

**UNIVERSITE DE BLIDA 1**

**Faculté des Sciences**

Département d'Informatique

**THESE DE DOCTORAT**

en Informatique

**APPROCHE POUR LA DÉTECTION D'OBJETS  
SAILLANTS BASÉE SUR LA CLASSIFICATION  
PAR GRAPHE ET LE RAFFINEMENT SPATIAL**

Par

**Idir FILALI**

Devant le jury composé de :

S. OUKID	MCA,	U. de Blida1	Présidente
S. LARABI	Professeur,	USTHB	Examineur
A. SERIR	Professeur,	USTHB	Examineur
N.F. CHIKHI	MCA,	U. de Blida1	Examineur
N. BENBLIDIA	Professeur,	U. de Blida1	Directeur de thèse
M. S. ALLILI	Professeur,	U. du Québec en Outaouais	Co-Directeur de thèse

Blida, Mai 2017

## RESUME

Nous proposons une méthode d'ordonnement sur un graphe multi-couches pour la détection d'objets saillants et un processus de raffinement itératif défini à la fois au niveau local et global. A partir d'une décomposition multi-échelles de l'image en superpixels, nous proposons une fonction objective optimisée sur une nouvelle structure graphique compacte à couches superposées qui utilise les informations de couleur, de texture, de contour et de contiguïté spatiale pour l'estimation des parties saillantes de l'image. Les différentes couches sont disposées de bas en haut de la résolution à plus forte granularité à la résolution à plus faible granularité avec un couplage fort inter et intra résolutions. Pour plus de pertinence et de meilleures performances, nous proposons ensuite un processus de raffinement itératif constitué de trois étapes majeures. D'abord, nous définissons une procédure de localisation spatiale guidée par la saillance qui délimite la zone spatiale ajustée aux contours de l'objet saillant. Appliquer notre modèle graphique sur cet espace améliore la précision des résultats tout en écartant plus efficacement les parties de l'arrière-plan. Ensuite, nous utilisons conjointement l'information de contour et l'évaluation de la pertinence des caractéristiques utilisées pour mieux discriminer l'objet saillant du fond de l'image avec une meilleure précision dans la localisation de ses contours. Enfin, pour garder la conformité avec les statistiques globales de l'image, nous définissons un raffinement de la saillance basé sur les forêts aléatoires et les mixtures de gaussiennes. Notre algorithme retourne des résultats pertinents en comparaison avec les méthodes de l'état de l'art spécialement dans des cas de fonds d'image chargés et de contraste faible entre l'objet saillant et l'arrière-plan. En outre, notre approche est capable de détecter des objets saillants apparaissant à des échelles variées. Les résultats expérimentaux sur de larges bases d'images ont montré la supériorité de notre approche par rapport à plusieurs méthodes récentes sur la détection d'objets saillants.

## **REMERCIEMENTS**

Mes vifs remerciements s'adressent en premier lieu à mes encadrateurs, Mme N.BENBLIDIA professeur à l'université Saad Dahlab de Blida et Mr M.S ALLILI professeur à l'université du Québec en Outaouais pour m'avoir proposé ce sujet, dirigé mes travaux, et permis de développer mon sens de la recherche. Qu'ils soient assurés de ma profonde gratitude et de ma haute considération.

Je remercie vivement Mme S.OUKID, maître de conférences à l'université Saad Dahlab de Blida, pour m'avoir fait l'honneur de présider le jury.

Mon profond respect et mes cordiaux remerciements à Mr S.LARABI professeur à l'USTHB , Mme A.SERIR professeur à l'USTHB et Mr N.F CHIKHI maître de conférences à l'université Saad Dahlab de Blida pour leur participation à mon jury en tant qu'examineurs.

## TABLE DES MATIERES

<b>INTRODUCTION</b> .....	13
1 Cadre générale.....	13
2 Position du problème et objectifs.....	15
3 Contributions.....	18
4 Plan de la thèse.....	20
<b>1. LA SAILLANCE VISUELLE</b> .....	23
1.1 Introduction.....	23
1.2 L'attention.....	24
1.3 L'attention visuelle et la sélectivité.....	26
1.4 Le système visuel humain.....	27
1.4.1 Le flou rétinien.....	27
1.4.2 Structure de La rétine.....	28
1.4.3 Types d'attentions visuelles.....	31
1.5 Détection de la saillance visuelle.....	32
1.5.1 Saillance orientée espace ou objet.....	32
1.5.2 Détection d'objets saillants.....	35
1.5.3 Critères d'évaluation de la détection d'objets saillants.....	36
1.5.4 Information de haut niveau pour la détection d'objets saillants.....	36
1.6 Domaines étroitement liés à la détection d'objets saillants.....	38
1.6.1 Prédiction des fixations du regard.....	38
1.6.2 Segmentation d'image.....	39
1.6.3 Localisation d'objets.....	40
1.7 Conclusion.....	41
<b>2. ETAT DE L'ART SUR LA DETECTION D'OBJETS SAILLANTS</b> .....	43
2.1 Introduction.....	43
2.2 Méthodes basées sur l'estimation du contraste local.....	43
2.3 Méthodes basées sur l'estimation du contraste global.....	45

2.3.1 Méthodes probabilistes.....	45
2.3.2 Méthodes graphiques.....	48
2.3.2.1 Méthode graphique basée sur les marches aléatoires.....	48
2.3.2.2 Méthode graphique basée sur la distance géodésique.....	49
2.3.3 Méthodes basées sur la distance colorimétrique.....	50
2.3.3 Méthode d'estimation de la saillance dans le domaine fréquentiel.....	51
2.4 Méthodes hybrides.....	53
2.4.1 Méthodes probabilistes.....	53
2.4.1.1 Méthode d'estimation de la saillance au niveau pixel: Analyse de la transformée en ondelettes.....	54
2.4.1.2 Estimation de la saillance au niveau bloc: Apprentissage de codes caractéristiques.....	55
2.4.1.3 Estimation de la saillance au niveau superpixel: Approche basée sur les multi-résolutions et les distributions normales.....	57
2.4.2 Méthodes basées sur les graphes.....	59
2.4.2.1 Méthodes basé sur une seule résolution.....	59
2.4.2.1.1 Marches aléatoires markoviennes.....	59
2.4.2.1.2 Connectivité avec l'arrière plan.....	63
2.4.2.1.3 Ordonnancement sur un graphe mono couche.....	65
2.4.2.2 Méthodes basées sur plusieurs résolutions.....	68
2.4.2.2.1 Structure arborescente pour la détection de la saillance.....	68
2.4.2.2.2 Structure hiérarchique pour la détection de la saillance.....	73
2.4.2.2.3 Multi-graphes pour la détection d'objets saillants....	76
2.4.3 Autres méthodes.....	78
2.4.3.1 Reconstruction dense et éparses d'erreur .....	78
2.4.3.2 Décomposition matricielle de l'image: partie redondante et partie saillante.....	81
2.4.3.3 Apprentissage de régression de type forêts aléatoires.....	82
2.4.3.4 Espace de haute dimension pour la détection de la saillance.....	83
2.4.3.5 Estimation de la saillance basée sur le calcul de la variance	85
2.5 Conclusion.....	89
<b>3. ORDONNANCEMENT SUR UN GRAPHE MULTI-COUCHES POUR LA DETECTION D'OBJETS SAILLANTS.....</b>	<b>91</b>
3.1 Introduction.....	91
3.2 Segmentation en superpixels.....	92
3.2.1 Intérêt de la segmentation en superpixels.....	92
3.2.2 La méthode SLIC.....	93
3.3 Extraction de caractéristiques.....	95

3.3.1 La couleur.....	96
3.3.2 La texture.....	96
3.3.3 Les contours.....	97
3.4 Graphe multi-couches pour la détection d'objets saillants.....	97
3.4.1 Construction du graphe.....	97
3.4.1.1 Structure du graphe.....	97
3.4.1.2 Ajustement des poids du graphe.....	99
3.4.2 Ordonnancement sur notre graphe multi-couches .....	103
3.4.3 Etapes de détection de la saillance.....	107
3.4.3.1 Ordonnancement avec des requêtes extraites de l'arrière-plan.....	107
3.4.3.2 Ordonnancement avec des requêtes extraites du noyau de l'objet saillant.....	109
3.5 Conclusion.....	112
<b>4. RAFFINEMENT SPATIAL DE LA SAILLANCE.....</b>	<b>114</b>
4.1 Introduction.....	114
4.2 Localisation spatiale de l'objet saillant.....	115
4.3 Raffinement de la saillance en utilisant les informations de contour et de pertinence des caractéristiques.....	120
4.3.1 Information de contour.....	120
4.3.2 Evaluation de la pertinence des caractéristiques.....	121
4.4 Raffinement global de la saillance.....	124
4.5 Notre algorithme.....	128
4.6 Conclusion.....	131
<b>5. EVALUATION EXPERIMENTALE.....</b>	<b>132</b>
5.1 Introduction.....	132
5.2 Bases de tests.....	132
5.3 Mesures d'évaluation.....	133
5.4 Evaluation expérimentale de notre algorithme.....	135
5.5 Etude comparative de notre algorithme sur les bases MSRA-B et ASD.....	138
5.6 Etude comparative de notre algorithme sur les bases SOD, SED1, SED2.....	144
5.7 Analyse du temps de calcul.....	149
5.8 Conclusion.....	153
<b>CONCLUSION GÉNÉRALE.....</b>	<b>154</b>
<b>REFERENCES.....</b>	<b>158</b>

## TABLE DES FIGURES

Figure 1	Exemple illustrant les résultats des meilleures méthodes de l'état de l'art dans le cas d'images complexes .....	17
Figure 2	Schéma général de notre méthode pour la détection d'objets saillants .....	19
Figure 1.1	Flux d'informations que le système visuel humain gère quotidiennement.....	24
Figure 1.2	Déplacement du regard en fonction de la saillance visuelle.....	27
Figure 1.3	Flou rétinien. Les images capturées par la rétine sont bien plus précises au centre.....	28
Figure 1.4	Structure de l'œil humain et organisation cellulaire de la rétine.....	30
Figure 1.5	Exemple illustrant la perception globale des formes.....	33
Figure 1.6	Illustration des différentes lois de la Gestalt.....	34
Figure 1.7	Quelques exemples d'objets d'intérêt.....	35
Figure 1.8	Résultats de traitement de différents modèles pour une même image.....	38
Figure 2.1	Structure pyramidale dont chaque niveau correspond à une échelle de représentation. La différence de gaussiennes est utilisée pour estimer le contraste centre/périphérie à travers la pyramide .....	44
Figure 2.2	Limitation des méthodes basées sur l'estimation du contraste local et des méthodes basées sur l'estimation du contraste global.....	53
Figure 2.3	Représentation de 200 fonctions de base à partir de patches de taille 8x8 sur le canal <i>L</i> de l'espace de couleur <i>LAB</i> .....	56
Figure 2.4	Des exemples illustrant les avantages du traitement de mise à jour de la saillance.....	62

Figure 2.5	Différentes étapes de la méthode présentée dans [82].....	65
Figure 2.6	Structure du graphe utilisé dans [9].....	65
Figure 2.7	Différentes étape de la méthode proposée dans [9].....	66
Figure 2.8	Exemples où l'objet saillant touche au moins un bord de l'image....	68
Figure 2.9	Différente étapes de génération de la carte de saillance initiale [10].....	71
Figure 2.10	Exemple illustrant la fusion des nœuds dans l'arbre proposé dans [10].....	72
Figure 2.11	Exemple où la méthode présentée dans [10] génère un taux de faux positifs important.....	72
Figure 2.12	Exemple illustrant les différentes résolutions du graphe proposé dans [18].....	73
Figure 2.13	Aperçu général de la méthode de [18].....	75
Figure 2.14	Exemple où la méthode présentée dans [18] estime mal la saillance.....	75
Figure 2.15	Exemple où l'objet saillant est connecté à un des bords de l'image et résultats retournées par la méthode présentée dans [15].....	77
Figure 2.16	Cartes de saillance basées sur la reconstruction dense et éparses de l'erreur [86].....	80
Figure 2.17	L'image en couleur peut être décomposée en deux parties : partie redondante et partie saillante [89].....	81
Figure 2.18	Exemple où l'objet saillant est mal détecté par la dans [16].....	83
Figure 2.19	Illustration de la saillance locale [90].....	85
Figure 2.20	Exemple de déroulement de l'algorithme de [91].....	87
Figure 3.1	Images segmentées en superpixels de taille approximative de 64, 256 et 1014 pixels en utilisant l'algorithme SLIC.....	94
Figure 3.2	Illustration de notre graphe multi-couches avec trois résolutions superposées pour la détection d'objets saillants.....	99
Figure 3.3	Comparaison des cartes de saillances retournées par notre méthode d'ordonnement sur un graphe multi-couches en considérant la couleur, LBP, l'entropie et Leur combinaison	



	respectivement.....	100
Figure 3.4	Exemples d'estimation de contour.....	103
Figure 3.5	Différentes étapes d'ordonnement effectué par notre méthode..	110
Figure 3.6	Comparaison entre les cartes de saillance retournées par l'ordonnement sur des graphes mono-couche et l'ordonnement sur un graphe multi-couches.....	111
Figure 3.7	Exemples qui comparent le résultat de notre ordonnancement et le résultat de l'ordonnement proposé dans [9].....	112
Figure 4.1	Figure4.1 Exemples de localisation spatiale d'objets saillants.....	123
Figure 4.2	Exemple montrant le résultat du classifieur <i>RF</i> en considérant le bord de gauche de l'image comme information à priori sur l'arrière-plan.....	125
Figure 4.3	Illustration du processus de fusion des différentes cartes de saillance (équations 4.17 et 4.18) sur l'exemple de la figure 4.2.....	127
Figure 4.4	Exemple illustrant notre processus de raffinement global.....	130
Figure 5.1	Évolution de la mesure $F_{\mu}$ en fonction du nombre croissant des résolutions sur 500 images de la base MSRA-B.....	137
Figure 5.2	Résultats comparatifs des différentes méthodes <i>SLR</i> , <i>WSLR</i> , <i>RF</i> , <i>MLR</i> , <i>WMLR</i> et <i>WMLR-RF</i> en utilisant les bases MSRA-B et ASD en utilisant la courbe <i>P-R</i> et les mesures <i>F-measure</i> , <i>P</i> et <i>R</i> .....	137
Figure 5.3	Résultats comparatifs des différentes méthodes <i>SLR</i> , <i>WSLR</i> , <i>RF</i> , <i>MLR</i> , <i>WMLR</i> et <i>WMLR-RF</i> en utilisant les bases MSRA-B et ASD en utilisant la courbe <i>ROC</i> et les mesures <i>AUC</i> et <i>MAE</i> .....	138
Figure 5.4	Courbes <i>P-R</i> obtenus pour 21 méthodes de l'état de l'art en comparaison avec notre méthode en utilisant la base MSRA-B.....	140
Figure 5.5	Valeurs obtenue pour les mesures <i>P</i> , <i>R</i> et <i>F-measure</i> pour 21 méthodes de l'état de l'art en comparaison avec notre méthode sur MSRA-B.....	140
Figure 5.6	Comparaison de notre méthode avec <i>DRFI</i> sur 2500 images de MSRA-B.....	141
Figure 5.7	Courbes <i>P-R</i> obtenus pour 22 méthodes de l'état de l'art en comparaison avec notre méthode en utilisant la base ASD.....	141
Figure 5.8	Valeurs obtenue pour les mesures <i>P</i> , <i>R</i> et <i>F-measure</i> pour 21 méthodes de l'état de l'art en comparaison avec notre méthode sur	

	ASD.....	142
Figure 5.9	Résultats comparatifs entre notre méthode et les onze meilleures méthodes de l'état de l'art sur la base MSRA-B en utilisant les mesure ROC, AUC et MAE.....	143
Figure 5.10	Résultats comparatifs entre notre méthode et <i>DRFI</i> sur 2500 image de la base MSRA-B en utilisant les mesure ROC, AUC et MAE.....	143
Figure 5.11	Résultats comparatifs entre notre méthode et les douze meilleures méthodes de l'état de l'art sur la base ASD en utilisant les mesures ROC, AUC et MAE.....	144
Figure 5.12	Résultats comparatifs entre notre méthode et les 12 meilleures méthodes de l'état de l'art sur la base de test SOD.....	145
Figure 5.13	Résultats comparatifs entre notre méthode et les 12 meilleures méthodes de l'état de l'art sur la base de test SED1.....	146
Figure 5.14	Résultats comparatifs entre notre méthode et les 12 meilleures méthodes de l'état de l'art sur la base de test SED1.....	147
Figure 5.15	Exemple montrant des cas où <i>DRFI</i> n'arrive pas à bien se généraliser.....	148
Figure 5.16	Comparaison visuelle entre notre méthode et les meilleures méthodes de l'état de l'art sur la détection d'objets saillants sur la base MSRA-B.....	151
Figure 5.17	Comparaison visuelle entre notre méthode et les meilleures méthodes de l'état de l'art sur la détection d'objets saillants sur les bases SOD, SED1 et SED2.....	152

