

# Thèse de Doctorat

## Junle Wang

*Mémoire présenté en vue de l'obtention du  
grade de Docteur de l'Université de Nantes  
sous le label de l'Université de Nantes Angers Le Mans*

**Discipline : Automatique et Informatique Appliquée**

**Spécialité : Traitement du signal et des images**

**Laboratoire : Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN)**

**Soutenue le 16 Novembre 2012**

**École doctorale : STIM**

**Thèse n° : ED 503-181**

## From 2D to Stereoscopic-3D Visual Saliency: Revisiting Psychophysical Methods and Computational Modeling

Saillance Visuelle, de la 2D à la 3D Stéréoscopique : Examen des  
Méthodes Psychophysique et Modélisation Computationnelle  
Vol 2: Résumé étendu en Français

### JURY

Rapporteurs : **M<sup>me</sup> Anne GUÉRIN DUGUÉ**, Professeur, Université Joseph Fourier  
**M<sup>me</sup> Ingrid HEYNDERICKX**, Professeur, Delft University of Technology

Examineurs : **M<sup>me</sup> Luce MORIN**, Professeur, Institut National des Sciences Appliquées  
**M. Frédéric DUFAUX**, Directeur de recherche CNRS, TELECOM ParisTech  
**M. Vincent RICORDEL**, Maître de conférences, Université de Nantes  
**M. Patrick LE CALLET**, Professeur, Université de Nantes

Directeur de thèse : **M. Patrick LE CALLET**, Professeur, Université de Nantes

Co-encadrant de thèse : **M. Vincent RICORDEL**, Maître de conférences, Université de Nantes



# Contents

<b>Contents</b>	<b>i</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Contexte . . . . .	1
1.2 Motivations . . . . .	1
1.3 Structure de la thèse . . . . .	3
<b>Vérités de terrain de l'attention visuelle</b>	<b>5</b>
<b>2 Une étude comparative des cartes de densité de fixation obtenues à partir de différentes expériences oculométriques</b>	<b>7</b>
2.1 Expériences avec l'oculomètre . . . . .	7
2.1.1 Stimuli . . . . .	8
2.1.2 Comparaison des procédures expérimentales . . . . .	8
2.2 Comparaison inter laboratoires . . . . .	9
2.2.1 Différences inter laboratoires . . . . .	9
2.2.2 Impact du point de fixation central sur le biais . . . . .	9
2.2.3 Dépendance au contenu . . . . .	13
2.3 Applications . . . . .	13
2.3.1 Modèles de saillance visuelle . . . . .	13
2.3.2 Evaluation de la qualité d'image . . . . .	15
2.3.3 Reciblage d'image basé sur la saillance . . . . .	16
2.4 Perspectives . . . . .	17
<b>3 Relation entre salience et importance visuelle</b>	<b>19</b>
3.1 Introduction . . . . .	19
3.2 Méthodes . . . . .	19
3.2.1 Expérience I : Importance visuelle . . . . .	19
3.2.2 Expérience II : Saillance visuelle . . . . .	20

3.3	Résultats et analyse . . . . .	20
3.3.1	Observations qualitatives des cartes d'importance et des cartes de saillance . . . . .	20
3.3.2	Prédire le sujet principal, les objets secondaires, et l'arrière plan .	23
3.3.3	Analyse temporelle . . . . .	24
3.4	Discussion et conclusion . . . . .	25
<b>Attention visuelle en 3D</b>		<b>27</b>
<b>4</b>	<b>Perception de profondeur et contenu 3D stéréoscopique</b>	<b>29</b>
4.1	Introduction . . . . .	29
4.1.1	Perception de la profondeur . . . . .	29
4.1.2	Conflits en vision stéréoscopique . . . . .	32
4.2	Quantifier comment la combinaison du flou et de la disparité affecte la profondeur perçue . . . . .	33
4.2.1	Expérimentation . . . . .	33
4.2.2	Résultats et analyse . . . . .	35
4.2.3	Conclusion . . . . .	35
<b>5</b>	<b>Influence de la profondeur sur l'attention visuelle 3D : une étude du biais de profondeur</b>	<b>39</b>
5.1	Introduction . . . . .	39
5.2	Expérimentation . . . . .	40
5.2.1	Participants . . . . .	40
5.2.2	Conditions d'affichage . . . . .	40
5.2.3	Stimuli . . . . .	40
5.2.4	Post-traitement des données de suivi oculaire . . . . .	43
5.3	Résultats . . . . .	43
5.3.1	Distribution des fixations en profondeur . . . . .	43
5.3.2	Variation de la profondeur de fixation en fonction de l'ordre temporel de fixation . . . . .	46
5.3.3	Dépendance temporelle de la distribution de la fixation en profondeur . . . . .	46
5.4	Discussion . . . . .	46
<b>6</b>	<b>Modèle de calcul de l'attention visuelle 3D</b>	<b>51</b>
6.1	Introduction . . . . .	51
6.2	Un modèle de calcul de l'attention visuelle 3D basé saillance de profondeur	52
6.2.1	Une approche bayésienne pour la génération d'une carte de saillance de profondeur . . . . .	52
6.2.1.1	Extraction des caractéristiques de profondeur . . . . .	53

---

6.2.1.2	Modélisation de la distribution de probabilité . . . . .	54
6.2.2	Un schéma pour le modèle calculatoire de l'attention visuelle 3D .	54
6.3	Une base de données oculométriques de contenu d'images naturelles 3D .	56
6.3.1	Stimuli . . . . .	56
6.3.2	Appareils et procédures . . . . .	56
6.3.3	Participants . . . . .	56
6.3.4	Création des cartes de densité de fixation . . . . .	56
6.4	Evaluation des performances . . . . .	59
6.4.1	Evaluation qualitative . . . . .	59
6.4.2	Métriques quantitatives d'évaluation . . . . .	59
6.4.3	Performance la carte de saillance de profondeur seule . . . . .	59
6.4.4	Valeur ajoutée de la carte de saillance de profondeur . . . . .	61
6.5	Conclusion et discussion . . . . .	62
<b>7</b>	<b>Conclusion et perspectives</b>	<b>63</b>
7.1	Résumé et contributions . . . . .	63
7.2	Limites et perspectives . . . . .	65
	<b>Bibliographie</b>	<b>67</b>



# Chapitre 1

## Introduction

### 1.1 Contexte

Dans notre vie quotidienne, nous recevons constamment une quantité abondante d'informations sollicitant nos différents sens. Parmi ces derniers, la vue est considérée comme dominante car elle est comparativement plus développée que les autres sens, comme par exemple l'odorat et le goût [Wandell 95]. Notre système sensoriel dédié à la vision, le système visuel humain (SVH), reçoit en permanence une quantité très importante de données visuelles ( $10^8 - 10^9$  bits par seconde) [Borji 12]. Cette masse de données est très importante et au-delà de la capacité de notre cerveau s'il devait en traiter l'ensemble. Pour faire face à cette masse d'informations, l'attention visuelle est l'un des plus importants mécanismes déployés par le SVH pour réduire la complexité de l'analyse de la scène visuelle [Wolfe 00]. Conduits par l'attention visuelle, les téléspectateurs peuvent de façon sélective se concentrer sur les zones spécifiques d'intérêt de la scène.

Durant les dernières décennies, des efforts considérables ont été consacrés à l'étude de l'attention visuelle. Des neurologues, des psychologues, des scientifiques spécialistes de la vision, et des informaticiens ont pris part et ont contribué à divers aspects de l'étude de l'attention visuelle. Ces efforts dans différentes disciplines ont fait que la recherche dans le domaine de l'attention visuelle est devenue très interdisciplinaire. Ces différentes disciplines du domaine oeuvrent donc à partir de différents points de vue, et chacune pour le profit de l'autre.

### 1.2 Motivations

Ces dernières années, la mise en oeuvre du mécanisme de l'attention visuelle par des systèmes de traitement d'image a eu un intérêt croissant. Prendre en compte les informations liées à l'attention visuelle devient un moyen efficace pour améliorer différents algorithmes existants en traitement d'image. Une large variété de domaines est concer-

née. Par exemple des applications de compression [Park 02], de reciblage [Wang 11a], de recherche d'images [Vu 03], ou d'évaluation de la qualité [Liu 11], ont tiré bénéfice de l'information de l'emplacement de la scène visuelle qui attire l'attention du spectateur.

Jusqu'à présent, de nombreuses études liées à l'attention visuelle s'appuient sur les cartes de densité de fixation (FDM). Il y a deux raisons à cela : (1) on estime qu'il y a un lien évident entre attention visuelle et mouvements des yeux, et (2) les FDM sont obtenues à partir d'expériences oculométriques et sont donc considérées comme étant une vérité de terrain fiable. Cependant, les expériences oculométriques ont été menées de manière indépendante dans différents laboratoires de différents pays. Les résultats de ces expériences pourraient donc dépendre de facteurs variés liés aux observateurs et aux démarches expérimentales. La question demeure posée au sujet de la fiabilité à accorder à ces vérités de terrain, ainsi qu'à l'impact de leurs différences sur les applications de traitement d'image.

Lorsque l'attention visuelle est prise en compte par la communauté du traitement du signal, les deux termes « saillance (visuelle) » et « importance (visuelle) » sont traditionnellement considérés comme synonymes. En effet les deux désignent (et représentent) les zones les plus "pertinentes" visuellement dans la scène. Cependant, du point de vue du spécialiste de la vision, ils sont deux concepts différents, car ils sont issus de deux mécanismes différents de l'attention visuelle : l'un "bottom-up" et l'autre "top-down". Ces deux mécanismes sont conduits par différents types de stimuli, et sont traités par des voies visuelles différentes sollicitant des zones cérébrales différentes. Par conséquent, il est intéressant de bien identifier ces deux termes dans le contexte du traitement d'image.

Un autre problème adressant les chercheurs dans le domaine de l'attention visuelle, est l'impact de la 3D. En effet, pour la visualisation de contenu 3D, la perception de la profondeur de la scène est renforcée. Ce changement de perception de la profondeur modifie aussi beaucoup le comportement de la vision humaine [Hakkinen 10, Huynh-Thu 11a]. En raison de l'émergence de contenus 3D et de la disponibilité d'équipements d'affichage en 3D, les études liées à l'attention visuelle 3D ont gagné en intérêt au cours des dernières années. Néanmoins, la communauté de l'attention visuelle 3D est confrontée à certains problèmes :

1. Comparé au grand nombre de modèles de calcul relatifs aux images et à la vidéo en 2D, seul un petit nombre de modèles de l'attention visuelle 3D peut être trouvé dans la littérature.
2. Une vérité du terrain manque encore. Une base de données oculométriques d'images 3D n'est pas encore disponible. Ce manque rend difficile l'évaluation quantitative et la comparaison des performances des modèles 3D existants.
3. L'impact du "renforcement" de la perception de la profondeur sur l'attention visuelle doit encore être étudié. Ceci peut aider à la modélisation informatique de l'attention visuelle 3D. Cependant, plusieurs études tirent des conclusions contradictoires à ce sujet.
4. De nouvelles applications de l'attention visuelle (en 3D) doivent être proposées et

vérifiées. Les chercheurs ont avancé qu'il serait possible de tirer parti de l'attention visuelle pour résoudre le problème d'inconfort visuel lié à la télévision 3D. Toutefois, la faisabilité et la méthode à mettre en œuvre demeurent des questions ouvertes.

Pour contribuer à la résolution des problèmes mentionnés ci-avant, nous avons conduit les travaux présentés dans cette thèse. Le contenu et la structure principale de la thèse sont donnés dans la partie qui suit.

## 1.3 Structure de la thèse

La première partie de la thèse décrit deux travaux en lien avec la vérité du terrain, à savoir la carte de densité de fixation (FDM) utilisée dans le contexte de la modélisation informatique de l'attention visuelle. Ces 2 travaux concernent : (1) la précision de la FDM, et (2) la signification de la FDM.

- Dans le chapitre 2, nous examinons la fiabilité de données oculométriques issues de différentes bases de données. Exactement, dans le contexte d'une coopération internationale, nous évaluons les similarités et différence entre les bases de données FDM obtenues à partir de trois expériences oculométriques menées dans différents laboratoires (et pays). Nous avons également étudié l'impact de la différence inter-laboratoires entre vérités de terrain sur des applications de traitement d'image.
- Dans le chapitre 3, nous présentons une étude psychophysique menée pour quantifier la relation entre données oculométriques et résultats d'une expérience de notation menée pour identifier les régions d'intérêt. Dans notre étude, les FDM sont créées via des expériences oculométriques (la tâche de vision est libre), elles sont censées fournir les informations de la saillance visuelle "bottom-up". D'autre part, la carte d'importance (ROI) est censée être liée aux mécanismes "top-down" davantage liés à la signification et au sens de la scène. Par conséquent, l'étude présentée dans ce chapitre peut être considérée comme une étude de la relation entre mécanismes "bottom-up" (ascendants) et "top-down" (descendants).

La deuxième partie de la thèse compte quatre travaux portant sur la modélisation informatique de l'attention visuelle, en particulier sur la modélisation de l'attention visuelle 3D.

- Dans le chapitre 4, premièrement nous présentons comment les indices de profondeur affectent la perception de la profondeur d'une scène 3D. Ensuite nous proposons une application potentielle du modèle de l'attention visuelle 3D. Nous quantifions la façon dont deux indices de profondeur, le flou de défocalisation et la disparité binoculaire, interagissent et affectent la profondeur perçue. Nous montrons que, connaissant la région d'intérêt, il est possible de flouter certaines parties de la scène afin de modifier la profondeur perçue de la scène, et d'améliorer encore la qualité de l'expérience de visualisation 3D.

- Dans le chapitre 5, nous nous concentrons sur l’impact de la profondeur perçue lors du déploiement de l’attention visuelle 3D. Nous essayons de déterminer s’il existe un « biais de profondeur » lors de la visualisation de contenu 3D affiché sur écran plat stéréoscopique.
- Dans le chapitre 6, nous proposons un nouveau modèle de calcul de l’attention visuelle sur image fixe stéréoscopique 3D. Le modèle proposé considère la profondeur comme une dimension visuelle supplémentaire. La mesure de saillance de profondeur exploite les données des mouvements oculaires qui sont obtenues à partir d’une expérience oculométrique utilisant des stimuli synthétiques. Deux manières différentes sont ensuite proposées et examinées, pour intégrer les informations de profondeur dans le modèle d’attention visuelle 3D. Nous avons également créé une base de données contenant dix-huit images stéréoscopiques 3D (il s’agit d’images naturelles), la base contient aussi leurs cartes de disparité correspondantes et les données oculométriques capturées lors de tâches de visualisation libre. Cette base de données aide à résoudre le problème de manque de vérité de terrain dans le domaine de la recherche sur la modélisation de l’attention visuelle 3D.
- Dans le chapitre 7, nous présentons une étude approfondie sur le biais central et ses différences qui apparaissent selon les conditions de visualisation 2D ou 3D. Nous proposons également d’évaluer deux façons différentes pour intégrer le biais central dans un modèle d’attention visuelle 3D. Enfin, nous proposons un modèle hybride de calcul de l’attention visuelle 3D, tenant compte à la fois de la saillance de profondeur ainsi que du biais central.

A la fin de la thèse, nous présentons sa conclusion au chapitre 8. Une synthèse de nos contributions, et des perspectives sont également données.

# **Vérités de terrain de l'attention visuelle**



## Chapitre 2

# Une étude comparative des cartes de densité de fixation obtenues à partir de différentes expériences oculométriques

Les cartes de densité de fixation (CDF) créées à partir d'expériences du suivi oculaire ont été largement utilisées pour les études de l'attention visuelle et des applications de traitement d'images. Depuis les CDF sont supposées être des vérités de terrain fiables sur le comportement attentionnel des observateurs humains, la précision des CDF a une importance particulière. De nos jours, la communauté scientifique dispose de nombreuses bases de données CDF fournies par différents laboratoires. Ces CDF sont proches, et elles sont toutes tenues de fournir la vérité de terrain. Cependant, aucune étude à ce jour n'a analysé les différences entre elles et leur impact sur les applications.

Puisque c'est le premier chapitre de cette thèse concernant l'attention visuelle, nous commençons notre travail par une étude portant sur l'exactitude de la vérité de terrain, c'est à dire des CDF. Dans ce chapitre, nous procédons à une comparaison approfondie de trois bases de données CDF, qui ont été créées par 3 laboratoires différents<sup>1</sup>. Nous nous concentrons sur l'effet du temps de présentation de l'image ainsi que de son contenu, et nous évaluons l'impact de la différence de CDF sur trois applications : la modélisation de la saillance visuelle, l'évaluation de la qualité d'images, et le "retargeting" d'images.

### 2.1 Expériences avec l'oculomètre

Nous avons effectué une expérience avec l'oculomètre à l'Université de Nantes (UN). En outre, deux autres expériences de ce type ont été réalisées dans deux autres laboratoires indépendants (1) à l'école d'informatique et de mathématiques à l'Université

---

1. Cette étude est le résultat d'une coopération internationale entre l'Université de Nantes (France), l'Université Tehnologique de Delft (Pays-Bas), l'Université de l'ouest de Sydney (Australie) et les laboratoires de recherche de Philips (Pays-Bas).

de l'ouest de Sydney en Australie (UWS)[Engelke 09], et (2) dans l'équipe étudiant les interactions homme-machine de l'Université technologique de Delft (TUD) au Pays-Bas[Liu 09]. Ces trois expériences sont référencées dans le reste de ce chapitre comme UN, UWS et TUD, respectivement.

### 2.1.1 Stimuli

Les images qui ont été utilisées dans les trois expériences provenaient de la même base de données de qualité d'image LIVE [Sheikh 05]. 29 images de référence ont été utilisées. Toutes les images et les CDF correspondantes sont présentées à la figure 2.1.1.

### 2.1.2 Comparaison des procédures expérimentales

Une vue d'ensemble des trois expériences est présentée au tableau 2.1. En outre, certaines différences, qui pourraient être une source de variabilité parmi les données enregistrées, sont introduites comme suit.

Details	UWS	TUD	UN
Number	15	18	21
Age range (average age)	20-60 (42)	-	18-42 (26)
Male/female	9/6	11/7	11/10
Non-experts/experts	12/3	18/0	21/0
Occupation	University staff/students	University students	University staff/students
Compensated	No	No	Yes
Environment	Laboratory		
Illumination	Low		
Viewing distance	60 cm	70 cm	70 cm
Task	Free-viewing: the observers were not instructed with any particular task but to view the images		
Make	Samsung SyncMaster	iiyama	DELL
Type	LCD	CRT	LCD
Size	19"		
Resolution [pixels]	1280 × 1024	1024 × 768	1280 × 1024
Make	EyeTech TM3 [27]	SMI iView X RED [28]	SMI iView X Hi-Speed [29]
Type	Infrared video-based		
Frequency	45 GP/s	50 GP/s	500 GP/s
Accuracy	< 1 dva	0.5-1 dva	0.25-0.5 dva
Mounting	Under the display	Under the display	Tower with head rest
Calibration	16 point screen	9 point screen	9 point screen
Order	Random		
Image duration	12 s	10 s	15 s
Grey-screen duration	3 s		
Max. visual angle [pixels/deg]	36	32.8	41.8
Central fixation point	Yes	No	No

TABLE 2.1: Vue d'ensemble des trois expériences avec oculomètre.

### Observateurs

Dans les trois expériences, le nombre d'observateurs variait de 15 (pour UWS) à 21 (pour UN). En outre, comme les expériences ont été menées dans trois pays différents

(France, Australie et Pays-Bas), les différences culturelles entre les observateurs pourraient également avoir une influence sur les résultats.

### Conditions d'observation

Parmi les trois expériences, les conditions d'observation diffèrent principalement en (1) la distance d'observation et (2) la taille de l'image physique ce qui est due à l'utilisation de différentes tailles d'écran variant en résolution. En outre, la présentation de l'image diffère essentiellement en deux facteurs : (1) la durée d'observation comprise entre 10 s et 15 s ; et (2) car un point de fixation central est présenté entre deux stimuli pour UWS.

## 2.2 Comparaison inter laboratoires

Dans cette étude, nous utilisons deux mesures de similarité : le coefficient de corrélation linéaire de Pearson (PLCC) et l'aire sous la courbe ROC (AUC).

### 2.2.1 Différences inter laboratoires

Nous affichons les résultats avec le PLCC et l'AUC (figure 2.2.1 et figure 2.2.2) en fonction du temps de présentation pour 3 comparaisons inter-laboratoires (UWS-TUD, UWS-UN, TUD-UN) et des comparaisons à une CDF aléatoire (UWS-RND, RND-TUD, UN-RND). La figure 2.2.1 montre que la progression de la PLCC moyenne avec le temps de présentation est similaire pour les trois expériences. Les CDF des expériences TUD et UN semblent être les plus semblables. Des observations similaires peuvent être observées avec l'AUC (voir la figure 2.2.2).

### 2.2.2 Impact du point de fixation central sur le biais

Nous voulons savoir si le point de fixation central utilisé par UWS affecte le biais au centre, et aussi par conséquent les comparaisons entre CDF. A la figure 2.2.3(a)-(c), nous présentons des exemples de CDF ayant des temps de présentation de  $t = 500$  ms avec l'image numéro 18 «parrots». Le point de fixation central utilisé par UWS aboutit à une répartition plus centrée des fixations.

Pour chaque image, nous avons créé une carte de biais au centre [Le Meur 06]. Une telle carte est montrée à la figure 2.2.3(d). Nous calculons ensuite le PLCC et l'AUC (pour  $\tau = 200$ ) entre ces cartes de biais et les CDF. Les résultats (voir la figure 2.2.4) montrent que le PLCC et l'AUC sont systématiquement plus élevés pour UWS par rapport à TUD et UN, et cela pour presque tous les temps de présentation.

