilibrium



- Chaotic behavior
  - Comes from discrete dynamic systems
  - Usually not a wanted property, but…
  - Allows emergence of original exploration path even in non salient area
  - Curiosity !

How do preys / predators systems work ?

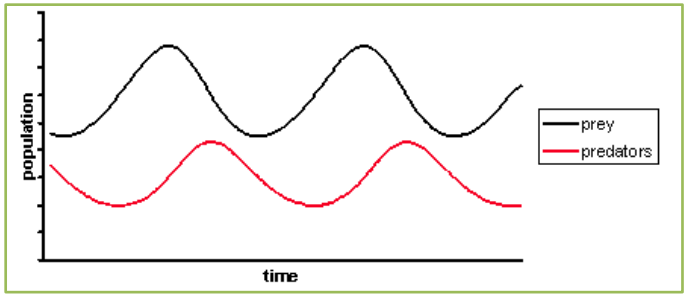
- Equations proposed independently by V. Volterra and A. J. Lotka in the 1920's.
- first-order, non-linear, differential equations
- describe the dynamics of biological systems in which two species interact
- Used originally to model fish catches in the Adriatic



- In it's simplest form : $\frac{dx}{dt} = x(\alpha - \beta y)$ and $\frac{dy}{dt} = -y(\gamma - \delta x)$

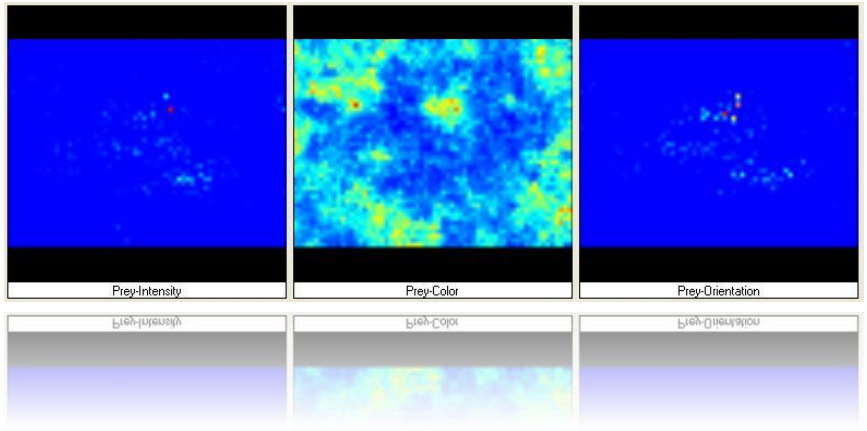
- Where
  - x is the number of preys
  - y is the number of predators
  - α is the prey's birth rate (exponential growth)
  - β is the predation rate
  - γ is the predators natural death rate
  - δ is the predators growth rate (linked to predation)

- In theory, solutions to the equations are periodic



Transposing preys / predators systems to visual attention

- General features
  - 2D preys / predators system (maps)
  - Preys and predators can move (diffusion)



- Metaphor
  - The system is comprised of
    - 3 types of resources
    - 3 types of preys
    - 1 type of predators
  - **Preys** represent the spatial distribution of **curiosity** generated by the 3 types of resources (conspicuity maps) : intensity, color and orientation
  - **Predators** represent the **interest** generated by the consumption of curiosity (preys)
  - The **global maximum of the predators** map (interest) is the **focus of attention** at time t