## LISTE DES TABLEAUX

Tableau 1.1	Deux structure de type center-surround existant dans la rétine.	31
Tableau 2.1	Différentes méthodes de l'état de l'art sur la détection d'objets saillants.....	88
Tableau 5.1	Temps de calcul moyen pour chaque tâche de notre modèle WMLR-RF.....	152
Tableau 5.2	Temps de calcul moyen de notre méthode en comparaison avec certaines méthodes de l'état de l'art.....	153

## LISTE DES SYMBOLES ET ABREVIATION

LBP	Local Binary Pattern
SLIC	Simple Linear Iterative Clustering
SC	approche séparation/combinaison
MML	Minimum message length
EMA	Expectation maximisation algorithm
GMM	Gaussians Mixture Model
PR	Precision-Recall
ROC	Receiver Operating Characteristic
AUC	Area Under the ROC Curve
MAE	Mean Absolute Error
$F_{\mu}$	F-Measure
SLR	Single Layer graph Ranking
WSLR	Window refinement for Single Layer graph Ranking
MLR	Multi Layer graph Ranking
WMLR	Window refinement for Multi Layer graph Ranking
RF	Random Forest based Saliency approach
WMLR-RF.	Window and random forest based refinement for Multi Layer graph Ranking
CS	Context and shape prior
VSAS	human Visual Sensitivity And Amplitude Spectrum
HDCT	High Dimensional Color Transform and local spatial support
RBD	saliency optimization from Robust Background Detection
SMD	Structured Matrix Decomposition
SLMR	Segmentation driven Low-rank Matrix Recovery
HSL	Hierarchical Saliency detection
SCC	Joint Spatial-Color Constraint and multi-scale segmentation
ST	Saliency Tree

LMR	Unified approach to salient object detection via Low rank Matrix Recovery
BS	Bayesian Saliency via low and mid level cues
PCA	Principal components analysis based method
CSP	Combining Simple Priors method
WT	low level features based on Wavelet Transform
MSN	Multi-size Superpixel approach based on multivariate Normal distribution estimation
LGC	saliency region detection via simple Local and Global Contrast
SIA	saliency region detection with Soft Image Abstraction
DSR	saliency detection via Dense and Sparse Reconstruction error
MC	saliency detection via absorb Markov Chain
EQCUT	Extended Quantum Cuts
DRFI	Discriminative Regional Feature Integration approach
MSRA	Microsoft Research Area dataset
ASD	Accurate human labeled images for Saliency Detection dataset
SOD	Salient Objects Dataset
SED1	Segmentation Evaluation Database on one salient object
SED2	Segmentation Evaluation Database on two salient objects

## INTRODUCTION GENERALE

### 1 Cadre générale

La vision par ordinateur est un domaine de recherche dédiée au développement d'algorithmes de traitement d'informations issues d'images numérisées par le biais de modèles fondés sur la géométrie, la biologie, les statistiques et la théorie d'apprentissage, etc. Les questions qui se posent alors concernent la nature de ces informations et leur représentation: quelle sorte d'information extraire de l'image? Comment décrire et/ou représenter cette information pour en faciliter l'interprétation? Il est clair que la nature et la représentation des informations dépendent de l'application envisagée. Cependant, dans tout processus d'analyse d'image, il faut pouvoir extraire certaines parties de l'image, mesurer les propriétés de ces parties ou les relations entre ces parties et utiliser les valeurs de ces propriétés pour interpréter le contenu de l'image. Ces trois étapes d'extraction, de mesure et d'interprétation, sont présentes dans presque toutes les applications. Des développements importants ont été réalisés depuis les années 40, mais la vision par ordinateur reste un champ d'investigation très actif avec de nombreux problèmes difficiles et non entièrement résolus.

La détection d'objets saillants est un problème important de vision par ordinateur qui consiste à localiser des objets d'intérêt ayant la particularité d'attirer l'attention dans les zones spatiales les plus informatives d'une scène visuelle. Une région est dite saillante si elle possède un contraste important par rapport aux autres régions avoisinantes ou bien par rapport à toute à la scène visuelle. Cette distinction est estimée en utilisant des mesures qui encodent les propriétés d'irrégularité, de rareté, de compacité, etc. Le système visuel humain traite un énorme flux d'informations visuelles de l'ordre de  $10^8$  bits/s à  $10^{10}$  bits/s [1]. Il possède cependant la faculté de localiser les régions saillantes sans aucun effort et avec une rapidité et efficacité étonnantes tout en associant une interprétation

géométrique à l'agencement de ces régions pour constituer des formes d'objets. L'attention, en tant qu'outil de gestion de l'information visuelle permet à l'être humain de construire une perception adaptée à ses capacités et ses besoins. L'efficacité de ce mécanisme passe par un processus de sélection des informations les plus pertinentes [1]. Sans l'attention, notre système visuel ne pourrait pas traiter en un temps raisonnable un problème si complexe. Cette propriété de gestion de la complexité fait de l'attention un mécanisme clé pour la création de systèmes de vision artificiels [2].

La détection d'objets saillants vise à construire des systèmes automatiques capables d'égaliser les performances de la vision humaine en termes d'efficacité et rapidité de détection des régions saillantes. Les approches classiques de détection d'objets saillants s'inspirent des modèles de fixation du regard humain pour estimer l'unicité d'une région de l'image par rapport à son entourage [3,4]. La difficulté majeure rencontrée dans ces approches est la sélection de la taille appropriée des régions voisines du moment que la surface de l'objet est inconnue a priori. Ce type de méthodes tend à retourner les contours de l'objet plutôt que l'objet lui-même.

D'autres types d'approches se sont basés sur l'estimation de l'unicité des régions en termes de statistiques globales de l'image pour la détection de la saillance. La théorie des probabilités [5,6] et l'analyse de fréquences [7,8] ont largement été utilisées dans ce contexte. L'inconvénient de ces approches est qu'elles ne sont pas adaptées pour l'encodage de l'information spatiale comme la contiguïté des régions adjacentes. Cela a comme conséquence de retourner des petites régions isolées de l'arrière-plan ayant un contraste élevé par rapport à toute l'image générant un bruit dans la carte de saillance. Une autre limitation majeure de ce type d'approches est que du moment qu'elles sont basées sur les statistiques globales de l'image, elles ne sont pas appropriées pour la détection d'objets saillants dont la surface occupe la partie majoritaire de l'image.

Comme solution à toutes ces limites, plus récemment, de nouvelles approches proposent de combiner l'information locale et globale pour améliorer l'efficacité de détection d'objets saillants. Parmi ces méthodes, nous pouvons par exemple dénombrer les méthodes basées sur les graphes [9,10] et les méthodes basées sur les probabilités [11,6]. En plus de leur simplicité, les graphes offrent un

cadre efficace pour l'encodage des informations à priori sur la position spatiale comme la contiguïté des régions d'objets saillants et l'information à priori sur la position de l'arrière-plan. L'unité de traitement pour certaines de ces méthodes est le bloc [12] (ensemble de pixels constituant une forme régulière rectangulaire). L'inconvénient majeure de ces méthodes est qu'elles ne respectent pas les contours des objets du moment que les blocs sont extraits d'une manière uniforme de l'image sans aucune analyse de leur contenu.

Pour une meilleure détection aux abords des contours des objets, plusieurs méthodes se sont penchées à considérer le superpixel [10,13] comme unité de traitement pour leurs algorithmes. L'avantage du superpixel est qu'il est constitué d'un ensemble de pixels voisins homogènes. Comme le contour forme un ensemble de pixels hautement contrasté avec son voisinage qui sépare deux régions adjacentes, il s'adhère généralement bien avec les frontières des superpixels. La préservation des contours contribue considérablement à l'efficacité du système de détection des objets saillants et permet d'augmenter le contraste entre l'objet saillant et le fond de l'image. Cependant, même si la segmentation en superpixels a prouvé qu'elle permet d'améliorer la qualité de la saillance retournée, le choix de l'échelle de segmentation est arbitraire car la surface de l'objet saillant est inconnu à priori. En effet, l'échelle de segmentation a un impact direct sur l'estimation de la saillance. D'une part, une segmentation à granularité fine retourne les régions isolées et les petits détails du fond de l'image comme saillants et cela génère un bruit dans la détection. D'une autre part, une segmentation à granularité grossière retourne des grandes régions et omet souvent d'inclure certains détails de l'objet saillant. Une solution adoptée par plusieurs méthodes est de considérer plusieurs résolutions sur lesquelles des fonctions d'estimation de la saillance sont optimisées. Les performances de ces méthodes améliorent nettement les résultats retournés quel que soit la surface occupée par l'objet saillant dans l'image.

## 2 Position du problème et objectifs

Malgré tous les efforts et progrès effectués dans la détection d'objets saillants, les performances de toutes les méthodes déjà proposées sont encore bien loin d'égaliser les facultés de la perception humaine dont les résultats de



détection se rapprochent étonnamment de la vérité terrain. Cette difficulté est surtout perçue dans les cas de scènes complexes suivantes: 1) le fond de l'image est chargé et doté d'une structure diversifiée en couleur et texture (cf. figure 1 première rangée); 2) l'objet saillant n'est pas très contrasté par rapport à l'arrière-plan, par exemple la similarité de distribution de couleur entre le fond et l'objet saillant est élevée (cf. figure 1, rangée 2); 3) présence de distracteurs dans l'image (cf. figure 1, rangée 3); 4) présence de plusieurs objets saillants dans l'image (cf. figure 1, rangée 2); 5) Variation d'échelles sous lesquelles les objets apparaissent dans l'image (cf. figure 1, rangée 4). La figure 1 illustre les cartes de saillance retournées sur des images complexes par trois des meilleures méthodes de l'état de l'art [14]: Extended Quantum Cuts (EQCUT) [15], Discriminative Regional Feature Integration Approach (DRFI) [16] et Low-Rank and Structured Sparse Matrix Decomposition (SMD) [17].

Construire des algorithmes de plus en plus performants pour la détection d'objets saillants est un objectif que s'est donc fixée la communauté de recherche en vision artificielle. Les algorithmes basés sur l'optimisation des graphes sur des structures à multi résolutions ont montré récemment une efficacité de détection remarquable comme mentionné dans la section 1. Certaines de ces méthodes construisent un graphe indépendant sur chaque résolution pour la génération d'une carte de saillance puis une opération de fusion de toutes ces cartes est effectuée [15]. Le couplage est donc défini uniquement sur chaque résolution indépendamment des autres. En d'autres termes, chaque nœud est connecté uniquement aux nœuds présents au sein du même graphe. D'autres méthodes [18,10] construisent une structure graphique multi-couches hiérarchique sur la base de la superposition des différentes résolutions. Les nœuds adjacents et similaires de chaque résolution sont fusionnés pour former un nœud appartenant à la résolution de niveau supérieur. Chaque nœud est alors connecté uniquement avec les nœuds de niveau inférieur qui ont contribué à sa formation ou bien au nœud de niveau supérieur dont il a contribué à sa formation. Nous notons qu'il n'y a aucun lien entre les nœuds d'une même résolution et que chaque résolution n'est connectée qu'aux résolutions de niveaux inférieur et supérieur uniquement. Toutes ces méthodes sont définies avec un couplage faible. Développer une

fonction sur une structure dont toutes les résolutions sont inter connectées tout en gardant les liens intra résolution pourrait sensiblement améliorer les résultats.

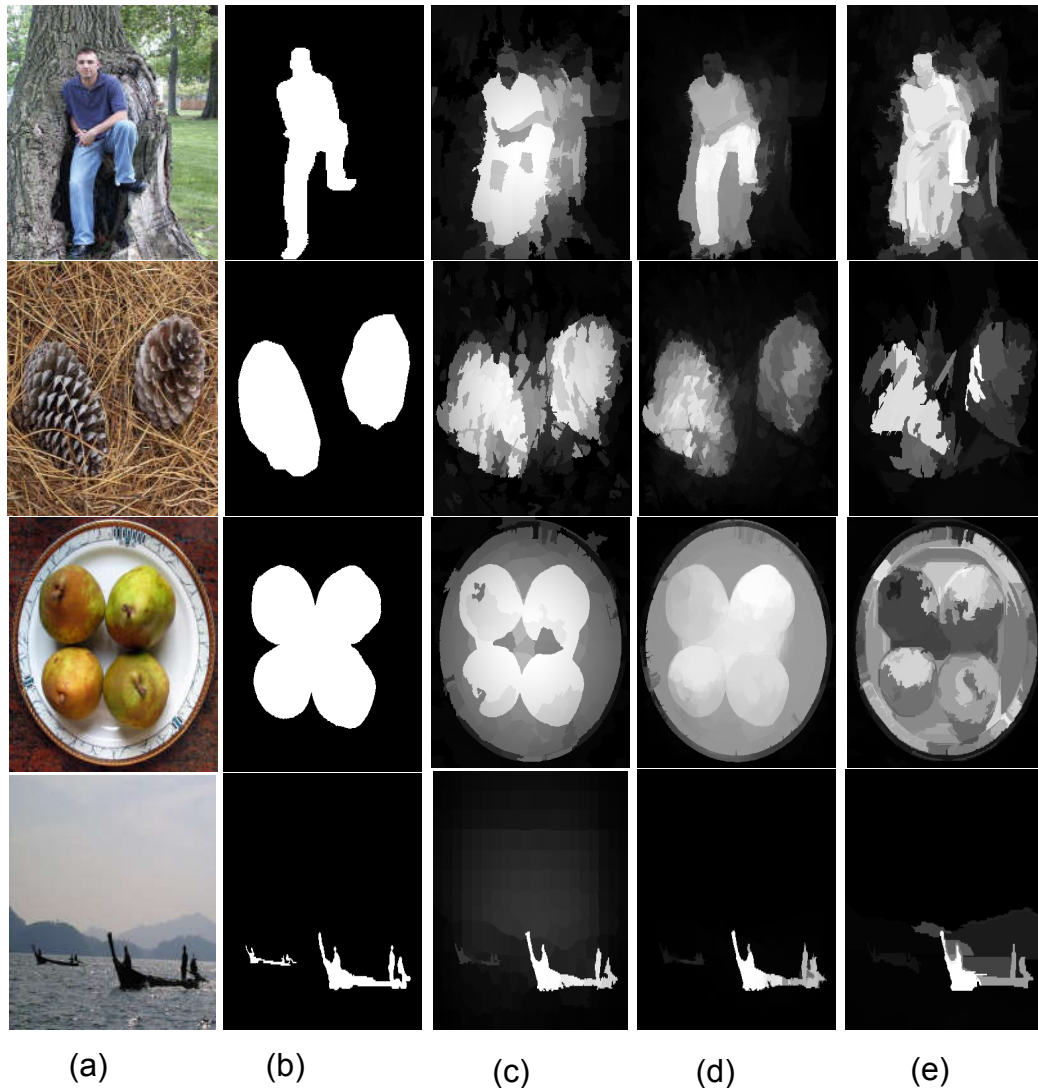


Figure 1 Exemple illustrant les résultats des meilleures méthodes de l'état de l'art dans le cas d'images complexes. De gauche à droite: (a) l'image en couleur, (b) Vérité terrain, cartes de saillances retournées par les méthodes (c) SMD, (d) EQCUT, (e) DRFI

Nous notons que la faiblesse de la majorité des méthodes précédentes réside également dans l'utilisation arbitraire des caractéristiques de bas niveau pour la détection de la saillance. En effet, les objets peuvent avoir différents niveaux de contraste avec l'arrière-plan et une caractéristique peut être discriminante dans certaines images et inappropriée dans d'autres. Par exemple, si l'objet saillant possède une distribution colorimétrique similaire à celle du fond

de l'image et qu'il est caractérisé par une texture bien différente de l'arrière-plan, l'utilisation de la texture est un choix plus judicieux que la couleur.

A notre connaissance, il n'existe dans la littérature qu'un seul travail [16] qui effectue une sélection des caractéristiques les plus discriminatives. Cependant, cette sélection est basée sur un apprentissage à partir d'une large base d'images. Hors, le pouvoir discriminatif d'une caractéristique varie d'une image à une autre. En effet, une caractéristique non incluse dans la sélection peut très bien convenir pour la discrimination de l'objet saillant dans certaines images. En outre, ce modèle peut bien souffrir du problème de généralisation et peut ne pas convenir à de nouvelles bases d'images avec des objets saillants différents de ceux de la base d'apprentissage. Il convient alors de définir une nouvelle manière de combiner les différentes caractéristiques au niveau locale de l'image de façon à améliorer les résultats de la saillance.

Une autre manière d'améliorer les résultats de la saillance est de combiner l'information régionale et de contours. En effet, les contours d'objets sont caractérisés par des discontinuités régionales importantes. Cette caractéristique peut alors contribuer considérablement à l'estimation de la saillance. Bien que l'information de contours soit largement utilisée dans la segmentation [19,20], elle est totalement négligée dans la détection d'objets saillants.

