Face perception: Influence of location and number in videos

Anis Rahman Denis Pellerin Dominique Houzet

GIPSA-lab, UMR 5216, Grenoble, France

May 25, 2012

Outline

Introduction

Faces are special

- Specialized face-selective regions
  - Faces are coded in a specific cortical area of the brain; fusiform face area (FFA).
  - Face processing is fast with neuronal responses as low as 70ms.
- Inherent preference for faces linked to exposure
  - Specific pattern with distinct regions
  - Social and biological importance

Motivation

- To study the influence of faces during free viewing of videos
- To study the impact of location and number of faces for videos
  - To help improve visual saliency model

Eye position experiment

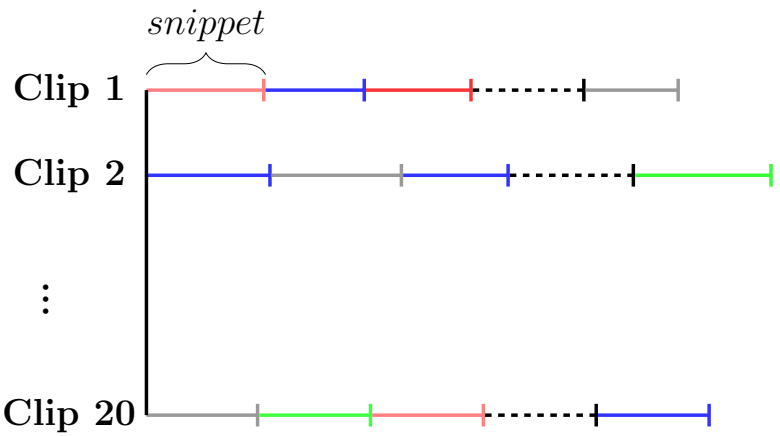
Designed to record eye movements for video stimuli that are then used to investigate the participants behavior.

- 53 videos from different sources i.e. movies, tv shows and news, animated movies, commercials, sports, music concerts etc.
- Each video is decomposed into extracts of 1-3s, resulting in 305 snippets[1]
- Total of 14000 frames of gray level stimuli (720×576 pixels per frame)
- 15 participants volunteered for the experiment

[1] R Carmi and L Itti. "Visual causes versus correlates of attentional selection in dynamic scenes". In: *Vision Res* 46.26 (2006), pp. 4333 –4345. ISSN: 0042-6989.

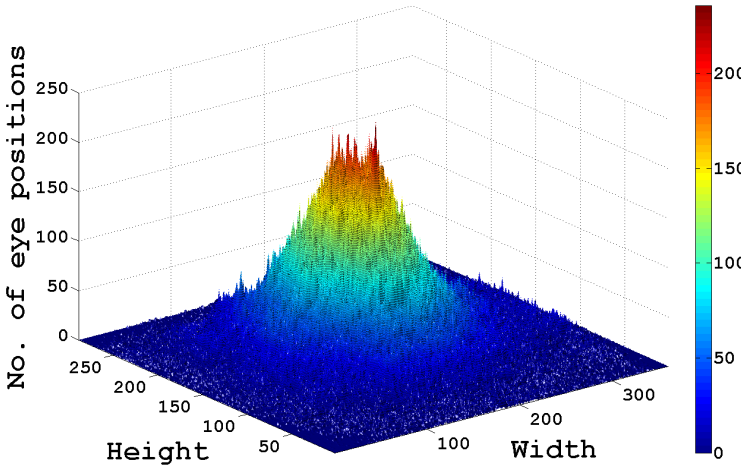
Video stimuli

- The snippets are strung together into 20 clips of 30s
- Eye-link II tracker is used to record eye movements of participants with FOV $40° \times 30°$

