Un modèle de saillance dépendant du temps combinant les biais centré et de profondeur pour la visualisation en 2D et 3D

J. Gautier O. Le Meur

IRISA / Université de Rennes 1

Campus de Beaulieu
Rennes, 35042 France

{jgautier, olemeur}@irisa.fr

Résumé

Le rôle de la disparité binoculaire comme élément d'acquisition de la profondeur par le Système Visuel Humain (SVH) est étudié. Les données oculométrique d'observateurs en conditions de visualisation "2D" et "3D" ont été comparées. Les résultats montrent l'influence de la disparité sur la saillance, les biais du centre et de profondeur. En particulier, les sujets regardent tout d'abord les zones les plus proches en 3D, puis dirigent de façon plus étendue leurs regards qu'en 2D. Afin d'améliorer les performances de modèles de saillance existants, les biais centré et de profondeur sont introduits dans ces modèles. Les résultats indiquent que suite au mécanisme de recentrage initial, l'avant-plan joue un rôle prépondérant dans les instants suivants de visualisation. Ce rôle est renforcé en condition stéréoscopique. Enfin, nous proposons un modèle computationnel de saillance sur image fixe, combinant les caractéristiques bas-niveau à ces biais centré et de profondeur. Les performances obtenues justifient la pertinence de l'approche.

Mots clefs

modèle de saillance, stéréoscopie, 3D, disparité binoculaire.

1 Introduction

Le système humain traite l'information de son environnement au travers d'organes sensoriels dédiés. Mais la quantité d'informations du monde visuel nous empêche de percevoir notre environnement "d'un clin d'œil". Ainsi, des mouvements oculaires exploratoires dirigent notre fovéa -aire de l'œil de très grande résolution optique- sur des zones particulières du champ visuel, c'est ce que l'on appelle l'attention "overt". Ce déploiement de l'attention visuelle implique deux mécanismes : un premier dépendant du stimulus, dit "ascendant" ou *"Bottom-Up"*, et un second singulier à l'observateur, "descendant" ou *"Top-Down"*. Le premier est piloté par des caractéristiques bas-niveaux [1, 2], alors que le second implique des processus cognitifs de haut-niveau, liés à la tâche et à la connaissance a-priori de l'observateur [3, 4]. Différents modèles computationnels de l'attention "overt" ont été proposés pour prédire notre attention sur un stimulus donné, au travers d'une représentation topographique sous forme de cartes de saillances, le plus souvent *"Bottom-Up"* [1, 2]. Puisque dans le SVH, le traitement de la profondeur suit ce traitement *"Bottom-Up"* le long des voies visuelles [5], il est intéressant d'évaluer si le signal profondeur améliore les performances des modèles de saillance. Cependant, le processus d'acquisition de la profondeur des objets dans le champ visuel par le SVH est intrinsèquement ambigu. Il s'agit d'une projection de signaux lumineux d'un monde tridimensionnel sur une surface rétinienne bidimensionnelle, qui peut être reprojetée à l'infini. Pour résoudre cette ambiguïté, le SVH s'appuie sur une combinaison de différents signaux de profondeurs : les signaux monoculaires tels que l'accommodation, le mouvement de parallax, la perspective, les ombres etc., ainsi que sur des signaux binoculaires : la convergence et la disparité binoculaire [6]. Alors que les premiers donnent une information relative à la distance des objets les uns par rapport aux autres, les seconds informent de la distance absolue aux objets.

Ainsi la perception de la profondeur implique une combinaison de multiples - mais potentiellement conflictuels- signaux de profondeur. Différentes propositions ont été faites pour inclure la profondeur ou la disparité stéréo comme caractéristique complémentaire à un modèle computationnel de l'attention visuelle. Maki et al.[7, 8] ont tout d'abord proposé un modèle basé sur les signaux image, profondeur et sur la détection de mouvements. La profondeur permet de sélectionner des "cibles" selon leur priorité : les plus hautes priorités sont données aux objets les plus proches. La principale limitation vient de cette hypothèse puisque quelque chose de plus proche n'implique pas nécessairement qu'elle soit la plus saillante. Ouerhani et al.[9] incluent également la profondeur directement comme caractéristique dans le modèle de Itti [2]. La profondeur est transformée en carte de visibilité par un mécanisme "centre-voisinage" au même titre que le proposait L.Itti sur l'intensité lumineuse par exemple. Le principe est uniquement illustré qualitativement. Plus récemment, Zhang et