Our preys / predators systems equations

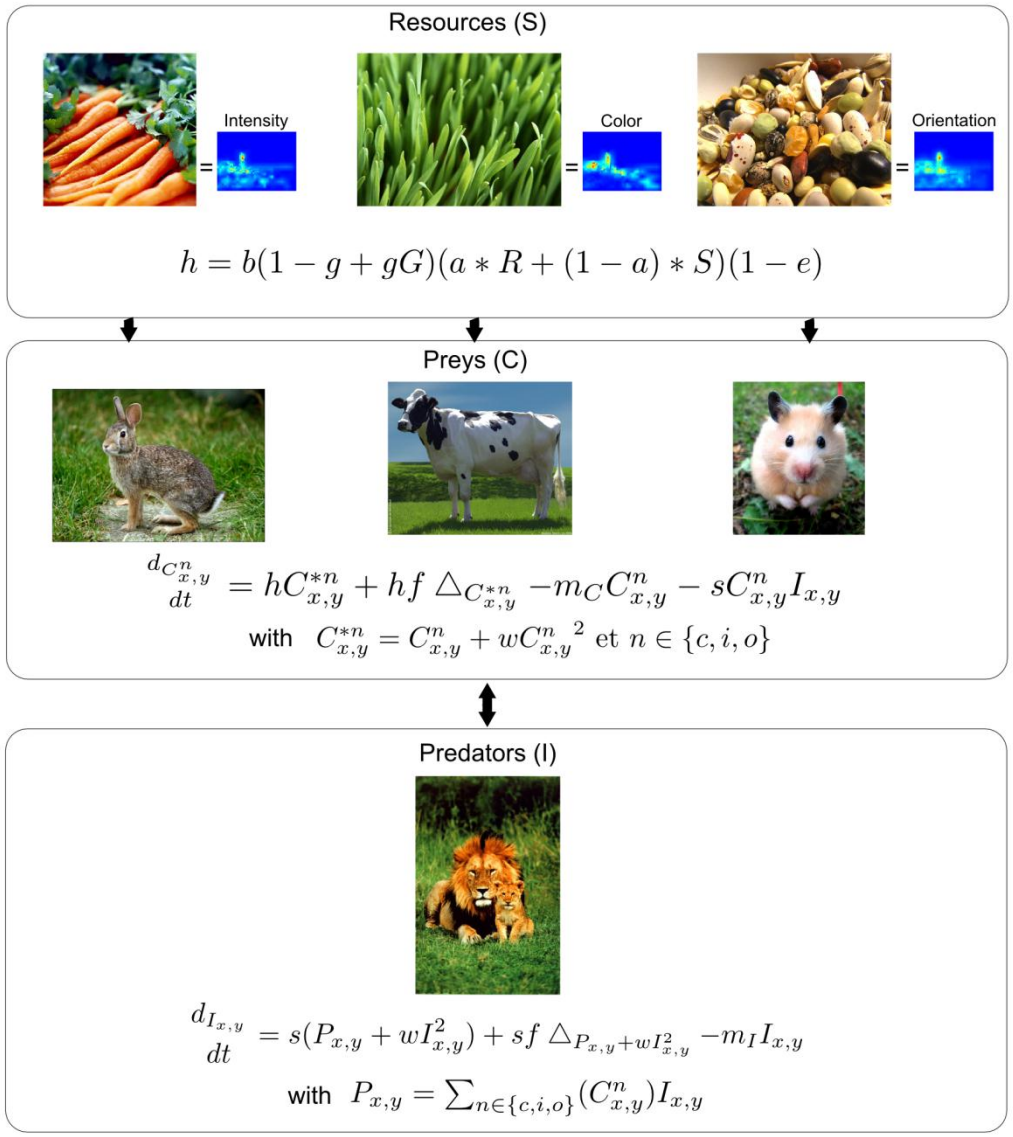
C: preys (curiosity)
I: predators (interest)

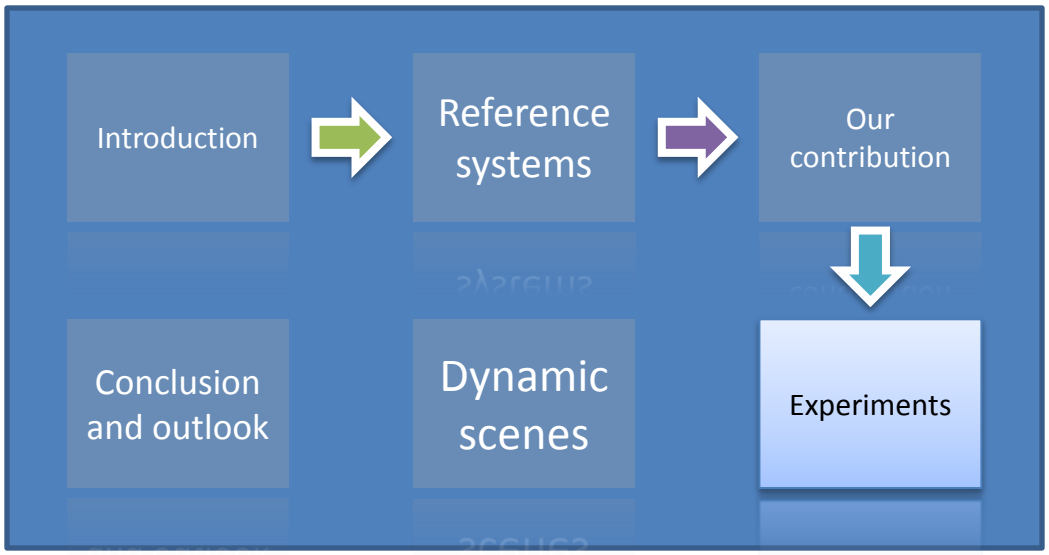
S: Image conspicuity
G: Gaussian map
R: random map
e: entropy of the conspicuity map