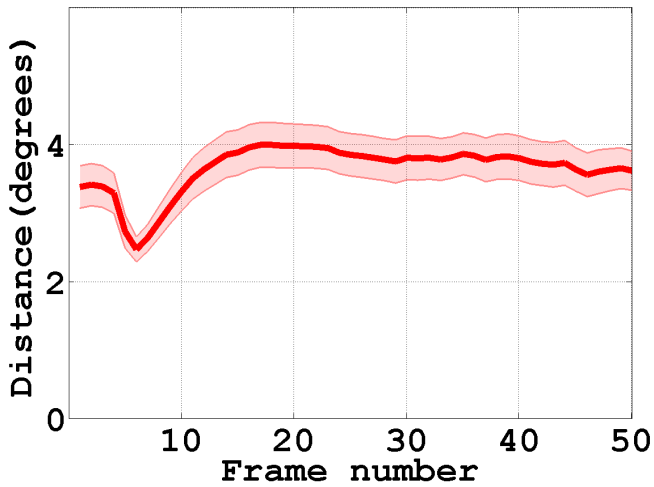


Video stimuli for the experiment

Experimental eye positions



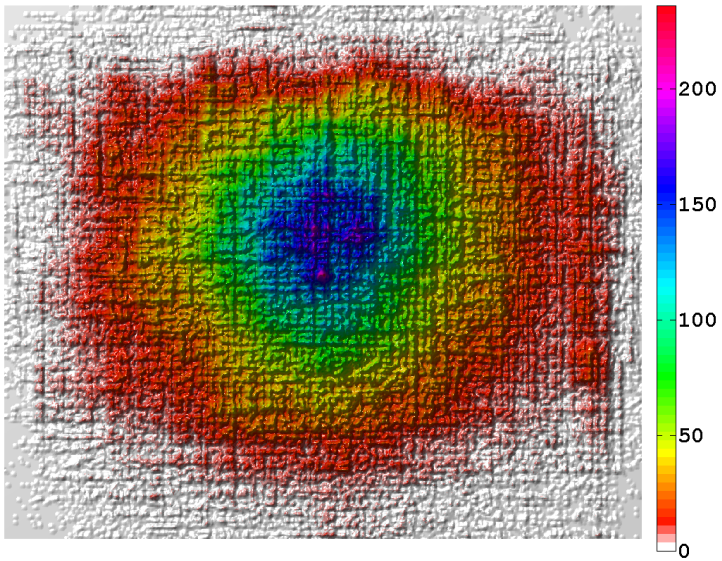
(a) Density map of the recorded eye-positions



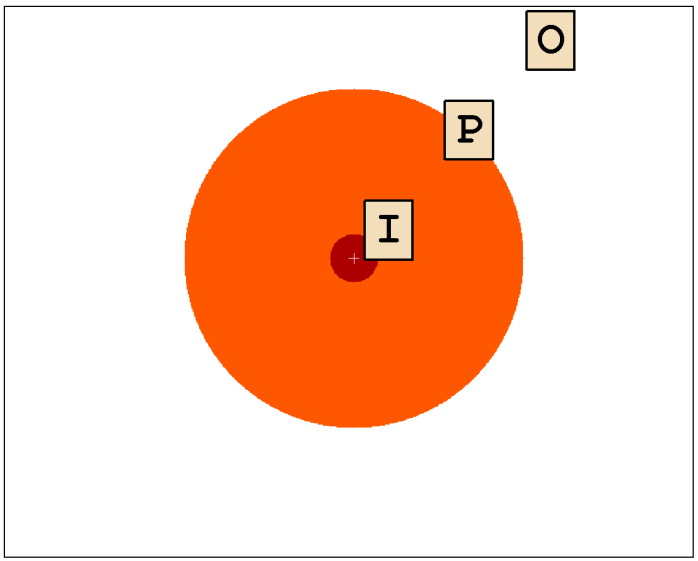
(b) Distance of eye positions from center

Definition of locations

- We define three locations around the center,
  - **Inside (I)** upto 2°
  - **Periphery (P)** 2° to 14°
  - **Outside (O)** 14° and above