al.[10] proposent un modèle d'attention stéréoscopique. La profondeur est combinée au mouvement et également au modèle de Itti. La fusion de ces 3 caractéristiques est réalisée par pondération arbitraire, mais il n'y pas d'étude des performances. Une des rares tentatives considérant la perception stéréoscopique à été proposé par Bruce et Tsotsos [11]. Un ajustement sélectif du modèle original de Tsotsos est étendu pour inclure la rivalité binoculaire qui se produit en vision stéréoscopique.
Un autre facteur qui impacte significativement notre déploiement visuel est le biais centré. On considère souvent qu'il résulte d'un biais oculomoteur du système saccadique ou de la distribution à tendance centrale des caractéristiques importantes sur un stimulus image. Pourtant Tatler [12] a montré que ce biais de fixation central est présent quel que soit la localisations des caractéristiques ou la tâche de l'observateur (malgré une implication différente du biais au cours du temps selon la tâche). Quoiqu'il en soit, l'inclusion de ce biais centré, dans les modèles existants, améliore significativement les performances [13, 14]. Une première proposition d'inclusion de signaux de profondeurs et de biais centré a été proposée par Vincent et al. [15]. Les contributions de l'avant-plan et du centre, ainsi que d'autres facteurs haut-niveau sont étudiées quantitativement. Les résultats soulignent le rôle potentiel de l'avant-plan et du biais central dans la prédiction de l'attention. Ho Phuoc et al. [16] suivent une méthodologie similaire, mais étudient le rôle de caractéristiques visuelles de bas-niveau au cours du temps. Suite à une analyse des pondérations des caractéristiques au cours du temps, un modèle de saillance est proposé, mais sur des poids fixes au cours du temps, comme dans [15].
Dans ce papier, nous développons un modèle d'attention dépendant du temps, afin de prédire où les observateurs regardent en conditions mono et stéréoscopiques. Suite à une analyse statistiques de biais au cours du temps et pour différentes conditions de visualisations, un modèle est proposé et ses performances évaluées. Ainsi, cet article vise à répondre à 2 questions :
– Comment modéliser les biais centré et de profondeur comme caractéristiques individuelles ?
– Comment inclure ces caractéristiques dans un modèle prenant en compte l'aspect temporel de l'attention ?

2 Rappels

2.1 Conditions expérimentales

La base de donnée oculométrique utilisée a été gracieusement fournie par Jansen et al. [17]. Nous résumons ici les conditions d'acquisition des images puis des fixations des observateurs. Les paires d'images stéréoscopiques ont été obtenues avec un banc composé de deux appareils photo. Pour acquérir l'information de profondeur, puis de disparité entre ces deux images, un scanner laser 3D a été utilisé. Ainsi, en projetant la profondeur acquise dans le référentiel de chaque image, puis en recherchant les correspondances stéréo, des cartes de disparité ont été générées. Plus d'informations sur l'acquisition peuvent être trouvées dans [18]. 28 paires stéréo de photographies de forêt, rectifiées et converties en niveau de gris, sont donc affichées sur un écran auto-stéréoscopique de 18,1". En condition 2D, deux copies de l'image gauche sont affichées à l'écran, alors qu'en condition 3D la paire gauche et droite est affichée simultanément, permettant une vision stéréoscopique et introduisant la disparité binoculaire. Les mouvements oculaires de 14 observateurs sont enregistrés à l'aide d'un oculomètre, mais seules les données de l'œil gauche sont gardées. Ainsi, puisque le stimulus image est identique quel que soit les conditions 2D et 3D, le facteur disparité binoculaire est isolé et observable.

2.2 Résultats précédents

Jansen et al. [17] ont montré que l'introduction de la disparité altère les propriétés basiques des mouvements oculaires, tels que le taux de fixation, la longueur des saccades et leurs dynamiques, mais pas la durée de fixation (sur des images naturelles). Ils montrent également que cette disparité influence l'attention visuelle en début de visualisation, où l'observateur tend à regarder les zones plus proches.
Nous avons réalisé une étude préliminaire et supplémentaire en trois points pour examiner si la disparité impacte : la position des zones fixées, la distance moyenne des fixations par rapport au centre de l'écran (biais centré) et la profondeur sur les fixations (biais de profondeur). Cette pré-étude souligne l'influence de la disparité binoculaire sur notre attention visuelle. Celle-ci existe, en particulier sur les toutes premières fixations. Ce facteur, induit par les conditions de visualisation stéréoscopiques, affecte donc notre stratégie visuelle et provoque une tendance à fixer des zone plus proches en début de visualisation, puis à élargir le champ de fixation.