h: preys birth rate
b: preys growth factor (0.005)
m_c: preys death factor
g: central bias factor (0.1)
a: randomness factor (0.3)

f: diffusion factor (0.2)
w: quadratic term (0.001)

s: predation / predators growth factor
m_i: predators death factor

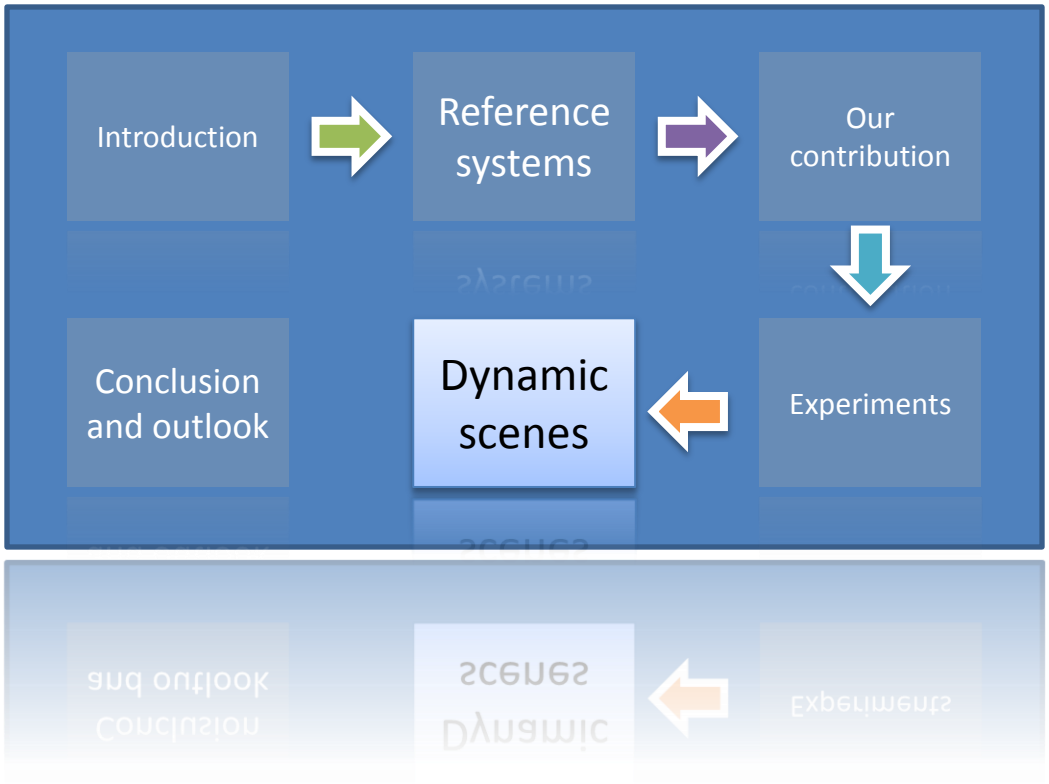




EXPERIMENTS

EXPERIMENTS

- Validation of subjective and objectives evaluation
  - *Evaluation of preys / predators systems for visual attention simulation*, in [VISAPP 2010 - International Conference on Computer Vision Theory and Applications, 275-282, INSTICC, Angers (2010).
  - *Objective Validation Of A Dynamical And Plausible Computational Model Of Visual Attention*, in IEEE European workshop on visual information processing, France (2011).
  - *Image Complexity Measure Based On Visual Attention*, in IEEE ICIP, 3342-3345 (2011).



DYNAMIC SCENES

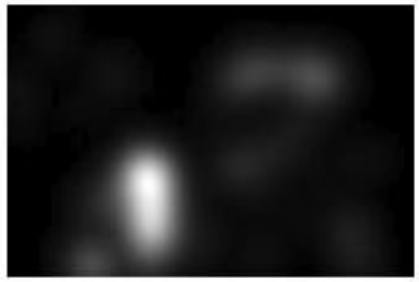
WHAT ABOUT DYNAMIC ?