### 3 Contributions

A la lumière des objectifs cités dans la section 2, nous avons proposé une approche efficace qui répond aux limites rencontrés par les méthodes de l'état de l'art spécialement dans le cas d'images complexes. Nos contributions s'établissent à différents niveaux comme suit:

Nous proposons une nouvelle structure graphique multi-couches compacte basée sur la décomposition multi-échelle de l'image en superpixels. Les différentes couches sont disposées de bas en haut de la résolution à plus forte granularité à la résolution à plus faible granularité et dotées d'un couplage fort inter et intra résolutions au niveau du voisinage de chaque nœud.

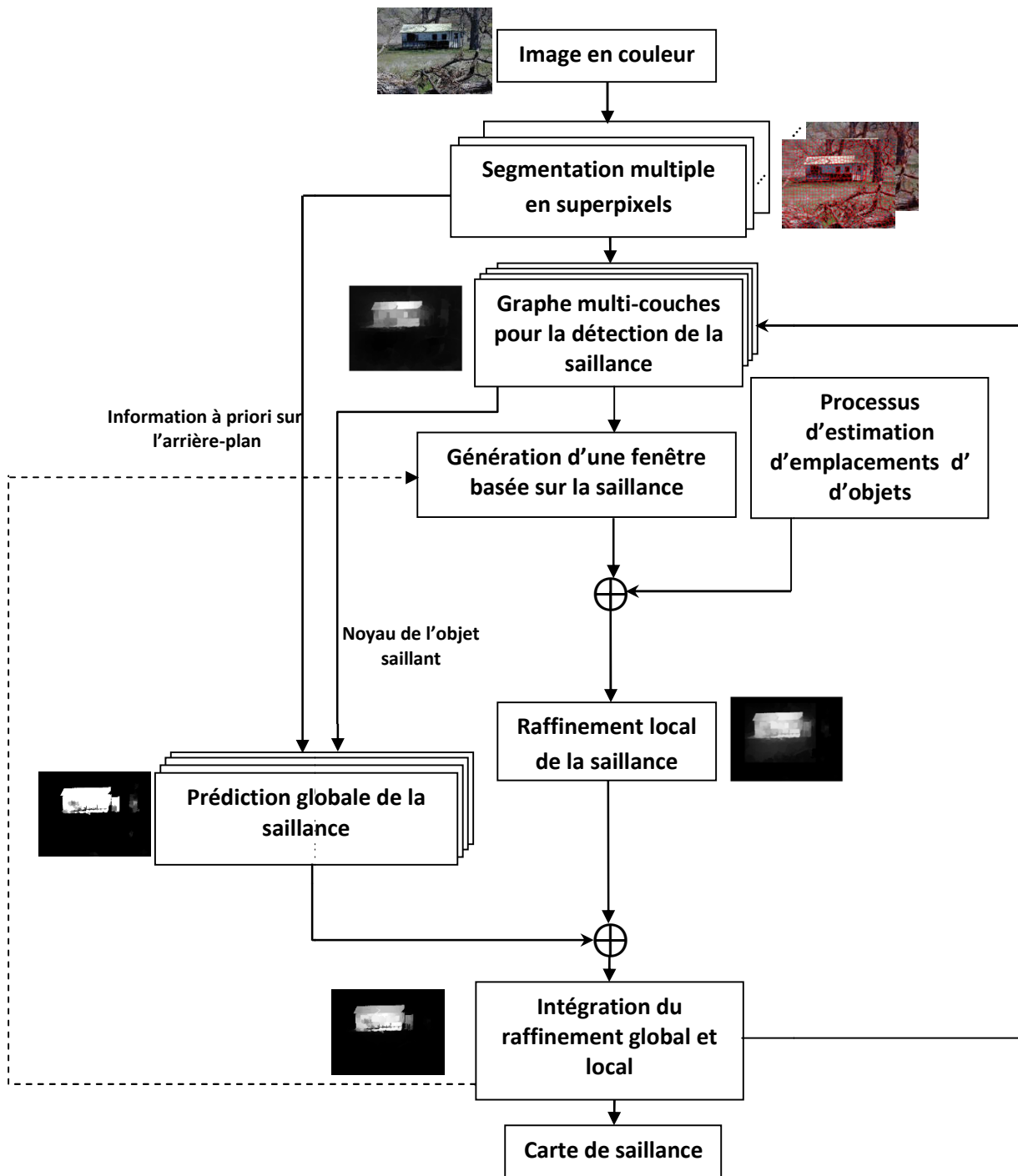


Figure 2 Schéma général de notre méthode pour la détection d'objets saillants

Pour une meilleure cohérence, nous raccordons chaque nœud avec ses voisins situés sur la même résolution (relations horizontales) ainsi qu'avec ses voisins situés sur toutes les autres résolutions (relations verticales). La structure de notre graphe multi-couches permet de combiner efficacement l'information de couleur, de texture, de distance spatiale et de contours à travers une analyse des différentes résolutions.

Nous proposons une nouvelle fonction objective optimisée sur notre graphe pour le calcul de la saillance. Toute la structure du graphe est utilisée à la fois dans cette optimisation et des valeurs de saillance sont assignées au niveau de toutes les résolutions tout en préservant les régions de l'objet présentées à différentes échelles.

Nous proposons une procédure qui raffine itérativement la localisation de l'objet saillant en combinant les informations locales et globales de l'image (cf. figure 2). D'une part, nous utilisons les forêts aléatoires et les mixtures de gaussiennes pour mieux séparer l'objet saillant de l'arrière-plan en termes de statistiques globales de l'image. Dans ce contexte, des informations à priori sont utilisées pour l'apprentissage du modèle. Le noyau de l'objet extrait à partir des parties les plus saillantes de l'image est utilisé comme information à priori sur l'objet saillant et les régions situées sur les bords de l'image comme information à priori sur l'arrière-plan. D'une autre part, nous définissons un processus de raffinement spatial basé sur les informations de texture, de couleur, de distance spatiale, de saillance, de localisation d'objets, de contours et d'évaluation de pertinence de chaque caractéristique utilisée. Ce raffinement consiste à délimiter le processus de détection de la saillance à un espace réduit aux alentours de l'objet saillant. L'estimation de la saillance est alors renforcée par l'analyse du pouvoir discriminatif de chaque caractéristique de bas niveau utilisée et la considération du paramètre du gradient.

#### 4 Plan de la thèse

Ce manuscrit est organisé en une introduction générale, cinq chapitres et une conclusion générale. Après avoir ressorti le schéma général de notre approche, nous allons retracer dans le chapitre 1 les origines et le fondement de la détection de la saillance. Initialement, les premiers travaux sur la détection de la saillance se sont inspirés des fixations du regard humain et des mécanismes de l'attention visuelle. En effet, la faculté de l'être humain à sélectionner les informations les plus pertinentes dans le flot de données auquel le cerveau est confronté reste inégalée de nos jours. Nous présenterons alors la structure du système visuel humain, les différents types d'attention visuelle et nous finirons par présenter la théorie de la forme qui affirme que l'attention joue un rôle dans les

phénomènes de groupement perceptuel permettant au système visuel d'être attiré par des forme particulière voir des objets dans les scènes visuelles. La majorité des travaux récents sur la détection de la saillance s'inspirent sur cette affirmation pour construire des détecteurs d'objets saillants.

Le chapitre 2 sera consacré à la présentation des différentes méthodes de détection d'objets saillants de l'état de l'art. Nous proposerons une taxonomie de ces méthodes en fonction du contexte spatiale sur laquelle l'estimation du contraste s'effectue. Les méthodes basées sur l'estimation du contraste local mesurent le contraste dans un contexte spatial réduit à une région ou un bloc et son entourage. Les méthodes basées sur l'estimation du contraste global mesurent le contraste dans un contexte spatial plus large englobant toutes les régions de la scène visuelle. Les méthodes hybrides se trouvent au confluent de ces deux types d'approches. Elles estiment le contraste à la fois dans un contexte spatial local et global. Parmi ces méthodes, on dénombre les méthodes basées sur les probabilités, les méthodes basées sur les graphes et d'autres méthodes basées la reconstruction dense et éparses d'erreur, la décomposition matricielle, apprentissage de forêts aléatoires, etc. Nous étudierons les limites de chaque type de méthodes avant de nous consacrer dans le chapitre suivant à la présentation de notre méthode

Nous présenterons dans le chapitre 3 notre approche basée sur l'ordonnancement à partir de notre graphe multi-couches. Nous avons définis un processus de diffusion de l'ordonnancement à partir des régions positionnées sur les bords de l'image vers tous les autres nœuds du graphe à travers toutes les couches pour constituer une première carte de saillance. Une deuxième diffusion de l'ordonnancement est effectuée à partir du noyau de l'objet obtenu à partir des parties les plus saillantes vers tous les autres nœuds de notre graphe pour constituer une carte de saillance intermédiaire.

Dans le chapitre 4, nous présenterons notre processus de raffinement de la saillance que nous définissons en trois étapes principales. Nous combinons l'information de saillance et de localisation d'objets pour estimer l'emplacement spatiale de l'objet saillant. Nous utilisons ensuite l'information de contour et l'évaluation de la pertinence de chaque caractéristique de couleur et de texture dans un cadre de raffinement régionale de la saillance. Ceci permet de mieux

ressortir l'objet saillant de l'arrière-plan et de mieux détecter ses contours. En parallèle, nous appliquons un algorithme basé sur l'apprentissage des forêts aléatoires et les mixtures de gaussiennes afin d'estimer la saillance dans un cadre global de l'image. Une stratégie de fusion des résultats de ces étapes est alors appliquée pour constituer la carte de saillance finale.

Le chapitre 5 sera consacré à la présentation de nos résultats expérimentaux en comparaison avec les meilleures méthodes de l'état de l'art, puis nous concluons ce mémoire de thèse en résumant les contributions scientifiques et les résultats obtenus. Nous présenterons également les perspectives de recherche.

## CHAPITRE 1 LA SAILLANCE VISUELLE

### 1.1 Introduction

La vision est le sens le plus utilisé dans la vie quotidienne. 80% de l'information que nous recevons chaque jour vient de notre système visuel: livres, télévision, internet, etc. Notre vie quotidienne (déplacements, orientation, coordination des mouvements) est également guidée par cette modalité. Durant chaque seconde, notre rétine peut recevoir jusqu'à 10 milliards de bits d'informations [22], tandis que le nombre total de neurones dans le cortex cérébral atteint seulement 20 milliards [23]. Ce fait indique que la quantité d'information visuelle que l'on reçoit est supérieure à la capacité de stockage de notre cerveau. En outre, la capacité de traitement de notre cerveau est également limitée et il est impossible d'effectuer simultanément une analyse complexe sur toute l'information visuelle d'entrée. Afin de remédier à ces problèmes, une tâche centrale du système de vision humaine est de détecter efficacement les parties les plus importantes du champ visuel. Ces parties sont rapidement sélectionnées et traitées en priorité par les ressources limitées du cerveau tandis que les autres parties sont souvent inhibées ou même ignorées afin d'augmenter l'efficacité du traitement.

Pour explorer davantage la capacité du cerveau humain dans le traitement de l'information, celui-ci est comparé avec les superordinateurs les plus rapides et les plus puissants dans le monde. En Novembre 2013, il est indiqué que le "K-Processeur" de Fujitsu est le 4<sup>ème</sup> plus puissant superordinateur dans le monde. En Août 2013, les scientifiques japonais et allemands utilisent 82501 processeurs et 1 pétaoctet de mémoire sur le "K-Processeur" pour simuler l'activité de 1,69 milliards de neurones reliés par 10,4 milliards de synapses. Les résultats ont montré qu'il a fallu 36 minutes pour reproduire une seconde d'activité cérébrale. Notons qu'un neurone peut approximativement traiter 200 signaux par seconde,



tandis qu'un processeur de l'ordinateur peut effectuer des milliards d'opérations par seconde (soit plusieurs millions de fois plus rapide qu'un neurone) [24]. Une question peut alors se poser: comment le réseau de neurones avec des milliards de neurones avec une activité lente peut surpasser remarquablement le supercalculateur? A cette question, une explication possible est que le flux massif de signaux d'entrée est traité efficacement, hautement résumé et faiblement représenté à un stade très précoce, avant qu'il entre dans les zones plus élevées du cerveau humain. Il s'agit de l'attention visuelle.



Figure 1.1 Flux d'informations que le système visuel humain gère quotidiennement [24]

## 1.2 L'attention

Dans [25], William James a décrit l'attention comme la prise de possession par l'esprit sous une forme claire et vive, d'un objet ou d'une suite de pensées parmi plusieurs qui semblent possibles. Elle implique le retrait de certains stimuli afin de traiter plus efficacement les autres. Par la suite, Alexander Romanovich

Luria s'est penchée à l'étude de la propriété sélective de l'attention [26] et que l'attention permet de se focaliser sur un stimuli bien particulier et ignorer les autres présents dans l'environnement. Selon Solso [27], l'attention se réfère au traitement d'une quantité limitée d'informations à partir de toutes les informations acquises à travers les sens, les souvenirs et autres processus cognitifs. De ce fait, l'attention a pour rôle d'accentuer certaines caractéristiques du stimulus (contraste, couleur, etc.), tout en filtrant ce qui pourrait distraire. Cette sélection facilite le traitement par le cortex visuel. L'attention contrôle donc l'allocation des ressources cognitives sur les informations acquises. Il existe à cet effet plusieurs types d'attention qui peuvent être classées en plusieurs catégories selon la manière dont les ressources cognitives sont allouées. Dans [28], cinq grandes familles d'attention sont définies:

*Attention ciblée*: Il s'agit de la capacité à répondre discrètement à des stimuli visuels, auditifs ou tactiles spécifiques (concentration sur la lecture de sous-titrage lors de la diffusion d'un film).

*Attention sélective* ou focalisée: est la capacité de maintenir d'une manière sélective la focalisation des ressources cognitives sur un stimuli spécifique tout en ignorant les autres stimuli concurrents (focalisation sur le protagoniste dans diverses scènes complexes d'un film tout en ignorant la présence des autres personnes dans ces mêmes scènes).

*Attention maintenue*: intervient après l'attention sélective, afin de rester focalisé sur la tâche / le *stimulus* en cours pendant un certain temps, évitant ainsi le passage trop rapide d'un *stimulus* ou d'une activité à l'autre ; (par exemple, rester attentif pendant le film).

*Attention partagée* ou divisée: permet de répondre à plusieurs stimuli simultanément et traiter plusieurs types d'informations (discuter avec des amis tout en regardant un film) ;

Dans le cadre de la vision (naturelle ou artificielle), l'attention est principalement sélective. Notons cependant que l'attention maintenue peut également jouer un rôle, par exemple, dans le suivi de cible et que l'attention divisée contribue à la réponse pour plusieurs stimuli visuels.

### 1.3 L'attention visuelle et la sélectivité

L'attention visuelle est la capacité de sélectionner des stimuli visuels parmi plusieurs et de les transmettre au cerveau pour leur traitement. Ceci implique l'habileté à rester en état d'alerte de façon à être prêt pour recevoir et traiter l'information visuelle; l'habileté à centraliser le regard sur les informations pertinentes parmi plusieurs stimuli situés dans le champ visuel en ignorant ceux non pertinents à la réalisation d'une tâche (attention sélective). L'attention visuelle inclut également l'effort mental utilisé pour concentrer le regard sur un stimulus, malgré les distractions, afin de poursuivre et compléter une tâche visuelle (la vigilance visuelle) et finalement la capacité de répondre à plusieurs stimuli visuels de façon simultanée pour réaliser deux tâches visuelles ou plus en même temps (l'attention divisée).

Dans de nombreux scénarios, la sélectivité de l'attention, en particulier l'attention dite ouverte, peut être caractérisée comme le mouvement du regard guidé par les différentes régions saillantes de la scène visuelle. Ces régions peuvent être modélisées sur une carte topographique 2D dite de saillance qui sert de référence pour mettre l'accent sur les régions saillantes de la scène visuelle. Nous notons que les modèles à carte de saillance se situent dans la continuité des travaux précurseurs d'Anne Treisman [29]. Selon la *Feature-Integration Theory*, l'attention est codée dans une carte centrale interne (dont le nom varie selon les théories: carte maîtresse, carte de saillance, etc.) permettant une représentation du champ visuel.

A titre d'exemple, nous considérons une peinture sur un tableau avec des contenus riches (cf. Figure 1.2). Il est particulièrement difficile de percevoir simultanément les détails de toute la scène. Le regard est donc d'abord attiré par les régions les plus saillantes puis déplacé vers les autres régions par ordre décroissant de saillance. Le déplacement du regard est donc guidé par la valeur d'importance de chaque région désignée par la valeur de saillance. Cette valeur de saillance est donc utilisée comme référence pour guider le déplacement du regard sur la scène visuelle. Généralement, plus la couleur affectée à une région dans la carte de saillance est chaude, plus elle est saillante (Figure 1.2). Les cartes de saillances sont plus généralement représentées par des niveaux de gris.

Dans ces cas, plus une région est représentée par un niveau de gris claire sur la carte, plus elle est considérée comme saillante.