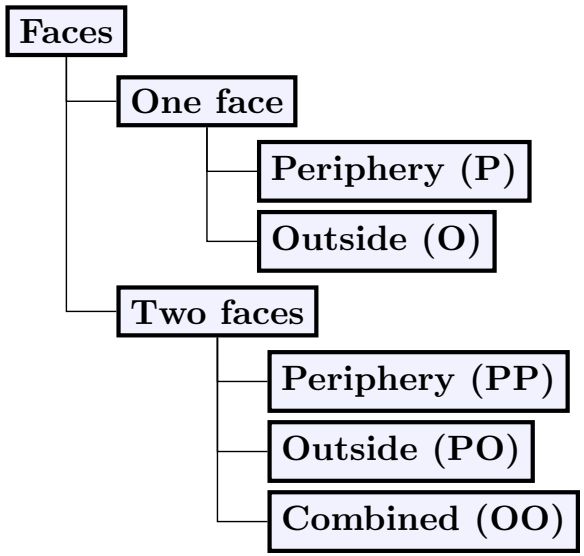
(c) Density map of the recorded eye-positions



(d) Definition of locations

Labeled face data

- Faces are hand-labeled for all video clips
- Total ∼18000 faces
- Discarded large faces i.e. greater than $(R_{periphery} - R_{inside})$



Different categories of face frames based on location and number

	One face		Two faces		
	P	O	PP	PO	OO
# of samples	2089	852	450	430	395

No. of samples or frames for different categories of number and location

Evaluation criterion

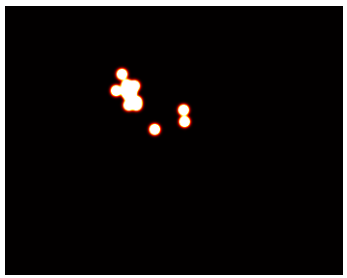
- **NSS:** Normalized Saliency Scanpath[2]

$$NSS_{(i,j)} = \frac{M_f \cdot M_h - \bar{x}_f}{s_f}$$

M_f : face map with 2D gaussians on labeled faces
M_h : human eye position density map
$\bar{x}_f$: empirical mean of saliency map M_m
s_f : empirical standard deviation of saliency map M_m

(a) Image frame

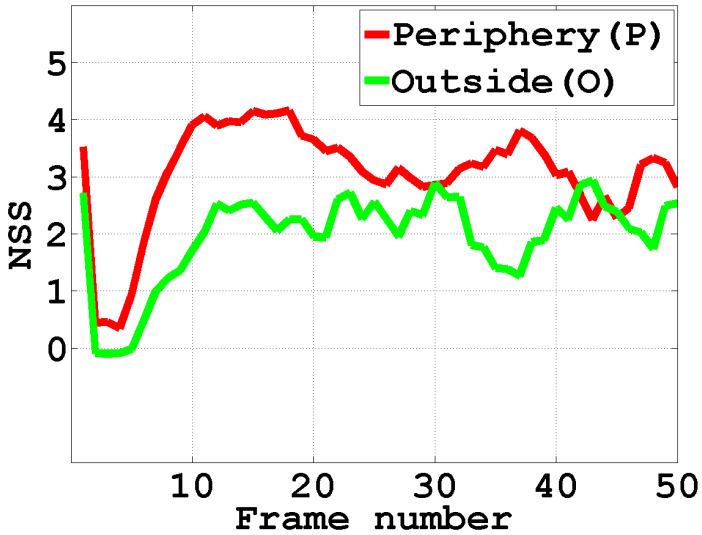
(b) Face map M_f



(c) Human eye position density map M_h

[2] R J Peters et al. "Components of bottom-up gaze allocation in natural images". In: *Vision Res* 45 (2005), pp. 2397–2416.