3 Modélisation de la saillance en fonction du temps

Les études récentes [12, 19] ont montré l'importance et l'influence des biais externes dans le déploiement de l'attention visuelle pré-attentive. En soit, la part de l'attention visuelle dirigée par des stimuli -exogène- ou par des facteurs propres au sujet et à son action -endogène- est un débat ouvert [20, 21]. Cependant, considérer leurs interactions et leurs implications au cours du temps est vraisemblablement nécessaire pour améliorer la prédictibilité des modèles de saillances existants [21, 12].

3.1 Analyse statistique

Notre pré-étude comportementale soulignait l'influence de deux biais au cours du temps. Nous proposons donc de les modéliser puis les quantifier afin d'évaluer plus précisément leur apport à des modèles existants. À cette fin, nous suivons l'approche de Vincent et al. [15].
Un modèle statistique de fonction de densité de fixation f(x,t) est exprimé par un mélange additif de différentes caractéristiques ou "modes", chacune associée à une proba-

bilité donnée. Ainsi, chaque mode est un facteur *a priori* de guidage quel que soit la scène. Une fonction de densité est définie sur l'ensemble des positions des fixations dans l'espace par la variable bidimensionnelle x ainsi :

$$f(x,t) = \sum_{k=1}^{K} p_k(t)\phi_k(x)$$

avec K le nombre de caractéristiques, $\phi_k(x)$ la fonction de densité de probabilité associée à chacune et $p_k(t)$ la contribution ou "poids" de chaque caractéristique k avec la contrainte $\sum_{k=1}^{K} p_k = 1$ par instant t.
L'analyse statistique vise ainsi à séparer la contribution de la caractéristique *"Bottom-Up"*, basée sur des mécanismes bas-niveau, des autres biais et donc attributs potentiels observés jusqu'alors. Ainsi, il va s'agir d'abord de modéliser ces autres attributs, puis de caractériser l'évolution temporelle des contributions de l'ensemble des attributs. Un algorithme d'Espérance-Maximisation (EM) est utilisé pour estimer les poids du maximum de vraisemblance de ce modèle paramétrique. Avant d'étayer cette méthode, nous décrivons la modélisation des biais centré et de profondeur.

Modélisation du biais centré. Le plus fort biais souligné par les expériences oculométriques est le biais centré. C'est un biais à part entière, expliquant une proportion importante des mouvements oculaires. Le fait qu'il soit une composante de la stratégie visuelle est encore en question. Tatler [12] étudia ce biais au cours du temps et selon la tâche de l'observateur. Il prouva que cette tendance à la fixation au centre de l'écran persiste tout au long de la visualisation en tâche libre, mais se dissipe rapidement en tâche de recherche visuelle. Ainsi, au delà de la 3ème fixation, ce biais est à peine détectable. Dans notre cas, il s'agissait d'une expérience de recherche de plans de profondeur, les sujets devant presser un bouton dès qu'ils distinguaient au moins 2 plans de profondeur différents. En cohérence avec les expériences avec tâche de recherche dans [12], le biais de fixation centré est non détectable dès la 3ème fixation.
Suite aux résultats de la littérature et à nos observations, le biais centré est modélisé par une gaussienne bidimensionnelle. Ceci est justifié empiriquement par la convergence des distributions de fixations. Comme proposé dans [13], les paramètres de la gaussienne sont prédéfinis et non estimés pendant l'apprentissage. Le biais centré est donc modélisé par une gaussienne, centrée sur le centre de l'écran : $N(0,\Sigma)$ avec $\Sigma = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$ la matrice de covariance et $\{\sigma_x^2$ et $\sigma_y^2\}$ la variance. Nous adaptons la gaussienne bidimensionnelle à la distribution des fixations sur la première fixation. Quelles que soient les conditions (2D ou 3D), les distributions sont similaires

Modélisation du biais de profondeur. Les résultats de [17] indiquent que la profondeur perçue dépend des conditions de visualisation. La façon dont la profondeur interagit pour moduler l'attention visuelle est un sujet de recherche actif. En particulier, le mécanisme de distinction forme/fond : un élément du signal de profondeur d'interprétation des contours [22], dirige l'attention de façon préattentive [5]. Ceci appuie notre choix de la modélisation de la profondeur par une organisation forme/fond, c'est-à-dire par une classification de la profondeur en cartes d'avant et d'arrière-plan comme illustrées sur la figure 1(a). Ces cartes sont obtenues par seuillage à la moitié de l'amplitude de profondeur et lissage par une fonction sigmoïde. Les valeurs d'arrière-plan sont complémentées de manière à ce que plus un point est éloigné, plus il contribue à l'arrière-plan, et inversement pour l'avant-plan.