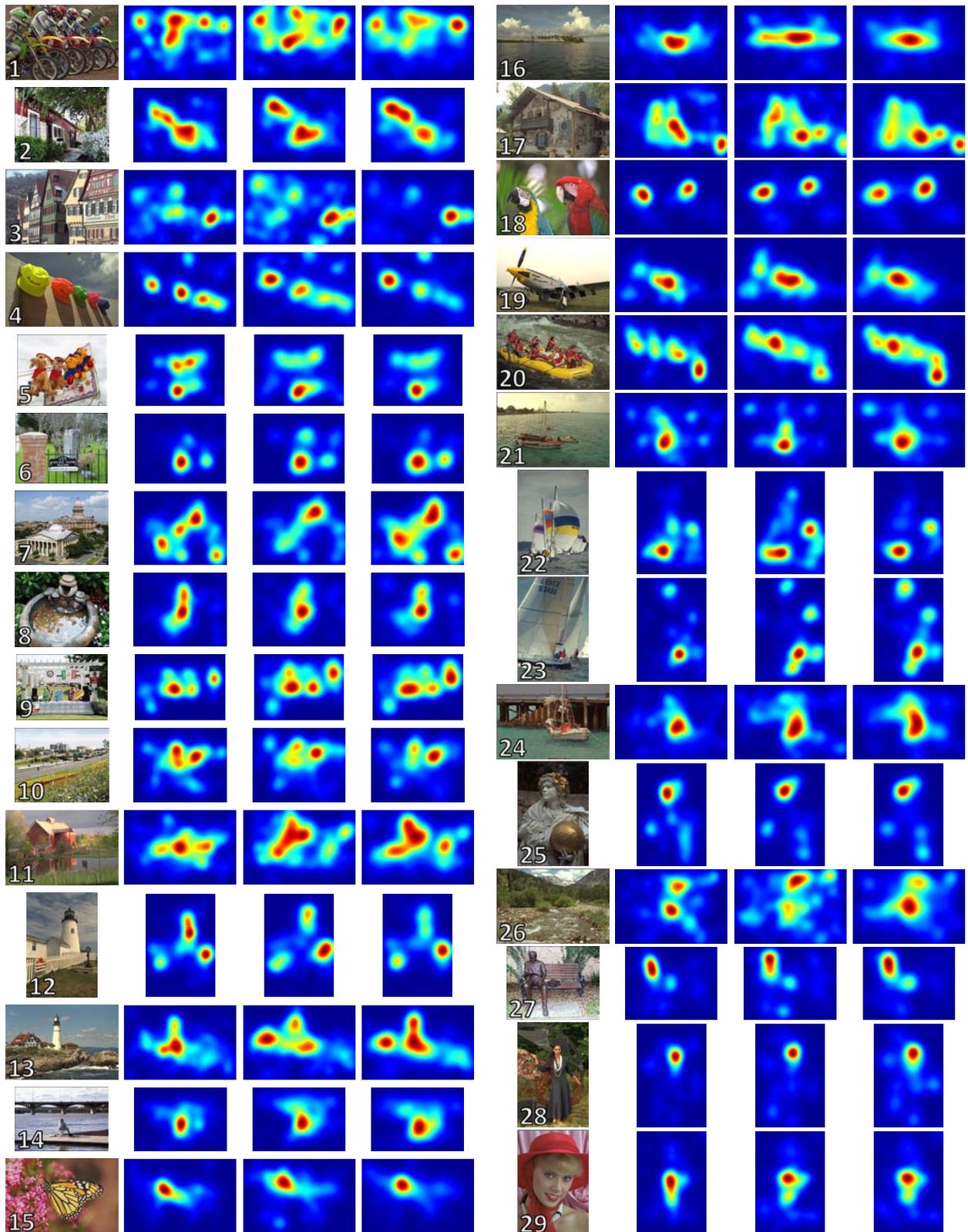


FIGURE 2.1.1: Exemples de CDF pour un temps de présentation de 10 s. De gauche à droite : image originale, UWS, TUD, UN.

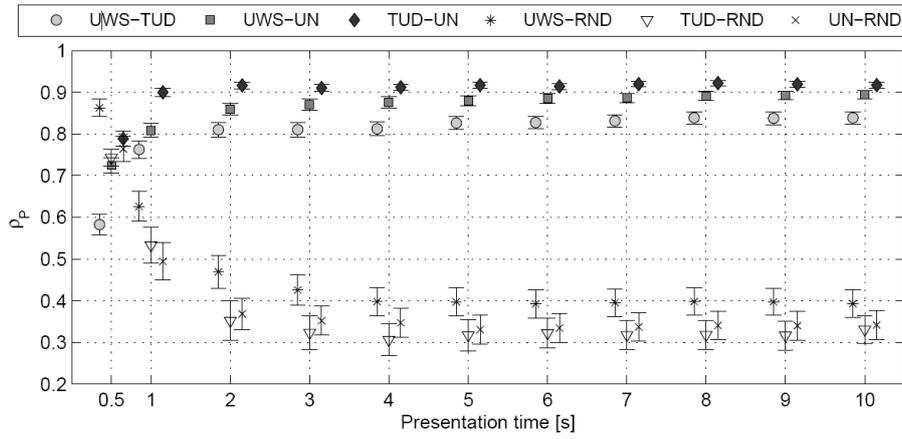


FIGURE 2.2.1: PLCC moyen et écart-type pour toutes les images et tout  $t$ .

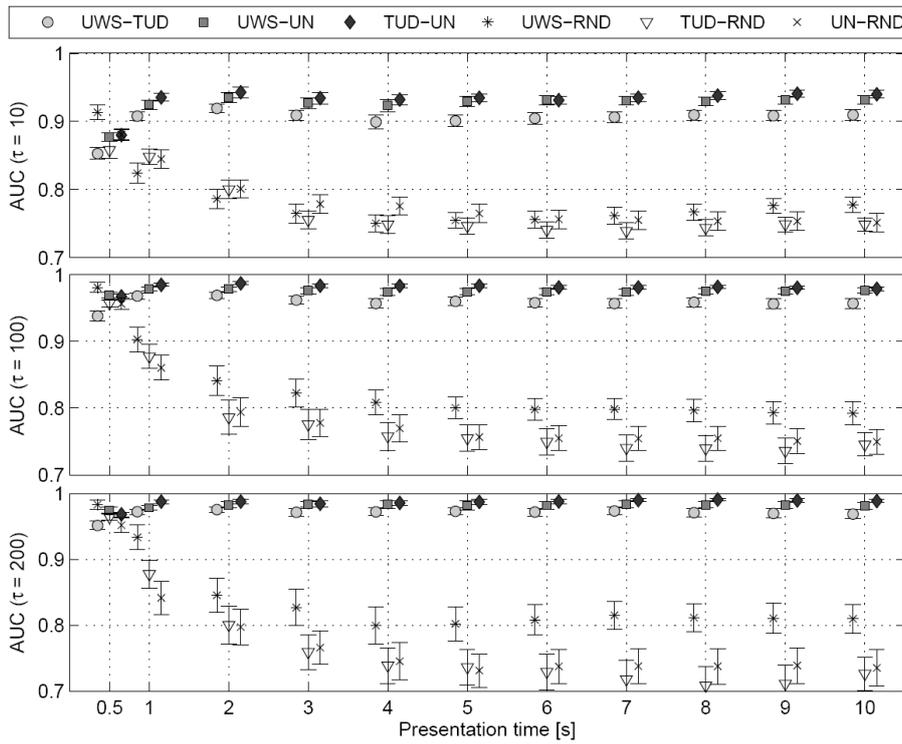


FIGURE 2.2.2: AUC moyen et écart-type pour toutes les images et tout  $t$ , et pour  $\tau = 10$  (en haut),  $\tau = 100$  (au milieu), et  $\tau = 200$  (en bas).

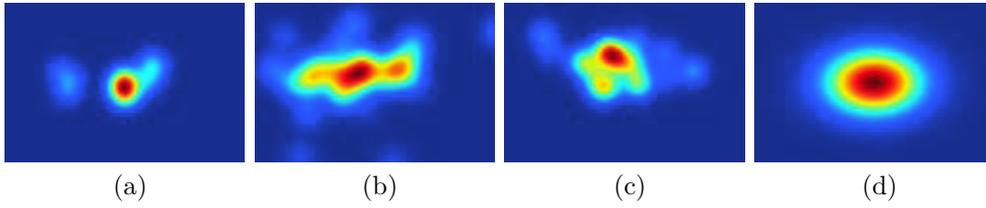


FIGURE 2.2.3: Comparaison des CDF pour  $t = 500 \text{ ms}$  : (a)  $M_{UWS}^{(0.5)}(18)$ , (b)  $M_{UN}^{(0.5)}(18)$ , (c)  $M_{TUD}^{(0.5)}(18)$ , et (d) carte de biais.

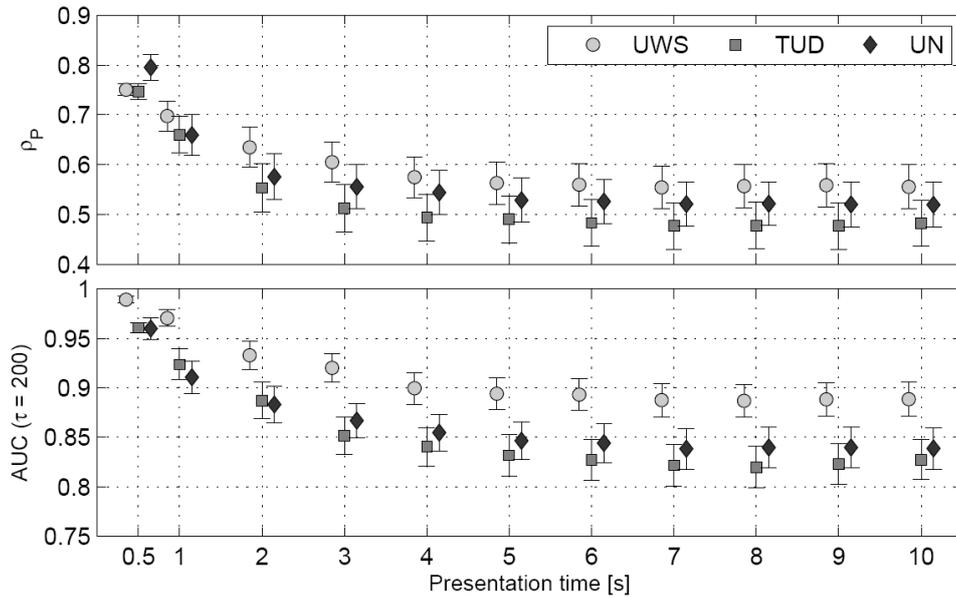


FIGURE 2.2.4: PLCC (en haut) et AUC avec  $\tau = 200$  (en bas) entre la CDF et sa carte de biais central pour tous les temps de présentation  $t$ .

### 2.2.3 Dépendance au contenu

Nous analysons ici la similitude entre CDF en relation avec le contenu des images. Nous considérons deux temps de présentation :  $t = 1s$  et  $t = 10s$ . A la figure 2.2.5, le PLCC et l'AUC pour les 29 images sont présentés. Il est constaté que la similarité entre CDF semble fortement dépendre du contenu des images. Pour les moyennes élevées de PLCC et l'AUC, les valeurs comparées sont très proches.

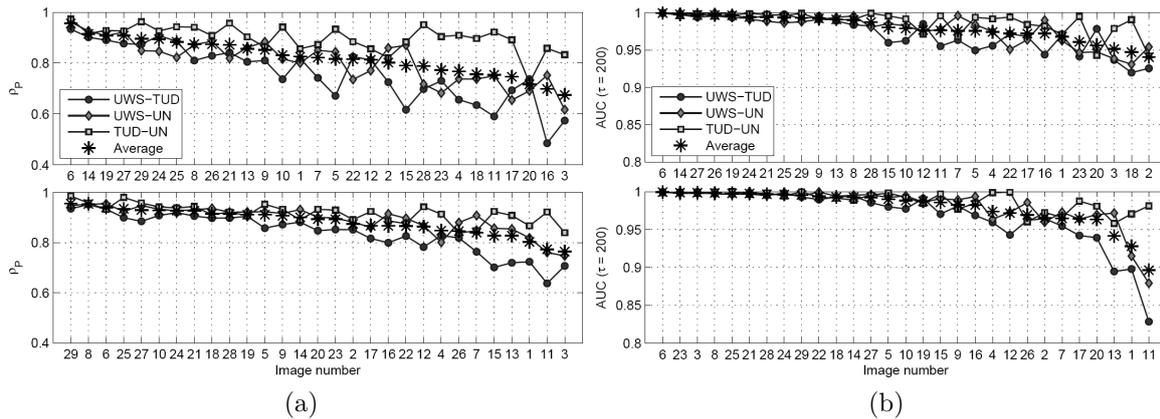


FIGURE 2.2.5: Impact du contenu de l'image, mesuré en utilisant (a) le PLCC et (b) l'AUC ( $\tau = 200$ ) pour  $t = 1s$  (en haut) et  $t = 10s$  (en bas).

## 2.3 Applications

Dans cette partie, nous avons donc déterminé la sensibilité à la CDF utilisée pour trois applications : la modélisation de la saillance visuelle, l'évaluation de la qualité d'image et le "retargeting" (reciblage) d'image.

### 2.3.1 Modèles de saillance visuelle

Nous analysons ici dans quelle mesure la validation des modèles de saillance dépend de la réalité de terrain, la CDF. Nous considérons cinq modèles de saillance bien connus, qui sont proposés par Itti et al. [Itti 98], Rajashekar et al. [Rajashekar 08], Bruce et al. [Bruce 09], Achanta et al. [Achanta 09], et Hou et al. [Hou 07]. Nous calculons les cartes de saillance pour toutes les images à l'aide de ces modèles. Les résultats pour le PLCC et l'AUC ( $\tau = 200$ ) entre cartes de saillance et CDF sont présentés au tableau 2.2. Nous avons constaté que les résultats diffèrent considérablement, plus entre les modèles de saillance visuelle qu'entre les CDF.

t	Database	Modèles d'attention visuelle					$\sigma_{SAL}$
		Itti	Rajashekar	Bruce	Achanta	Hou	
1s	UN	0.099	0.372	0.272	0.254	0.32	0.103
	UWS	0.096	0.288	0.218	0.202	0.241	0.071
	TUD	0.097	0.348	0.244	0.242	0.300	0.094
	$\sigma_{DB}$	0.001	0.044	0.027	0.027	0.041	—
	RND	0.040	0.282	0.164	0.120	0.152	0.087
10s	UN	0.150	0.449	0.376	0.335	0.421	0.118
	UWS	0.147	0.435	0.371	0.312	0.384	0.111
	TUD	0.152	0.448	0.369	0.333	0.415	0.115
	$\sigma_{DB}$	0.003	0.008	0.004	0.013	0.020	—
	RND	0.072	0.297	0.208	0.135	0.171	0.084

(a)

t	Database	Modèles d'attention visuelle					$\sigma_{SAL}$
		Itti	Rajashekar	Bruce	Achanta	Hou	
1s	UN	0.648	0.801	0.749	0.681	0.758	0.061
	UWS	0.624	0.733	0.683	0.651	0.676	0.040
	TUD	0.620	0.786	0.713	0.693	0.748	0.062
	$\sigma_{DB}$	0.015	0.036	0.033	0.022	0.044	—
	RND	0.621	0.717	0.633	0.603	0.650	0.044
10s	UN	0.658	0.802	0.773	0.690	0.760	0.060
	UWS	0.660	0.797	0.758	0.687	0.737	0.055
	TUD	0.671	0.803	0.772	0.692	0.771	0.057
	$\sigma_{DB}$	0.007	0.003	0.008	0.002	0.017	—
	RND	0.616	0.682	0.624	0.554	0.606	0.046

(b)

TABLE 2.2: (a) PLCC et (b) AUC ( $\tau = 200$ ) entre les cartes de saillance prédite et CDF.

### 2.3.2 Evaluation de la qualité d'image

Cartes de saillance et CDF sont souvent intégrées dans des mesures de qualité d'image dans le but d'améliorer les performances de prédiction de la qualité [Hantao 11]. Nous analysons dans quelle mesure l'amélioration de trois métriques de qualité d'image, PSNR ([Wang 06]), SSIM ([Wang 04]) et VIF ([Sheikh 06]), varie en fonction de la CDF utilisée. Nous intégrons la carte de saillance dans le modèle de qualité objective par pondération multiplicative et locale de la carte distorsion. Nous comparons ensuite les performances des métriques en tenant compte ou non de la CDF. Les résultats de chaque mesure de qualité sont présentés dans les tableaux 2.3, 2.4 et 2.5. Les résultats montrent que, pour toutes classes de la distorsion et considérant leur moyenne, les améliorations sont très similaires pour les trois expériences. La cohérence entre CDF est meilleure que la cohérence entre modèles de prédiction de la qualité.

		JPEG 1	JPEG 2	J2K 1	J2K 2	Gaussian blur	White noise	Fast fading	Moyenne
PSNR	UWS	0.004	0.028	0.002	0.037	0.006	0	0.009	0.012
	TUD	0.008	0.031	0.006	0.037	0.020	0	0.016	0.017
	UN	0.006	0.029	0.003	0.036	0.018	0	0.015	0.015
	RND	0	0.001	-0.001	0.001	0.01	0	0.002	-0.001

TABLE 2.3: Gain de performance de prédiction du PSNR (basé sur le PLCC).

		JPEG 1	JPEG 2	J2K 1	J2K 2	Gaussian blur	White noise	Fast fading	Moyenne
SSIM	UWS	0.017	0.041	0.022	0.041	0.075	0.008	0.029	0.033
	TUD	0.019	0.039	0.019	0.038	0.070	0.008	0.023	0.031
	UN	0.014	0.036	0.019	0.037	0.070	0.009	0.025	0.030
	RND	-0.002	-0.003	-0.003	-0.001	-0.003	0.001	0	-0.002

TABLE 2.4: Gain de performance de prédiction du SSIM (basé sur le PLCC).

		JPEG 1	JPEG 2	J2K 1	J2K 2	Gaussian blur	White noise	Fast fading	Moyenne
VIF	UWS	0.022	0.008	0	0.007	0.017	0.010	0.012	0.011
	TUD	0.022	0.008	0	0.009	0.017	0.012	0.009	0.011
	UN	0.026	0.008	0.004	0.009	0.021	0.010	0.008	0.012
	RND	0.004	0.003	-0.019	0.006	0.019	0.005	0.002	0.003

TABLE 2.5: Gain de performance de prédiction du VIF (basé sur le PLCC).

### 2.3.3 Reciblage d'image basé sur la saillance

Les algorithmes de reciblage d'image [Avidan 07] redimensionnent les images en ôtant les bandes verticales de plus basse "énergie", préservant ainsi les régions les plus importantes dans les images. Nous avons utilisé la CDF dans un algorithme de reciblage l'image basé saillance [Wang 11a] et étudié les similarités entre images qui résultent du reciblage. Des exemples sont présentés à la figure 2.3.1 pour les images 27, 29, et 13 de la base de données LIVE. Nous constatons que le reciblage échoue lorsqu'est utilisée une CDF aléatoire (RND). Toutefois, les régions les plus pertinentes sont bien conservées lorsqu'est utilisée la CDF expérimentale, quelle que soit la CDF (UWS, TUD ou UN).

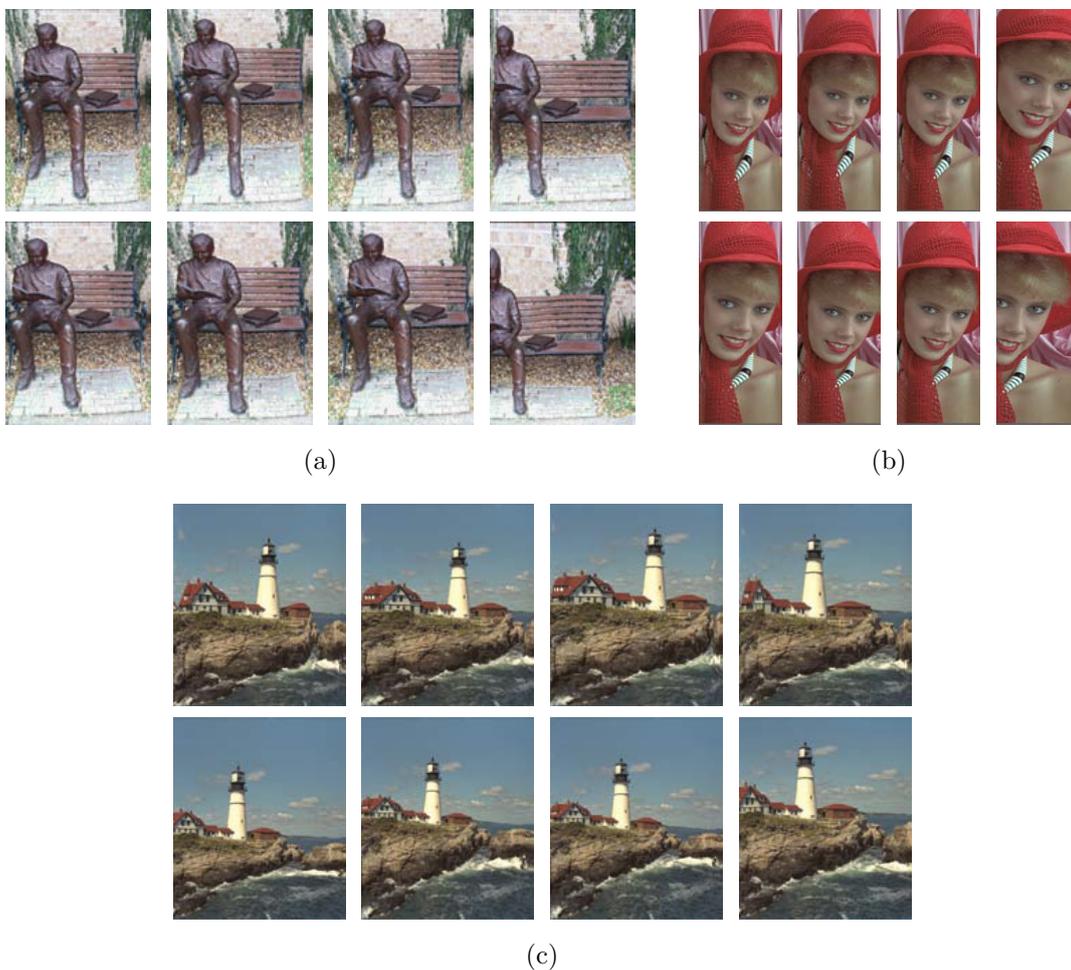


FIGURE 2.3.1: Résultats de reciblages basés CDF avec  $t = 1 s$  (rangée du haut) et  $t = 10 s$  (rangée du bas) pour (a) l'image 27, (b) l'image 28, et (c) l'image 13. De gauche à droite : UWS, TUD, UN, RND.