One face

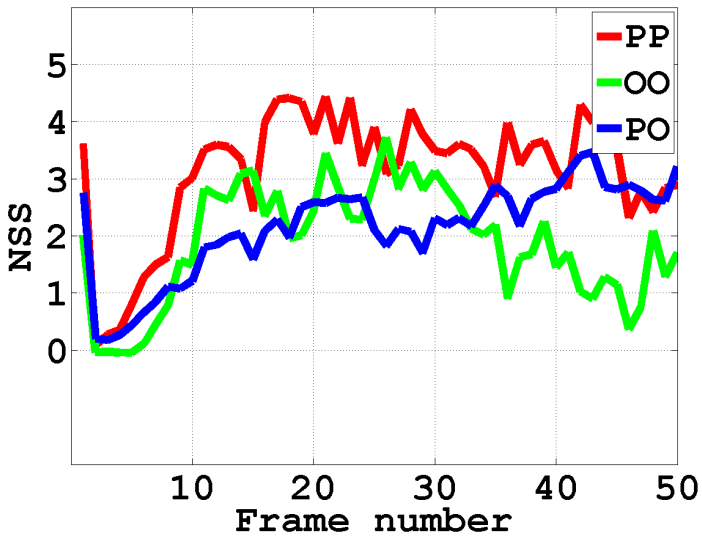


NSS evolution

		Periphery	Outside
NSS	$\bar{x}$	**3.72**	1.96
	$SE_{\bar{x}}$	0.160	0.187

Scores for one face; mean values for 'first fixation' (frames 8 to 16)

Two faces

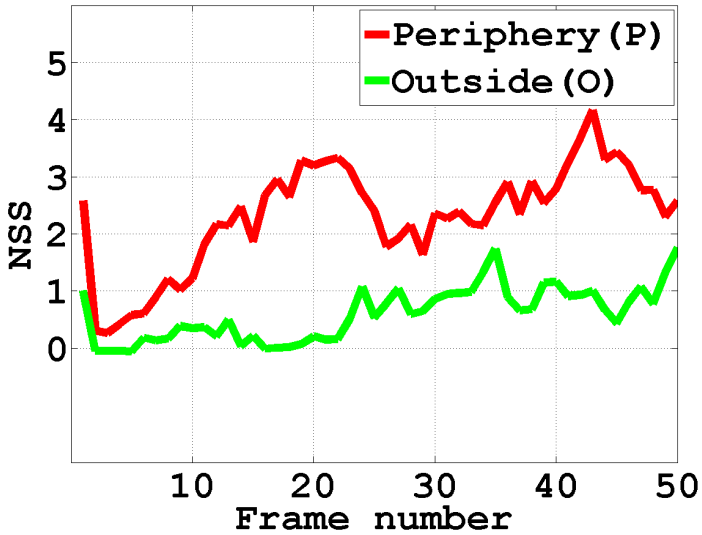


NSS evolution

		PP	OO	PO
NSS	$\bar{x}$	**2.94**	2.10	1.56
	$SE_{\bar{x}}$	0.268	0.306	0.145

Scores for two faces; mean values for 'first fixation' (frames 8 to 16)

Two faces: Mixed locations (PO)



NSS evolution

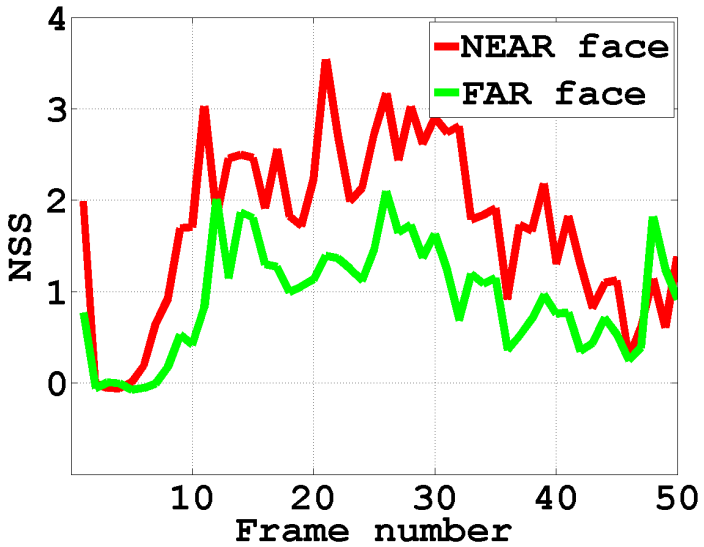
		P	O
NSS	$\bar{x}$	**1.47**	0.06
	$SE_{\bar{x}}$	0.095	0.063