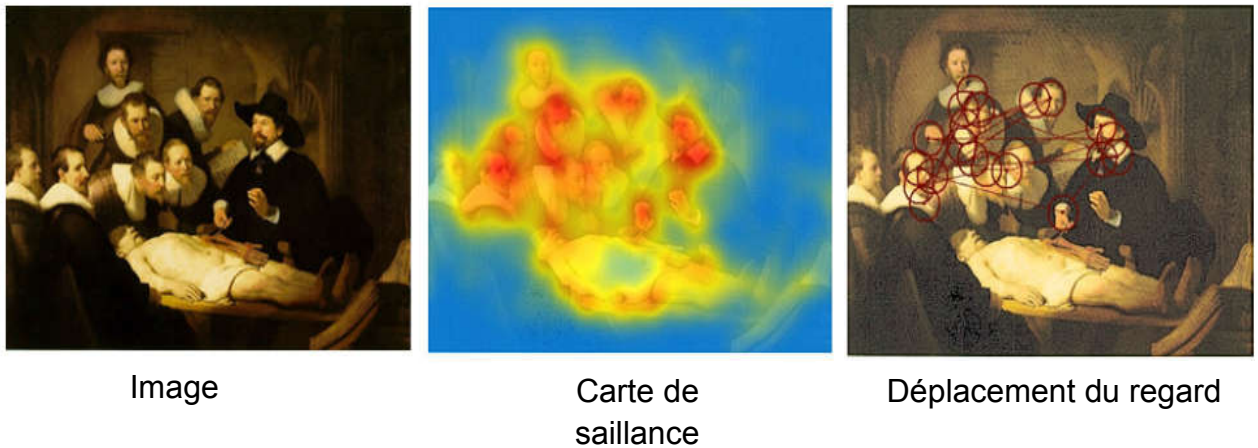


Figure 1.2 Déplacement du regard en fonction de la saillance visuelle [30].

#### 1.4 Le système visuel humain

Derrière un simple regard, notre cerveau met en œuvre une mécanique complexe, mais efficace, qui sollicite des milliards de neurones. Nous allons aborder dans cette section la perception du monde par l'œil humain, puis nous nous attarderons sur la structure et le fonctionnement de la rétine.

##### 1.4.1 Le flou rétinien

Notre vision du monde semble précise, continue et cohérente. Pourtant, l'étude des différents composants de notre système visuel laisse apparaître une situation bien différente. L'œil et la rétine ne capturent pas une image fidèle du monde. La répartition des différentes cellules photosensibles sur la rétine n'est pas homogène, le centre de celle-ci (la fovéa) contient beaucoup plus de récepteurs que la périphérie. Il en découle une représentation bien plus précise au centre de notre champ de vision (figure 1.3). Ainsi, il est nécessaire de déplacer nos yeux afin de capturer tous les détails de la scène et en construire une représentation cohérente [31]. Notre cortex visuel décompose le signal qu'il reçoit en différentes caractéristiques de complexité croissante (intensité, couleur, orientation, coins, lignes, intersections, etc.), générées dans des zones séparées (mais interconnectées) du cerveau [32]. Cette spécialisation se retrouve également dans le traitement de la reconnaissance des objets et leur localisation.

Notre cerveau doit alors reconstruire une représentation cohérente à partir d'un ensemble d'informations relativement hétérogènes. Comme nous le constatons, pour pouvoir interpréter ce que nous voyons, notre système visuel a mis en place des mécanismes de sélection et d'optimisation du traitement de l'information. C'est le cas de l'attention, qui voit l'explication de son origine.



Figure 1.3 Flou rétinien. Les images capturées par la rétine sont bien plus précises au centre [33]

#### 1.4.2 Structure de La rétine

L'œil est un dispositif optique assez classique. Les rayons lumineux traversent d'abord la cornée qui agit comme une première lentille, permettant aux rayons lumineux de se concentrer afin de traverser la pupille, dont le diamètre dépend de l'ouverture de l'iris. L'iris permet de réguler la quantité de lumière qui traversera le cristallin pour être projetée sur la rétine. Au centre de la rétine, dans le prolongement de l'axe optique, on trouve la fovéa. C'est dans cette zone, où

l'on retrouve la plus grande quantité de photorécepteurs, que notre acuité visuelle est la plus grande. La circulation de l'information dans la rétine est résumée en trois étapes principales [24]:

*Les photorécepteurs (cônes et bâtonnets):* Leur répartition n'est pas uniforme, plus on s'éloigne de la fovéa et moins les récepteurs sont nombreux. Les cônes, principalement utilisés pour la vision diurne, sont concentrés au niveau de la fovéa alors que les bâtonnets, principalement utilisés pour la vision nocturne, sont absents de cette zone. De par leur sensibilité importante, les bâtonnets permettent une grande aptitude à réagir aux mouvements. Cependant, cette capture du mouvement n'est pas très précise spatialement. Le rôle des photorécepteurs est de transformer les signaux lumineux en signaux électriques dans le réseau neuronal. Certains de ces signaux sont ensuite transmis aux cellules horizontales qui effectuent le calcul du signal moyen.

*Cellules de transmission et transformation de l'information:* Le signal provenant des cônes et bâtonnets est transformé par les cellules horizontales, amacrines et bipolaires avant de traverser différents types de cellules ganglionnaires. Les cellules bipolaires calculent la différence entre le signal moyen résultant du traitement des cellules horizontales et le signal provenant des photorécepteurs afin de garder uniquement les signaux hautement contrastés. Ces signaux sont ensuite transmis aux cellules amacrines qui effectuent un second calcul du signal moyen.

*Cellules ganglionnaires:* Elles sont généralement situées à proximité de la surface interne de la rétine. Elles reçoivent des informations visuelles à partir de cellules bipolaires et les cellules amacrines et envoient des informations visuelles à différentes parties du cerveau via les fibres du nerf optique.

Comme montré ci-dessus, deux mécanismes de prétraitement opèrent dans la rétine: différence et moyenne. Avec ces opérations, la rétine permet de compresser et coder efficacement l'information visuelle. Il y a environ 124 millions de photorécepteurs dans la rétine, y compris environ 120 millions de bâtonnets et près de 8 millions de cônes [34], pour capturer les signaux lumineux. Cependant,

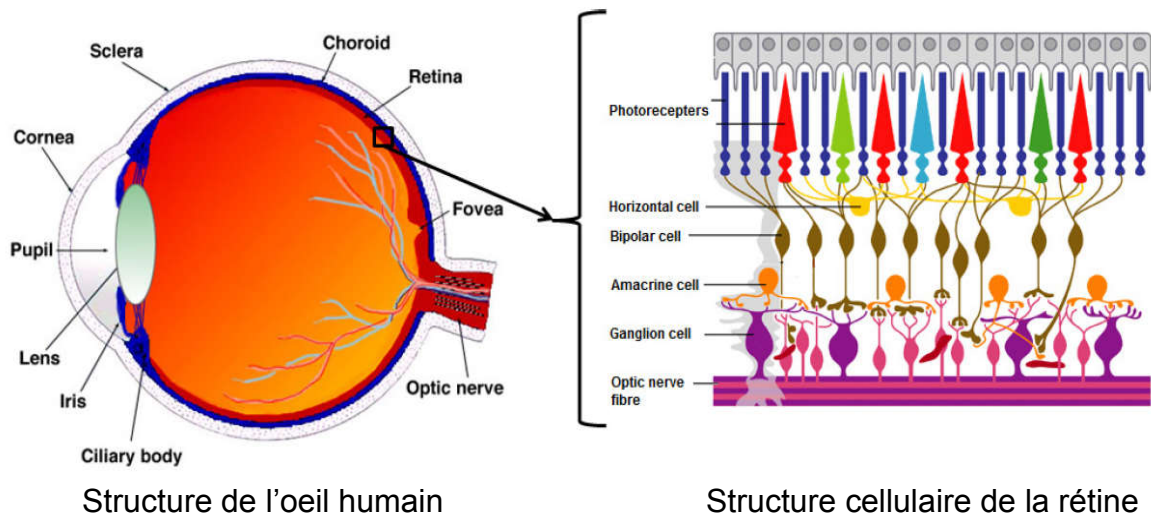


Figure 1.4 Structure de l'œil humain et organisation cellulaire de la rétine [35].

il y a seulement environ un million de nerfs optiques pour transmettre l'information visuel au cerveau. Raichle [22] estime que la rétine peut recevoir jusqu'à 10 milliards de bits d'informations par seconde, tandis que 6 millions de bits par seconde peuvent être transmis par la rétine au cerveau. Par conséquent, nous pouvons supposer qu'il existe certaines structures dans la rétine pour la compression de données qui utilisent les opérations de moyenne et de différence.

La répartition et les caractéristiques des cônes et bâtonnets ont des conséquences importantes sur le fonctionnement de notre vision. En condition diurne, L'image que nous percevons est de résolution variable: précise au centre et approximative en périphérie. Par contre, en condition de lumière faible, notre vision est principalement périphérique et achromatique. Des études expérimentales ont bien montré que les différentes interconnexions entre les cellules forment des champs récepteurs. La plupart de ceux-ci sont de type centre-périphérie (*center-surround*) [24]. Deux type de cette structure sont mises en œuvre dans la rétine: on-center and off-center. Comme montré sur le tableau 1.1, les cellules de type on-center ont une réponse maximale aux signaux provenant des régions centrales et insensibles aux signaux provenant des régions périphériques et vice versa pour les cellules off-center. De telles propriétés peuvent s'avérer efficace pour la détection d'irrégularités locales qui constitue un centre d'intérêt captivant et plus informatif que les régions avoisinantes. De cette manière, les signaux visuels reçus peuvent être

compressés considérablement pour pouvoir être traités par les ressources limitées du cerveau.

Tableau 1.1: Deux structures de type center-surround existantes dans la rétine

[24]

Structure logique	Stimulus & réponse			
	Pas de stimulus	Uniquement au centre	Uniquement à la périphérie	Dans les deux régions
<p>On-center</p>	Pas de réponse	Forte réponse	Pas de réponse	Réponse faible
<p>Off-center</p>	Pas de réponse	Pas de réponse	Forte réponse	Réponse faible

#### 1.4.3 Types d'attentions visuelles

Il existe deux types d'attentions visuelles: l'attention ouverte (*overt*) et l'attention couverte (*covert*) qui se différencie sur la manière dont l'œil réagit à des zones captivantes (objets) d'une scène visuelle.



L'attention couverte est la plus simple à observer. Cela suppose que lorsque nous portons notre attention sur un objet, nos yeux se déplacent afin de fixer cet objet. C'est à partir de cette constatation, somme toute triviale, que les premières ébauches de la notion d'attention ont été définies.

*[While the sense organs are occupied with one object, they cannot be simultaneously be moved by another so that an image of both arises. There cannot therefore be two images of two objects but one put together from the action of both.]* Hobbes (1612) [36].

L'attention ouverte correspond quant à elle à notre faculté à focaliser notre attention sur une cible (objet ou position) sans déplacer nos yeux. Puisqu'elle n'implique aucun mouvement oculaire, son observation est particulièrement difficile. Posner [37] a effectué le lien entre attention ouverte et couverte en mettant en place des expériences ingénieuses d'orientation spatiale signalée (*attention cueing*). Le principe de ces expériences est de mesurer le temps de réaction d'un sujet lors d'une tâche de recherche de cible. Avant l'affichage de la cible, on présente au sujet un indice (valide ou non) permettant au système d'attention couverte d'anticiper la position d'apparition de la cible. Ces expériences montrent que lorsque l'indice est valide, le temps de réaction du sujet est plus rapide que sans indice. A l'opposé, lorsque l'indice n'est pas valide, le temps de réaction du sujet est plus lent. Posner conclut que l'attention est un processus séquentiel: l'attention couverte est allouée avant l'attention ouverte afin de faciliter le déplacement des yeux vers des *stimuli intéressants*.

## 1.5 Détection de la saillance visuelle

### 1.5.1 Saillance orientée espace ou objet

Il est important de déterminer l'unité de base provoquant la saillance visuelle. Notre attention peut être orientée vers des emplacements spatiaux bien définis ou bien sur des objets saillants entiers. Posner [38], montre que l'attention peut être dirigée dans des zones plus ou moins grandes de la scène visuelle en se basant sur la prédiction des fixations du regard humain. Ces zones sont alors considérés saillantes dû à leur contraste élevé par rapport à leur environnement. D'autres études [39] suggèrent que l'allocation de l'attention pourrait être orientée vers des formes correspondant à des objets entiers, qui seraient le résultat d'un

regroupement (*grouping*) effectué pré-attentivement. La théorie de la forme suggère que l'attention pourrait jouer un rôle dans les phénomènes de groupement perceptuel (*perceptual grouping*) permettant à notre système visuel de résoudre le problème de la segmentation objet / fond. Elle soutient donc directement que l'attention visuelle est plutôt orientée objet.



Figure 1.5 Exemple illustrant la perception globale des formes. Le dalmatien est bien distingué bien qu'il ne soit constitué que d'un ensemble de tâches [33].

Élaborée par les psychologues gestaltistes [40], cette théorie suppose que les objets sont perçus comme un tout (une forme), et non comme la somme d'un ensemble de parties. La perception consiste alors en une "segmentation" fond / forme (objets). La structuration des formes amenant à leur perception est réalisée selon un ensemble de règles (cf. figure 1.6):

*Continuité*: les éléments d'une forme tendent à être perçus comme un tout lorsqu'ils sont dans le prolongement les uns des autres;

*Proximité*: les éléments proches les uns des autres ont tendance à être regroupés ensemble en une forme unique ;

*Similitude*: nous regroupons ensemble les éléments similaires;

*Destin commun*: des éléments se déplaçant avec la même trajectoire sont perçus comme faisant partie de la même forme;

*Fermeture*: notre esprit peut regrouper des éléments ensemble, s'ils contribuent à construire une forme fermée;

*Symétrie*: des éléments formant une forme symétrique sont perçus comme un tout, indépendamment de leur distance.

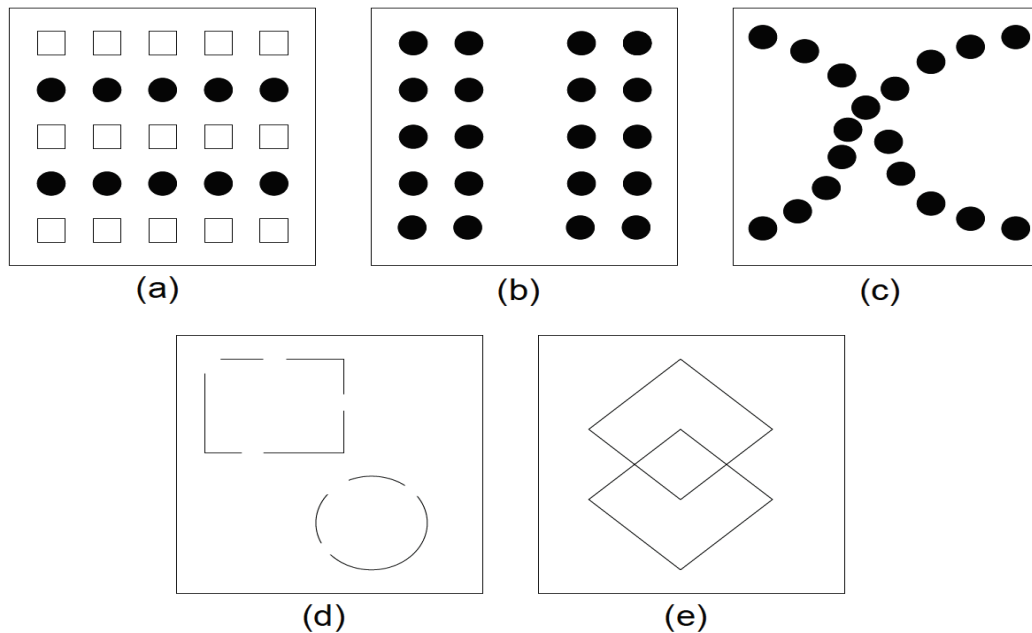


Figure 1.6 Illustration des différentes lois de Gestalt. (a) similarité: nous percevons des lignes de ronds noirs et carrés blancs; (b) proximité: nous percevons deux blocs verticaux; (c) continuité: la figure semble représenter deux courbes qui se croisent; (d) fermeture: les différents segments et arcs de cercle sont perçus comme un rectangle et un cercle; (e) symétrie: nous percevons deux losanges superposés plutôt que trois polygones [33].

Cependant cette théorie est principalement descriptive, et peine à mettre à jour les mécanismes permettant la mise en œuvre de ces règles. De plus, la mise en application des différentes règles est problématique, car celles-ci entrent régulièrement en conflit. Des études théoriques et expérimentales continuent cependant à être menées afin d'affiner ces règles, établir une priorité entre elles (pour gérer les conflits) et essayer de révéler les processus cognitifs sous-jacents

[41, 42]. Certains de ces travaux théoriques effectuent un lien entre attention et groupement perceptuel [43].

D'un point de vue computationnel, différents travaux ont tenté de mettre en application les lois de la Gestalt dans des applications variées: analyse de mouvement [44], traitement d'image [45], vision par ordinateur [46], ou encore analyse d'images [47]. D'un point de vue plus général, [48] propose une formulation neurogéométrique de certaines lois gestaltistes, calculable à très bas niveau. Enfin, Zou [49] utilise un système attentionnel afin de guider un algorithme de groupement perceptuel; le modèle ainsi obtenu se rapproche alors d'un modèle d'attention objet.