## 2.4 Perspectives

Dans ce chapitre, nous avons analysé la similarité de CDF pour trois expériences oculométriques indépendantes. Les CDF ont été jugées très semblables, et cette similarité a été jugée très dépendante du contenu de l'image. Les images qui contiennent une région saillante distincte montrent une plus grande similarité de CDF comparées aux images ayant plusieurs régions saillantes ou aucune. Toutefois, la distribution des fixations est traditionnellement considérée comme étant liée aux stimuli "bottom-up". D'autre part, ce sont les mécanismes "top-down" qui déterminent le contenu de l'image. Déterminer dans quelle mesure la distribution des fixations peut refléter l'attention "top-down" demeure une question ouverte. Une étude pour quantifier la relation entre répartition de fixation et attention "top-down" sera présentée dans le prochain chapitre.



# Chapitre 3

## Relation entre saillance et importance visuelle

### 3.1 Introduction

Saillance visuelle [Koch 85, Itti 98] et importance visuelle [Osberger 98, Maeder 95, Etz 00, Kadiyala 08] sont issues de deux mécanismes différents de l'attention visuelle, les mécanismes "bottom-up" et "top-down", respectivement. Les deux fournissent des informations importantes sur la façon dont le système biologique visuel résout le problème d'analyse d'images. Ils sont également connus pour désigner les parties de la scène les plus visuellement "pertinentes". Cependant, malgré les différences dans la façon dont ils sont déterminés en termes de traitement visuel humain, ils sont traditionnellement considérés comme synonymes par la communauté du traitement du signal. Dans ce chapitre, nous présentons les résultats de deux expériences psychophysiques et l'analyse associée qui démontrent les similitudes et les différences entre saillance visuelle et l'importance visuelle.

### 3.2 Méthodes

#### 3.2.1 Expérience I : Importance visuelle

Une expérience psychophysique a été réalisée afin d'obtenir des cartes d'importance visuelle pour une grande base de données d'images. La carte d'importance visuelle associée à une image indique l'importance visuelle de chaque objet de cette image par rapport à ses autres objets (y compris ce qui est normalement considéré comme l'arrière-plan). Les objets qui reçoivent la plus grande importance visuelle sont traditionnellement considérés comme sujet principal de l'image. Nous avons utilisé des images de la base de Berkeley (qui fournit pour chaque image, sa carte de segmentation), et avons recueilli des

notes d'importance pour chaque objet de 300 images de cette base. Ces notes d'importance sont généralement considérées comme résultant d'un mécanisme "top-down" par le système visuel puisque les décisions faites pour évaluer chaque objet impliquent typiquement l'interprétation de la scène, la reconnaissance d'objets, et souvent aussi des considérations d'ordre artistique.

### 3.2.2 Expérience II : Saillance visuelle

Dans une seconde expérience, les trajectoires oculaires ont été enregistrées pour 80 images de la même base de Berkeley (voir figure 3.1). A l'aide d'un oculomètre, les trajectoires du regard ont été enregistrées pour des tâches d'observation libre. Alors que les cartes d'importance sont obtenues principalement via un mécanisme "top-down", les trajectoires oculaires sont généralement considérées pour être le résultat d'un mécanisme "bottom-up", au moins en début d'observation. La saillance "bottom-up" [Koch 85] est un attribut particulier basé signal et qui a montré une bonne corrélation avec les emplacements du regard (en début d'observation). Une zone d'image est considérée comme visuellement saillante si elle "se distingue" de son arrière-plan en termes d'un ou de plusieurs attributs (par exemple, le contraste, la couleur, l'orientation). Lorsque les trajectoires visuelles sont mesurées pour des tâches d'observation libre, on peut considérer que les fixations désignent les régions saillantes de l'image. Ainsi, à partir des trajectoires oculaires, on peut construire une carte de saillance expérimentale.

## 3.3 Résultats et analyse

### 3.3.1 Observations qualitatives des cartes d'importance et des cartes de saillance

Une comparaison qualitative des cartes de saillance et des cartes importance révèle des similitudes et des différences distinctes. La figure 3.3.1 montre quelques exemples représentatifs.

Les cartes importance suggèrent que la catégorie de l'objet joue un rôle plus important que la plupart des autres facteurs. En général, nous avons constaté que les observateurs avaient tendance à noter comme plus importants les objets contenant des visages humains et / ou des animaux. D'autre part, les cartes de saillance suggèrent que généralement les régions qui possèdent une forme, une couleur, un contraste, ou une autre caractéristique spatiale locale remarquable attirent l'attention. Cependant, les observateurs regardent toujours le sujet principal de l'image, c'est à dire qui représente l'essentiel de l'image. L'arrière-plan attire toujours moins l'attention.

Malgré ces similitudes, les cartes de saillance et les cartes importance ne correspondent pas toujours. Les cartes de saillance ne séparent pas en objet, ce qui évident avec les

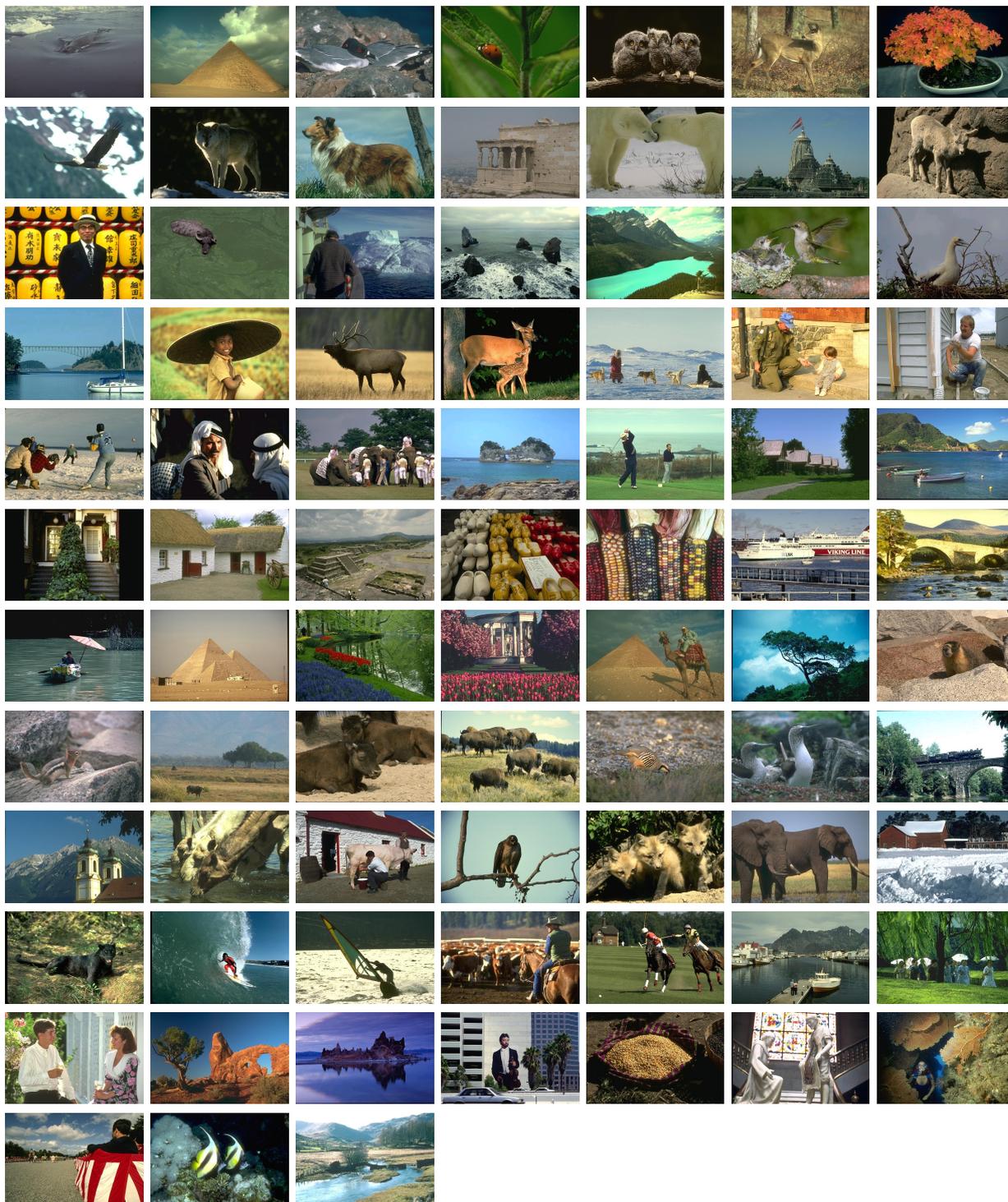


TABLE 3.1: Les 80 images utilisées lors de nos 2 expériences pour constituer les cartes d'importance visuelle et les cartes de saillances visuelle.

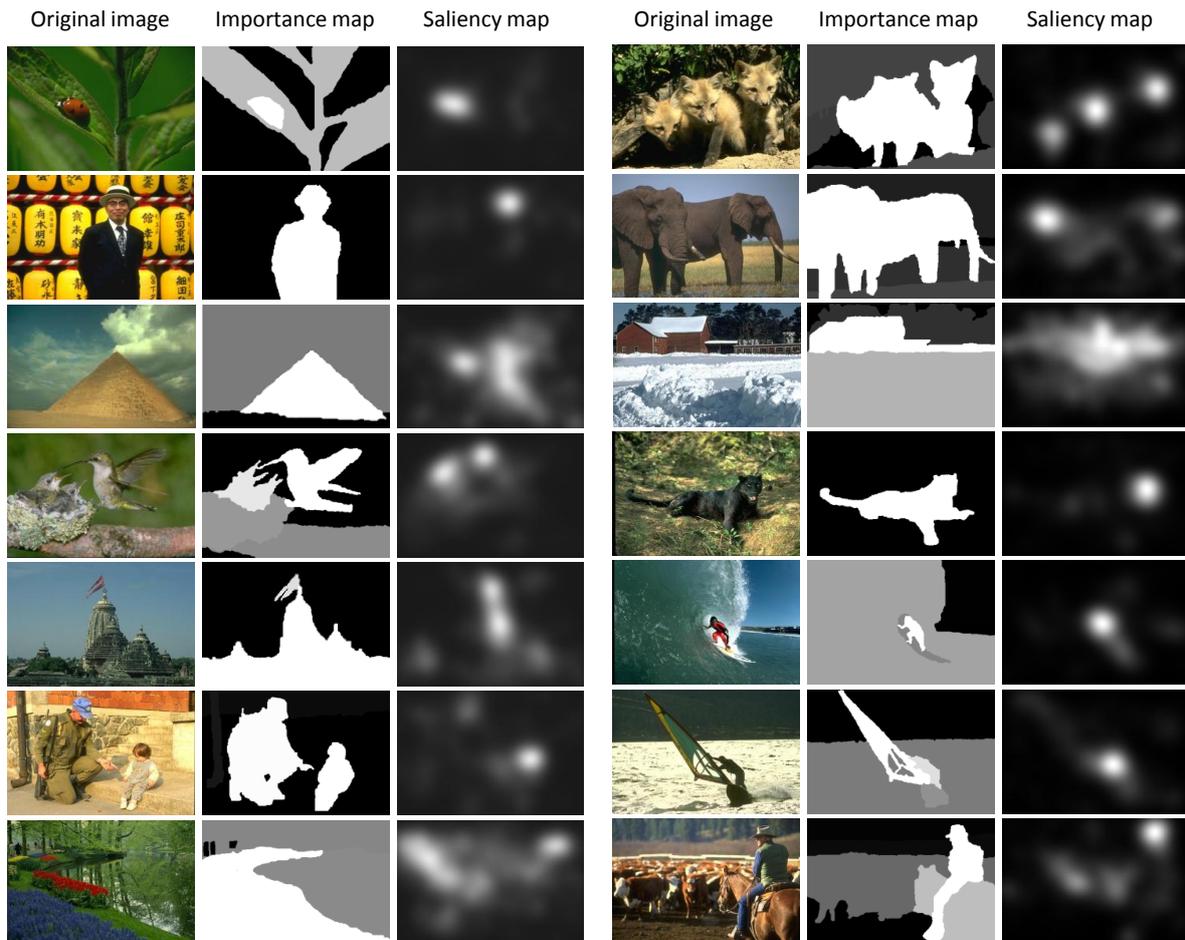


FIGURE 3.3.1: Résultats représentatifs issus des expériences.

cartes d'importance. Par exemple, chaque fois qu'un visage est présent dans une image, le regard de l'observateur ne se porte que sur le visage.

### 3.3.2 Prédire le sujet principal, les objets secondaires, et l'arrière plan

Une façon de procéder pour une analyse quantitative est de prédire les cartes importance à partir des cartes de saillance et en utilisant aussi les cartes de segmentation. Pour y parvenir, les deux approches suivantes ont été testées :

- Par *saillance moyenne* : Pour chaque objet, nous avons additionné les valeurs de saillance des pixels qui sont à l'intérieur, puis nous avons divisé cette valeur par le nombre total de pixels de l'objet.
- Par *saillance de couverture* : Pour chaque objet, nous avons additionné les valeurs de saillance des pixels qui sont à l'intérieur, puis nous avons divisé cette valeur par le nombre de pixels ayant une valeur supérieure à 0.

Afin de faciliter la prédiction, les objets sont classés en trois catégories fonction des valeurs d'importance :

- *Sujet principal*, qui a une valeur d'importance allant de  $2/3$  à  $1$ ,
- *Objet secondaire*, ayant une valeur d'importance entre  $1/3$  et  $2/3$ ,
- *Arrière plan*, objet ayant une valeur d'importance entre  $0$  et  $1/3$ .

		Prédiction		
		Sujet principal	Objet secondaire	Arrière plan
Vérité	Sujet principal	<b>80.5%</b>	29.8%	12.6%
	Objet secondaire	12.5%	<b>42.6%</b>	40.7%
	Arrière plan	7.1%	27.6%	<b>46.7%</b>

(a)

		Prédiction		
		Sujet principal	Objet secondaire	Arrière plan
Vérité	Sujet principal	<b>56.5%</b>	38.6%	8.2%
	Objet secondaire	13.0%	<b>40.4%</b>	24.7%
	Arrière plan	30.5%	21.1%	<b>67.1%</b>

(b)

TABLE 3.2: Matrices “de confusion” obtenues en prédisant l'importance de chaque objet à partir des trajectoires oculaires.

Les résultats de la prédiction sont fournis dans le tableau 3.2 sous la forme de “matrices de confusion”. Chaque ligne de la matrice représente la véritable classe d'importance, et chaque colonne représente la classe prédite. Une prédiction idéale donnerait une matrice diagonale avec des valeurs de 100%. Comme le montre le tableau 3.2(a), la saillance

moyenne peut prédire avec succès le sujet principal dans environ 81% des cas. D'autre part, la saillance de couverture (voir le tableau 3.2(b)) donne de moins bonnes performances avec les sujets principaux, mais des performances légèrement meilleures pour les objets d'arrière-plan.

### 3.3.3 Analyse temporelle

Nous calculons le nombre d'échantillons de positions oculaires par classe d'importance qui se sont produits pendant chaque intervalle de 100 ms des 15 secondes du temps de visualisation. Les trois courbes résultantes, cumulées en considérant tous les observateurs, sont présentées à la figure 3.3.2.

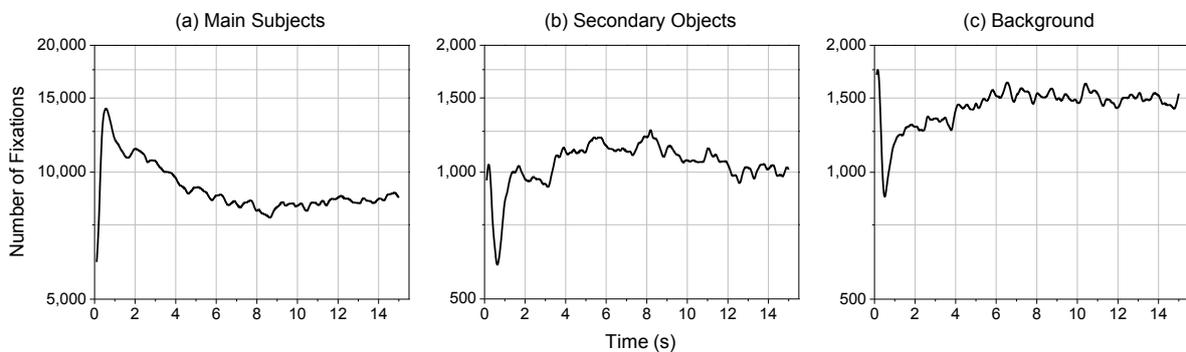


FIGURE 3.3.2: Nombre total d'échantillons de positions oculaires sur (a) les sujets principaux, (b) les objets secondaires, et les objets d'arrière-plan (c) calculés tous les intervalles de 100 ms de la durée d'observation de 15 secondes. Notez que l'échelle de l'axe vertical du premier graphique est 10 fois celle des deux autres graphiques.

Les courbes indiquent que les classes d'objets attirent des quantités considérablement différentes de l'attention. Pendant l'intervalle 0-500 ms, le nombre d'échantillons de positions oculaires pour les sujets principaux était déjà 4-6 fois plus élevé que pour des objets secondaires et d'arrière-plan. Cette observation suggère que des mécanismes "bottom-up" peuvent être effectifs pour repérer les sujets principaux de ces images, ce qui pourrait résulter du fait que les photographes ont tendance à augmenter la saillance des sujets principaux par, par exemple, retouche, mise au point sélective, ou d'autres techniques photographiques. Entre 500-2000 ms, il y a une forte augmentation du nombre d'échantillons de positions oculaires pour les sujets principaux. Ces changements pourraient indiquer l'influence de mécanismes "top-down" forçant les observateurs vers les sujets principaux. Après ce processus, le nombre d'échantillons de positions oculaires pour les sujets principaux a légèrement diminué et ceux des deux autres classes ont légèrement augmenté. Ce dernier changement peut indiquer que les observateurs tentent d'explorer la totalité de l'image, mais leur attention est toujours retenue par les sujets principaux.

Ces trois courbes suggèrent que la relation entre la saillance visuelle et l'importance visuelle peut être dépendante du temps. En particulier, le fait que les sujets principaux attirent le plus l'attention pendant les 0-2000 ms indique que ces premiers échantillons de positions oculaires pourraient être un meilleur prédicteur d'importance visuelle pour ces sujets principaux que celui obtenu en utilisant tous les échantillons. Par conséquent, nous avons prédit les cartes importance en utilisant les échantillons prélevés à partir seulement des premiers 0-2000 ms d'observation. En utilisant uniquement ces données oculaires initiales, une meilleure prédiction est obtenue pour les sujets principaux (voir le tableau 3.3 et la figure 3.3.3).

		Prédiction		
		Sujet principal	Objet secondaire	Arrière plan
Vérité	Sujet principal	<b>89.0%</b>	43.5%	12.4%
	Objet secondaire	3.3%	<b>43.5%</b>	27.2%
	Arrière plan	7.7%	13.0%	<b>60.5%</b>

TABLE 3.3: Matrice de confusion obtenue en prédisant l'importance à partir des 2 premières secondes d'échantillons des trajectoires oculaires et en utilisant la saillance moyenne.

### 3.4 Discussion et conclusion

Ce chapitre a présenté les résultats de deux expériences psychophysiques et l'analyse associée conçue pour quantifier la relation entre saillance visuelle et importance visuelle. Nous avons constaté que les cartes de saillance et celles d'importance sont liées, mais peut-être moins qu'on ne pourrait s'y attendre. Les cartes de saillance ont montré être efficaces pour prédire les sujets principaux. Cependant, les cartes de saillance ont été moins efficaces pour prédire les objets d'importance secondaire et ceux sans importance. Nous avons également constaté que la grande majorité des premiers échantillons de positions oculaires (0-2000 ms) se porte sur les sujets principaux. Ceci suggère qu'une stratégie possible du système visuel humain est de localiser rapidement le ou les sujets principaux de la scène.

Les implications de ces résultats initiaux en traitement de l'image peuvent être très importantes. Plusieurs algorithmes ont été publiés pour prédire avec succès où se porte le regard (par exemple, réf. [Itti 98, Le Meur 06]). Nos résultats suggèrent que ces schémas prédictifs peuvent être aussi utilisés pour prédire les cartes importance lorsqu'ils sont couplés avec des cartes de segmentation. À leur tour, les cartes importance peuvent être utilisées pour effectuer des traitements basés importance comme le cadrage automatique, l'amélioration d'images, la compression, la protection inégale d'erreur, ou l'évaluation de la qualité.

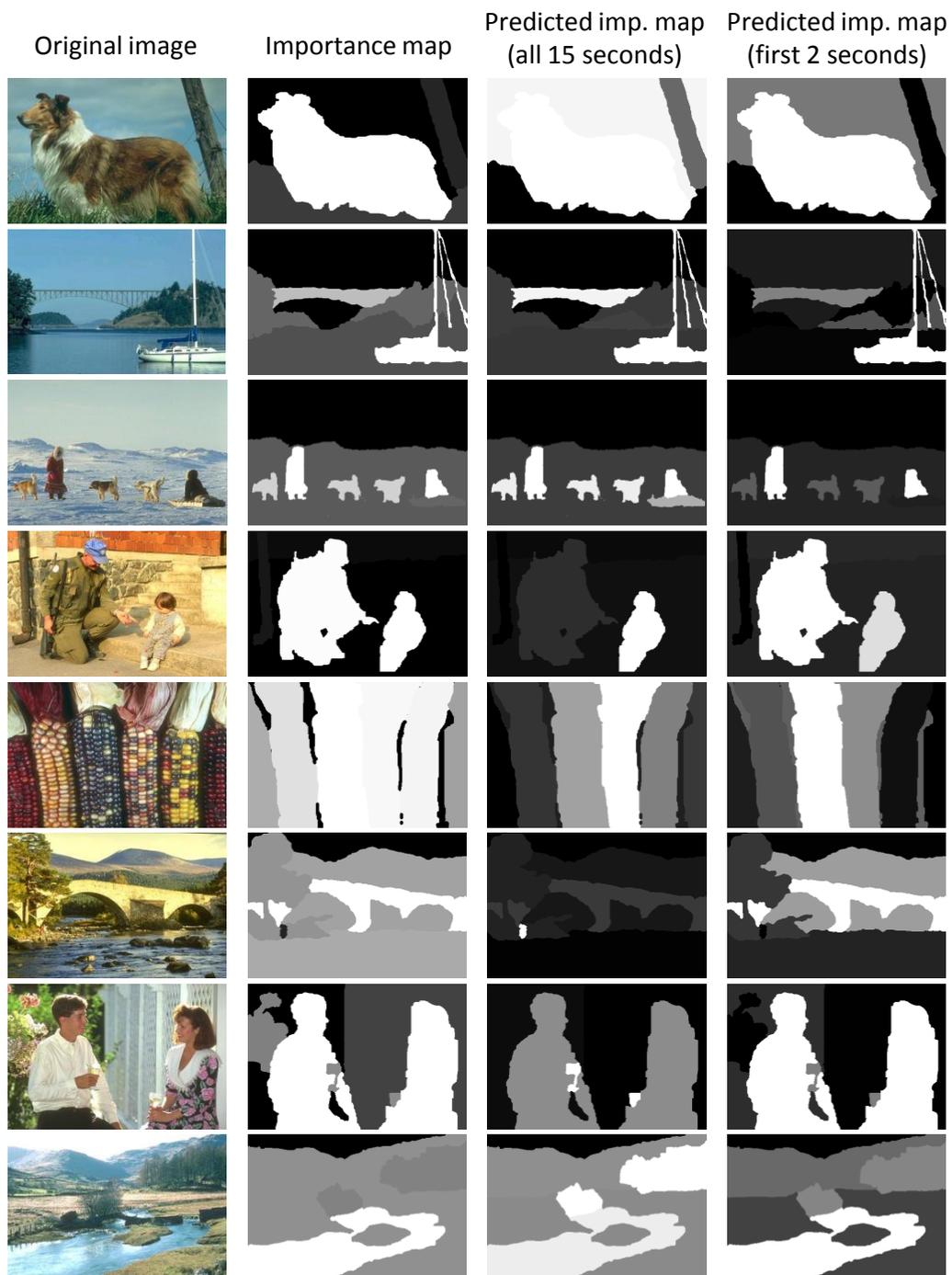


FIGURE 3.3.3: Résultats représentatifs utilisant tous les échantillons oculaires vs. seulement ceux des 2 premières secondes pour prédire les cartes d'importance.