Autres biais potentiels. Le modèle vise à prédire où notre regard se déploie en mono- et en stéréoscopie. La prédiction est basée sur une combinaison linéaire de caractéristiques bas-niveau, de biais centré et de profondeur. Pourtant, d'autres facteurs autrement plus complexes se manifestent au cours du temps. Ainsi, les mécanismes *"Top-Down"* interagissent, en particulier dès la période "tardive". Afin de les prendre en compte, une carte caractéristique est ajoutée, dont la contribution à chaque fixation est spatialement équiprobable. Ceci modélise l'influence de facteurs endogènes tel que la connaissance et l'expérience a priori, etc. Évidemment, la contribution de cette carte uniforme doit être aussi faible que possible, ce qui signifie que les autres attributs sont les plus significatifs et les plus contributeurs à la prédiction de l'attention. En résumé, cinq caractéristiques sont extraites en cartes comme illustrées sur la figure 1(a) :

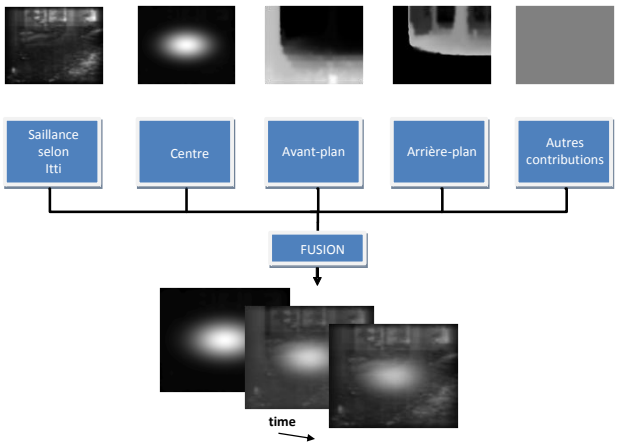

Figure 1 – *(a) 1ère ligne : Illustration d'une carte de saillance de Itti sur une des 24 images, du biais centré en 2D et des cartes d'avant et d'arrière-plan correspondantes. (b) 2ème ligne : Description du modèle proposée dépendant du temps. (c) Dernière ligne : Illustration des cartes de saillance finales dépendantes du temps, pour les 1ère, 10ème, et 20ème fixations en condition 2D*

La saillance "bas-niveau", les caractéristiques avant-plan et arrière-plan dépendent du contenu visuel. Les caractéristiques centré et uniforme représentent des biais haut-niveau et sont fixes au cours du temps et invariantes aux stimuli. Ainsi le modèle par mélange additif est :

$$f(x,t) = p_{Ms}(t)\phi_{Ms}(x) + p_{bC}(t)\phi_{bC}(x) + p_{Av}(t)\phi_{Av}(x) + p_{Ar}(t)\phi_{Ar}(x) + p_{Un}(t)\phi_{Un}(x)$$