- Top-Down feedback & adaptation mechanisms
  - global weighting of feature maps: allows a bias of the attentional system in favor of the distinctive features of a target object
  - local weighting of feature maps: allows specifying prior knowledge about the target localization

Original image

(a)

(b)



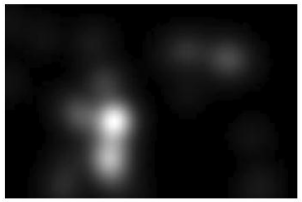
(c)

a) heatmap generated with default parameters, b) heatmap generated with lower color weights,c) heatmap generated with high color weight

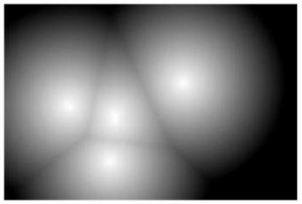
WHAT ABOUT DYNAMIC ?

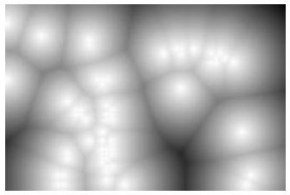
- Scene exploration: different scenario
  - scene exploration maximization : the attentional system will favor unvisited areas;
  - localization stability : the attentional system will favor already visited areas

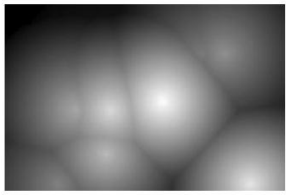
(a) Original image

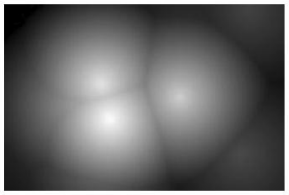


(b) *Heatmap.*









(c) Visited areas, $F_{forget} = 1.0$.

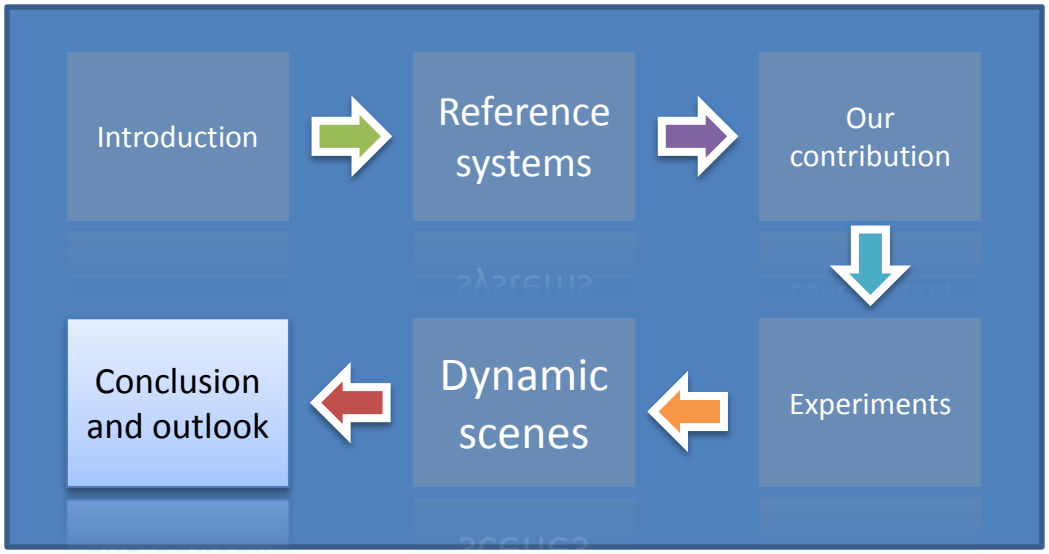
(d) Visited areas, $F_{forget} = 0.95$

(e) Visited areas, $F_{forget} = 0.90$.

DEMONSTRATION

- Let’s try a little demo …

- Please start a little prey for me :o)



CONCLUSION AND OUTLOOK

Conclusion

- Better conspicuity maps *fusion*

  - Normalization + linear combination are difficult to adjust in the absence of prior knowledge

  - Maps fusion is a *competition* between different information to gain attention… why not using *"existing"* preys / predators models ?

- Dynamical scene analysis

Conclusion

- Fast but efficient conspicuity maps generation
- Dynamical systems based conspicuity maps fusion architecture
  - Generates real time attentional focus
  - Seems efficient
- A validated method
- Dynamic may be included in several ways

OUTLOOK

- Ongoing projects
  - Integrate complex feature and conspicuity maps
  - Integrate top-down maps
  - Video streams and depth flow
    - Working but not evaluated yet
- Outlook
  - *Attention as a decision in information space* – [J.Gottlieb2010]


Thank you for your attention
vincent.courboulay@univ-lr.fr