# **Attention visuelle en 3D**



# Chapitre 4

## Perception de profondeur et contenu 3D stéréoscopique

De nos jours, le contenu 3D stéréoscopique augmente la sensation de présence à travers le renforcement de la perception de la profondeur. Les études relatives à la 3D stéréoscopique ont récemment gagné en intérêt en raison de l'émergence de contenus 3D (au cinéma et à la maison) et de la disponibilité d'équipements pour l'acquisition et l'affichage de la haute définition en 3D. A partir de ce chapitre, nous nous concentrons sur l'étude de sujets liés au déploiement de l'attention visuelle en condition de visualisation en trois dimensions. Avant d'entrer dans la modélisation informatique de l'attention visuelle 3D stéréoscopique, des études relatives à la perception de la profondeur dans des conditions de visualisation 3D stéréoscopique sont tout d'abord introduites.

### 4.1 Introduction

#### 4.1.1 Perception de la profondeur

Pour percevoir dans le monde en trois dimensions, le système visuel utilise des indices de profondeur pour estimer les informations de profondeur d'une scène. Il existe différentes façons pour classer les indices de profondeur [McAllister 93]. Selon le nombre de yeux impliqués, les indices de profondeur peuvent être divisés en deux catégories (voir les tableaux 4.1, [Mather 09]) : (1) les indices monoculaires, qui sont considérés si un seul œil est utilisé, ils sont également disponibles lorsque les deux yeux sont utilisés, (2) les indices binoculaires, qui sont disponibles uniquement lorsque les deux yeux sont utilisés conjointement. Nous fournissons également des illustrations de certains des indices de profondeur.

Indices monoculaires	Indices binoculaires
Taille de l'image rétinienne (fig. 4.1.1)	Vergence
Hauteur dans le champ visuel	Disparité binoculaire
Gradient de la texture (fig. 4.1.2)	
Flou de défocalisation	
Perspective atmosphérique (fig. 4.1.3)	
Accommodation	
Parallaxe de mouvement	
Ombres et dégradés (fig. 4.1.4)	
Occlusion et interposition (fig. 4.1.5)	

TABLE 4.1: Liste des indices de profondeur.

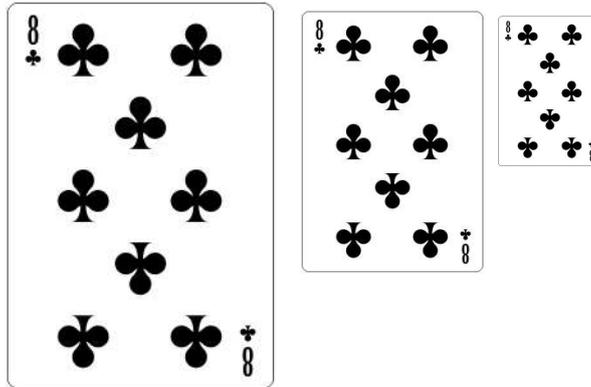


FIGURE 4.1.1: Exemple de l'effet de l'image rétinienne sur la profondeur perçue.

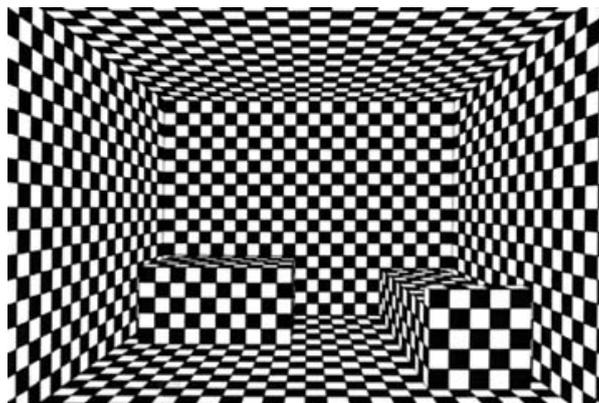


FIGURE 4.1.2: Exemple de l'effet du gradient de la texture sur la profondeur perçue.

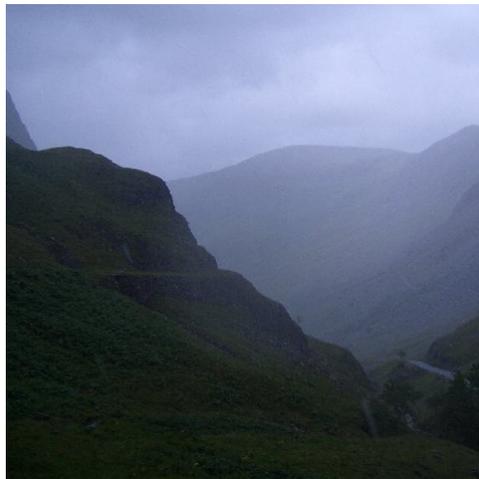


FIGURE 4.1.3: Exemple de l'effet de la perspective atmosphérique sur la profondeur perçue.

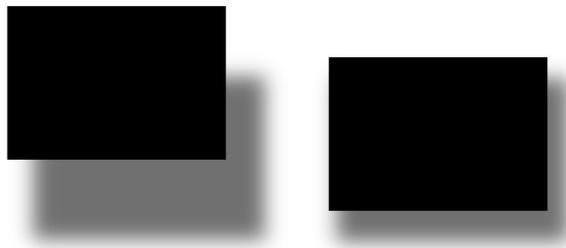


FIGURE 4.1.4: Exemple de l'effet des ombres et des dégradés sur la profondeur perçue.

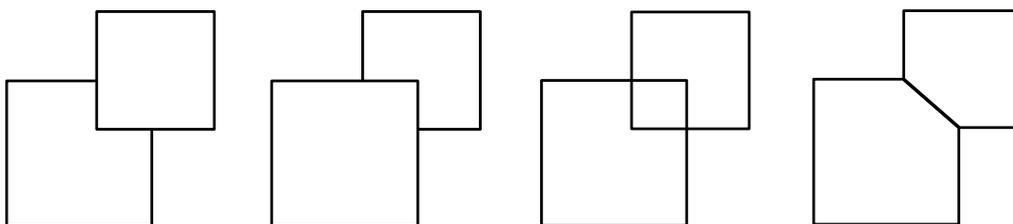


FIGURE 4.1.5: Exemple de l'effet d'occlusions et d'interpositions sur la profondeur perçue. Dans la première et la seconde image, l'ordre des profondeurs des deux carrés peut être facilement perçu. Dans la troisième et la quatrième image, il est difficile de déterminer la profondeur des objets.

### 4.1.2 Conflits en vision stéréoscopique

Il est souvent prétendu que regarder un contenu 3D stéréoscopique provoque fatigue visuelle et gêne visuelle. Il est considéré que le problème d'inconfort et de fatigue visuelle est causé par des conflits entre indices de profondeur. Le conflit entre accommodation et vergence [Hoffman 08] est généralement considéré comme la principale raison de l'inconfort et de la fatigue visuelle lorsqu'il est regardé un contenu 3D sur un écran 3D stéréoscopique.

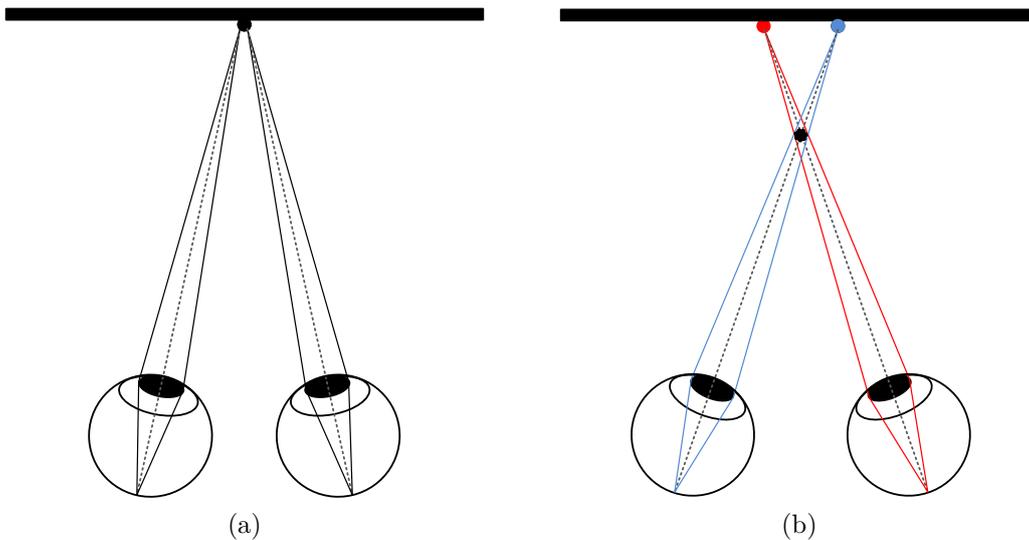


FIGURE 4.1.6: Illustration du conflit entre accommodation et vergence.

Un exemple du conflit entre l'accommodation et la vergence est illustré à la figure 4.1.6. Lorsqu'est regardé un contenu 2D, accommodation et vergence sont normalement couplés [Fincham 57]. Les deux yeux s'accommodent et convergent vers le même emplacement (voir la figure 4.1.6a). Aucun conflit n'est donc introduit.

Lorsqu'est regardé un écran 3D (comme illustré à la figure 4.1.6b), la distance focale est différente de la distance de convergence. Accommodation et convergence peuvent donc être découplées. Lorsque la différence entre la distance focale et la distance de convergence est en dehors de la plage de tolérance, inconfort visuel et fatigue visuelle apparaissent. Pour les raisons exposées ci-avant, la disparité du contenu stéréoscopique doit être comprise dans une certaine plage, qui est appelée "zone de vision confortable" [Chen 10].

## 4.2 Quantifier comment la combinaison du flou et de la disparité affecte la profondeur perçue

Pour résoudre le problème causé par le conflit entre gagner de la profondeur apparente et éviter l'inconfort visuel causé par la disparité, une voie possible est de diminuer la disparité binoculaire tout en renforçant les indices monoculaires pour compenser la perte de profondeur perçue. Comme présenté précédemment, plusieurs indices monoculaires de la profondeur ont une incidence sur la perception de la profondeur. Parmi ces indices monoculaires de la profondeur, le flou est connu pour être important. Cependant, il n'existe toujours pas de consensus fort déterminant la façon dont la défocalisation du flou, qui agit sur la disparité binoculaire, influe sur la profondeur perçue. Certaines études antérieures ont montré clairement les contributions du flou à la perception de profondeur [Pentland 87, Watt 05, Hoffman 10, Held 10], tandis que d'autres ont montré que le flou n'a soit aucun effet, soit seulement certains effets qualitatifs sur l'ordonnement de la profondeur perçue [Marshall 96, Mather 96, Palmer 08].

Dans la suite de ce chapitre, nous présentons une étude quantifiant comment la combinaison de flou et de la disparité affecte la profondeur perçue. Dans la partie 4.2.1, nous décrivons les méthodes expérimentales. La partie 4.2.2 présente les résultats et l'analyse des expériences. Quelques conclusions générales sont présentées dans la partie 4.2.3.

### 4.2.1 Expérimentation

#### Appareils et stimuli

Les stimuli ont été affichés sur un écran LCD Samsung de 22,5 pouces ( $1680 \times 1050$  à 120 Hz). Chaque pixel de l'écran avait  $65,32$  secondes d'arc. Les stimuli ont été visualisés binoculairement à travers des lunettes à obturation active (kit de vision 3D NVIDIA) à une distance d'environ 90 cm.

Chaque stimulus se composait d'un seul objet au premier plan et d'un arrière-plan. Une image de  $400 \times 400$  pixels représentant un papillon a été utilisée comme objet de premier plan. L'arrière-plan est une photo d'un parterre de fleurs. Le cas échéant, le fond a été spatialement flouté en le filtrant avec un noyau de flou gaussien (voir la figure 4.2.1).

#### Conception du test et procédure

Pour chaque test, une paire de stimuli a été présentée à l'observateur. Un stimulus (BB-stimulus) contenait un arrière-plan flouté, l'autre stimulus (SB-stimulus) contenait un fond net (voir la figure 4.2.1). Les objets au premier plan de ces deux stimuli sont toujours nets.

Lorsque les deux stimuli ont été conçus, les arrière-plans des deux stimuli ont été positionnés à la même profondeur. Deux paramètres des stimuli ont variés : la position

absolue ( $D_a$ ) du plan du fond, et la distance relative ( $D_r$ ) de l'objet au premier plan par rapport à l'arrière-plan. Une fois que  $D_a$  et  $D_r$  ont été sélectionnés pour un BB-stimulus, alors on les associe, un par un, avec un ensemble de 7 ou 8 SB-stimuli. L'expérience consiste à faire 155 essais au total (les  $D_a$  et  $D_r$  pour les deux stimuli BB et SB sont présentés dans le tableau 4.2). Cette configuration est illustrée à la figure 4.2.2.

$D_a$ (cm)	$D_{r_{BB}}$ (cm)	$D_{r_{SB}}$ (cm)							
-19.7	0	0	1.2	2.3	3.4	4.5	5.6	6.6	
-19.7	6.6	3.4	6.6	7.6	8.6	9.6	10.6	11.5	13.4
-19.7	13.2	11.1	13.2	14.3	15.6	16.9	18.1	19.7	21.7
-19.7	19.7	17.3	19.7	20.9	22	23.5	24.9	26.6	28.3
-19.7	26.3	24.9	26.3	27	28	29.3	30.5	32	33.4
-19.7	32.8	30.5	32.8	34	35.1	36.2	37.5	38.7	39.9
-13.2	0	0	1	2	3	4	4.9	5.9	
-13.2	6.6	4.5	6.6	7.7	8.6	9.8	11.1	12.7	14.3
-13.2	13.2	11.5	13.2	14.3	15.4	16.5	17.6	18.7	19.7
-13.2	19.7	18.3	19.7	20.7	21.4	22.3	23.3	24.5	26
-13.2	26.3	24.5	26.3	27.1	27.9	29	30.1	31.4	15.4
-6.6	0	0	0.9	1.8	2.7	3.5	4.4	5.2	
-6.6	6.6	3.5	6.6	8	9.1	10.2	11.6	13.4	15.4
-6.6	13.2	12	13.2	14.4	15.4	16.3	17.3	18.5	19.6
-6.6	19.7	16.9	19.7	20.8	21.6	22.7	23.8	25.3	26.8
0	0	0	0.8	1.6	2.3	3.4	4.5	5.6	
0	6.6	4.5	6.6	7.6	8.6	9.5	10.8	12.3	14.3
0	13.2	11.4	13.2	13.7	14.3	15.1	15.9	17.2	18.7
6.6	0	0	1	2	2.9	3.9	4.8	6.3	
6.6	6.6	4.2	6.6	7.1	8	8.8	9.9	11.2	12.9

TABLE 4.2: La sélection des paramètres  $D_a$  et  $D_r$  pour les deux stimuli BB et SB. Notez qu'il y a seulement 7 sélections possibles dans le cas  $D_{r_{BB}} = 0$  cm, car mettre le premier plan derrière le fond provoquerait des troubles de la fusion.

A la figure 4.2.2, les plans bleus représentent les plans de profondeur (en avant) où le papillon se trouve. Le plan rouge représente le plan de profondeur où le fond est situé. La première figure montre le cas BB-stimulus, tandis que la seconde montre le cas SB-stimulus. Chaque BB-stimulus est associé à l'un des 7 ou 8 SB-stimuli afin de créer un seul test.

L'expérience subjective qui a été réalisée était une tâche de choix forcé avec deux alternatives. Trente-cinq observateurs ont participé à l'expérience. Les observateurs ont été invités à déterminer si le BB-stimulus présentait un intervalle de profondeur plus large entre le papillon et le fond, par rapport au cas du SB-stimulus.



FIGURE 4.2.1: Exemple de stimuli : les vues gauche d'un SB-stimulus et d'un BB-stimulus.

### 4.2.2 Résultats et analyse

Pour chaque condition, certains observateurs ont considéré que le BB-stimulus avait un intervalle de profondeur plus grand (entre le premier plan et l'arrière-plan), tandis que les autres observateurs ont choisi le SB-stimulus. Nous mesurons la proportion de réponses "Le BB-stimulus contient un intervalle de profondeur plus large». Le résultat est tracé comme une fonction de la différence de  $D_r$  entre les stimuli BB et SB.

La fonction cumulative de Weibull est utilisée comme fonction psychométrique. La différence de disparité correspondant au point à 50% est considérée comme le point d'égalité subjective (PSE). La mesure de la différence de disparité en ce point, donne l'augmentation de la profondeur perçue. Un exemple est illustré à la figure 4.2.3.

Nous avons tracé également la courbe du PSE en fonction de l'intervalle de profondeur (en unité de longueur) entre l'objet au premier plan et l'arrière-plan (voir la figure 4.2.4a). A la figure 4.2.4a, l'axe des Y montre de combien la disparité peut être compensée par le flou (par augmentation de la profondeur perçue). Comme le montre la figure, l'augmentation de la profondeur perçue est approximativement égale à 300 secondes d'arc. Nous constatons également que l'augmentation de la profondeur perçue est similaire quel que soit le plan de profondeur auquel on ajoute flou. En outre, nous pouvons également constater que le flou peut mieux compenser la disparité si le flou est ajouté à un plan de profondeur plus loin de l'avant-plan (voir la figure 4.2.4b).

### 4.2.3 Conclusion

L'influence du flou de défocalisation sur la profondeur perçue dans une scène stéréoscopique et sa relation avec la disparité binoculaire ont été étudiées. Le résultat expérimental indique que le floutage d'image contribue à la perception de la profondeur des images stéréoscopiques. Cette possibilité d'améliorer la profondeur perçue par flou de défocali-

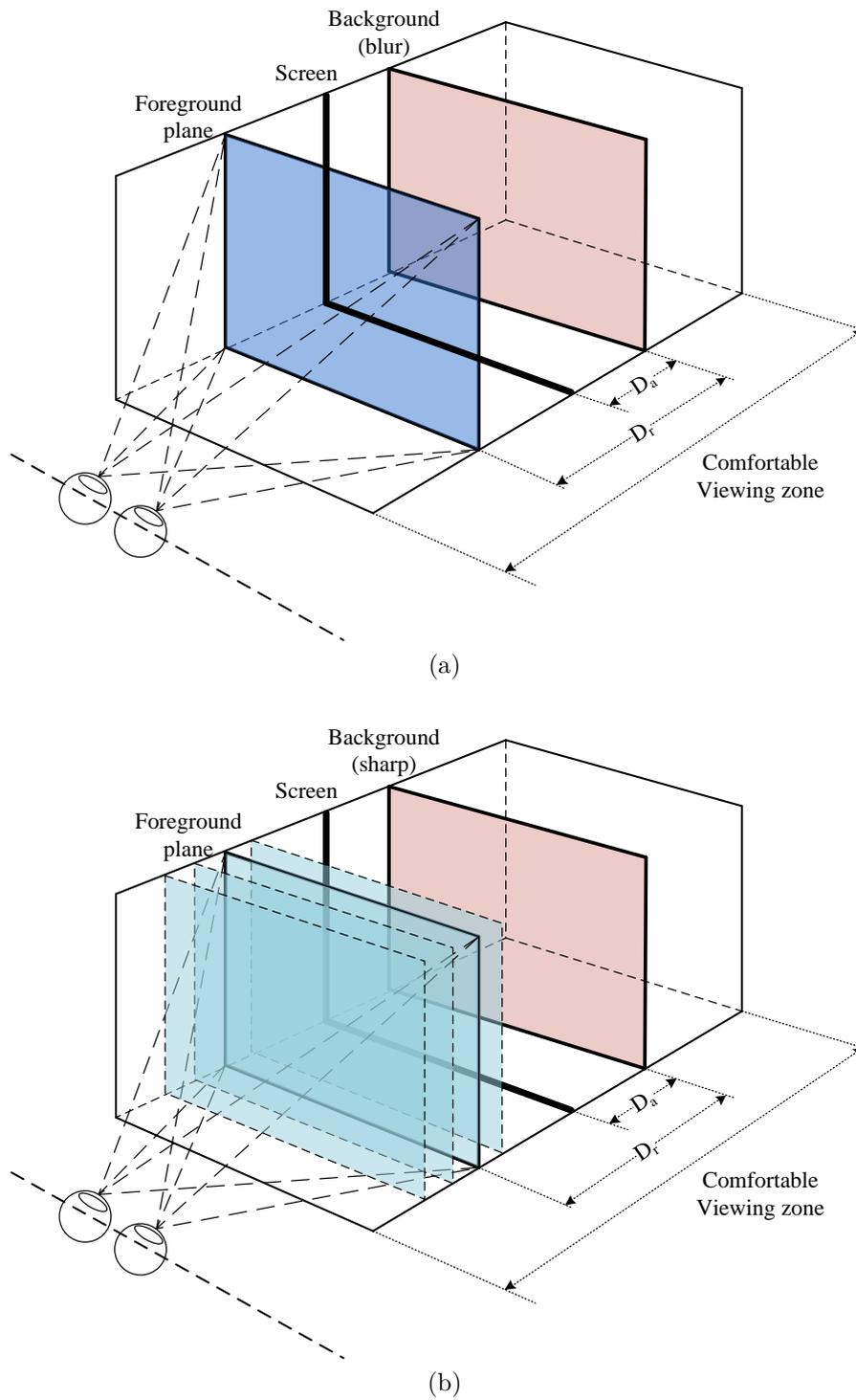


FIGURE 4.2.2: Schéma de principe du test expérimental.