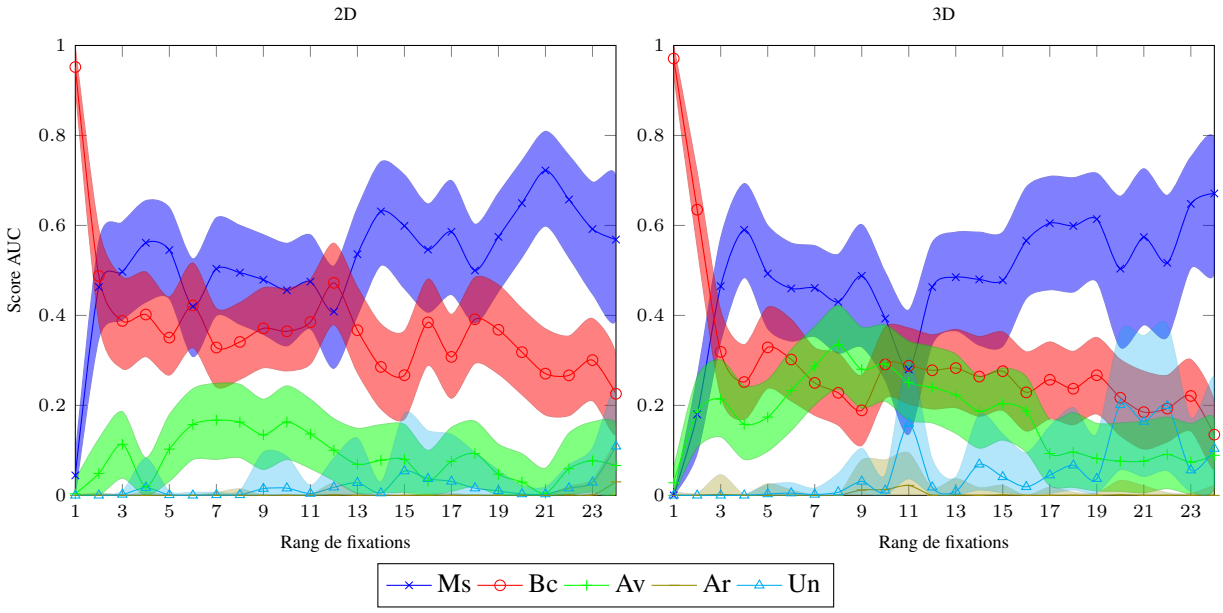

Figure 2 – *Contributions temporelles (poids) des 5 caractéristiques à partir de fixations 2D (gauche) et 3D (droite) en fonction du rang de fixation. Les intervalles de confiances à 95% sont calculées par une estimation "bootstrap" (1000 réplications).*

avec ϕ_{Ms} la saillance du modèle de Itti , ϕ_{Bc} la fonction de biais centré, ϕ_{Av} et ϕ_{Ar} les fonctions respectivement d'avant et d'arrière plan, et ϕ_{Un} fonction de densité uniforme. Chaque caractéristique est homogène à une fonction de densité de probabilité. $p_{Ms}, p_{Bc}, p_{Av}, p_{Ar}$ et p_{Un} sont les poids dépendant du temps, à estimer, leur somme étant égale à l'unité. La Figure 1(a) donnent une illustration des caractéristiques impliquées. Les poids $p_k^{(m)}(t)$ sont les seuls paramètres estimés à chaque itération m dans l'algorithme EM (en pratique, un nombre M de 50 itérations est un bon compromis estimation de qualité/complexité).

3.2 Résultats

Les contributions temporelles des caractéristiques retenues de l'attention visuelle sont évaluées. Le modèle basé EM apprend les poids fixation par fixation (de la 1ère à la 25ème), sur la moitié de la base d'images et de fixations associées : chaque fixation des observateurs est projetée sur les cartes caractéristiques associées à un stimulus image. Il y a 14 participants et donc au plus 14 fixations par indice de fixation, par image. À la convergence, l'EM donne une estimation des poids maximisant la combinaison additive de différentes caractéristiques par rapport à la distribution des fixations originales. L'apprentissage par fixation est répété à chaque rang de fixation, en 2D et 3D. L'estimation des contributions en 2D et 3D est illustrée à la figure 2.

Le meilleur prédicteur dans les deux conditions est la saillance prédite (appelé "Ms" et provenant du modèle de Itti). Comme attendu, le biais de fixation central est fortement impliqué sur les deux premières fixations mais s'estompe rapidement à un niveau intermédiaire entre la saillance bas-niveau et les autres contributions. Cette contribution est significativement (t-test pairé, p<0.001) plus importante en condition 3D qu'en 2D. La contribution de l'avant-plan est également significativement (t-test pairé, p<0.001) plus importante en 3D qu'en 2D. Ainsi le biais centré est partiellement compensé en 3D tout d'abord par la contribution de l'avant-plan de la 3ème à la 18ème fixation, mais également par l'augmentation progressive de la saillance. Enfin, les contributions arrière-plan et uniformes restent constamment faibles en 2D, mais augmentent progressivement en période tardive en 3D.

Ainsi le biais central observé est cohérent avec les observations de Tatler [12] en tâche de recherche : l'effet de recentrage n'est très probablement pas dû au marqueur de préfixation avant affichage du stimulus, mais à une tendance systématique et vraisemblablement stratégique pour le SVH.