### 1.5.2 Détection d'objets saillants

Notre environnement regorge d'objets animés ou inanimés qui capture l'attention humaine. Ces objets sont dotés de caractéristiques captivantes comme la nouveauté, la rareté et la distinctivité par rapport à l'entourage. La figure 1.7 montre quelques exemples de tels objets. La problématique consiste donc à extraire de l'image les informations ou propriétés nécessaires pour distinguer l'objet du fond.



Figure 1.7 Quelques exemples d'objets d'intérêt

Il existe deux grandes familles de repères visuels exploitées pour la détection d'objets saillants: les indices intrinsèques et les indices extrinsèques. Les indices intrinsèques font référence à des caractéristiques comme la couleur, la texture et les contours qui sont extraits sur la même image à traiter fin d'estimer l'objet saillant et supprimer les distracteurs. Cependant, certaines approches de

détection d'objets saillants prétendent que certains objets possèdent des attributs visuels communs avec le fond ou les distracteurs. De ce fait, les indices intrinsèques s'avèrent insuffisantes pour distinguer l'objet du fond. Ils incorporent alors les indices extrinsèques qui sont extraites de sources d'informations externes à l'image. Nous pouvons citer comme exemple l'annotation de l'image par un utilisateur [50,51] et l'étude statistique sur des images similaires [52].

### 1.5.3 Critères d'évaluation de la détection d'objets saillants

La détection d'objets saillants est généralement interprétée en vision par ordinateur comme un processus qui comprend deux étapes: la localisation des objets les plus saillants et leur segmentation aussi précise que possible à partir de leurs contours. Dans la même voie que les travaux précurseurs d'Itti et al. [4] et Liu et al [5], beaucoup de modèles adoptent le concept de saillance pour effectuer ces deux étapes simultanément. En général, il est convenu que, pour une bonne détection de saillance, un modèle devrait satisfaire au moins les quatre critères suivants [50]:

1. Précision de détection: Le taux de faux positifs détecté ainsi que le taux de faux négatifs doivent être bas.
2. La carte de saillance à générer doit avoir une haute résolution pour localiser avec précision les objets saillants et conserver les informations d'image d'origine.
3. Complexité algorithmique réduite: Etant donné que la détection d'objets saillants peut servir comme étape préliminaire pour différentes applications, ces modèles doivent détecter les régions saillantes avec un temps d'exécution bas.
4. Contraste élevé entre le fond et les objets sur la carte de saillance.

### 1.5.4 Information de haut niveau pour la détection d'objets saillants

Souvent, l'information de bas niveau extraite à partir des caractéristiques comme la couleur ou la texture peut s'avérer insuffisante pour bien discerner l'objet saillant du fond de l'image. Beaucoup de travaux ce sont alors intéressés à utiliser des informations à priori définies comme règles ou suppositions dont le rôle est de guider l'estimation de la saillance vers les régions qui sont les plus

susceptible de contenir les objets d'intérêt. Nous pouvons distinguer trois grandes familles d'informations à priori.

1. Informations photométriques: se basent sur les propriétés de luminance de l'image. Par exemple, l'information à priori sur le contraste correspond à l'information qui est commune à toutes les méthodes d'estimation de la saillance. Il s'agit de considérer que les objets saillants ont un contraste considérable par rapport à l'arrière-plan dans le domaine spatial ou fréquentiel [1,2]. L'information à priori sur la couleur considère que certaines couleurs (par exemple les couleurs chaudes) sont plus attirantes (saillantes) que d'autres [79].

2. Informations spatiales: se basent sur la disposition spatiale pour inférer des informations à priori sur l'estimation de la saillance des différentes régions de l'image. Par exemple, les informations sur l'arrière-plan [9] et sur le centre de l'image [53] infèrent des suppositions sur l'emplacement spatial de l'arrière-plan et des objets saillants dans l'image respectivement. L'information de contiguïté [9,54] assume que les régions voisines partagent des valeurs de saillance similaires.

3. Informations géométriques: infèrent des suppositions géométriques sur les objets (resp. l'arrière-plan). Par exemple, l'information à priori sur la compacité [56] assume que l'arrière-plan possède une distribution de couleur ou de texture plus complexe que l'objet saillant. Ce dernier possède une distribution de couleur ou texture plus uniforme par rapport à l'arrière-plan. L'information sur la connectivité [57] assume que les parties qui composent un objet saillant sont fortement connectés entre elles, par contre elles sont faiblement connecté avec les parties de l'arrière-plan (et vice versa). L'information sur la concavité [55] suggère qu'une région du côté convexe d'un bord courbé est plus susceptible d'appartenir à un objet saillant.

Nous notons que récemment, des méthodes qui combinent plusieurs de ces informations ont été proposées. Nous pouvons prendre comme exemple la mesure de localisation d'objets [21] qui propose un certain nombre de fenêtres qui sont susceptibles de contenir des objets potentiels. La Prédiction des fixations du regard [4,58] est un autre exemple qui détermine les régions de l'image les plus attractives pour la vision humaine et qui est lui même considéré comme information à priori. Celui-ci est souvent utilisé par les modèles de détection

d'objets saillants qui estiment les régions saillantes à partir des points retournés par les fixations du regard.

### 1.6 Domaines étroitement liés à la détection d'objets saillants

Nous abordons dans cette section les domaines étroitement liés à la détection d'objets saillants qui sont la prédiction des fixations du regard, la segmentation d'image et la localisation d'objets. La figure 1.8 montre un exemple de traitement de chacun de ces domaines pour une même image. Nous allons étudier dans cette section la relation, les similitudes et les différences de ces domaines par rapport à la détection d'objets saillants et la manière dont ils peuvent se compléter.

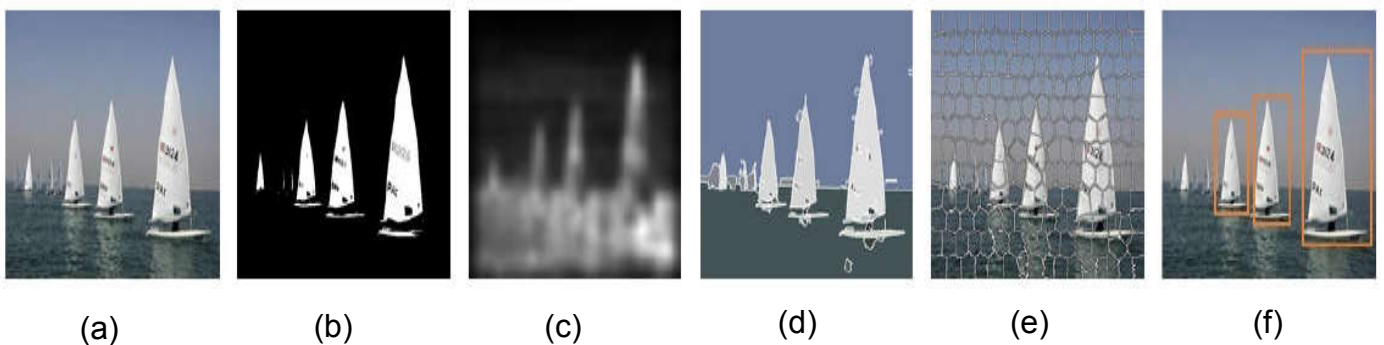


Figure 1.8 Résultats de traitement de différents modèles pour une même image. (a) image en couleur, (b) détection d'objets saillant [59], (c) prédiction des fixations du regard [4], (d) segmentation d'image (régions avec différentes surfaces) [60], (e) segmentation d'image en superpixels [61] et (f) mesure de localisation d'objets [21].

#### 1.6.1 Prédiction des fixations du regard

Le développement et l'émergence des modèles de détection d'objets saillants est particulièrement motivé par les applications basées sur l'estimation de la saillance comme la segmentation d'objet [62,63] ou le résumé de l'image et de la vidéo [64]. Les modèles de prédiction des fixations du regard sont par contre construits initialement pour étudier et comprendre le fonctionnement de l'attention visuelle humaine et la prédiction des mouvements de l'œil [4,65]. Les objectifs fixés par chaque type de modèle est très différent de ceux de l'autre. Les modèles de détection d'objets saillants ont comme objectif de localiser, détecter et

segmenter les objets les plus saillants en entier en respectant et préservant le plus possible leurs contours. Les modèles de prédiction des fixations du regard s'intéressent par contre uniquement à la prédiction des points où les gens portent leur attention pendant les 3 à 5 premières secondes durant lesquelles ils examinent visuellement une scène naturelle.

En théorie, Chaque type de modèle possède sa propre problématique et l'application d'un modèle à la résolution de la problématique de l'autre s'avère inadéquate. Par exemple, les modèles de détection d'objets saillants retournent des surfaces correspondantes à des objets entiers avec une préservation des contours. Cette détection génère alors un taux de faux positifs très élevé quand elle est évaluée avec les fixations du regard humain. D'une autre part, les modèles de fixations du regard omettent de détecter beaucoup de points à l'intérieur des objets saillants. Cette détection génère alors un taux de faux négatifs quand elle est évaluée avec des masks d'objet saillants. Dans la pratique, les deux types de modèles peuvent être complémentaires. Chaque type de modèle peut constituer une étape préliminaire pour l'autre. Comme mentionné dans la section 1.5.4, la prédiction des fixations du regard peut être considérée comme information a priori pour la détection des objets saillants. En outre, du moment que la détection d'objets saillants génère des surfaces larges qui respectent les contours et la forme de l'objet saillant et que le regard se porte généralement sur ces surfaces, la détection d'objets saillant peut être une étape préliminaire pour prédire les fixations du regard humain et réduire les zones de recherche de celles-ci.

### 1.6.2 Segmentation d'image

La segmentation d'image est l'un des domaines de recherche les plus étudiés en vision artificielle. Son objectif est de changer et simplifier la structure d'une image en une représentation plus significative et plus facile à analyser et étudier. Il s'agit de partitionner l'image en plusieurs segments différents. Chaque segment représente un ensemble de pixels ayant des caractéristiques similaires comme la couleur et la texture. De ce fait, les segments adjacents ont des caractéristiques sensiblement différentes. Dans la segmentation sémantique, une étiquette est associée à chaque région désignant l'objet ou l'arrière-plan auquel elle appartient. Par contre, les modèles de détection d'objets saillants s'intéressent



uniquement aux objets les plus saillants et traitent la tâche de segmentation comme un problème d'étiquetage binaire objets saillants/ arrière-plan. L'objectif est de déterminer si un pixel/région appartient à un objet saillant ou pas.

En pratique, il est possible de tirer profit des modèles de segmentation d'objets afin d'analyser chaque segment /région (qui correspond soit à l'objet soit à l'arrière plan), déterminer leur niveau de confiance et détecter ceux qui sont les plus saillants [50]. Cette approche n'a cependant pas été adoptée dans le passé à cause de deux raisons majeures:

1. Les algorithmes de segmentation d'objets de l'état de l'art ne sont pas assez performants pour bien définir correctement d'une manière précise les frontières des objets. Hors, une grande précision dans la segmentation est requise afin de bien estimer la saillance des objets. Autrement, Les erreurs générées durant la segmentation affecteront considérablement le processus d'estimation de la saillance.

2. La segmentation d'objets génère un temps d'exécution assez élevé alors que la détection d'objets saillants requiert une vitesse d'exécution raisonnable car elle est utilisée comme pré-processus préliminaire pour beaucoup de champs d'application.

Afin de remédier à ces deux problèmes, récemment, les modèles de détection d'objets saillants se sont intéressés à la segmentation en superpixels [62] comme représentation intermédiaire utile de l'image. Cette représentation permet de diviser l'image en des segments avec un niveau de granularité bien déterminé. Contrairement aux modèles de segmentation d'objets, les superpixels couvrent des surfaces homogènes plus petites dont la particularité est de bien préserver les contours. Un autre avantage majeur des superpixels est leur rapidité d'exécution tout à fait convenable pour la détection d'objets saillants.

### 1.6.3 Localisation d'objets

Les modèles basés sur la mesure de localisation d'objets ont comme objectif de générer des fenêtres indépendantes délimitant des positions spatiales susceptibles de contenir des objets génériques dans une image. L'ensemble de ces fenêtres doivent couvrir tous les objets de l'image sans s'intéresser à leurs

catégories [21,66]. La mesure de localisation d'objets possède trois avantages majeurs par rapport aux approches par fenêtres glissantes classiques [67,64]:

1. Se rapproche plus au comportement du système visuel humain qui perçoit d'abord rapidement les objets avant de les identifier [68].
2. Accélère considérablement le temps d'exécution en réduisant le nombre d'emplacements à générer de quelques millions à quelques centaines.
3. Les fenêtres retournées par la mesure de localisation d'objets sont plus précises et ajustées aux contours des objets.

La détection d'objets saillants et la localisation d'objets sont étroitement liées. D'une part, les modèles basés sur la mesure de localisation d'objets considèrent la saillance comme un indice important pour estimer l'appartenance d'une région à un objet [21]. Autrement dit, un objet dans une scène visuelle est plus susceptible d'être saillant qu'une autre région de l'arrière-plan. Ceci est basé sur la constatation que le fond de l'image est généralement plus structuré et homogène (donc moins saillant) que les objets [69]. D'une autre part, les modèles de détection d'objets saillants considèrent la localisation d'objets comme un indice important qui renseigne sur les positions d'objets auxquelles une plus grande valeur de saillance serait susceptible d'être assignée.

### 1.7 Conclusion

La détection d'objets saillants a fait objet de plusieurs travaux de recherche. Elle représente une étape cruciale et indispensable pour beaucoup de champs d'application comme la segmentation d'objets, résumés d'images et de vidéos, etc. La détection de la saillance s'est inspirée du fonctionnement du système visuel humain. Certains travaux sur la vision humaine ont montré que l'attention visuelle est orientée vers des espaces caractérisés par des irrégularités par rapport au voisinage. D'autres travaux ont montré que l'allocation de l'attention pourrait être orientée directement vers des formes correspondant à des objets entiers, d'où la motivation de la détection d'objets saillants. Celle-ci est souvent guidée par des informations de haut niveau. Il s'agit d'informations à priori qui peuvent être de type photométrique, spatiales ou géométrique. La détection d'objets saillants se trouve au confluent de plusieurs disciplines comme la prédiction des fixations du regard, la localisation d'objets et la segmentation

d'objets. Elle est considérée comme un domaine proche et complémentaire à chacune d'elles. Elle peut être exploitée soit comme étape préliminaire de base ou bien comme un post-traitement pour les autres domaines.

## **CHAPITRE 2**

### **ETAT DE L'ART SUR LA DETECTION D'OBJETS SAILLANTS**

#### 2.1 Introduction

Les êtres humains possèdent la faculté d'identifier rapidement des objets saillants dans des scènes visuelles complexes avec une précision remarquable. La saillance d'une région dans une image est généralement caractérisée par sa rareté ou bien par le contraste local de ses attributs comme la couleur, la texture et les contours par rapport à son entourage [70]. Les premiers modèles de saillance ont été développés principalement à base de la prédiction de fixations oculaires dans les images naturelles où le but est de comprendre l'attention visuelle humaine [8,3,7]. Récemment, plusieurs méthodes ont été proposées pour détecter les objets saillants qui se distinguent de leur environnement [70,50,62]. Les approches de détection de saillance définissent l'unicité (rareté) d'une région dans un contexte spatiale local ou global. L'irrégularité locale se mesure dans un contexte spatial réduit à une région et son entourage. Par contre, l'irrégularité globale se mesure dans un contexte spatial plus large englobant toutes les régions de la scène visuelle. Nous allons dans ce chapitre nous baser sur ces contraintes spatiales pour classer les méthodes de détection d'objets saillants en trois grandes classes majeures: les méthodes basées sur l'estimation du contraste local, les méthodes basées sur l'estimation du contraste global et les méthodes hybrides qui utilisent une estimation de la saillance basée sur le contraste global et le contraste local à la fois.

#### 2.2 Méthodes basées sur l'estimation du contraste local

Ces méthodes sont généralement liées à la prédiction des fixations du regard en explorant la rareté des régions d'images par rapport à leur voisinage. Rappelons que les cellules bipolaires et ganglionnaires de la rétine peuvent former deux structures logiques comprenant les parties on-center et off-center (cf.

chapitre 1, section 1.4.2). Ces deux structures peuvent être utilisées pour extraire les parties les plus informatives d'une scène visuelle. L'une des méthodes les plus influentes est celle de Laurent Itti [4]. Dans cette approche, la saillance visuelle pour un emplacement est quantifiée comme sa différence à partir des emplacements voisins en considérant de multiples caractéristiques et des échelles multiples. L'architecture de ce modèle est basée sur les étapes suivantes. L'image source est décomposée en différents canaux perceptuels (dans la version de base de l'algorithme: intensité, couleur et orientation). Puis, une représentation multi-échelles pyramidale de l'image est construite à partir de chaque canal des caractéristiques utilisées.