Scores for PO case for 'first fixation' (frames 8 to 16)

Two faces: Outside location (OO)

The pair of faces in outside location are divided into near and far faces based on their distance from center.

$$d_f = \sqrt{(c_x - f_x)^2 + (c_y - f_y)^2}$$



NSS evolution

		NEAR	FAR
NSS	$\bar{x}$	**1.91**	1.01
	$SE_{\bar{x}}$	0.232	0.231

Scores for NEAR and FAR faces in outside location for 'first fixation' (frames 8 to 16)

Results summarized

	One face		Two faces				
	P	O	PP	PO		OO	
NSS	3.72	1.96	2.94	1.56		2.10	
				P	O	NEAR	FAR
				1.47	0.06	1.91	1.01

Summary of NSS scores for different cases of face number and location for 'first fixation' (frames 8 to 16)

Conclusion and Perspectives

Conclusion

- Impact of face decreases with increasing eccentricity
- Impact of face decreases with increasing number of faces
- Crucial to understand the influence of complex objects on eye movements for videos

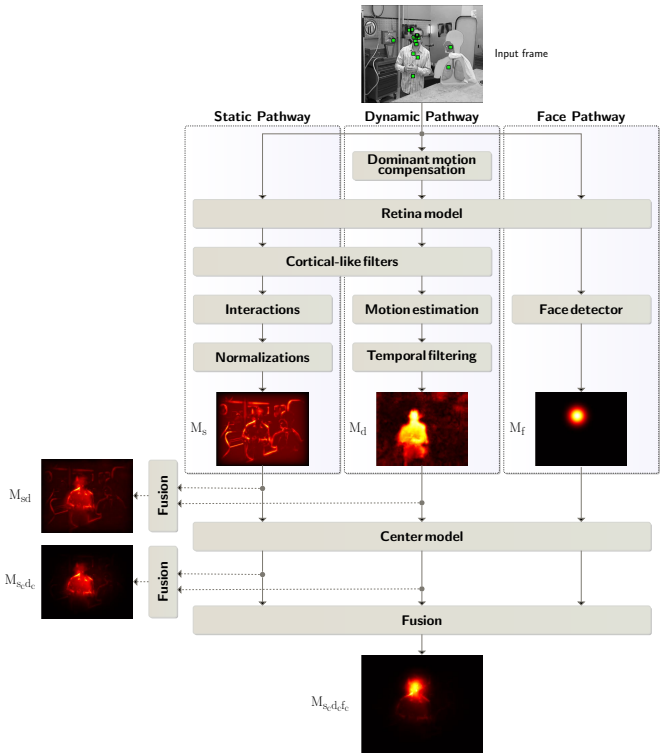
Perspectives

- Incorporate into visual saliency model
- Evaluate against other databases

Thank you for your attention!