La disparité binoculaire rehausse la composante avant-plan jusqu'à la 17ème fixation, ceci suggère qu'elle participe à la discrimination des plans de profondeur. Alors que l'avant-plan participait déjà à la prédiction des zones saillantes en 2D, il contribue d'autant plus en présence de la disparité. Ainsi cette caractéristique de profondeur contribue à la prédiction en conditions monoscopiques (car la profondeur peut être inférée par de nombreux signaux de profondeur monoscopiques), mais d'autant plus en stéréoscopie. À l'inverse, l'arrière-plan contribue peu à l'attention, lorsque c'est le cas (dès la 23ème et 19ème fixation en 2D et 3D respectivement), ceci est combiné à une contribution du biais uniforme, et relativise l'impact réelle de l'arrière-plan. La composante uniforme simule l'influence d'autres facteurs haut-niveau non décrits par notre modèle. Puisque celle-ci reste faible jusqu'à la 20ème fixation, ceci justifie les choix et la modélisation des biais. Au delà, la contribution uniforme suggère que les biais centré et d'avant-plan ne sont pas suffisants à l'explication des mouvements oculaires.

Dans la section suivante, nous utilisons ces poids appris au cours du temps pour prédire où regardent les observateurs. Les performances sont donc testées sur la moitié restante de la base de données images et fixations, et sur un intervalle temporel où la contribution uniforme est stable

et faible.

4 Modèle de saillance dépendant du temps

Les pondérations des caractéristiques permettent de calculer une carte de saillance globale prenant en compte des caractéristiques bas-niveau, et les biais centré et de profondeur. La même fusion additive des caractéristiques est utilisée que pour l'analyse précédente. Pour chaque fixation, les poids appris varient, ce qui produit des cartes de saillance adaptées temporellement. Ce modèle adapté est ainsi comparé aux cartes de saillance originales du modèle de Itti en 2D et en 3D. Notre modèle est également comparé à un modèle à pondérations égales : la carte de saillance est la moyenne des cinq cartes de caractéristiques (les poids $p_k(t)$ sont alors fixés à $1/K$, avec K le nombre de caractéristiques valant 5). Par la suite, deux critères sont

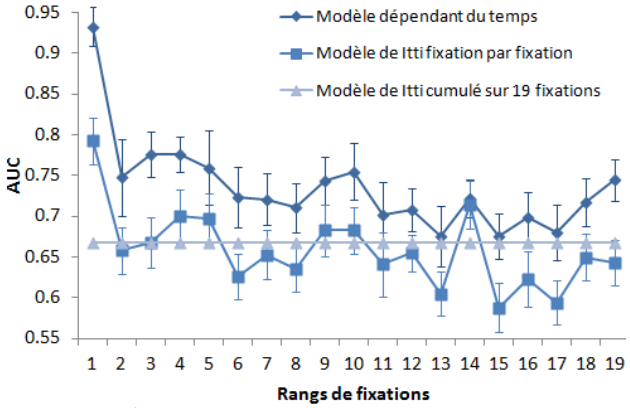

Figure 3 – *Évolution temporelle des performances du modèle dépendant du temps, du modèle de Itti fixation par fixation, et de Itti cumulés sur 19 fixations.*

utilisés pour l'étude et la comparaison des performances des modèles. L'aire sous la courbe ("AUC") de la caractéristique de fonctionnement du récepteur ("ROC") est utilisée pour quantifier le degré de similarité entre une distribution de fixations humaines (2D ou 3D) et une carte de saillance prédite. La mesure AUC est non-paramétrique et encadrée par la borne basse 0.5 en cas de correspondance au niveau aléatoire, et la borne haute 1 pour une correspondance parfaite. Ainsi pour chaque couple "image x fixation", à chaque rang de fixation, une valeur AUC est obtenue. Les résultats sont alors moyennés sur l'ensemble des images tests pour un rang de fixation donné. Les performances AUC au cours des fixations du modèle de Itti, et de notre modèle basé Itti adapté temporellement, sont illustrées à la figure 3. La valeur AUC fixe par image, telle qu'on la calcule habituellement, est également tracée. Les résultats illustrent un gain constant des performances au cours du temps, et prouvent l'importance de considérer le temps dans l'évaluation des modèles.

La mesure de Parcours Visuel Saillant Normalisé (NSS) est également employée pour évaluer les performances des cartes de saillance prédites sur les points de fixation. À la différence de l'AUC, le NSS n'est pas invariant à une transformation monotone croissante de la carte de saillance, est non borné et n'utilise pas de points aléatoires. Une valeur NSS est donnée pour chaque couple "image x fixation / participant". Les résultats sont moyennés comme avec l'AUC. La figure 4 illustre côte à côte les performances NSS et AUC du modèle original de Itti, du modèle à pondérations égales et du modèle adapté temporellement, en 2D et 3D, moyennées sur le temps.