4.2 Quantifier comment la combinaison du flou et de la disparité affecte la profondeur perçue

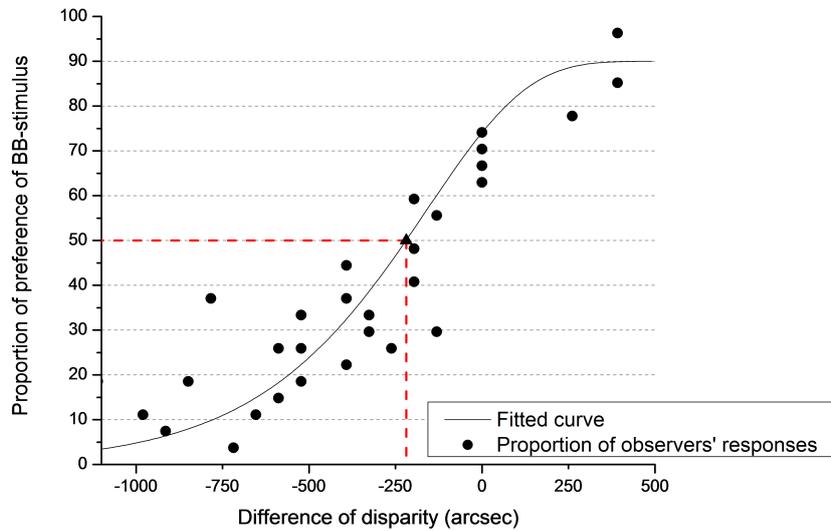


FIGURE 4.2.3: Un exemple de proportion de réponses d’observateurs avec la fonction psychométrique associée. Dans cette étude, nous considérons  $D_{r_{BB}} = 6.6$  cm et  $D_a = -19.7$  cm,  $-13.2$  cm,  $-6.6$  cm,  $0$  cm,  $6.6$  cm. Une profondeur apparente égale est atteinte à  $-220$  secondes d’arc, ce qui correspond au deuxième point de la figure 4.2.4b.

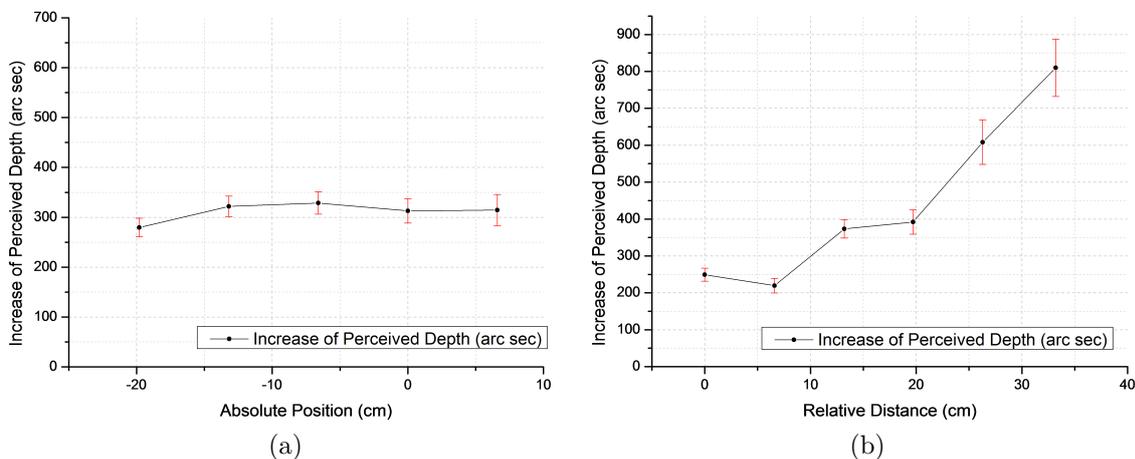


FIGURE 4.2.4: Résultats de l’expérience. La figure de gauche montre l’augmentation de la profondeur perçue comme une fonction de la position absolue, tandis que la figure de droite montre l’augmentation de la profondeur perçue comme une fonction de la distance relative. Notez que chaque point d’une courbe peut être obtenu à partir d’un nombre différent d’observations.

sation pourrait fournir un moyen intéressant de traiter le conflit entre accommodation et vergence. Nos résultats montrent qu'il est possible de diminuer la disparité des objets sans perdre l'effet "pop-out" en ajoutant un effet de flou sur l'arrière-plan.

## Chapitre 5

# Influence de la profondeur sur l'attention visuelle 3D : une étude du biais de profondeur

### 5.1 Introduction

Dans les études liées au déploiement de l'attention visuelle sur écran plat, il a été constaté que les fixations des observateurs présentent un biais marqué vers certaines zones de l'écran. Lors de la visualisation d'images 2D ou de vidéos, un "biais au centre" (ou "biais de fixation centrale") a été mis en évidence : les fixations du regard sont biaisées vers le centre de la scène [Tseng 09, Tatler 07]. Cependant, lors de la visualisation de contenus 3D (sur écran plat stéréoscopique), il a été récemment montré que les fixations des observateurs sont biaisées, non seulement vers la zone centrale de l'écran, mais aussi vers certains plans de profondeur de la scène [Jansen 09, Wang 11b, Ramasamy 09]. Il est donc raisonnable de supposer l'existence d'un soi-disant «biais de profondeur».

Dans le domaine de l'élaboration de modèles calculatoires de l'attention visuelle en visualisation stéréoscopique, des hypothèses de biais de profondeur ont été proposées (voir [Maki 96, Maki 00, Zhang 10, Chamaret 10]). Cependant, des études psychophysiques de ce biais lors de l'affichage sur écran plat stéréoscopique font encore défaut.

Dans ce chapitre, nous présentons une expérience oculométrique binoculaire. Cette expérience a été menée en montrant des stimuli synthétiques sur un écran stéréoscopique récent. Vingt-sept observateurs ont participé à l'expérience. Les stimuli ont été conçus de façon à éviter le plus grand nombre possible de caractéristiques visuelles 2D. Nos résultats montrent que les objets proches de la scène attirent plus de fixations, surtout au tout début de l'observation. Le nombre de fixations varie en fonction de la profondeur de l'objet. Ce type de distribution de fixation est aussi dépendant du temps.

## 5.2 Expérimentation

### 5.2.1 Participants

Vingt-sept sujets ont participé à cette expérience. Les sujets étaient âgés de 18 à 44 ans, et tous étaient “naïfs” (quant aux fins de l’expérience). Tous avaient une acuité normale (sans ou avec correction).

### 5.2.2 Conditions d’affichage

Les stimuli ont été affichés sur un écran LCD stéréoscopique Panasonic BT-3DL2550 de 26-pouces ( $552 \times 323$  mm). L’écran avait une résolution de  $1920 \times 1200$  pixels. La distance d’observation a été fixée à 93 cm. Un oculomètre SMI RED 500 a été utilisé pour enregistrer les mouvements oculaires. La précision de cette oculomètre est de 0,4 degré. Une mentonnière a été utilisée pour stabiliser la tête de l’observateur. Les observateurs ont été invités à «regarder n’importe où sur l’écran comme ils le voulaient».

Les 118 scènes ont été présentées dans un ordre aléatoire. Chaque scène a été présentée pendant 3 secondes. Après chaque scène, un point situé au centre de l’écran a été présenté pendant 500 ms et sans disparité.

### 5.2.3 Stimuli

L’expérience a consistée à la présentation de scènes stéréoscopiques dans lesquelles certains objets identiques ont été présentés à différents plans de profondeur.

L’arrière-plan est une image plane composée d’un bruit blanc (voir la figure 5.2.1a) qui a été placé à une profondeur de -20 cm. Dans chaque scène, les objets consistent en un ensemble de disques noirs de même diamètre  $S$ . Ils ont été affichés à différentes valeurs de profondeurs choisies au hasard parmi  $\{-20, -15, -10, -5, 0, 5, 10, 15, 20\}$  cm (voir la figure 5.2.1c).

Pour l’affichage stéréoscopique, la profondeur perçue est obtenue par déplacement horizontal de l’objet dans les vues gauche et droite afin de simuler la disparité binoculaire. La plage de profondeur (de -20 cm à 20 cm) a été choisie afin de correspondre à la zone de vision confortable [Chen 10, Hoffman 08].

Pour générer les différents stimuli, trois paramètres ont variés indépendamment d’une scène à l’autre :

- Le nombre d’objets,  $N \in \{5, 6, 7, 8, 9\}$ .
- Le rayon  $R$  du cercle sur lequel les objets sont projetés sur le plan de l’écran,  $R \in \{200, 250, 300\}$  pixels.
- La taille des objets, qui est fixée par le diamètre du disque  $S$  variant de 48 pixels à 168 pixels par pas de 12 pixels.

La figure 5.2.2 donne des exemples de différentes scènes.

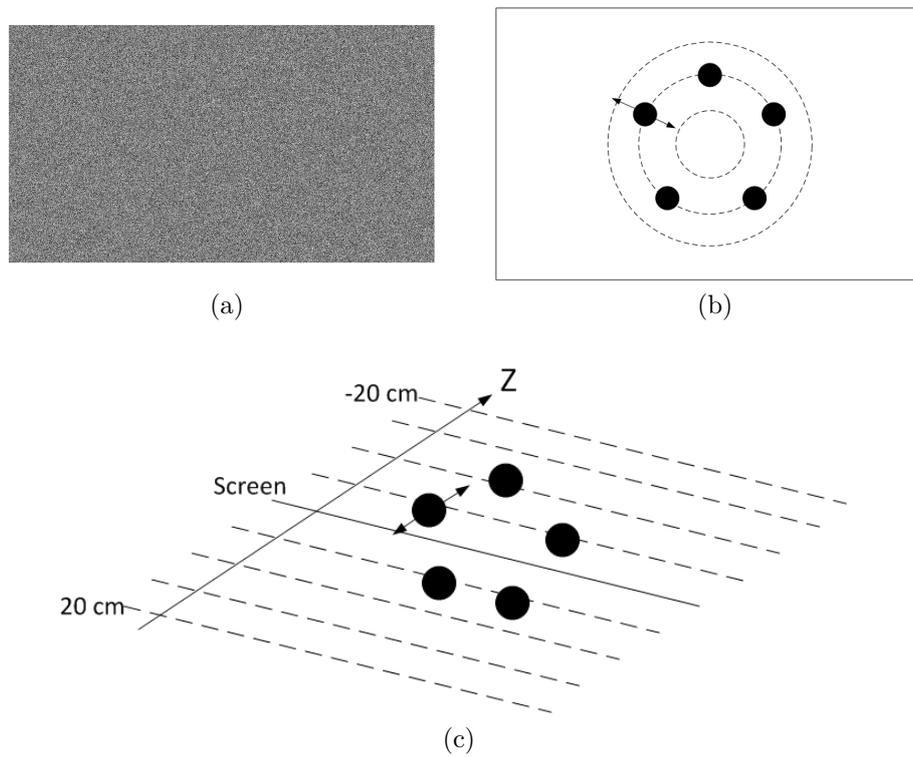


FIGURE 5.2.1: Composition du stimuli. (a) Arrière-plan. (b) Positions des projections des objets sur l'écran. (c) Distribution des objets en profondeur.

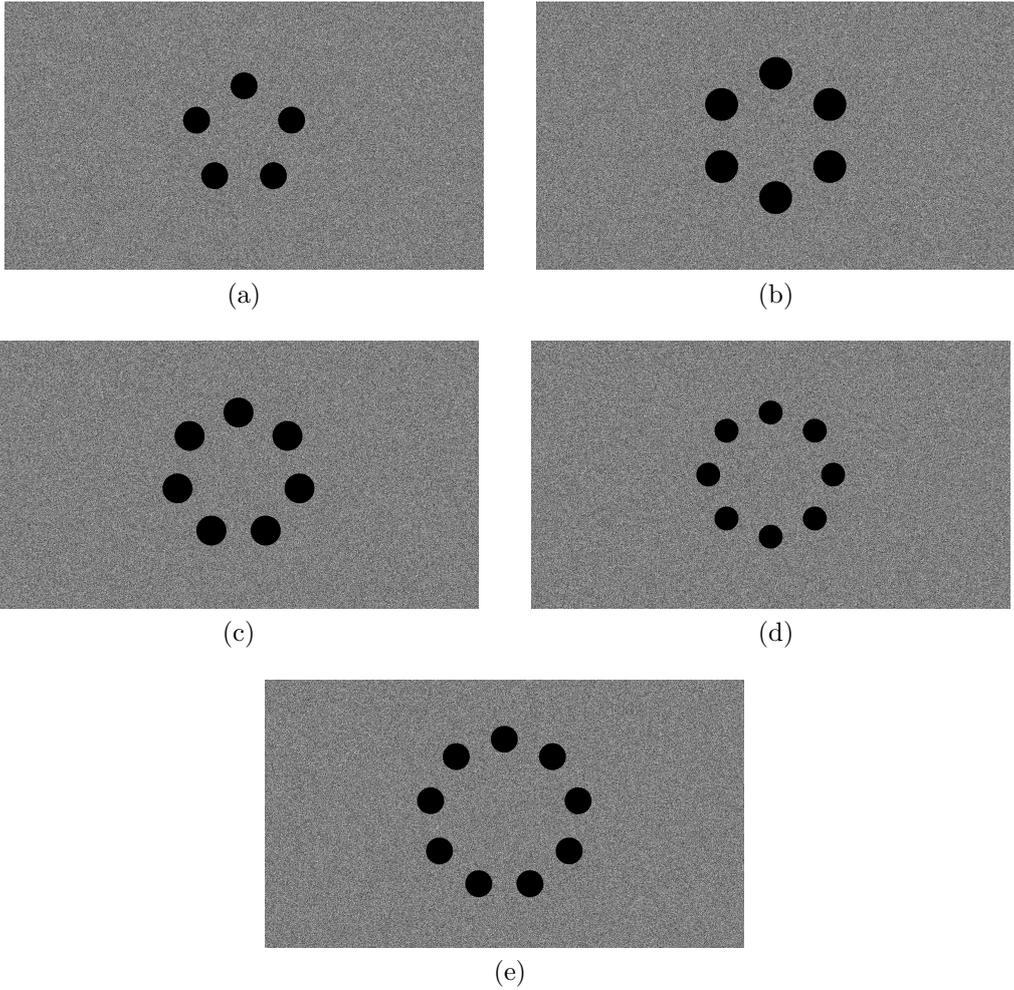


FIGURE 5.2.2: Exemples de cinq types de stimuli (en termes de nombre d'objets contenus) utilisés dans l'expérience oculométrique.

Il y avait plusieurs avantages à utiliser ce type de stimuli synthétiques stéréoscopiques pour l'étude du biais de profondeur :

- Il a permis un contrôle précis de l'influence des caractéristiques visuelles 2D sur l'attention visuelle. Dans notre expérience, tous les objets sont de forme constante, de taille constante, et ont été positionnés à une distance constante par rapport au centre de l'écran. Ce procédé permet de se débarrasser du plus grand nombre possible de caractéristiques "bottom-up" de l'attention visuelle.
- La disparité était le seul indice de profondeur utilisé dans cette expérience. Par conséquent, cette configuration a permis un contrôle précis de la perception de la profondeur.
- Le bruit blanc en arrière-plan et l'allocation simple d'objets ont limité la complexité des scènes.

#### 5.2.4 Post-traitement des données de suivi oculaire

La première étape du traitement était d'identifier les fixations et de filtrer les saccades. Les données enregistrées des mouvements oculaires ont été traitées par le logiciel de détection "BeGazed" fourni par SMI.

La deuxième étape consistait à déterminer la position spatiale de chaque fixation afin de la relier aux objets présents dans la scène. Le calcul a été fait indépendamment pour les deux yeux. Les positions de l'œil gauche (resp. droite) ont été appariées au stimuli gauche (resp. à droite). Il a ensuite été vérifié si une fixation était située sur l'un des objets ou non (voir la figure 5.2.3).

Les fixations des deux yeux ont ensuite été fusionnées selon la règle suivante : un objet donné a été considéré comme étant fixé si la fixation d'au moins un œil était à l'intérieur de cet objet (figure 5.2.3c). La profondeur de chaque objet étant connue, la profondeur d'une fixation pouvait être déduite de sa position.

## 5.3 Résultats

### 5.3.1 Distribution des fixations en profondeur

Les nombres de fixations situées sur chaque objet ont été calculés pour chaque observateur et chaque scène. Le résultat a ensuite été transformé en une distribution de fréquence. Nous avons considéré la distribution de probabilité uniforme  $P_r = 1/N$  comme référence. La figure 5.3.1 montre comment les fixations sont réparties sur des objets situés sur des plans de profondeur différents dans une scène.

Nous avons trouvé que, quel que soit le nombre d'objets contenus dans la scène, l'objet le plus proche de l'observateur attire toujours la plupart des fixations. La fréquence des fixations sur les objets situés dans la partie avant de la scène est sensiblement plus élevée que celle donnée par la loi de distribution uniforme.

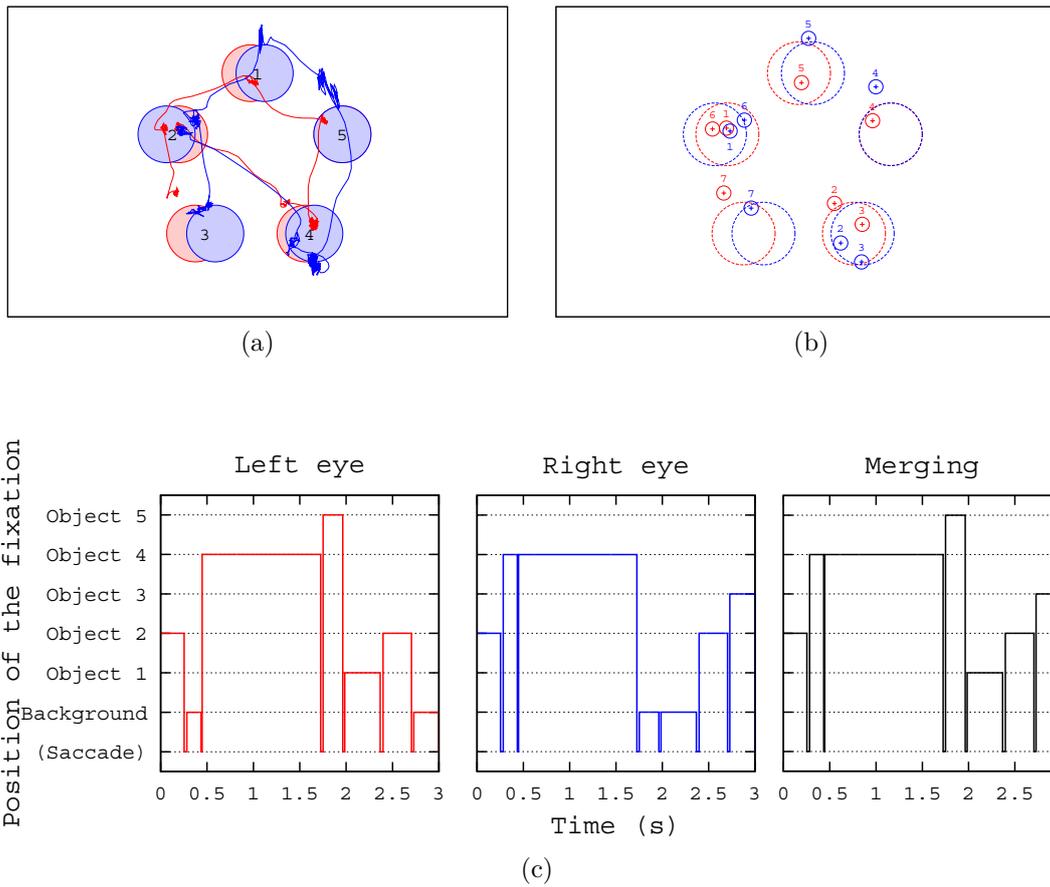


FIGURE 5.2.3: Illustration du processus de fixation d'un objet. En (a), la couleur rouge représente le chemin de l'œil gauche et le stimuli sur la vue gauche, tandis que la couleur bleue représente le chemin de l'œil droit et le stimuli sur la vue droite. (b) montre les fixations de chaque œil. Les chiffres indiquent l'ordre temporel des fixations. (c) représente la position de chaque fixation en fonction du temps.