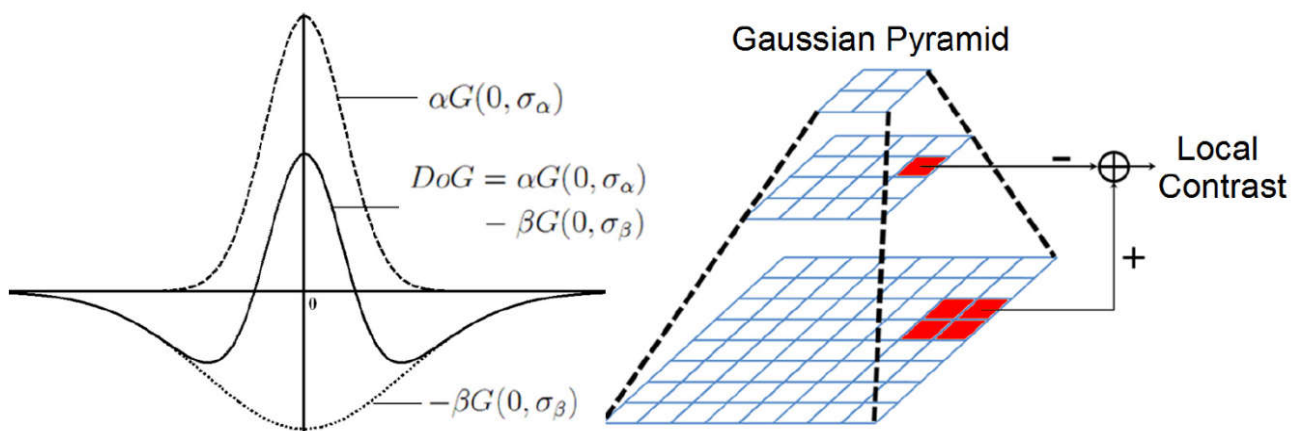


Figure 2.1 Structure pyramidale dont chaque niveau correspond à une échelle de représentation. La différence de gaussiennes est utilisée pour estimer le contraste centre/périphérie à travers la pyramide [4].

Une opération de filtrage centre-périphérie est alors effectuée par soustraction de certains niveaux des pyramides multi-échelles afin d'obtenir différentes cartes de caractéristiques (cf. figure 2.1). Celles-ci sont ensuite normalisées par un opérateur permettant de renforcer les cartes ne contenant qu'un petit nombre de pics d'activité (zones saillantes) puis sommées afin d'obtenir 3 cartes de singularité (intensité, couleur et orientation). Ces cartes sont également normalisées, puis sommées afin d'obtenir une carte de saillance.

Le contraste local peut donc bien simuler le fonctionnement neurobiologique de la vision humaine. Par conséquent, beaucoup de modèles de saillance ont adopté une architecture similaire. Par exemple, Hu et al. [71] utilisent la différence de texture entre une région et ses régions voisines comme critère

pour estimer le contraste local. Gao et al. [72] ont proposé une méthode mesurant la saillance comme le pouvoir discriminatif des caractéristiques utilisées d'une région par rapport aux régions voisines. Ils utilisent l'information mutuelle pour estimer le contraste local. Le modèle d'Olivier Le Meur [73] pousse l'inspiration biologique bien plus loin que le modèle original. Il modélise finement de nombreux phénomènes psycho-visuels: fonctions de sensibilité aux contrastes, décomposition en canaux perceptuels complexes, phénomène de masquage visuel, interaction centre/ périphérie, décomposition du flux perceptuel, etc. Cette approche montre qu'il est possible d'obtenir une modélisation très fine avec une approche plutôt macroscopique (le comportement individuel des neurones du cortex visuel n'étant pas du tout abordé). Kadir et al. [74] ont étudié la relation entre la saillance, l'échelle de présentation d'une région et ses caractéristiques. Ils proposent un modèle multi échelles basée sur l'entropie locale. Dans le même contexte, Les auteurs dans [75,5,76] utilisent la distance de  $\chi^2$  entre les histogrammes de couleur et de texture entre une région (ou bloc) et les régions (ou blocs) voisines pour l'estimation de la saillance locale.

### 2.3 Méthodes basées sur l'estimation du contraste global

Ces méthodes mesurent la rareté et l'unicité d'une région ou d'un bloc par rapport à l'ensemble de l'image. Plusieurs types d'approches ont été proposé pour mesurer le contraste global d'une région de l'image à traiter. La manière la plus triviale est d'utiliser la théorie des probabilités et statistiques en assignant une valeur de saillance élevée aux régions irrégulières. L'information a priori sur la compacité de couleur a été largement utilisé dans ce contexte. Cela suppose que plus une couleur possède une distribution de couleur élevée dans l'image, moins elle est susceptible d'être contenue dans l'objet saillant. Dans ce contexte, l'utilisation de la théorie des probabilités semble être un outil efficace. Outre les probabilités, la distance colorimétrique entre régions, la théorie des graphes et l'utilisation du domaine fréquentiel ont été largement utilisés dans ce contexte.

#### 2.3.1 Méthodes probabilistes

Dans [5,6], les pixels d'une image  $I$  sont modélisés par des mixtures de

gaussiennes (GMM)  $\{w_c, \mu_c, \Sigma_c\}_{c=1}^C$ , avec  $\{w_c, \mu_c, \Sigma_c\}$  représente le poids, la moyenne colorimétrique et la matrice de covariance de la composante  $c$  respectivement. Chaque pixel  $x$  est affecté à une composante  $c$  avec la probabilité:

$$P(c|I_x) = \frac{w_c \mathcal{N}(I_x | \mu_c, \Sigma_c)}{\sum_c w_c \mathcal{N}(I_x | \mu_c, \Sigma_c)} \quad (2.1)$$

La variance spatiale horizontale de la composante  $c$  est définie comme suit:

$$V_h(c) = \frac{1}{|P|_c} \sum_x P(c|I_x) \|x_h - M_h(c)\|^2 \quad (2.2)$$

$$M_h(c) = \frac{1}{|P|_c} \sum_x P(c|I_x) x_h \quad (2.3)$$

avec  $x_h$  représente les coordonnées dans l'axe des abscisses du pixel  $x$  et  $|P|_c = \sum_x P(c|I_x)$ . Nous notons que la variance spatiale verticale  $V_v(c)$  est définie d'une manière similaire. La distribution colorimétrique spatiale d'une composante  $c$  est alors définie comme suit:

$$DCS(c) = (1 - V(c)) \cdot (1 - D(c)) \quad (2.4)$$

avec  $V(c)$  la variance spatiale de la composante  $c$ , définie comme la somme des variances spatiales horizontales et verticales ( $V(c) = V_h(c) + V_v(c)$ ) et  $D(c)$  correspond à un paramètre de normalisation pondéré par la distance du pixel au centre de l'image  $d_x$  calculé comme suit:

$$D(c) = \sum_x P(c|I_x) \cdot d_x \quad (2.5)$$

La saillance finale d'un pixel  $x$  est alors calculée par rapport à toutes les composantes comme suit:

$$s(x) = \sum_c P(c|I_x).DCS(c) \quad (2.6)$$

Le formalisme bayésien est également utilisé pour calculer la saillance dans [77,78]. Il s'agit d'estimer la probabilité à postériori que chaque pixel  $x$  appartienne à l'objet saillant dans une image  $I$ . Soit  $s_x$  une variable binaire qui renseigne sur l'appartenance du pixel  $x$ . la valeur 1 est attribuée à l'objet saillant et la valeur 0 à l'arrière-plan. La probabilité à postériori dans [77] est calculée comme suit:

$$P(s_x = 1|I_x) = \frac{P(s_x = 1)P(I_x|s_x = 1)}{P(s_x = 1)P(I_x|s_x = 1) + P(s_x = 0)P(I_x|s_x = 0)} \quad (2.7)$$

avec  $P(s_x = 0|I_x) = 1 - P(s_x = 1|I_x)$ . Afin d'estimer la saillance à priori, une *enveloppe convexe*  $H$  est d'abord estimée autour des points d'intérêt de l'image. A partir d'une segmentation de l'image en superpixels, les auteurs regroupent les régions similaires en clusters. Supposons que le pixel  $x$  appartienne au cluster  $r$ , la saillance à priori  $P(s_x = 1)$  est alors définie comme suit:

$$P(s_x = 1) = \frac{|r \cap H|}{|r|}, x \in r \quad (2.8)$$

L'enveloppe convexe  $H$ , qui divise l'image  $I$  en deux régions: la région intérieure ( $R_I$ ) et la région extérieure ( $R_E$ ) retourne une estimation grossière de l'objet saillant ainsi que de l'arrière-plan et peut donc être employé dans le calcul des distributions colorimétriques dans les deux classes. Assumant que chaque canal de couleur soit indépendant, la distribution colorimétrique d'un pixel  $x$  dans l'espace  $Lab$  est calculée comme suit:

$$P(I_x|s_x = 1) = \prod_{v \in \{l,a,b\}} \frac{|\{x'|x' \in R_I, v(x') = v(x)\}|}{|\{x'|x' \in R_I\}|} \quad (2.9)$$

$$P(I_x|s_x = 0) = \prod_{v \in \{l,a,b\}} \frac{|\{x'|x' \in R_E, v(x') = v(x)\}|}{|\{x'|x' \in R_E\}|} \quad (2.10)$$



Les méthodes basées sur les probabilités sont efficaces dans les cas de la présence d'un seul objet saillant dominant dans l'image mais peuvent faillir sensiblement dans le cas de scènes complexes spécialement lorsque l'arrière-plan possède une distribution colorimétrique similaire à celle de l'objet saillant. La présence de plusieurs objets distracteurs dans l'image limite aussi l'efficacité de telles méthodes.

### 2.3.2 Méthodes graphiques

#### 2.3.2.1 Méthode graphique basée sur les marches aléatoires markoviennes

Comme réponse aux limites des méthodes probabilistes, Harel et al. [79] proposent un modèle basé sur une représentation graphique pour la détection de la saillance. L'image est représentée sous forme d'un graphe non orienté et pleinement connecté dont les nœuds correspondent à des blocs et les arcs sont des liens pondérés en fonction du degré de similarité et la distance spatiale entre les nœuds. Les marches aléatoires sont appliquées sur le graphe dont la transition entre un nœud et un autre est contrôlé par les liens du graphe. Durant le processus de marches aléatoires, les nœuds les moins visités sont considérés les plus saillants dû à leur unicité ou rareté dans un contexte global. Nous supposons que  $\{\mathcal{F}_k\}_{k=1}^K$  représente l'ensemble des caractéristiques utilisées. Pour la  $k^{\text{ième}}$  caractéristique, le degré de similarité entre deux nœuds positionnés dans les emplacements  $(i, j)$  et  $(p, q)$  respectivement est défini comme suit:

$$d_k((i, j) \parallel (p, q)) = \left| \log \frac{\mathcal{F}_k(i, j)}{\mathcal{F}_k(p, q)} \right| \quad (2.11)$$

Notons que dans certains cas, la similarité logarithmique peut être remplacée par une simple différence  $\mathcal{F}_k(i, j) - \mathcal{F}_k(p, q)$  qui est une opération plus simple avec des performances comparables. Les poids entre les nœuds sont par la suite définis par la mesure de similarité et la distance spatiale comme suit:

$$w_k((i, j), (p, q)) = d_k((i, j) \parallel (p, q)) \cdot \exp\left(-\frac{(i-p)^2 + (j-q)^2}{2\sigma^2}\right) \quad (2.12)$$

avec  $\sigma$  est un paramètre de régularisation ajusté approximativement à un dixième ou un cinquième de la largeur de l'image. Une chaîne de Markov est alors appliquée sur le graphe où les nœuds correspondent à des états et les arcs à des transitions. Un nœud irrégulier possède des valeurs de similarité basse par rapport aux autres nœuds, il est donc moins visité dans la chaîne de Markov. Ce nœud possède une valeur de saillance élevée. Nous notons que dans les cas d'images avec plusieurs objets saillants et plusieurs distracteurs, les parties qui correspondent aux régions saillantes peuvent être visitées aussi fréquemment que les distracteurs. La performance de l'estimation de la saillance est alors considérablement affectée.

### 2.3.2.2 Méthode graphique basée sur la distance géodésique

Dans [57], un modèle graphique basé sur la distance géodésique est proposée pour estimer la saillance. Les auteurs considèrent un template noté  $B$  défini à partir des contours de l'image comme information à priori sur l'arrière-plan. Ils effectuent ensuite une segmentation de l'image en régions sur laquelle ils construisent un graphe  $G = (\mathcal{V}, \mathcal{E})$  pleinement connecté unidirectionnel où l'ensemble des nœuds ( $\mathcal{V}$ ) correspond aux régions de l'image et au template  $B$  comme région virtuelle ( $\mathcal{V} = \{r_i\} \cup B$ ) et  $\mathcal{E}$  dénote l'ensemble des arcs du graphe. Il existe deux types d'arcs dans le graphe: les arcs internes qui relient les régions adjacentes et les arcs du contour de l'image qui relient les nœuds du contour à  $B$ .

$$\mathcal{E} = \{(r_i, r_j) | r_i \text{ est adjacent à } r_j\} \cup \{(r_i, B) | r_i \text{ est sur le contour de l'image}\} \quad (2.16)$$

Un arc du graphe est pondéré en fonction de la distance géodésique entre les deux nœuds qu'il relie ( $r_i$  et  $r_k$ ). Il s'agit de la somme des poids sur le plus court chemin reliant  $r_i$  à  $r_k$  défini comme suit:

$$d_{geo}(r_i, r_k) = \min_{r_{i1}=r_i \dots r_{in}=r_k} \sum_{j=1}^{n-1} D_r(r_{ij}, r_{i j+1}) ; (r_{ij}, r_{i j+1}) \in \mathcal{E} \quad (2.17)$$

Le score de saillance de  $r_i$  est considéré comme sa distance géodésique par rapport à  $B$  sur le graphe  $G$  comme suit:

$$s(r_i) = d_{geo}(r_i, B) \quad (2.18)$$

### 2.3.3 Méthodes basées sur la distance colorimétrique

D'autres travaux [62,57] ont eu recours à la distance colorimétrique pour estimer le contraste d'une région par rapport à toutes les autres régions de l'image. Dans ce contexte, M.M. Cheng et al. [62] proposent un modèle basé sur la distance spatiale et la distance entre histogrammes de couleur des différentes régions de l'image pour estimer la saillance. D'abord, l'image est segmentée en régions. La saillance pour une région donnée  $r_k$  est calculée par rapport à toutes les autres régions de l'image comme suit:

$$S(r_k) = \sum_{r_k \neq r_i} w(r_i) D_r(r_k, r_i) \quad (2.13)$$

avec  $D_r(.,.)$  est la distance colorimétrique et  $w(r_i)$  correspond au nombre de pixels dans  $r_i$  qui favorise les grandes régions à avoir une saillance importante. La distance colorimétrique entre deux régions  $r_1$  et  $r_2$  est calculée comme suit:

$$D_r(r_1, r_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f(c_{1,i}) f(c_{2,j}) D(c_{1,i}, c_{2,j}) \quad (2.14)$$

avec  $f(c_{k,i})$  la probabilité de la  $i^{\text{ème}}$  couleur  $c_{k,i}$  parmi toutes les  $n_k$  couleurs de la  $k^{\text{ème}}$  région  $r_k$ ,  $k = \{1,2\}$ . Cette probabilité est obtenue à partir de l'histogramme normalisé de  $r_k$ .  $D(.,.)$  est la distance de couleur dans l'espace *CIE Lab*. Afin de favoriser la similarité entre les régions proches, les auteurs incorporent dans le processus d'estimation de la saillance la distance spatiale entre comme suit:

$$S(r_k) = \sum_{r_k \neq r_i} \exp(-D_s(r_k, r_i)/\sigma_s^2) w(r_i) D_r(r_k, r_i) \quad (2.15)$$

avec  $D_s(r_k, r_i)$  est la distance spatiale euclidienne entre les centroides des régions  $r_k$  et  $r_i$  et  $\sigma_s$  un paramètre qui contrôle la contribution de la pondération spatiale. Plus la valeur de  $\sigma_s$  est importante, plus l'effet de la pondération spatiale est réduit

et plus le contraste avec les régions éloignées contribuent dans l'estimation de la saillance.