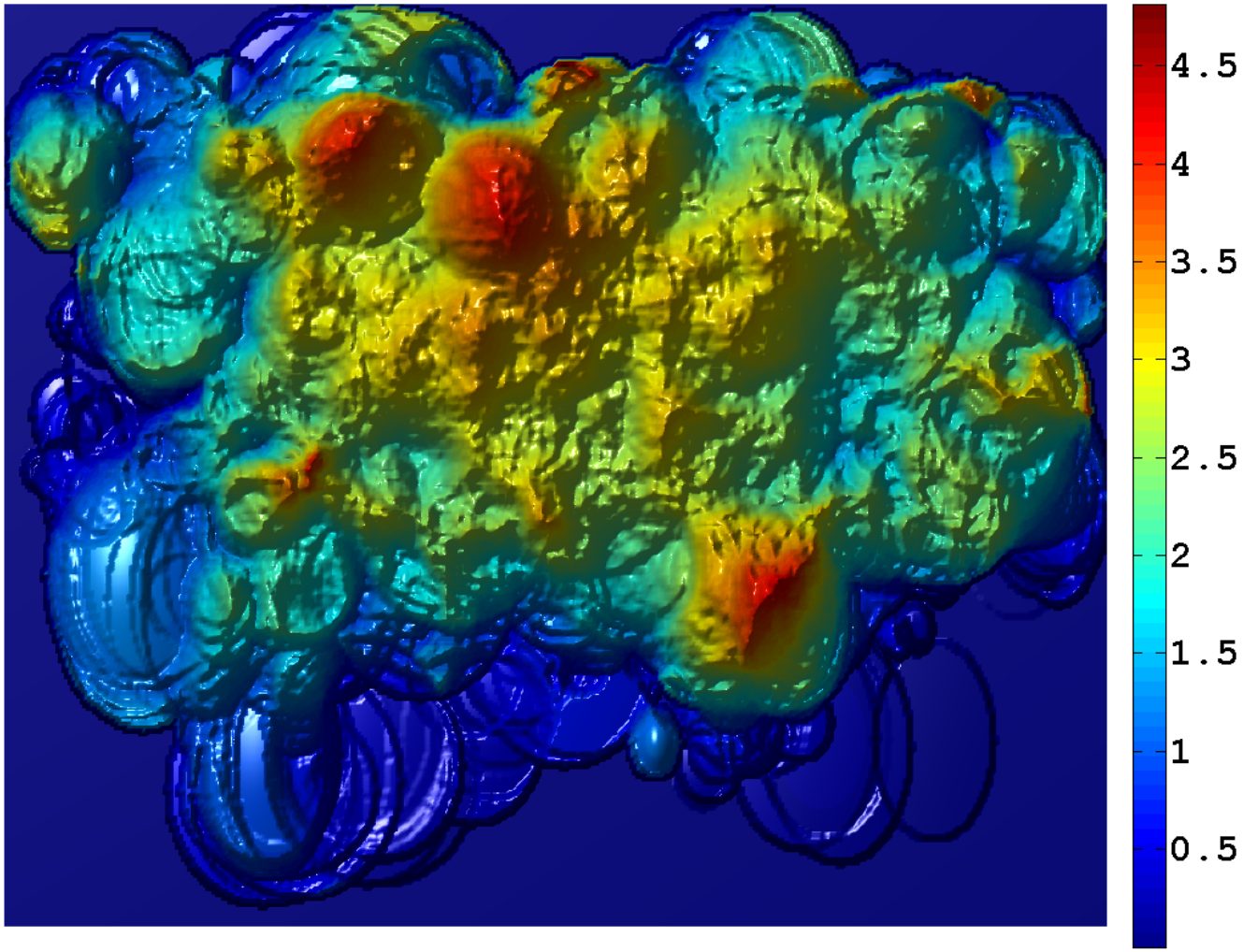
Visual saliency model

- Type of visual attention model[3]
- To find the spotlight of focus
- Based on human visual system
- Bottom-up model
- Implements three pathways



[3] S. Marat et al. "Improving visual saliency by adding 'face feature map' and 'center bias'". In: *Cognitive Computation* (2012). DOI: 10.1007/s12559-012-9146-3.

NSS Scores



Surface map with NSS scores for all faces