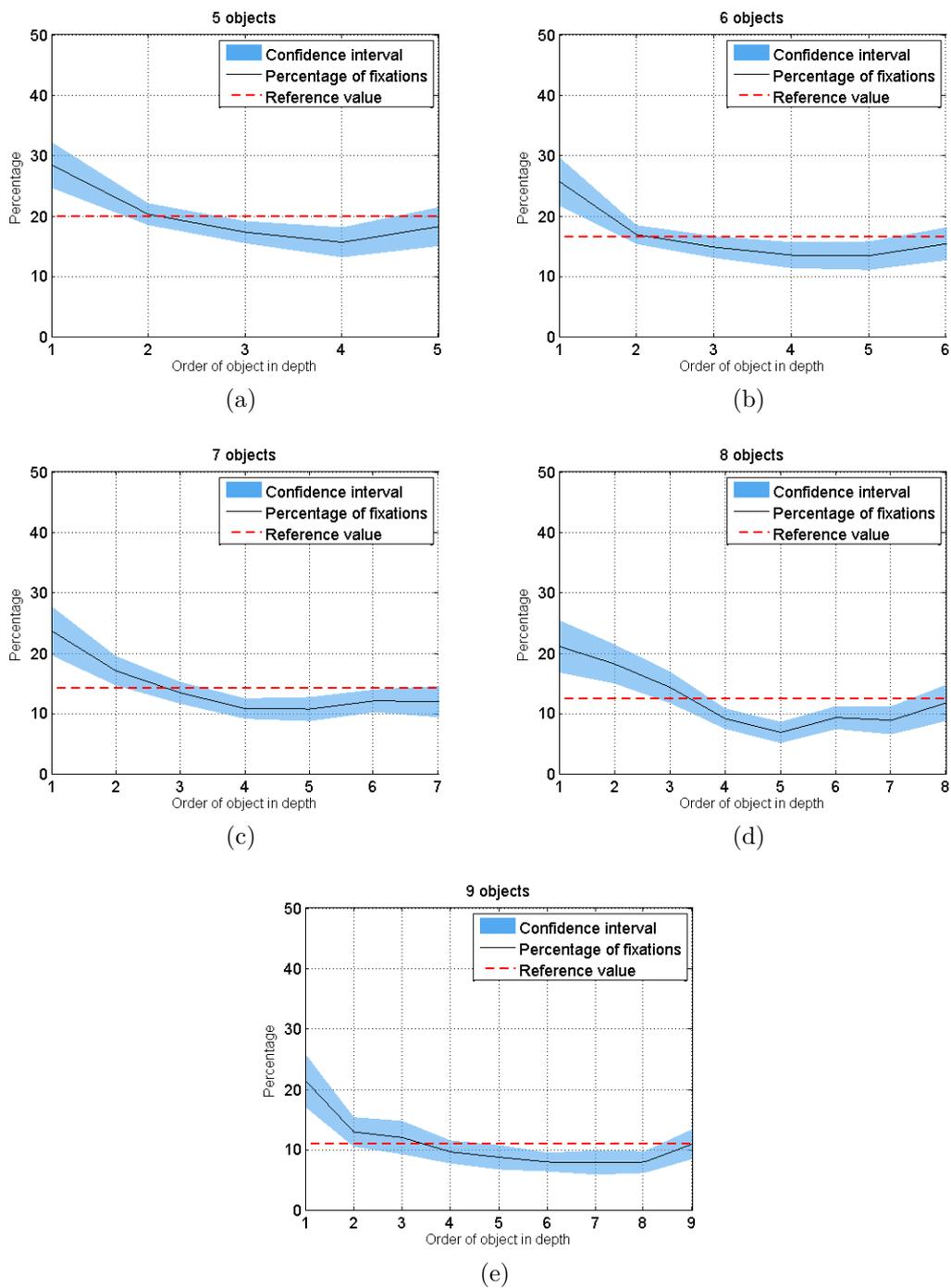


FIGURE 5.3.1: Distribution des fixations (toutes les fixations sont considérées) en fonction de l'ordre de profondeur d'objets pour des scènes contenant un nombre différent d'objets ( $N \in 5, 6, 7, 8, 9$ ).

### 5.3.2 Variation de la profondeur de fixation en fonction de l'ordre temporel de fixation

Les courbes de la figure 5.3.2 montrent comment la première fixation pour tous les observateurs est répartie sur les objets de chaque type de stimuli. Ces courbes ont été calculées de la même façon que celle présentée dans le paragraphe précédent, sauf que seule la première fixation de chaque observation a été prise en compte. Ces courbes indiquent le degré de biais de profondeur pour une durée de visionnage bref au tout début de l'observation.

Afin de mieux évaluer l'évolution temporelle de la profondeur moyenne des fixations, nous avons étudié la façon dont la profondeur moyenne des fixations varie en fonction du temps. La position de profondeur relative de chaque fixation a été calculée par l'équation 5.3.1

$$D_{r_i} = (D_i - D_{min}) / (D_{max} - D_{min}) \quad (5.3.1)$$

où  $D_i$  est la profondeur absolue de la  $i^{\text{e}}$  fixation,  $D_{min}$  et  $D_{max}$  sont respectivement la profondeur absolue minimale et maximale des objets dans la scène. Les résultats sont présentés à la figure 5.3.3. A la figure 5.3.3, une réponse frontale initiale à une nouvelle scène est observée .

### 5.3.3 Dépendance temporelle de la distribution de la fixation en profondeur

Afin de vérifier la dépendance temporelle de la distribution de fixation, nous avons séparé uniformément le temps d'observation de 3 secondes en six tranches, puis la distribution de fixation en fonction de la profondeur a été calculée pour chaque tranche de temps. Ce traitement a été répété pour l'ensemble des cinq types de stimuli. Les résultats sont montrés à la figure 5.3.4.

Comme le montre la figure 5.3.4, pour les cinq types de scène, un biais de profondeur évident apparait dans les 1000 premiers ms du temps d'observation. Avec l'augmentation de temps d'observation, la distribution des fixations sur tous les objets d'une scène devient plus uniforme.

## 5.4 Discussion

Dans cette étude, nous avons mené une expérience utilisant un système performant mettant en oeuvre affichage stéréoscopique et oculomètre. Les résultats expérimentaux montrent clairement que les observateurs prêtent plus et plus tôt attention aux objets les plus proches d'eux. Ce phénomène pourrait être causé par une stratégie visuelle où les gens préfèrent explorer une scène à partir des objets les moins distants, en effet ces objets peuvent représenter dans la nature un danger potentiel [Bowler 89]. Nous avons trouvé

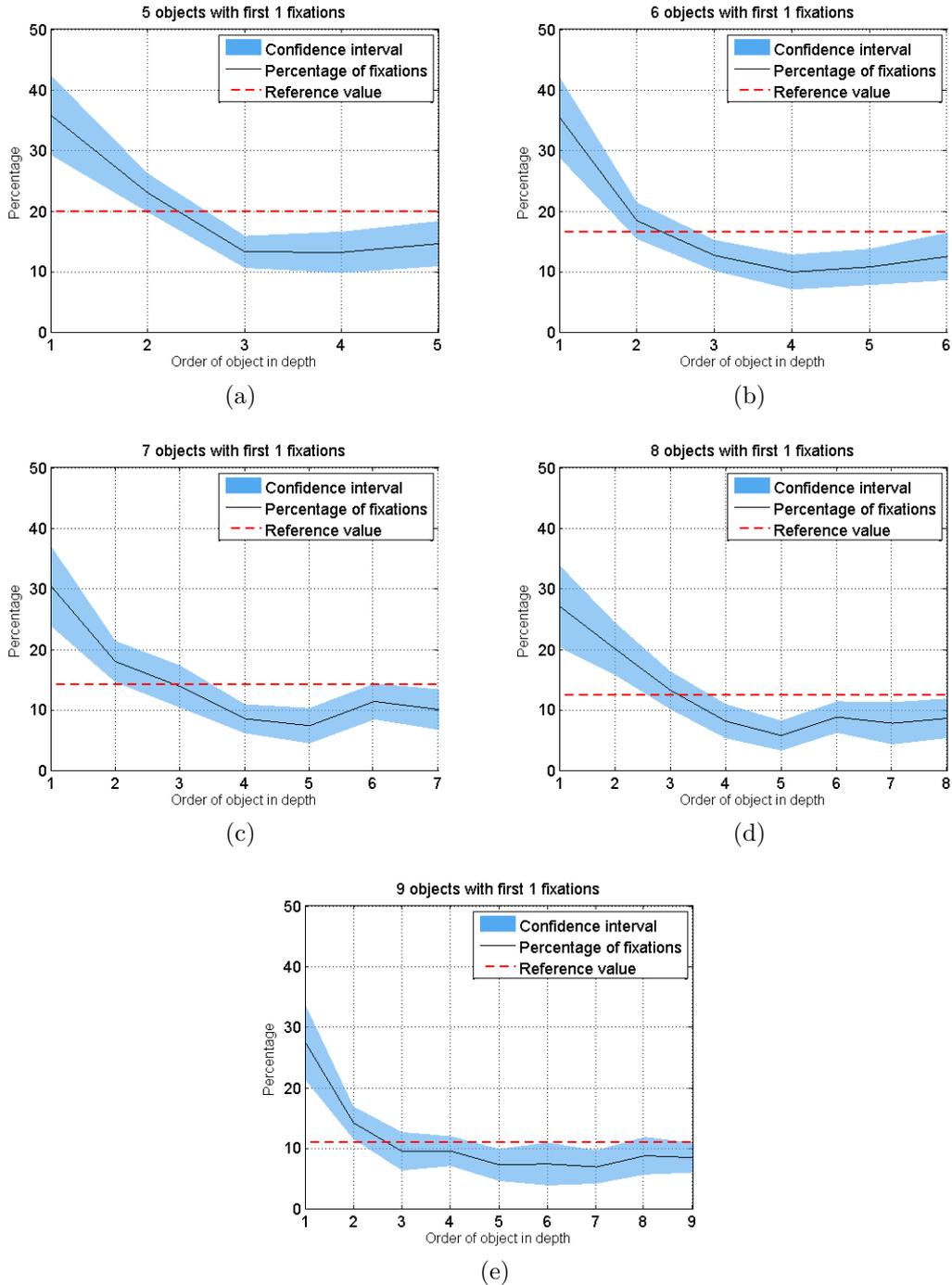


FIGURE 5.3.2: Distribution des fixations (seules les premières fixations sont considérées) en fonction de l'ordre de profondeur d'objets pour des scènes contenant un nombre différent d'objets. La zone bleue représente l'intervalle de confiance à 95%. La ligne en pointillés représente la distribution de probabilité uniforme ( $1/N$ ).

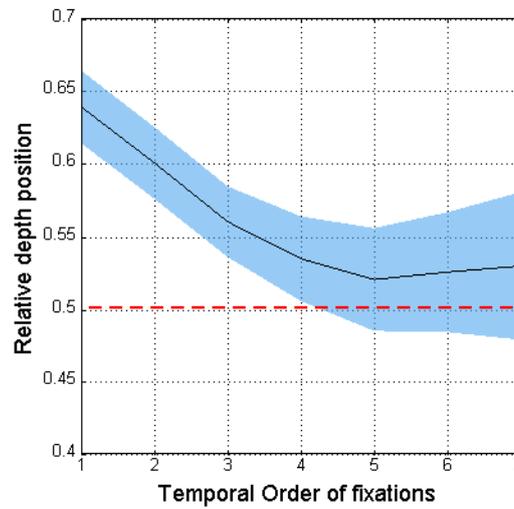


FIGURE 5.3.3: Position des profondeurs relatives des fixations fonction de l'ordre temporel. La ligne en pointillés rouge indique la valeur moyenne de la profondeur relative de tous les objets affichés au cours de l'expérience. La région bleue représente l'intervalle de confiance à 95%.

que le biais de profondeur était évident dès le début de l'observation (i.e. la première fixation de chaque observation). Ce biais initial et le court temps de réaction impliquent que le biais de profondeur pourrait être le résultat d'un mécanisme "bottom-up".

Comme il a été démontré l'existence de neurones sélectifs en disparité dans le cortex visuel primaire (V1) [Neri 99, Barlow 67, Nikara 68, Poggio 77], la valeur absolue de la disparité binoculaire peut être un autre facteur important qui influe sur le biais de profondeur.

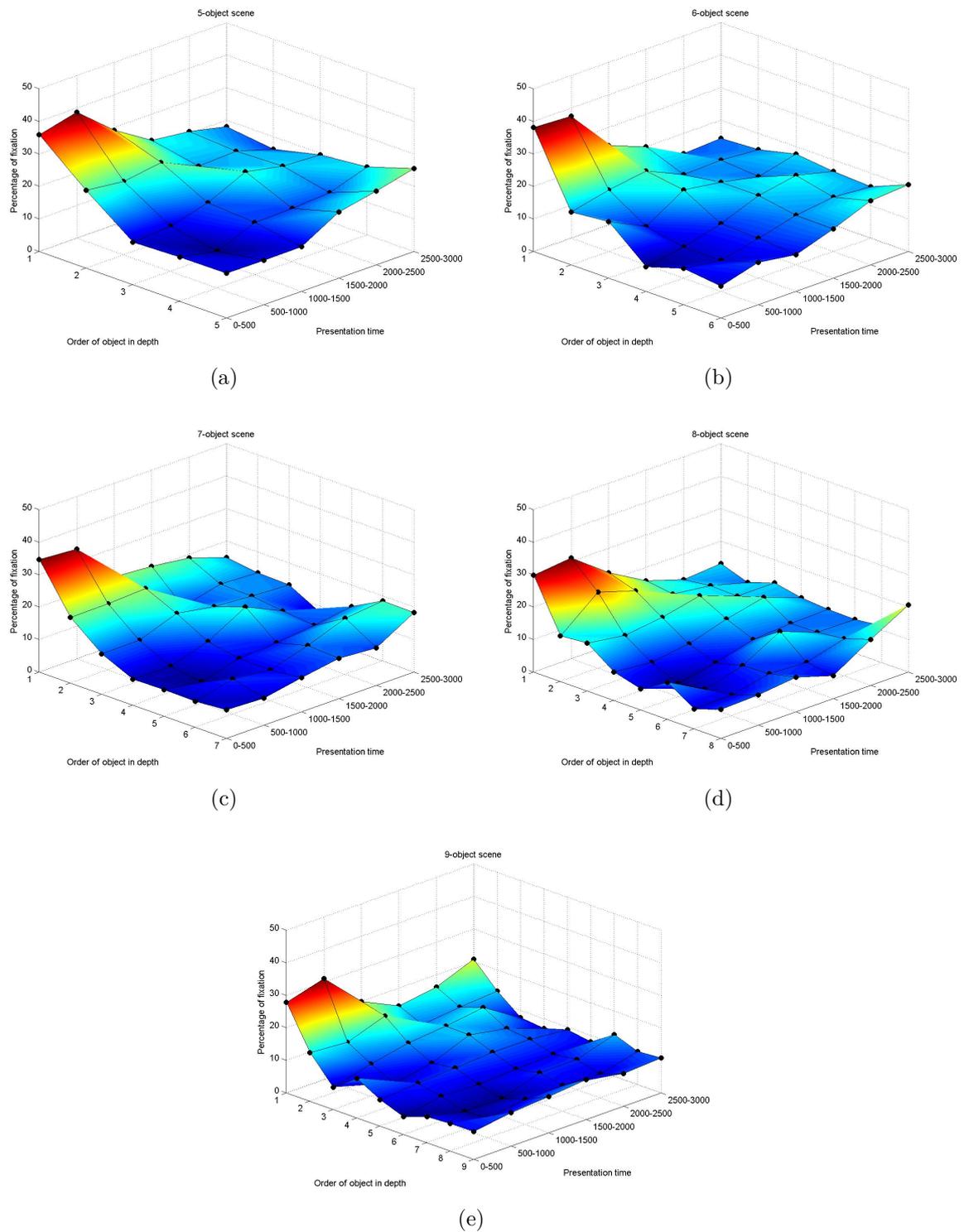


FIGURE 5.3.4: Variation de la distribution de fixation comme une fonction du temps. Les cinq sous-figures représentent cinq types de scènes qui contiennent des nombres différents d'objets.



# Chapitre 6

## Modèle de calcul de l'attention visuelle 3D

Un point essentiel pour les études concernant les modèles de calcul de l'attention visuelle 3D, est le développement de modèles efficaces qui peuvent bien prédire la distribution de saillance. Après avoir présenté les études concernant (1) la “réalité de terrain” de l'attention visuelle 3D et, (2) l'impact de la perception de la profondeur sur l'attention visuelle, nous mettrons particulièrement l'accent dans ce chapitre sur la modélisation de l'attention visuelle 3D. Un nouveau modèle de l'attention visuelle 3D s'appuyant à la fois sur les caractéristiques visuelles 2D et les caractéristiques extraites de l'information de profondeur est proposé et évalué.

Dans ce chapitre, nous présentons un modèle de l'attention visuelle 3D, modèle basé sur la saillance de profondeur. Afin de tirer bénéfice des études psychophysiques, nous avons proposé d'appliquer le théorème de Bayes sur le résultat d'une expérience oculométrique utilisant des stimuli synthétiques, pour modéliser la corrélation entre les caractéristiques de profondeur et le niveau de saillance de profondeur. Nous avons également présenté et discuté deux façons de combiner la carte de saillance de profondeur avec la carte de saillance 2D. A la fin de ce chapitre, une conclusion et une discussion sont présentées. 6.5.

### 6.1 Introduction

Pendant la visualisation de contenu 3D stéréoscopique, le renforcement de la perception de la profondeur affecte beaucoup l'attention visuelle humaine [Hakkinen 10, Huynh-Thu 11b]. Prédire les zones saillantes d'une scène 3D devient alors une tâche difficile en raison de l'information supplémentaire de profondeur. Cependant, la demande croissante d'applications liées au contenu 3D et basées sur l'attention visuelle, accroît l'importance des modèles de calcul de l'attention visuelle 3D.

Deux questions doivent figurer lors du développement d'un modèle de l'attention visuelle 3D :

- Quelle est l'influence des caractéristiques visuelles 2D (par exemple, la couleur, l'intensité, l'orientation, le biais central) en condition de visualisation 3D ?
- Quelle est l'influence de la profondeur sur le déploiement de l'attention visuelle en condition de visualisation 3D ? Par exemple, il est nécessaire de comprendre comment le biais de fixations en profondeur (ou "biais de profondeur"), et les caractéristiques visuelles basées sur l'information de profondeur (par exemple, l'orientation de la surface, le contraste en profondeur), affectent le déploiement de l'attention visuelle humaine.

Dans la dernière décennie, un grand nombre de modèles de l'attention visuelle 2D ont été étudiés. Par conséquent, une première possibilité est d'adapter ce grand nombre de modèles 2D existant au cas 3D. Une autre possibilité est de trouver le moyen de prendre en compte les informations additionnelles disponibles (celles liées à la profondeur).

La recherche en modélisation de l'attention visuelle 3D est également confrontée à un autre problème : la disponibilité d'une base de données oculométriques sur images fixes 3D fait encore défaut. En outre, il n'y a toujours pas de conclusion forte sur les moyens d'utiliser les informations de profondeur pour la modélisation de l'attention visuelle 3D : la profondeur doit-elle être utilisée pour pondérer les cartes de saillance 2D, ou bien être considérée comme une dimension visuelle supplémentaire où extraire des caractéristiques propres et créer une carte de saillance de profondeur.

## 6.2 Un modèle de calcul de l'attention visuelle 3D basé saillance de profondeur

Dans cette partie, nous introduisons une nouvelle méthode pour créer une "carte de saillance de profondeur". Cette carte est calculée en fonction du contraste de profondeur et en utilisant un cadre bayésien.

### 6.2.1 Une approche bayésienne pour la génération d'une carte de saillance de profondeur

Nous proposons donc une nouvelle approche appliquant le théorème de Bayes au calcul de cartes de saillance de profondeur en fonction des caractéristiques extraites de cartes de profondeur : la saillance de profondeur ( $S$ ) en chaque point (pixel de la carte) correspond à la probabilité de ce point d'être regardé, étant données les caractéristiques de profondeur observées en ce point et la localisation spatiale (3D) de ce point :

$$S = P(C = 1 | \bar{f}_{dep}, l_z) \quad (6.2.1)$$

où  $C$  est une variable binaire aléatoire indiquant si le point est regardé ou non. Le vecteur aléatoire  $\bar{f}_{dep}$  indique les caractéristiques de profondeur observées en ce point, et  $l_z$  indique sa position dans la scène 3D.

Si nous faisons l'hypothèse : 1) que les caractéristiques de profondeur de chaque point sont indépendantes de sa distance à l'observateur, et 2) que  $P(C = 1)$  est une simple constante, alors la probabilité peut être transformée :

$$\begin{aligned} S &= \frac{P(\bar{f}_{dep}, l_z | C = 1) \cdot P(C = 1)}{P(\bar{f}_{dep}, l_z)} \\ &= \frac{P(\bar{f}_{dep} | C = 1) \cdot P(l_z | C = 1)}{P(\bar{f}_{dep}) \cdot P(l_z)} \cdot P(C = 1) \\ &= P(C = 1 | \bar{f}_{dep}) \cdot P(C = 1 | l_z) \cdot const. \end{aligned} \quad (6.2.2)$$

Le premier terme de l'équation 6.2.2,  $P(C = 1 | \bar{f}_{dep})$ , représente la probabilité d'un point d'être regardé, étant données uniquement les caractéristiques extraites à partir de l'information de profondeur en ce point. Le second terme de cette équation,  $P(C = 1 | l_z)$ , représente la probabilité d'un point d'être regardé étant donnée sa distance à l'observateur, c'est à dire le biais de profondeur. Cependant, l'étude de ce biais de profondeur n'est pas l'objet de cette partie. Dans ce qui suit, nous nous concentrons que sur la modélisation de  $P(C = 1 | \bar{f}_{dep})$ .

En utilisant la règle de Bayes, nous obtenons :

$$S = P(C = 1 | \bar{f}_{dep}) = \alpha \cdot \frac{P(\bar{f}_{dep} | C = 1)}{P(\bar{f}_{dep})} \quad (6.2.3)$$

où  $\alpha$  est une valeur constante représentant la probabilité  $P(C = 1)$ . Par conséquent, on peut considérer que la saillance prédite est proportionnelle à la distribution de caractéristiques au point regardé, normalisée par la probabilité des caractéristiques (voir l'équation 6.2.3).

Pour réaliser le calcul de la carte de saillance de profondeur, l'approche proposée procède en deux étapes : (1) l'extraction des caractéristiques de profondeur, et (2) la modélisation de la distribution de probabilité.

### 6.2.1.1 Extraction des caractéristiques de profondeur

Le modèle proposé repose sur le contraste de profondeur. Un filtre de différence de gaussiennes (DOG) a été appliqué à la carte de profondeur pour extraire le contraste de profondeur. Le filtre DoG est donné par :

$$f(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) - \frac{1}{2\pi K^2\sigma^2} \exp\left(-\frac{x^2 + y^2}{2K^2\sigma^2}\right) \quad (6.2.4)$$

où  $(x, y)$  est la position du filtre.  $\sigma$  et  $K$  sont utilisés pour contrôler l'échelle du filtre DOG, et le rapport entre la "zone au centre du filtre" et celle "autour".

### 6.2.1.2 Modélisation de la distribution de probabilité

Dans le modèle proposé, la fonction  $P(C = 1|\bar{f}_{dep})$  modélise la relation entre la caractéristique de profondeur de chaque point et la probabilité que ce point soit regardé. Nous proposons de modéliser cette fonction en se basant sur une probabilité calculée à partir de données expérimentales (les mouvements oculaires recueillies lors d'une expérience oculométrique). Le détail de cette expérience oculométrique est présenté dans le chapitre précédent.