#### 2.3.4 Méthodes d'estimation de la saillance dans le domaine fréquentiel

Certaines autres méthodes proposent d'estimer la saillance globale dans le domaine fréquentiel. Hou et Zhang [7] considèrent que les similarités entre régions impliquent des redondances dans l'image et développent un modèle de saillance basé sur le calcul du résidu spectral. Ils supposent que les singularités statistiques dans le spectre peuvent être responsables de régions anormales dans l'image, où les objets deviennent visibles. D'abord, ils dérivent l'amplitude  $\mathcal{A}(f)$  et la phase  $\mathcal{P}(f)$  à partir d'une image  $I(x)$  en appliquant la transformée de Fourier ( $\mathcal{F}$ ). Le log-spectrum  $\mathcal{L}(f)$  est ensuite calculé à partir de l'amplitude.

$$\mathcal{A}(f) = \mathcal{R}(\mathcal{F}[I(x)]), \quad \mathcal{P}(f) = \varphi(\mathcal{F}[I(x)]), \quad \mathcal{L}(f) = \log(\mathcal{A}(f)) \quad (2.19)$$

A partir de  $\mathcal{L}(f)$ , le résidu spectral ( $\mathcal{R}(f)$ ) peut être obtenu en multipliant  $\mathcal{L}(f)$  avec un filtre  $n \times n$  local ( $h_n(f)$ ) puis en soustrayant le résultat de  $\mathcal{L}(f)$  comme suit:

$$\mathcal{R}(f) = \mathcal{L}(f) - h_n(f) * \mathcal{L}(f) \quad (2.20)$$

En utilisant la transformée de Fourier inverse, ils construisent la carte de saillance  $S$  dans le domaine spatial comme suit:

$$S(x) = g(x) * \mathcal{F}^{-1}[\exp(\mathcal{R}(f) + \mathcal{P}(f))]^2 \quad (2.21)$$

avec  $g(x)$  un filtre gaussien qui assure un meilleur effet visuel. Guo et al. [80] ont démontré que l'incorporation de la phase de la transformée de Fourier au lieu de l'amplitude améliore les résultats sur la carte de saillance.

Achanta et al. [8] ont implémenté une approche dans le domaine fréquentiel pour la détection de régions saillantes en utilisant les caractéristiques de couleur et de luminance dans l'espace de couleur CIE Lab. La saillance en un point  $x$  est estimée comme la différence (distance euclidienne) entre le vecteur représentant

la moyenne des caractéristiques de l'image  $I_\mu$  et le vecteur des caractéristiques de  $x$  après lissage de l'image par un filtre gaussien comme suit:

$$s(x) = \|I_\mu - I_w(x)\|^2 \quad (2.22)$$

où  $I_w$  représente le résultat du lissage de l'image  $I$ .

Au sein des méthodes basées sur l'estimation du contraste locale, l'unicité d'une région/pixel par rapport à toute l'image n'est pas prise en compte et sa saillance n'est pas détectée si celle-ci n'est pas contrastée par rapport à son entourage. En outre, comme l'échelle de l'objet dans l'image est inconnue à priori, il est particulièrement difficile de déterminer la taille des régions centrales et les régions voisines à comparer. Dans ce contexte, souvent, les contours qui sont les parties les plus contrastées par rapport à leur entourage sont retournés. Ces méthodes connaissent donc un taux de faux négatifs assez élevé et ne sont pas adaptées à une détection efficace et entière de tout l'objet saillant dans sa globalité (cf. figure 2.2 (c)).

Les méthodes basées sur le contraste global sont par contre plus orientées pour la séparation de l'objet saillant entier du reste de l'image. Cependant, ce type de méthode affecte souvent des valeurs de saillance élevées pour certaines régions de l'arrière-plan, spécialement dans le cas d'images avec un fond complexe. En effet, comme ces méthodes ne sont pas adaptées pour l'encodage de la contiguïté spatiale, certaines régions de l'arrière plan ne sont pas efficacement écartées. Un exemple est montré sur la figure 2.2 (d)-première rangée, où certaines régions de l'arrière-plan sont retournées comme saillantes. Ces méthodes sont également particulièrement limitées dans le cas de la similarité entre les distributions colorimétriques de l'objet saillant et de l'arrière-plan (cf. figure 2.2 (d)- deuxième et troisième rangées). En outre, comme ces méthodes se basent sur les statistiques globales de l'image, elles sont dépendantes de la surface occupée par l'objet saillant. Lorsque l'objet saillant occupe un espace majoritaire dans l'image, l'estimation de la saillance retourne un taux important de faux négatifs et l'objet saillant dans sa globalité est mal détecté (cf. figure 2.2 (d)-quatrième figure).

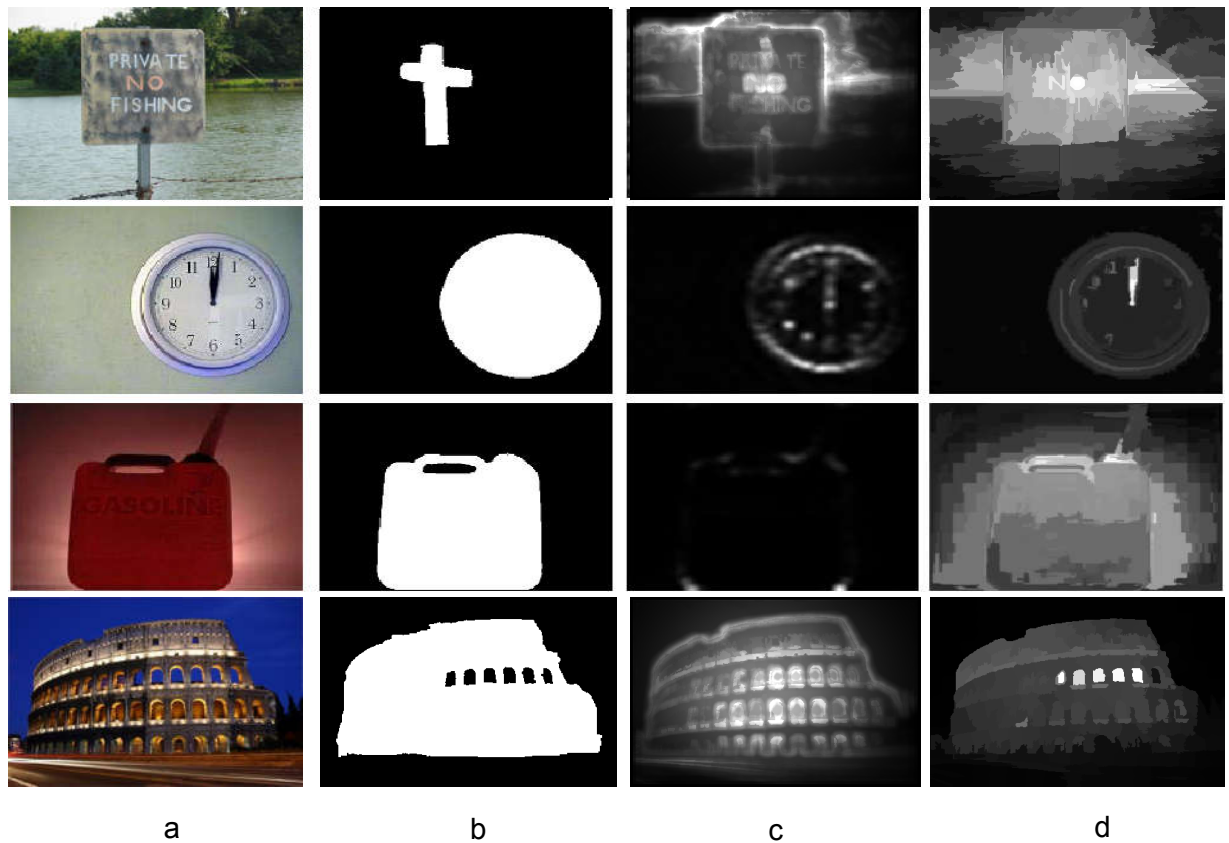


Figure 2.2 Limite des méthodes basées sur l'estimation du contraste local et celles basées sur l'estimation du contraste global. (a) Image en couleur, (b) vérité terrain, (c) carte de saillance retournée par une méthode basée sur l'estimation du contraste local [71] et (d) carte de saillance retournée par une méthode basée sur l'estimation du contraste global [8].

## 2.4 Méthodes hybrides

Comme solution aux inconvénients des méthodes basées sur l'estimation du contraste local et les méthodes basées sur l'estimation du contraste global, plusieurs méthodes proposent de combiner les informations issues des deux types de contraste pour la détection d'objets saillants. Parmi ces méthodes, nous pouvons distinguer trois types d'approches: méthodes probabilistes, méthodes basées sur les graphes et les autres méthodes.

### 2.4.1 Méthodes probabilistes

La saillance d'un objet pouvant être définie comme sa singularité par rapport aux autres, il est assez naturel d'utiliser une théorie probabiliste ou

statistique afin de relier la saillance aux éléments les moins probables d'une scène ou d'une partie de la scène. Dans ce cadre, différentes approches sont possibles.

#### 2.4.1.1 Méthode d'estimation de la saillance au niveau pixel: Analyse de la transformée en ondelettes

Dans [81], l'analyse des ondelettes est utilisée sur l'image à traiter avec différentes échelles afin de générer  $N$  cartes de caractéristiques pour chaque canal de l'espace de couleur *CIE LAB*. Chaque carte  $f_s^l$  correspond à un niveau de décomposition en ondelettes  $s$ , ( $s \in \{1, 2, \dots, N\}$ ) sur l'image dans le canal  $l$  ( $l \in \{L, a, b\}$ ). Un pixel à la position spatiale  $(x, y)$  peut alors être représenté par le vecteur de caractéristiques  $f(x, y)$  avec une longueur de  $N \times 3$  (Il existe trois canaux dans CIE LAB). La carte de saillance globale est générée au niveau pixel en utilisant une fonction de densité de probabilité avec une distribution gaussienne notée  $p(f(x, y))$  avec  $(x, y)$  la disposition spatiale d'un pixel comme suit:

$$S_G(x, y) = \left( \log \left( p(f(x, y))^{-1} \right) \right)^{\frac{1}{2}} \times l_{k \times k} \quad (2.23)$$

avec  $l_{k \times k}$  un filtre passe-bas gaussien  $k \times k$  et  $p(f(x, y))$  est calculé comme suit:

$$P(f(x, y)) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \times \exp \left( -\frac{1}{2} (f(x, y) - \mu)^T \Sigma^{-1} (f(x, y) - \mu) \right) \quad (2.24)$$

avec  $\mu$  le vecteur contenant la valeur moyenne de chaque carte de caractéristique et  $\Sigma$  la matrice de covariance  $n \times n$ , ( $n = N \times 3$ ). Les auteurs génèrent également une carte de saillance locale construite en prenant les valeurs maximales entre les différentes cartes  $f_s^l$  comme suit:

$$S_L(x, y) = \left( \sum_{s=1}^N \arg \max \left( f_s^L(x, y), f_s^a(x, y), f_s^b(x, y) \right) \right) \times l_{k \times k} \quad (2.25)$$

Une fois les deux cartes générées, un processus de fusion leur est appliqué comme suit:

$$S(x, y) = \frac{(S_L(x, y) \times \exp(S_G(x, y)))^{\ln(\sqrt{2})}}{\sqrt{2}} \quad (2.26)$$

Afin d'améliorer le contraste entre l'objet et l'arrière-plan, les auteurs réévaluent l'estimation de la saillance autour des points les plus saillants de l'image. Ils considèrent que les points les plus saillants correspondent au focus d'attention de l'œil humain (notés *foc*). Les régions avoisinantes doivent alors avoir une valeur de saillance plus importante que les régions éloignées. La saillance  $S$  est alors mise à jour comme suit:

$$S'(x, y) = S(x, y) \cdot (1 - d_{foc}(x, y)) \quad (2.27)$$

avec  $d_{foc}(x, y)$  est la distance euclidienne entre le pixel à la position  $(x, y)$  vers le point *foc* le plus proche.

#### 2.4.1.2 Estimation de la saillance au niveau bloc: Apprentissage de codes caractéristiques

Borji et Itti [70] ont proposé une approche pour estimer la saillance en utilisant des coefficients de projection pour calculer la différence du contraste local et la rareté globale de chaque patche (bloc) de l'image. En particulier, les espaces de couleurs *RVB* et *Lab* ont été utilisés simultanément pour apprendre les codes caractéristiques ainsi que pour le calcul de saillance visuelle. Les auteurs utilisent 500 000 patches de taille 8x8 extraits à partir de 1500 images en couleur pour alimenter leur dictionnaire. Ils construisent pour chaque patche sa fonction de base (représentation éparsée) en minimisant la fonction suivante:

$$\alpha^*(x, D) = \arg \min_{\alpha \in \mathbb{R}^n} \frac{1}{2} \|x - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (2.28)$$

avec  $\lambda$  un paramètre de régularisation et  $D = [d_1, d_2, \dots, d_n] \in \mathbb{R}^{m \times n}$  le dictionnaire des données. Nous notons que le processus d'estimation de la saillance est effectué pour chaque canal des deux espaces de couleurs considérés (RGB et



LAB). La valeur de  $m$  est alors ajustée à 1. La figure 2.3 montre un exemple de 200 coefficients construits à partir du dictionnaire.

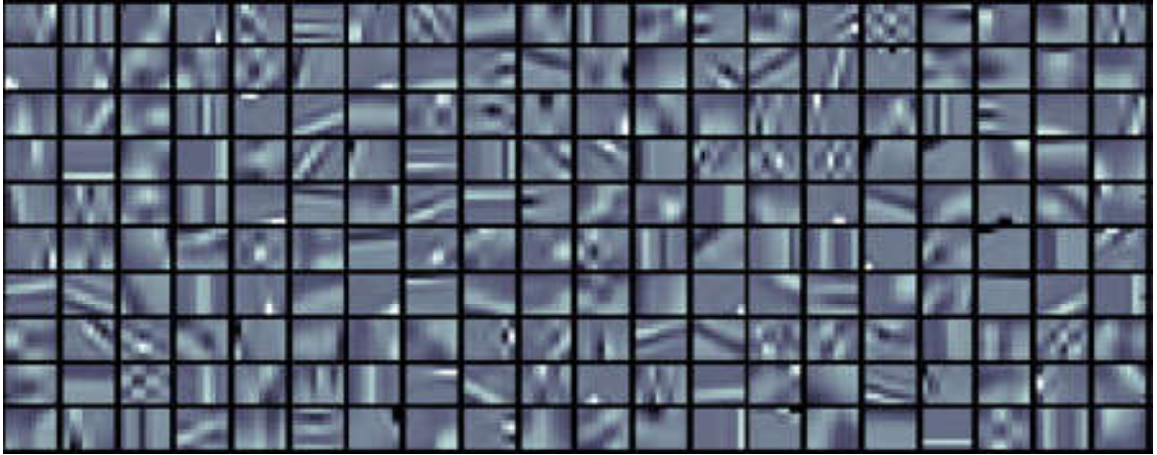


Figure 2.3 Représentation de 200 fonctions de base à partir de patches de taille 8x8 sur le canal  $L$  de l'espace de couleur  $LAB$  [70].

Soient  $P = [P_1, P_2, \dots, P_n]$  l'ensemble des  $n$  patches de l'image à traiter. La saillance locale d'un patche est calculée comme la moyenne pondérée de ses distances avec ses  $L$  patches voisins selon l'équation suivante:

$$S_i^c(P_i) = \frac{1}{L} \sum_{j=1}^L W_{ij}^{-1} D_{ij}^c \quad (2.29)$$

avec  $W_{ij}$  la distance euclidienne entre le centre de  $P_i$  et le centre de  $P_j$ . Les centres des patches les plus proches auront alors une plus grande influence sur la saillance locale.  $D_{ij}$  représente la distance euclidienne entre  $P_i$  et  $P_j$  dans l'espace des caractéristiques dénotée par les fonctions de base  $\alpha_i$  et  $\alpha_j$  et  $c \in \{l, a, b, r, g, b\}$ . Le contraste global d'un patche  $P_i$  est estimé en calculant l'inverse de sa probabilité par rapport à toute l'image comme suit:

$$S_g^c = P(p_i)^{-1} = \left( \prod_{j=1}^n P(\alpha_{ij}) \right)^{-1} \quad (2.30)$$

avec  $\alpha_{ij}$  représente les coefficients de la fonction de base relative au patche  $p_i$ . La probabilité  $P(\alpha_{ij})$  du coefficient  $\alpha_{ij}$  est obtenue en estimant sa densité de

probabilité (*pdf*). Pour chaque coefficient  $\alpha_{ij}$ , l'histogramme de tous les patches de l'image est construit dans l'espace des caractéristiques avec 100 bins. La densité du coefficient  $\alpha_{ij}$  est alors estimée par la division de la valeur de  $\alpha_{ij}$  retournée par l'histogramme sur la somme de toutes les valeurs de l'histogramme. Les cartes de saillance locales sont alors fusionnées après normalisation par l'opérateur *max* afin de générer une carte de saillance locale finale. La même stratégie de fusion est appliquée aux cartes de saillances globales. La carte de saillance finale est obtenue par la somme de la carte de saillance locale normalisée et la carte de saillance globale normalisée.

L'inconvénient majeur des méthodes qui opèrent au niveau pixels ou bloc est qu'elles ne sont pas stables dans le cas d'images complexes. Elles ne préservent généralement pas les contours de l'objet saillant dans certains cas ou bien ne retournent pas toujours des régions pleinement saillantes. Pour palier à ce problème, certaines méthodes récentes se sont intéressées à la segmentation en superpixels et définissent chaque région comme unité de base pour l'estimation de la saillance. Les superpixels ont comme particularités la complexité temporelle d'exécution basse et l'ajustement de la segmentation sur les contours.