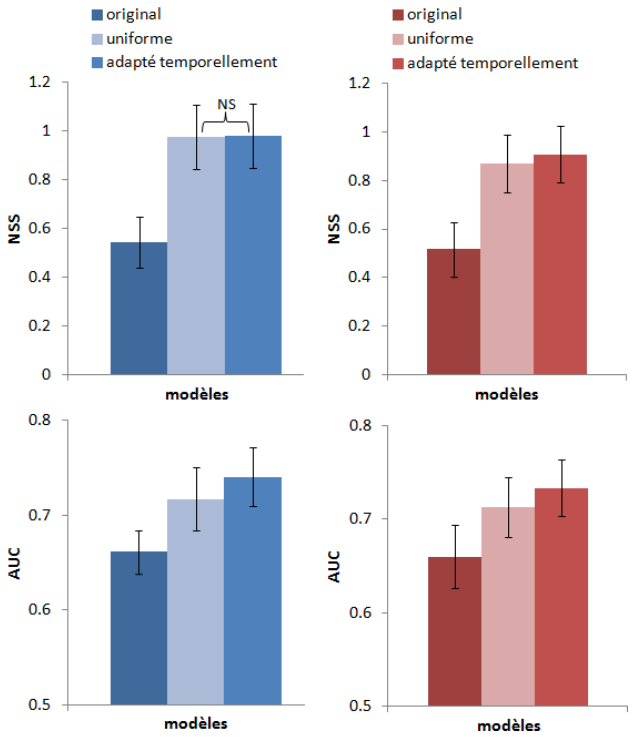

Figure 4 – *Comparaison des 3 modèles de saillances en conditions 2D (gauche) et 3D (droite). 1ère ligne : le critère NSS, seconde ligne le critère AUC. Les barres d'erreurs correspondent à l'erreur type. La mention NS signifie une différence Non-Significative.*

Tout d'abord, nous notons que les résultats sont tous au-dessus de la chance (0 pour le NSS et 0.5 pour l'AUC). Les 2 modèles incluant les 4 caractéristiques visuelles de saillance bas-niveau, de biais centré, avant et arrière-plan (ainsi que la contribution uniforme) dépassent significativement les performances du modèle original et ceci pour les deux métriques. Les différences entre modèles sont toutes statistiquement significatives, (t-test pairé, $p<0.05$) pour les 2 critères dans les 2 conditions (sauf dans le cas noté "NS"). Le modèle proposé à réellement amélioré les performances entre modèles de saillance. Alors que le modèle utilisant des poids uniformes améliore déjà significativement les performances, le modèle adapté temporellement augmente d'autant plus cette prédictibilité.

La base expérimentale contenait un nombre réduit (24) d'images avec différentes orientations, échelles de profondeur et contraste. L'apprentissage EM a été réalisé sur la

moitié de cette base, et l'évaluation des modèles sur la moitié restante. Par l'intégration de différentes contributions "externes" et "haut-niveau", la pertinence des cartes a été augmentée quelles que soient les conditions de vue, et quel que soit l'instant temporel. On constate néanmoins deux limitations : premièrement des stimuli images en luminance ont été présentés lors des expériences. Même si la couleur semble être un contributeur faible de l'attention relativement à la luminance [16], l'intégration de la couleur dans les modèles améliore leurs résultats. Ainsi, la contribution des caractéristiques bas-niveau incluant la couleur aurait pu être plus importante. Une seconde limitation provient du contenu de l'image en elle-même, limité ici à des scènes de forêts. Malgré tout, cela ne remet pas en cause la démarche, qui montre que considérer un *"pooling"* de caractéristiques bas-niveau avec des caractéristiques amont est plausible et efficace en image fixe. Enfin, en accord avec [5], il apparait que l'avant-plan est un bon prédicteur et une caractéristique visuelle pertinente pour l'attention visuelle.