Les probabilités  $P(\bar{f}_{dep})$  et  $P(\bar{f}_{dep}|C = 1)$  sont obtenues à partir des cartes de caractéristiques de profondeur de stimuli synthétiques, et des positions des points regardés. La vraisemblance  $P(C = 1|\bar{f}_{dep})$  peut donc être calculée par l'équation 6.2.3. Pour prédire la carte de saillance de profondeur d'une image, on applique le modèle de  $P(C = 1|\bar{f}_{dep})$  sur sa carte de caractéristiques de profondeur. La valeur de saillance de profondeur de chaque pixel de l'image peut être ainsi calculée.

## 6.2.2 Un schéma pour le modèle calculatoire de l'attention visuelle 3D

Dans cette partie, nous présentons un schéma qui intègre la carte de saillance de profondeur aux cartes de saillance calculées sur la base de caractéristiques visuelles 2D, et qui réalise la prédiction de la carte de saillance 3D finale. L'architecture générale du schéma proposé est présentée à la figure 6.2.1.

Puisque le développement d'un modèle entièrement nouveau pour le calcul de l'attention visuelle 2D n'est pas ici notre objectif, nous confions la tâche de détection des caractéristiques visuelles et de création des cartes de saillance 2D à trois modèles existants : le modèle de Itti [Itti 98], celui de Bruce [Bruce 09] et celui de Hou [Hou 07].

Une étape de combinaison des cartes de saillance est nécessaire afin de fusionner les cartes de saillance obtenues à partir des différentes dimensions visuelles (par exemple, l'information de profondeur et les caractéristiques visuelles 2D). Pour le modèle proposé, nous adoptons une approche directe qui est la même que celle utilisée dans [Potapova 11] pour fusionner la carte de saillance de profondeur ( $SM_{dep}$ ) et la carte de saillance 2D ( $SM_{2D}$ ) : la carte de saillance finale  $SM_S$  est égale à la somme des deux cartes (voir l'équation 6.2.5) :

$$SM_S(i, j) = \omega_1 SM_{dep} + \omega_2 SM_{2D} \quad (6.2.5)$$

où  $\omega_1 = \omega_2 = 0.5$ .

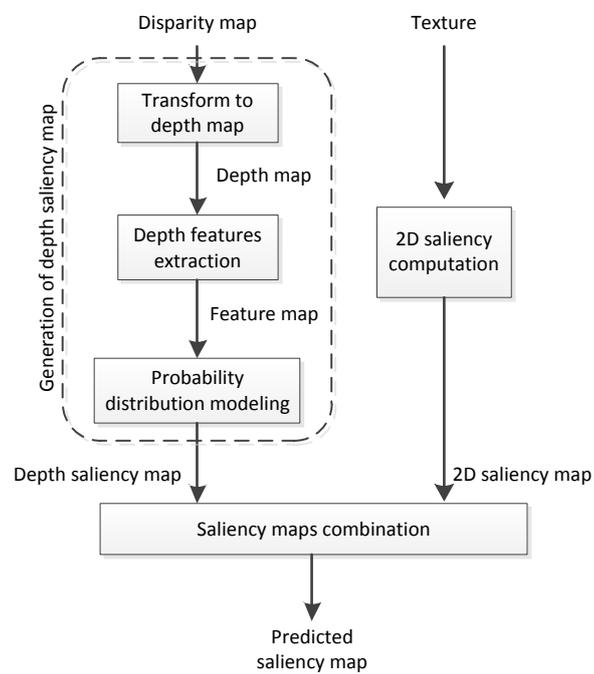


FIGURE 6.2.1: Schéma d'ensemble du modèle proposé.

## 6.3 Une base de données oculométriques de contenu d'images naturelles 3D

Jusqu'à présent, le manque de vérité de terrain a limité les études de modèles calculatoires de l'attention visuelle 3D. Pour évaluer la performance du modèle de cette étude, nous avons créé et publié une nouvelle base de données oculométriques contenant dix-huit images naturelles stéréoscopiques, les cartes de disparité correspondantes, et les données relatifs aux mouvements oculaires des deux yeux.

### 6.3.1 Stimuli

Les images stéréoscopiques utilisées pour la base de données proposée ont été acquises à partir de deux sources : (1) la base de données image 3D de Middlebury 2005/2006, et (2) la base de données image 3D de l'équipe IVC. Nous avons sélectionné 10 images de la base Middlebury 2005/2006 [Scharstein 07] et 8 images de la base IVC. Les images sont montrées aux figures 6.3.1 et 6.3.2.

### 6.3.2 Appareils et procédures

Les stimuli étaient affichés sur un écran LCD Panasonic BT-3DL2550 de 26 pouces (552×323 mm), ayant une résolution de 1920×1200 pixels. Un oculomètre SMI RED 500 a été utilisé pour enregistrer les mouvements oculaires. Une mentonnière a été utilisée pour stabiliser la tête de l'observateur.

Les dix-huit scènes (10 de la base Middlebury et 8 de la base IVC) ont été présentées dans un ordre aléatoire. Le temps de présentation de chaque scène était de 15 secondes. Entre deux scènes, un point a été affiché pendant 500 ms au centre de l'écran et sans disparité. Les sujets étaient invités à effectuer une visualisation libre de la scène.

### 6.3.3 Participants

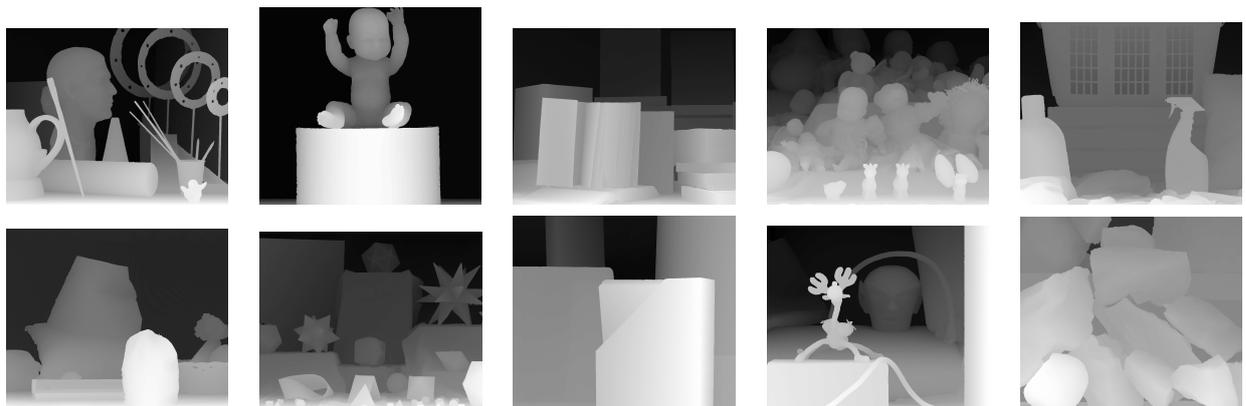
Trente-cinq personnes ont participé à l'expérience. Les sujets étaient âgés de 18 à 46 ans. L'âge moyen des sujets était de 24,2 ans. Tous les sujets avaient une acuité visuelle normale (avec ou sans correction).

### 6.3.4 Création des cartes de densité de fixation

Tous les points enregistrés par l'oculomètre, à la fois pour l'œil gauche et pour l'œil droit, ont été utilisés pour créer les cartes de densité de fixation. Les fixations de l'œil gauche et de l'œil droit ont été fusionnées en suivant la procédure décrite par [Huynh-Thu 11a]. La carte de fixation résultante a ensuite été filtrée par un noyau gaussien bidimensionnel pour tenir compte 1) de la diminution de l'acuité visuelle lorsque



(a)

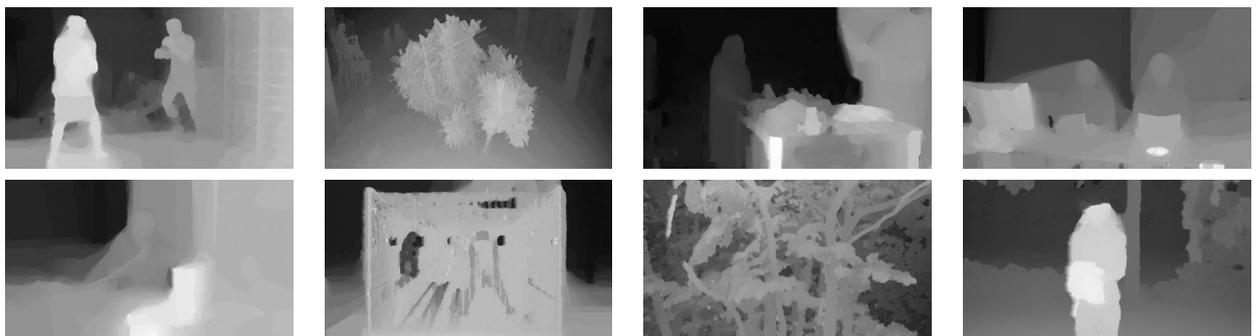


(b) The corresponding disparity maps

FIGURE 6.3.1: Exemple d'images 3D de la base de données Middlebury. (a) Images originales. Le numéro dans le coin en bas à droite indique l'index de l'image. (b) Cartes de disparité correspondantes.



(a)



(b)

FIGURE 6.3.2: Exemple d'images 3D de la base de données IVC. (a) Images originales. Le numéro dans le coin en bas à droite indique l'index de l'image. (b) Cartes de disparité correspondantes.

augmente l'excentricité depuis la fovéa, et 2) de la diminution de la précision de l'oculomètre. L'écart-type du noyau gaussien utilisé pour créer nos cartes de saillance était égal à 2 degrés d'angle visuel.

## 6.4 Evaluation des performances

Afin d'évaluer dans quelle mesure la carte de saillance de profondeur peut influencer la prédiction de la carte de saillance, et connaître la performance globale du modèle de calcul proposé, des comparaisons qualitatives et quantitatives entre la carte de densité de fixation et la sortie du modèle proposé sont réalisées.

### 6.4.1 Evaluation qualitative

La figure 6.4.1 donne des exemples de performance de cartes de saillance de profondeur (cartes de saillance prédites uniquement sur la base de cartes de profondeur) et des cartes de saillance prédites sur la base des seules caractéristiques visuelles 2D (à partir des trois modèles d'attention visuelle 2D introduits précédemment). La carte de densité de fixation générée à partir des données oculométriques est également fournie, c'est la vérité de terrain pour l'évaluation qualitative.

Sur le plan qualitatif, l'approche proposée crée des cartes de saillance de profondeur qui prédisent bien les zones saillantes des images 3D. Toutes les zones potentiellement saillantes sont reconnues comme telles dans les cartes de saillance de profondeur. Comparée à la carte de saillance de profondeur, la contribution des caractéristiques visuelles 2D pour la prédiction de la carte de saillance d'une image 3D, dépend en grande partie du modèle choisi.

### 6.4.2 Métriques quantitatives d'évaluation

Trois mesures de similarité sont utilisées pour comparer quantitativement les cartes de saillance avec les cartes de densité de fixation (la vérité de terrain) : (1) le coefficient de corrélation [Le Meur 06, Engelke 10], (2) la divergence de Kullback-Leibler [Le Meur 06, Bruce 09], et (3) l'aire sous la courbe ROC [Zhang 08, Zhao 11].

### 6.4.3 Performance la carte de saillance de profondeur seule

Les créations de cartes de saillance de profondeur et de cartes de saillance 2D sont les principales étapes du modèle proposé. Afin d'évaluer dans quelle mesure ces deux types de cartes de saillance peuvent prédire les zones saillantes, les performances de la carte de saillance de profondeur et des cartes de saillance 2D issues des trois modèles d'attention visuelle 2D, sont évaluées. Les résultats sont présentés au tableau 6.1, ils démontrent la grande influence de la caractéristique de profondeur.

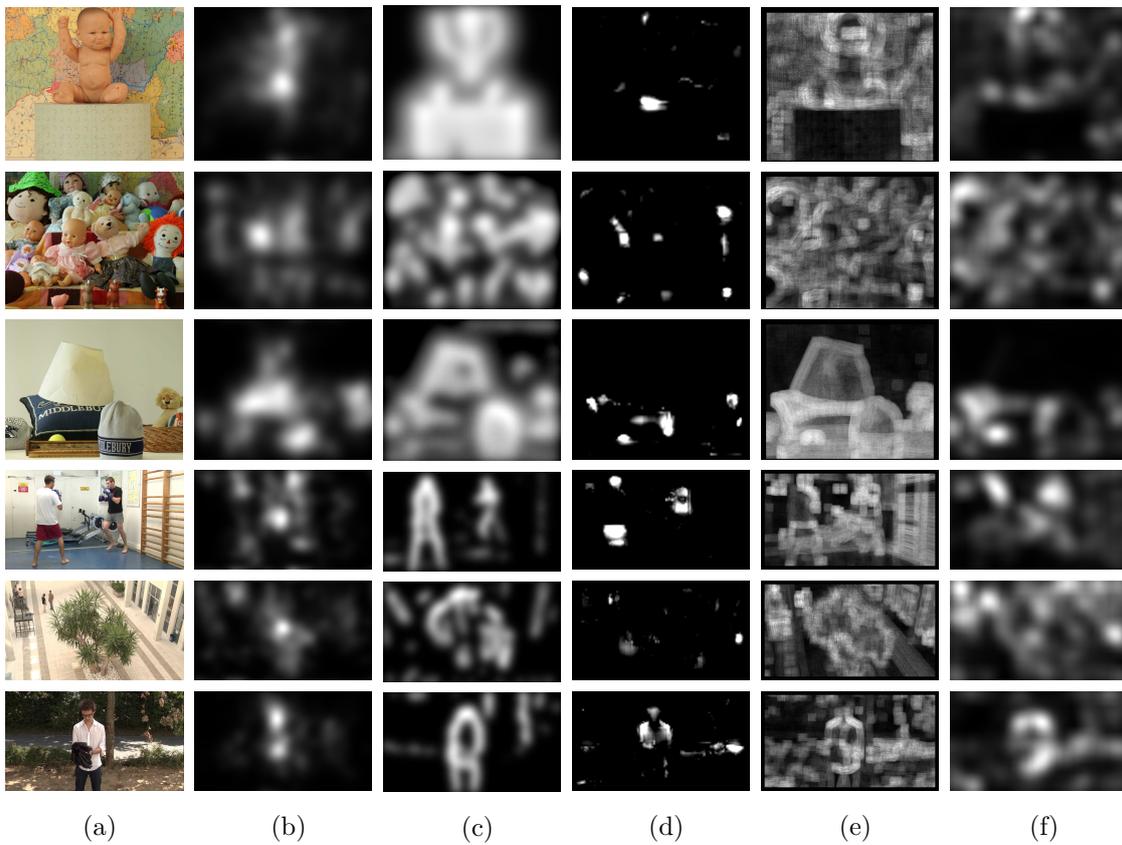


FIGURE 6.4.1: Exemples des performances des différents modèles, avec de gauche à droite : (a) l'image originale, (b) la carte de saillance obtenue à partir des données oculométriques, puis les cartes de saillance créées selon, (c) le modèle proposé, (d) le modèle de Itti, (e) le modèle de Bruce, et (f) le modèle de Hou.

	PLCC	KLD	AUC
DSM seul	0.368	0.708	0.656
Modèle de Itti	0.141*	2.809*	0.540*
Modèle de Bruce	0.325	0.735	0.638
Modèle de Hou	0.291	0.802*	0.630

TABLE 6.1: Performance de la carte de saillance de profondeur seule (notée DSM dans le tableau) et de trois modèles 2D de l'état de l'art. Notez qu'un plus petit score avec le KLD signifie une meilleure performance. \* indique un résultat significativement différent de la performance de la DSM (t-test par paire,  $p < 0.1$ ).

#### 6.4.4 Valeur ajoutée de la carte de saillance de profondeur

Le modèle proposé appartient à la catégorie des “modèles de saillance de profondeur” qui mettent donc en avant l’exploitation de la carte de saillance de profondeur. Pour comparer les deux façons de tirer parti des informations de profondeur, les performances des méthodes suivantes ont été mesurées et comparées :

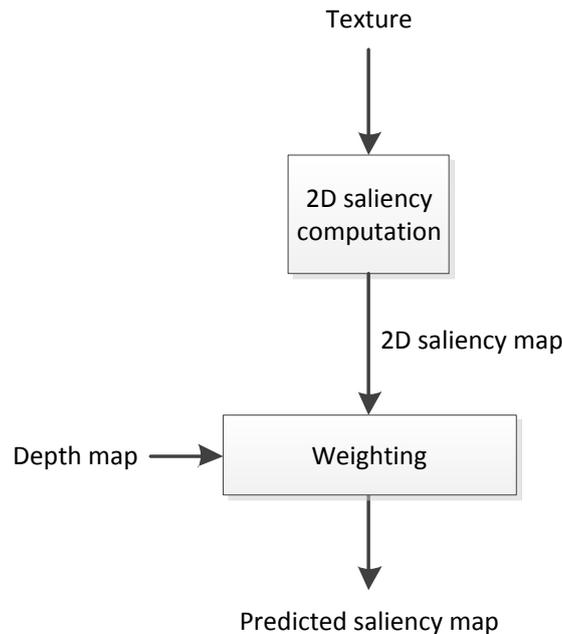


FIGURE 6.4.2: Schéma de la modélisation l’attention visuelle 3D selon la méthode décrite par Chamaret.

- Méthode n’utilisant pas la profondeur. Il s’agit d’une application directe du modèle de calcul 2D, aucune information de profondeur n’est prise en compte.
- Méthode de pondération fonction de la profondeur (DW). Nous adoptons la méthode de Chamaret [Chamaret 10], qui pondère chaque carte de saillance 2D en multipliant chaque valeur de pixel de la carte par sa profondeur (voir la figure 6.4.2). Nous n’avons pas codé exactement le même modèle de calcul que celui utilisé dans l’article, car la carte de saillance 2D que nous utilisons provient du modèle de Itti, de Bruce, ou de Hou.
- Méthode de saillance de profondeur (DS), c’est le modèle de calcul que nous proposons. Il crée d’abord la carte de saillance de profondeur et celle de saillance 2D, puis les combine pour obtenir la carte de saillance finale.

Les performances de ces trois méthodes sont présentées au tableau 6.2. La valeur ajoutée de la carte de saillance de profondeur est mise en évidence pour les trois modèles d’attention visuelle 2D.

		PLCC	KLD	AUC
Modèle de Itti	Modèle 2D seul	0.141	2.809	0.540
	Méthode DW (Chamaret)	0.140	2.892	0.540
	Méthode DS (proposée)	0.356*	0.701*	0.656*
Modèle de Bruce	Modèle 2D seul	0.325	0.735	0.638
	Méthode DW (Chamaret)	0.311	0.810	0.639
	Méthode DS (proposée)	0.423*	0.615	0.674
Modèle de Hou	Modèle 2D seul	0.291	0.802	0.630
	Méthode DW (Chamaret)	0.290	0.878	0.633
	Méthode DS (proposée)	0.409	0.603*	0.669

TABLE 6.2: Apport de l'information de profondeur aux modèles 2D. Notez qu'un plus petit score de KLD signifie une meilleure performance. \* signifie une différence significative par rapport à la performance du modèle 2D correspondant (t-test par paire,  $p < 0.1$ ).

## 6.5 Conclusion et discussion

Des études concernant les modèles de calcul de l'attention visuelle 3D sont présentées dans ce chapitre. Pour des images fixes 3D, nous proposons un modèle de calcul de l'attention visuelle 3D basé sur la saillance de profondeur. Deux façons différentes d'appliquer l'information de profondeur au sein du modèle de l'attention visuelle 3D sont également comparées dans notre étude. Nos résultats montrent que, en créant une carte de saillance de profondeur basée sur des caractéristiques de profondeur, on atteint une performance supérieure par rapport à une simple méthode de pondération par la profondeur (une multiplication de la carte de saillance 2D par la carte de profondeur). Ce résultat montre l'importance de la carte de saillance de profondeur pour la modélisation de l'attention visuelle 3D.

# Chapter 7

## Conclusion et perspectives

### 7.1 Résumé et contributions

Dans cette thèse, nous présentons des études portant sur plusieurs aspects de la recherche sur l'attention visuelle. Le cadre de la thèse peut être divisé en deux parties. La première partie concerne les problématiques liées aux vérités terrain de l'attention visuelle. Dans la deuxième partie, nos études portent sur la modélisation de l'attention visuelle dans des conditions de visualisation 3D. Le résumé de notre travail et de nos contributions suit.

#### **Vérification de la fiabilité de la vérité de terrain**

Notre travail débute par l'étude de la fiabilité de FDM calculées à partir de différentes bases de données oculométriques. Afin de comparer des FDM issues d'expériences indépendantes, une collaboration internationale a été lancée entre trois pays différents. Nous avons particulièrement étudié l'influence du contenu visuel et du temps de présentation des images. La fiabilité des FDM considérées comme vérité de terrain, a aussi été validée sur trois applications de traitement d'image : la modélisation de la saillance visuelle, l'évaluation de la qualité d'image et le reciblage d'image. Toujours à partir de la même base de données, nous avons constaté que les FDM des trois laboratoires étaient très proches. Les impacts sur les trois applications considérées étaient donc faibles. Ces résultats suggèrent que des FDM issues d'expériences oculométriques indépendantes peuvent en effet être considérées comme des vérités de terrain fiables pour ces trois applications de traitement d'image.

#### **Quantification de la relation entre deux types de vérité de terrain**

Notre deuxième étude vise à quantifier les similarités et les différences entre la carte de saillance visuelle (soit la carte de densité de fixation) et la carte importance, qui sont

deux vérités de terrain largement utilisées par des applications liées à l'attention. Nous avons commencé ce travail sur les deux mécanismes ("bottom-up" et "top-down") de l'attention, en comparant les cartes importance avec les cartes de saillance, nous avons constaté que les deux types de cartes sont liés, mais peut-être moins qu'on ne pourrait s'y attendre. Les cartes de saillance ont montré leur efficacité à prédire les principaux sujets d'intérêt.

## **Création d'une nouvelle base de données oculométriques pour le contenu 3D**

Pour résoudre le problème du manque de vérité de terrain pour la communauté de la modélisation de l'attention visuelle 3D, nous avons procédé à une expérience oculométrique binoculaire. Nous avons créé une nouvelle base de données oculométriques contenant dix-huit images naturelles 3D (ainsi que leurs versions 2D), les cartes de disparité correspondantes, et les données des mouvements oculaires des deux yeux (en incluant les cartes de densité de fixation pour les conditions 2D et 3D).

## **Influence de la profondeur sur l'attention visuelle 3D : étude du biais de profondeur**

Notre objectif est de déterminer s'il existe un « biais de profondeur » lors de la visualisation de contenu 3D sur écran plat stéréoscopique. Nous avons effectué une expérience oculométrique mettant en oeuvre un système d'affichage stéréoscopique et un oculomètre. Un grand nombre de stimuli synthétiques a été réuni pour l'expérience, ces stimuli ont été conçus afin d'écarter l'effet des caractéristiques visuelles 2D, et de telle sorte que l'attention visuelle des observateurs ne soit influencée que par l'information de profondeur. Nous avons examiné comment le rang (en profondeur) et la profondeur relative des objets influencent le comportement de visualisation. Les résultats expérimentaux montrent clairement que les observateurs prêtent plus d'attention et d'abord aux objets les plus proches d'eux.

## **Un modèle de l'attention visuelle 3D basé sur des caractéristiques de profondeur**

Nous avons proposé un nouveau modèle d'attention visuelle 3D basé sur la saillance de profondeur. Le théorème de Bayes est appliqué pour intégrer les données oculométriques précédentes, et calculer la carte de saillance de profondeur. Les résultats démontrent la grande valeur ajoutée de la carte de saillance de profondeur, et la bonne performance du modèle de saillance de profondeur que nous proposons. Deux façons différentes, pour appliquer l'information de profondeur au modèle de l'attention visuelle 3D, sont comparées dans notre étude. Les résultats montrent qu'en créant la carte de saillance

de profondeur à partir des caractéristiques de profondeur, on atteint une performance supérieure par rapport à une méthode simple de pondération avec la profondeur (une multiplication de la carte de saillance 2D par la carte de profondeur).

### **Intégration du biais central au modèle d'attention 3D**

Nous avons étudié les différences entre le biais central dans des conditions de visualisation 2D et celui dans des conditions de visualisation 3D, et la façon d'intégrer ce biais dans un modèle d'attention visuelle 3D. Notre travail démontre que le biais central dans des conditions de visualisation 3D est légèrement plus faible que celui dans des conditions de visualisation 2D. Nous proposons également un modèle hybride intégrant la saillance 2D, la saillance de profondeur et le biais central. Les évaluations qualitatives et quantitatives montrent la bonne performance du modèle hybride proposé sur des images 3D.

### **Vers une application potentielle de l'attention visuelle 3D**

L'attention visuelle est supposée améliorer la qualité d'expérience de la TV3D avec l'exploitation du "flou". Nous avons donc mené une expérience psychophysique pour étudier l'influence du flou de défocalisation sur la perception de la profondeur d'une scène stéréoscopique, et sa relation avec la disparité binoculaire. Nos résultats montrent que le floutage d'image accroît la profondeur perçue des images stéréoscopiques, si on compare à la profondeur perçue d'images stéréoscopiques sans flou additionnel.

## **7.2 Limites et perspectives**

Une large part du travail présenté dans cette thèse est relative à l'étude de l'attention visuelle par le biais d'expériences psychophysiques. Lors de notre recherche, nous avons aussi rencontré des limites, certaines questions restent ouvertes et devront être étudiées par les travaux futurs :

- En ce qui concerne la vérification de la fiabilité des FDM, il est encore difficile de quantifier le degré avec lequel chaque facteur affecte la FDM. Il serait par conséquent intéressant de poursuivre notre travail en réalisant les expériences tout en tenant particulièrement compte de la variation de certains facteurs. Il doit en plus être vérifié si un plus grand nombre de participants conduit à des FDM encore plus stables, et donc à une plus grande similarité entre les expériences. Des seuils doivent être déterminés afin de préciser le nombre minimum de participants aboutissant à des FDM suffisamment proches (des niveaux de similarité choisis et contrôlés).
- Il y a encore possibilité d'améliorer l'algorithme prédisant, en se basant sur la carte de saillance, la catégorie des objets. Les données temporelles suggèrent qu'une prédiction en plusieurs étapes serait plus efficace que nos essais en une seule étape.

Dans la première étape, les positions oculométriques des premières 0 à 2000 ms (ou d'une autre tranche de temps) seraient utilisées pour déterminer les objets les plus importants (les sujets principaux de la scène 3D). La connaissance de ces sujets guiderait ensuite les prédictions des objets suivants.

- Au niveau de la construction de vérité de terrain pour la modélisation de l'attention visuelle 3D, une base de données oculométriques contenant seulement dix-huit images n'est pas suffisante. Une base de données possédant un grand nombre d'images et de vidéos 3D reste à réunir. L'introduction de nouvelles techniques de calibration, par exemple la "calibration volumétrique" [Huynh-Thu 11a], est également nécessaire lors des expériences oculométriques 3D.
- Notre étude du biais de profondeur ne tient pas compte de l'influence de la profondeur absolue (par exemple, la disparité). Cette profondeur absolue pourrait être un facteur important influençant le biais de profondeur, il a en effet été démontré l'existence de neurones sélectifs en disparité dans le cortex visuel primaire (V1) [Neri 99, Barlow 67, Nikara 68, Poggio 77].

Le contribution principale de la thèse est le nouveau modèle d'attention visuelle 3D basé sur la saillance de profondeur. Le modèle proposé a encore des limites, par exemple il ne prend en compte qu'une seule caractéristique de profondeur (le contraste de profondeur), et met en oeuvre une stratégie simple pour combiner la carte de saillance de profondeur et celle de saillance 2D. Pour de travaux futurs, on pourrait envisager d'inclure plus de caractéristiques de profondeur (par exemple, la courbure de surface ou le gradient de profondeur). Une stratégie de combinaison de cartes plus sophistiquée pourrait donc sûrement améliorer les performances du modèle.

# Bibliographie

- [Achanta 09] R. Achanta, S. Hemami, F. Estrada & S. Susstrunk. *Frequency-tuned salient region detection*. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 1597–1604. IEEE, 2009.
- [Avidan 07] S. Avidan & A. Shamir. *Seam carving for content-aware image resizing*. In ACM Transactions on Graphics (TOG), volume 26, page 10. ACM, 2007.
- [Barlow 67] H.B. Barlow, C. Blakemore & JD Pettigrew. *The neural mechanism of binocular depth discrimination*. The Journal of physiology, vol. 193, no. 2, page 327, 1967.
- [Borji 12] A. Borji & L. Itti. *State-of-the-art in Visual Attention Modeling*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012.
- [Bowler 89] P.J. Bowler. *Evolution : the history of an idea*. Univ of California Press, 1989.
- [Bruce 09] N.D.B. Bruce & J.K. Tsotsos. *Saliency, attention, and visual search : An information theoretic approach*. Journal of Vision, vol. 9, no. 3, 2009.
- [Chamaret 10] C. Chamaret, S. Godeffroy, P. Lopez & O. Le Meur. *Adaptive 3D rendering based on region-of-interest*. In Proceedings of SPIE, volume 7524, page 75240V, 2010.
- [Chen 10] W. Chen, J. Fournier, M. Barkowsky & P. Le Callet. *NEW REQUIREMENTS OF SUBJECTIVE VIDEO QUALITY ASSESSMENT METHODOLOGIES FOR 3DTV*. In Video Processing and Quality Metrics 2010 (VPQM), Scottsdale, USA, 2010., 2010.
- [Engelke 09] U. Engelke, A. Maeder & H.J. Zepernick. *Visual attention modelling for subjective image quality databases*. In Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on, pages 1–6. IEEE, 2009.
- [Engelke 10] U. Engelke, A. Maeder & H.J. Zepernick. *Analysing inter-observer saliency variations in task-free viewing of natural images*. In Image Pro-

- cessing (ICIP), 2010 17th IEEE International Conference on, pages 1085–1088. IEEE, 2010.
- [Etz 00] Stephen P. Etz & Jiebo Luo. *Ground truth for training and evaluation of automatic main subject detection*. volume 3959, pages 434–442. SPIE, 2000.
- [Fincham 57] EF Fincham & J. Walton. *The reciprocal actions of accommodation and convergence*. The Journal of Physiology, vol. 137, no. 3, pages 488–508, 1957.
- [Hakkinen 10] Jukka Hakkinen, Takashi Kawai, Jari Takatalo, Reiko Mitsuya & Gote Nyman. *What do people look at when they watch stereoscopic movies?* volume 7524, page 75240E. SPIE, 2010.
- [Hantao 11] LIU Hantao. *Modeling Perceived Quality for Imaging Applications*. PhD thesis, Delft University of Technology, The Netherlands, 2011.
- [Held 10] R.T. Held, E.A. Cooper, J.F. O’Brien & M.S. Banks. *Using blur to affect perceived distance and size*. ACM Transactions on Graphics, vol. 29, 2010.
- [Hoffman 08] D.M. Hoffman, A.R. Girshick, K. Akeley & M.S. Banks. *Vergence–accommodation conflicts hinder visual performance and cause visual fatigue*. Journal of Vision, vol. 8, no. 3, 2008.
- [Hoffman 10] D.M. Hoffman & M.S. Banks. *Focus information is used to interpret binocular images*. Journal of vision, vol. 10, no. 5, 2010.
- [Hou 07] X. Hou & L. Zhang. *Saliency detection : A spectral residual approach*. In Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on, pages 1–8. Ieee, 2007.
- [Huynh-Thu 11a] Q. Huynh-Thu, M. Barkowsky, P. Le Callet *et al.* *The Importance of Visual Attention in Improving the 3D-TV Viewing Experience : Overview and New Perspectives*. IEEE Transactions on Broadcasting, vol. 57, no. 2, pages 421–431, 2011.
- [Huynh-Thu 11b] Q. Huynh-Thu & L. Schiatti. *Examination of 3D visual attention in stereoscopic video content*. In Proceedings of SPIE, volume 7865, page 78650J, 2011.
- [Itti 98] L. Itti, C. Koch & E. Niebur. *A model of saliency-based visual attention for rapid scene analysis*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 20, no. 11, pages 1254–1259, 1998.
- [Jansen 09] L. Jansen, S. Onat & P. König. *Influence of disparity on fixation and saccades in free viewing of natural scenes*. Journal of Vision, vol. 9, no. 1, 2009.

- 
- [Kadiyala 08] V. Kadiyala, S. Pinneli, E. C. Larson & D. M. Chandler. *Quantifying the perceived interest of objects in images : effects of size, location, blur, and contrast*. In Thrasyvoulos N. Rogowitz Bernice E. ; Pappas, editeur, Human Vision and Electronic Imaging XIII, volume 6806, pages 68060S–68060S–13, 2008.
- [Koch 85] C. Koch & S. Ullman. *Shifts in selective visual attention : towards the underlying neural circuitry*. Hum Neurobiol, vol. 4, no. 4, pages 219–27, 1985.
- [Le Meur 06] O. Le Meur, P. Le Callet, D. Barba & D. Thoreau. *A coherent computational approach to model bottom-up visual attention*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 28, no. 5, pages 802–817, 2006.
- [Liu 09] H. Liu & I. Heynderickx. *Studying the added value of visual attention in objective image quality metrics based on eye movement data*. In Image Processing (ICIP), 2009 16th IEEE International Conference on, pages 3097–3100. IEEE, 2009.
- [Liu 11] H. Liu & I. Heynderickx. *Visual Attention in Objective Image Quality Assessment : Based on Eye-Tracking Data*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 21, no. 7, pages 971–982, 2011.
- [Maeder 95] A.J. Maeder. *Importance maps for adaptive information reduction in visual scenes*. In Intelligent Information Systems, 1995. ANZIIS-95. Proceedings of the Third Australian and New Zealand Conference on, pages 24–29. IEEE, 1995.
- [Maki 96] A. Maki, P. Nordlund & J.O. Eklundh. *A computational model of depth-based attention*. In Pattern Recognition, 1996., Proceedings of the 13th International Conference on, volume 4, pages 734–739. IEEE, 1996.
- [Maki 00] A. Maki, P. Nordlund & J.O. Eklundh. *Attentional scene segmentation : integrating depth and motion*. Computer Vision and Image Understanding, vol. 78, no. 3, pages 351–373, 2000.
- [Marshall 96] J.A. Marshall, C.A. Burbeck, D. Ariely, J.P. Rolland & K.E. Martin. *Occlusion edge blur : a cue to relative visual depth*. JOSA A, vol. 13, no. 4, pages 681–688, 1996.
- [Mather 96] G. Mather. *Image blur as a pictorial depth cue*. Proceedings of the Royal Society of London. Series B : Biological Sciences, vol. 263, no. 1367, pages 169–172, 1996.
- [Mather 09] G. Mather. Foundations of sensation and perception, volume 2. Psychology Press, 2009.

- [McAllister 93] D.F. McAllister. Stereo computer graphics and other true 3d technologies. Princeton University Press, 1993.
- [Neri 99] P. Neri, A.J. Parker & C. Blakemore. *Probing the human stereoscopic system with reverse correlation*. Nature, vol. 401, no. 6754, pages 695–698, 1999.
- [Nikara 68] T. Nikara, P. O. Bishop & J. D. Pettigrew. *Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex*. Experimental Brain Research, vol. 6, pages 353–372, 1968.
- [Osberger 98] W. Osberger & A.J. Maeder. *Automatic identification of perceptually important regions in an image*. In Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on, volume 1, pages 701–704. IEEE, 1998.
- [Palmer 08] S.E. Palmer & J.L. Brooks. *Edge-region grouping in figure-ground organization and depth perception*. Journal of Experimental Psychology : Human Perception and Performance, vol. 34, no. 6, page 1353, 2008.
- [Park 02] K. Park & H.W. Park. *Region-of-interest coding based on set partitioning in hierarchical trees*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 12, no. 2, pages 106–113, 2002.
- [Pentland 87] A.P. Pentland. *A new sense for depth of field*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, no. 4, pages 523–531, 1987.
- [Poggio 77] G.F. Poggio, B. Fischer *et al.* *Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey*. J Neurophysiol, vol. 40, no. 6, pages 1392–1405, 1977.
- [Potapova 11] E. Potapova, M. Zillich & M. Vincze. *Learning what matters : combining probabilistic models of 2D and 3D saliency cues*. Computer Vision Systems, pages 132–142, 2011.
- [Rajashekar 08] U. Rajashekar, I. van der Linde, A.C. Bovik & L.K. Cormack. *GAFPE : A gaze-attentive fixation finding engine*. Image Processing, IEEE Transactions on, vol. 17, no. 4, pages 564–573, 2008.
- [Ramasamy 09] C. Ramasamy, D.H. House, A.T. Duchowski & B. Daugherty. *Using eye tracking to analyze stereoscopic filmmaking*. In SIGGRAPH'09 : Posters, page 28. ACM, 2009.
- [Scharstein 07] D. Scharstein & C. Pal. *Learning conditional random fields for stereo*. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pages 1–8. IEEE, 2007.

- 
- [Sheikh 05] HR Sheikh, Z. Wang, L. Cormack & AC Bovik. *LIVE image quality assessment database release 2 (2005)*, 2005.
- [Sheikh 06] H.R. Sheikh & A.C. Bovik. *Image information and visual quality*. Image Processing, IEEE Transactions on, vol. 15, no. 2, pages 430–444, 2006.
- [Tatler 07] B.W. Tatler. *The central fixation bias in scene viewing : Selecting an optimal viewing position independently of motor biases and image feature distributions*. Journal of Vision, vol. 7, no. 14, 2007.
- [Tseng 09] P.H. Tseng, R. Carmi, I.G.M. Cameron, D.P. Munoz & L. Itti. *Quantifying center bias of observers in free viewing of dynamic natural scenes*. Journal of Vision, vol. 9, no. 7, 2009.
- [Vu 03] K. Vu, K.A. Hua & W. Tavanapong. *Image retrieval based on regions of interest*. Knowledge and Data Engineering, IEEE Transactions on, vol. 15, no. 4, pages 1045–1049, 2003.
- [Wandell 95] B.A. Wandell. *Foundations of vision*, volume 21. Sinauer Associates, 1995.
- [Wang 04] Z. Wang, A.C. Bovik, H.R. Sheikh & E.P. Simoncelli. *Image quality assessment : From error visibility to structural similarity*. Image Processing, IEEE Transactions on, vol. 13, no. 4, pages 600–612, 2004.
- [Wang 06] Z. Wang & A.C. Bovik. *Modern image quality assessment*. Synthesis Lectures on Image, Video, and Multimedia Processing, vol. 2, no. 1, pages 1–156, 2006.
- [Wang 11a] D. Wang, G. Li, W. Jia & X. Luo. *Saliency-driven scaling optimization for image retargeting*. The Visual Computer, pages 1–8, 2011.
- [Wang 11b] J. Wang, P Le Callet, V. Ricordel & S Tourancheau. *Quantifying depth bias in free viewing of still stereoscopic synthetic stimuli*. 16th European Conference on Eye Movements, Marseille, France, 2011.
- [Watt 05] S.J. Watt, K. Akeley, M.O. Ernst & M.S. Banks. *Focus cues affect perceived depth*. Journal of Vision, vol. 5, no. 10, 2005.
- [Wolfe 00] J.M. Wolfe. *Visual attention*. Seeing, vol. 2, pages 335–386, 2000.
- [Zhang 08] L. Zhang, M.H. Tong, T.K. Marks, H. Shan & G.W. Cottrell. *SUN : A Bayesian framework for saliency using natural statistics*. Journal of Vision, vol. 8, no. 7, 2008.
- [Zhang 10] Y. Zhang, G. Jiang, M. Yu & K. Chen. *Stereoscopic visual attention model for 3D video*. Advances in Multimedia Modeling, pages 314–324, 2010.
- [Zhao 11] Q. Zhao & C. Koch. *Learning a saliency map using fixated locations in natural scenes*. Journal of Vision, vol. 11, no. 3, 2011.



# Thèse de Doctorat

**Junle Wang**

**From 2D to Stereoscopic-3D Visual Saliency: Revisiting Psychophysical Methods and Computational Modeling**

Saillance Visuelle, de la 2D à la 3D Stéréoscopique : Examen des Méthodes Psychophysique et Modélisation Computationnelle  
Vol 2: Résumé étendu en Français

## Résumé

L'attention visuelle est l'un des mécanismes les plus importants mis en oeuvre par le système visuel humain (SVH) afin de réduire la quantité d'information que le cerveau a besoin de traiter pour appréhender le contenu d'une scène. Un nombre croissant de travaux est consacré à l'étude de l'attention visuelle, et en particulier à sa modélisation computationnelle. Dans cette thèse, nous présentons des études portant sur plusieurs aspects de cette recherche. Nos travaux peuvent être classés globalement en deux parties. La première concerne les questions liées à la vérité de terrain utilisée, la seconde est relative à la modélisation de l'attention visuelle dans des conditions de visualisation 3D. Dans la première partie, nous analysons la fiabilité de cartes de densité de fixation issues de différentes bases de données oculométriques. Ensuite, nous identifions quantitativement les similitudes et les différences entre carte de densité de fixation et carte d'importance visuelle, ces deux types de carte étant les vérités de terrain communément utilisées par les applications relatives à l'attention. Puis, pour faire face au manque de vérité de terrain exploitable pour la modélisation de l'attention visuelle 3D, nous procédons à une expérimentation oculométrique binoculaire qui aboutit à la création d'une nouvelle base de données avec des images stéréoscopiques 3D.

Dans la seconde partie, nous commençons par examiner l'impact de la profondeur sur l'attention visuelle dans des conditions de visualisation 3D. Nous quantifions d'abord le "biais de profondeur" lié à la visualisation de contenus synthétiques 3D sur écran plat stéréoscopique. Ensuite, nous étendons notre étude avec l'usage d'images 3D au contenu naturel. Nous proposons un modèle de l'attention visuelle 3D basé saillance de profondeur, modèle qui repose sur le contraste de profondeur de la scène. Deux façons différentes d'exploiter l'information de profondeur par notre modèle sont comparées. Ensuite, nous étudions le biais central et les différences qui existent selon que les conditions de visualisation soient 2D ou 3D. Nous intégrons aussi le biais central à notre modèle de l'attention visuelle 3D. Enfin, considérant que l'attention visuelle combinée à une technique de floutage peut améliorer la qualité d'expérience de la TV-3D, nous étudions l'influence de flou sur la perception de la profondeur, et la relation du flou avec la disparité binoculaire.

## Mots clés

Attention visuelle, modèle computationnel, oculométrie, 3DTV, saillance visuelle, importance visuelle, expérience subjective, stéréoscopie.

## Abstract

Visual attention is one of the most important mechanisms deployed in the human visual system to reduce the amount of information that our brain needs to process. An increasing amount of efforts are being dedicated in the studies of visual attention, particularly in computational modeling of visual attention. In this thesis, we present studies focusing on several aspects of the research of visual attention. Our works can be mainly classified into two parts. The first part concerns ground truths used in the studies related to visual attention; the second part contains studies related to the modeling of visual attention for Stereoscopic 3D (S-3D) viewing condition. In the first part, our work starts with identifying the reliability of FDM from different eye-tracking databases. Then we quantitatively identify the similarities and difference between fixation density maps and visual importance map, which have been two widely used ground truth for attention-related applications. Next, to solve the problem of lacking ground truth in the community of 3D visual attention modeling, we conduct a binocular eye-tracking experiment to create a new eye-tracking database for S-3D images. In the second part, we start with examining the impact of depth on visual attention in S-3D viewing condition. We firstly introduce a so-called "depth-bias" in the viewing of synthetic S-3D content on planar stereoscopic display. Then, we extend our study from synthetic stimuli to natural content S-3D images. We propose a depth-saliency-based model of 3D visual attention, which relies on depth contrast of the scene. Two different ways of applying depth information in S-3D visual attention model are also compared in our study. Next, we study the difference of center-bias between 2D and S-3D viewing conditions, and further integrate the center-bias with S-3D visual attention modeling. At the end, based on the assumption that visual attention can be used for improving Quality of Experience of 3D-TV when collaborating with blur, we study the influence of blur on depth perception and blur's relationship with binocular disparity.

## Key Words

Visual attention, computational model, eye-tracking, 3DTV, visual saliency, visual importance, subjective experiment, stereoscopy.