5 Conclusion

L'objet de cette étude était d'analyser les différences dans le déploiement visuel en conditions mono et stéréoscopiques, et d'évaluer l'apport de caractéristiques pertinentes pour l'attention visuelle au cours du temps. L'analyse temporelle souligne des contributions successives des caractéristiques du centre, puis de l'avant-plan avec une implication constante de la saillance bas-niveau dès la 3ème fixation. La contribution importante de l'avant-plan, renforcée en stéréoscopie, en fait un prédicteur fiable de l'attention visuelle en début de visualisation. Ainsi un nouveau modèle de saillance a été défini pour lequel une fusion de caractéristiques, adaptée temporellement, permet de prédire l'attention sur images fixes. La fusion additive, aussi simple soit-elle, est une première approche pour combiner des mécanismes issus de l'aire corticale V1 à des mécanismes liés à V2. Enfin, l'intégration de différentes caractéristiques indépendantes au cours du temps est une manière pertinente de modéliser l'attention visuelle.

Références

[1] C. Koch et S. Ullman. Shifts in selective visual attention : towards the underlying neural circuitry. *Hum Neurobiol*, 4(4) :219–27, 1985.

[2] L. Itti, C. Koch, et E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11) :1254–1259, 1998.

[3] A. L. Yarbus. *Eye movements and vision*. Plenum press : New York, 1967.

[4] A. Torralba, A. Oliva, M. S Castelhano, et J. M Henderson. Contextual guidance of eye movements and attention in real-world scenes : The role of global features in object search. *Psychological review*, 113(4) :766–786, 2006.

[5] F. T. Qiu, T. Sugihara, et R. von der Heydt. Figure-ground mechanisms provide structure for selective attention. *Nature neuroscience*, 10(11) :1492–1499, 2007.

[6] J. E. Cutting et P. M. Vishton. Perceiving layout and knowing distances : The integration, relative potency, and contextual use of different information about depth. *Perception of space and motion*, 5 :69–117, 1995.

[7] A. Maki, P. Nordlund, et J. O Eklundh. A computational model of depth-based attention. Dans *Proceedings of the 13th International Conference on Pattern Recognition, 1996.*, volume 4, pages 734–739, 1996.

[8] A. Maki, P. Nordlund, et J. O. Eklundh. Attentional scene segmentation : integrating depth and motion. *Computer Vision and Image Understanding*, 78(3) :351–373, 2000.

[9] N. Ouerhani et H. Hugli. Computing visual attention from scene depth. Dans *15th International Conference on Pattern Recognition, 2000.*, volume 1, pages 375–378, 2000.

[10] Y. Zhang, G. Jiang, M. Yu, et K. Chen. Stereoscopic visual attention model for 3D video. *Advances in Multimedia Modeling*, pages 314–324, 2010.

[11] N.D.B. Bruce et J.K. Tsotsos. An attentional framework for stereo vision. Dans *The 2nd Canadian Conference on Computer and Robot Vision, 2005. Proceedings.*, pages 88–95. IEEE, 2005.

[12] 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*, 7(14), 2007.

[13] Q. Zhao et C. Koch. Learning a saliency map using fixated locations in natural scenes. *Journal of vision*, 11(3), 2011.

[14] T. Judd, K. Ehinger, F. Durand, et A. Torralba. Learning to predict where humans look. Dans *IEEE 12th International Conference on Computer Vision*, pages 2106–2113, 2009.

[15] B. T. Vincent, R. Baddeley, A. Correani, T. Troscianko, et U. Leonards. Do we look at lights ? using mixture modelling to distinguish between low-and high-level factors in natural image viewing. *Visual Cognition*, 17(6) :856–879, 2009.

[16] T. Ho-Phuoc, N. Guyader, et A. Guerin-Dugue. A functional and statistical Bottom-Up saliency model to reveal the relative contributions of Low-Level visual guiding factors. *Cognitive Computation*, 2(4) :344–359, 2010.

[17] L. Jansen, S. Onat, et P. Konig. Influence of disparity on fixation and saccades in free viewing of natural scenes. *Journal of Vision*, 9(1), 2009.

[18] J.M. Steger. Fusion of 3d laser scans and stereo images for disparity maps of natural scenes. *Publications of the Institute of Cognitive Science*, 14, 2010.

[19] L. Zhaoping, N. Guyader, et A. Lewis. Relative contributions of 2D and 3D cues in a texture segmentation task, implications for the roles of striate and extrastriate cortex in attentional selection. *Journal of vision*, 9(11), 2009.

[20] D. Parkhurst, K. Law, et E. Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1) :107–123, 2002.

[21] B. W. Tatler, R. J Baddeley, et I. D Gilchrist. Visual correlates of fixation selection : Effects of scale and time. *Vision Research*, 45(5) :643–659, 2005.

[22] S. Palmer. *Vision : From photons to phenomenology*. Cambridge, MA : MIT Press, 2000.