#### 2.4.1.3 Estimation de la saillance au niveau superpixel: Approche basée sur les multi résolutions et les distributions normales

Dans [11], les auteurs définissent plusieurs résolutions pour l'image à traiter en appliquant une segmentation en superpixels multi-échelles. Les auteurs utilisent la distance de Wasserstein pour mesurer ensuite le contraste local et global d'un superpixel. Ils supposent que chaque superpixel  $S_i^t$  (où  $i$  l'indice du superpixel et  $t$  l'indice de la résolution considérée) est caractérisé par sa distribution colorimétrique normale  $\mathcal{N}_S(\mu, \Sigma)$  et les coordonnées spatiales de son centroïde  $\vec{c}_S(x, y)$  ( $S_i^t = \{\mathcal{N}_S(\mu, \Sigma), \vec{c}_S(x, y)\}_i^t$ ). Les auteurs utilisent l'information a priori sur la compacité pour estimer le contraste globale. Ils regroupent les superpixels selon la similarité entre leurs distributions normales calculées grâce à la distance de Wasserstein en clusters. Chaque cluster  $C_k^t$  où  $k$  correspond à l'indice du cluster et  $t$  à la résolution considérée est caractérisé par sa distribution normal  $\mathcal{N}_c$ , la position spatiale de son centroïde  $\vec{c}_{C_k^t}$  et son superpixel le plus représentatif ayant la distribution normale qui lui est la plus similaire. Les

coordonnées spatiales du centroïde d'un cluster  $C_k^t$  sont obtenues par la combinaison linéaire des coordonnées des centroïdes de tous les superpixels pondérées par la probabilité de leur appartenance à  $C_k^t$  comme le montre l'équation suivante:

$$\overrightarrow{c_{C_k}^t} = \frac{\sum_{i=1}^{M(t)} P(C_k^t | S_i^t) \cdot \overrightarrow{c_{S_i}^t}}{\sum_{i=1}^{M(t)} P(C_k^t | S_i^t)} \quad (2.31)$$

avec  $M(t)$  dénote le nombre de superpixels dans la résolution  $t$  et  $P(C_k^t | S_i^t)$  la probabilité qu'un superpixel  $S_i^t$  appartienne au cluster  $C_k^t$ . Cette probabilité est calculée grâce à la distance de Wasserstein entre les distributions normales du superpixel le plus représentatif de  $C_k^t$  et  $S_i^t$ . La saillance globale d'un cluster  $C_k^t$  est calculée comme suit:

$$P_g(sal | C_k^t) = 1 / \sum_{j=1}^{K(t)} \frac{\sum_{i=1}^{M(t)} P(C_k^t | S_i^t) \cdot \|\overrightarrow{c_{S_i}^t} - \overrightarrow{c_{C_j}^t}\|^2}{\sum_{i=1}^{M(t)} P(C_k^t | S_i^t)} \quad (2.32)$$

avec  $Sal = \{sal, \overline{sal}\}$  est une variable binaire indiquant si une région (cluster ou superpixel) appartient à l'objet saillant ou non et  $K(t)$  l'ensemble des clusters dans la résolution  $t$ . La saillance global au niveau superpixel est calculée ensuite comme la somme des produits de la saillance de chaque cluster et de la probabilité d'appartenance du superpixel à celui-ci. Elle est calculée comme suit:

$$P_g(sal | S_i^t) = \sum_{k=1}^{K(t)} P(sal | C_k^t) \cdot P(C_k^t | S_i^t) \quad (2.33)$$

D'une autre part, les auteurs calculent le contraste local d'un superpixel en fonction de la distance entre distributions et de la distance spatiale comme suit:

$$P_l(S_i^t) = \frac{\sum_{j=1}^{M(t)} W(\mathcal{N}_{S_i^t}^t, \mathcal{N}_{S_j^t}^t) \cdot g(i, j, t)}{\sum_{j=1}^{M(t)} g(i, j, t)} \quad (2.34)$$

avec  $W$  la distance de Wasserstein et  $g(i, j, t)$  la distance spatiale normalisée entre superpixels adjacents. Cette valeur est mise à 0 pour les superpixels non adjacents. Le niveau d'adjacence est défini au préalable. Les auteurs définissent ensuite un post processus de raffinement de la saillance locale basée sur les marches aléatoires sur les graphes en incluant un template défini sur les bords de l'image comme information à priori sur l'arrière-plan. Les auteurs combinent par la suite la saillance locale et la saillance globale par simple multiplication normalisée par les valeurs de saillance de tous les superpixels de l'image. La carte de saillance finale est obtenue en effectuant la moyenne des cartes de saillances obtenues aux niveaux des différentes résolutions.

### 2.4.2 Méthodes basées sur les graphes

Récemment, les méthodes basées sur les graphes ont émergé comme un outil efficace pour la détection d'objets saillants. Ces méthodes permettent de représenter l'image sous forme d'un graphe dont les nœuds sont des régions de l'image et les arcs des liens pondérés reliant les différents nœuds. En plus de leur simplicité, ces méthodes permettent d'incorporer facilement les informations à priori sur la position spatiale comme la contiguïté des parties de l'objet saillant ainsi que du fond de l'image et l'information à priori sur l'arrière-plan.

#### 2.4.2.1 Méthodes basées sur une seule résolution

Ces méthodes se basent généralement sur un graphe mono-couche construit sur la base d'une segmentation de l'image en régions (généralement des superpixels) qui sont considérés comme les nœuds du graphe.

##### 2.4.2.1.1 Marches aléatoires markoviennes

Dans [54], les auteurs proposent un modèle basé sur les marches aléatoires markoviennes pour détecter les objets saillants construit sur un graphe mono-couche. Ils segmentent d'abord l'image en superpixels, puis construisent un graphe  $G(V, E)$  où les nœuds  $V$  sont représentés par les superpixels et les arcs  $E$  des liens pondérés reliant les nœuds. Comme les objets saillants occupent rarement les bords de l'image, les auteurs considèrent les nœuds qui touchent les bords comme absorbants (arrière-plan virtuel). Notons que les arcs relient uniquement chaque nœud à ses voisins transitoires. Les nœuds absorbants ne

sont donc pas interconnectés. Le poids reliant deux nœud  $i$  et  $j$  est défini comme suit:

$$w_{i,j} = e^{-\frac{\|x_i - x_j\|}{\sigma^2}}; \quad i, j \in V \quad (2.35)$$

avec  $x_i$  et  $x_j$  les moyennes des nœuds  $i$  et  $j$  respectivement dans l'espace de couleur CIE Lab et  $\sigma$  une constante qui contrôle la stabilité du poids. Les nœuds sont ensuite réorganisés de façon que les  $t$  premiers nœuds soient transitoires et les  $r$  derniers soient absorbants. Ils définissent alors la matrice d'affinité  $A$  comme suit:

$$A: \begin{cases} w_{ij} & j \in N(i), \quad 1 \leq i \leq t \\ 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases} \quad (2.36)$$

avec  $N(i)$  l'ensemble des nœuds connectés à  $i$ . La matrice de transition  $P$  est alors calculée comme suit:

$$P = D^{-1} \times A \quad (2.37)$$

avec  $D$  la matrice des degrés qui représente la somme des poids connectés à chaque nœud  $D = \text{diag}(\sum_j a_{ij}), j \in t + r$ . Comme le graphe ne relie un nœud qu'à ses voisins, la matrice  $P$  est donc éparsée. Cela restreint les marches aléatoires à un seul mouvement pour chaque étape dans le graphe. La matrice  $P$  possède la forme canonique suivante  $P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$  avec  $Q \in [0,1]^{t \times t}$  la matrice contenant les probabilités de transition entre chaque paire de nœuds transitoires et  $R \in [0,1]^{t \times r}$  la matrice contenant les probabilités de transition entre chaque état transitoire et chaque état absorbant. La matrice  $0$  représente une matrice de  $0$  de taille  $r \times t$  et  $I$  la matrice identité de taille  $r \times r$ . La matrice fondamentale  $N$  est définie comme suit:  $N = (I_{t \times t} - Q)^{-1}$  où l'élément  $n_{ij}$  présente le nombre de transitions sur le nœud  $j$  sachant que la chaîne a commencé à partir de  $i$ . La somme  $\sum_j n_{ij}$  reflète le

nombre de transitions nécessaire avant l'absorption sachant que la chaîne commence à partir du nœud  $i$ . La normalisation de ce nombre dans l'intervalle  $[0,1]$  correspond au niveau de saillance du nœud  $i$ . Autrement dit, plus il y a de transitions entre  $i$  et des nœuds transitoires avant d'atteindre un nœud absorbant, plus il est différent de l'arrière plan et plus il est saillant. La valeur de saillance de tous les nœuds peut donc être estimée comme suit:

$$\bar{y} = N \times c \quad (2.38)$$

où  $c$  est un vecteur de taille  $t$  dont les éléments sont tous égaux à 1 et  $\bar{y}$  une valeur normalisée dans  $[0,1]$ . Etant donné que le nombre de transitions à partir du centre de l'image jusqu'à ses bords peut être élevé, les parties de qui se situent au centre de l'image sont plus susceptibles d'être considérées comme saillantes. D'une part, cette affirmation n'est pas toujours correcte et cela peut sensiblement générer un bruit sur l'image surtout dans le cas de parties d'arrière-plan à surfaces larges et à caractère homogène. D'une autre part, une mise à jour de l'estimation de la saillance de telle manière à réduire les valeurs de saillance au centre de l'image peut affecter le taux de vrais positifs et considérer ainsi des parties saillantes comme arrière-plan. Les auteurs se basent donc sur l'observation qu'une bonne carte de saillance est caractérisée par un contraste élevé entre l'objet saillant et l'arrière-plan (cf. figure 2.4). Pour remédier à cela, les auteurs calculent un score pour estimer le degré de ce contraste comme suit:

$$score = \sum_{b=1}^{10} g(b) \times \min(b, (11 - b)) \quad (2.39)$$

avec  $g$  l'histogramme du niveau de gris de la carte de saillance et  $b \in \{1,2, \dots, 11\}$  représente un bin de l'histogramme. Si la valeur du score est supérieure à un certain seuil  $\gamma$ , alors une mise à jour de l'estimation de la saillance est requise, autrement aucune mise à jour n'est nécessaire. La mise à jour s'opère en considérant une chaîne de Markov ergodique. Soit  $\pi$  la distribution d'équilibre pour la chaîne de Markov ergodique considérée. Celle-ci est calculée à partir de la matrice d'affinité  $A$  comme suit:

$$\pi_i = \frac{\sum_j a_{ij}}{\sum_{ij} a_{ij}} \quad (2.40)$$

avec  $i$  et  $j$  les indexes des nœuds transitoires. La valeur de la somme  $\sum_j a_{ij}$  est particulièrement élevée dans les zones homogènes de l'image. Le temps récurrent dans ce type de régions est donc faible. Pour cette raison, les auteurs utilisent le temps récurrent moyen  $h_j = 1/\pi_j$  pour chaque nœud  $j$  afin de pondérer chaque élément  $n_{ij}$  correspondant dans la matrice fondamentale  $N$ .

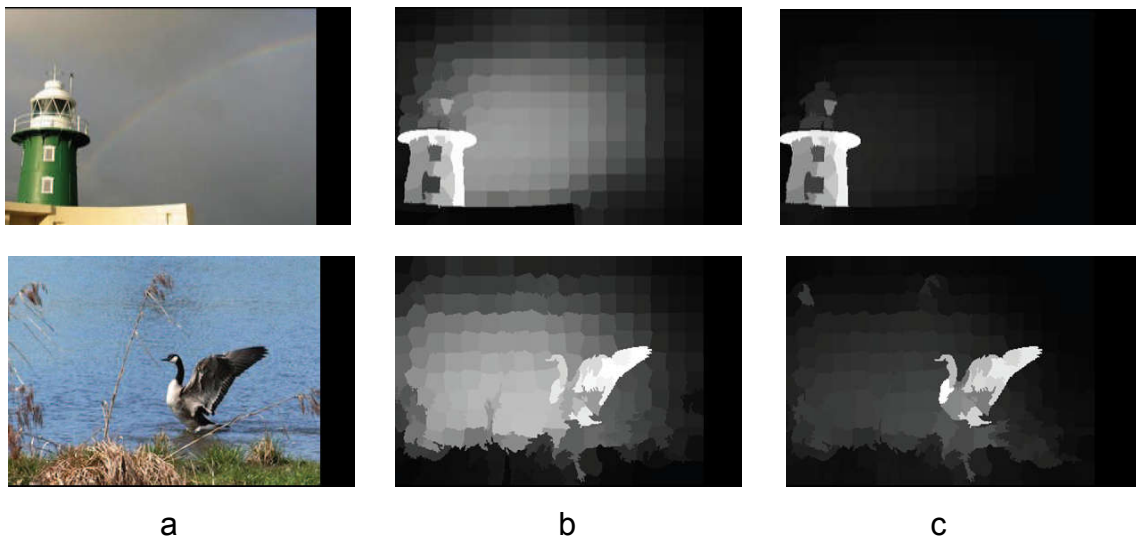


Figure 2.4 Des exemples illustrant les avantages du traitement de mise à jour de la saillance [54]. (a) Image en couleur, (b) carte de saillance générée sans le traitement de mise à jour, (c) carte de saillance générée avec le traitement de mise à jour.

Ils substituent ensuite le vecteur  $c$  dans l'équation 2.38 par le vecteur  $u$  défini comme suit:  $u_j = h_j / \sum_k h_k$  avec  $k$  et  $j$  des indexes des nœuds transitoires sur le graphe. Cela permet d'atténuer le niveau de saillance de ces zones au centre de l'image. Un inconvénient majeure de cette méthode réside dans le module de mise à jour qui peut affecter considérablement le taux de vrais positifs au centre de l'image dans le cas d'objets saillants définis sur une structure largement homogène avec des valeurs de saillance moyennes.

### 2.4.2.1.2 Connectivité avec l'arrière plan

Dans [82], les auteurs segmentent l'image à traiter en superpixels, puis définissent un graphe non orienté dont chaque nœud correspond à un superpixel. Le poids reliant deux nœuds adjacents du graphe est défini comme la distance euclidienne  $d_{euc}$  entre leur moyenne colorimétrique dans l'espace CIE LAB. La distance géodésique entre deux nœuds  $p$  et  $q$  est calculée comme la somme des poids du chemin le plus court qui les sépare comme suit:

$$d_{geo}(p, q) = \min_{p_1=p, p_2, \dots, p_n=q} \sum_{i=1}^{n-1} d_{euc}(p_i, p_{i+1}) \quad (2.41)$$

Nous notons que cette distance a été utilisée dans [53] comme seule mesure pour estimer la saillance dans un graphe. A partir de cette mesure, les auteurs définissent la zone que couvre l'influence d'un nœud  $p$  comme suit:

$$Area(p) = \sum_{i=1}^N \exp\left(-\frac{d_{geo}^2(p, p_i)}{2\sigma^2}\right) = \sum_{i=1}^N S(p, p_i) \quad (2.42)$$

avec  $N$  le nombre de nœuds du graphe et  $\sigma$  un paramètre de stabilisation fixé empiriquement à 10. Ils définissent ensuite une mesure basée sur les informations à priori sur la connectivité et l'arrière-plan pour estimer le taux d'overlapping (connectivité) d'une région avec les contours de l'image. Celle-ci est calculée comme suit:

$$Over = \frac{\sum_{i=1}^N S(p, p_i) \cdot \delta_i}{\sqrt{Area(p)}} \quad (2.43)$$

avec  $\delta_i = 1$  si le nœud  $i$  se trouve sur un des bords de l'image et  $\delta_i = 0$  sinon. A partir de cette mesure, ils formalisent la connectivité à l'arrière-plan pour chaque superpixel  $i$  sous forme d'une probabilité comme suit:

$$w_i^{bg} = 1 - \exp\left(-\frac{Over^2(p_i)}{2\varepsilon^2}\right) \quad (2.44)$$



avec  $\varepsilon$  une constante fixée empiriquement à 1. Le contraste d'un nœud est alors calculé en fonction de sa distance euclidienne ( $d_{euc}$ ), sa distance géodésique par rapport à tous les autres nœuds et sa connectivité par rapport à l'arrière-plan comme suit:

$$w_j^{fg} = \sum_{i=1}^N d_{euc}(p_j, p_i) \cdot d_{geo}(p_j, p_i) \cdot w_i^{bg} \quad (2.45)$$

Un post-processus d'optimisation est alors effectué afin d'améliorer le contraste entre les objets saillants et l'arrière-plan en minimisant la fonction suivante:

$$\sum_{i=1}^N w_i^{bg} s_i^2 + \sum_{i=1}^N w_i^{fg} (s_i - 1)^2 + \sum_{i,j} w_{ij} (s_i - s_j)^2 \quad (2.46)$$

avec  $\{s_i\}_{i=1}^N$  les valeurs de saillance des  $N$  nœuds du graphe. Pour chaque paire de nœuds adjacents  $(i,j)$ , correspond le poids  $w_{ij}$  défini comme suit:

$$w_{ij} = \exp\left(-\frac{d_{euc}(p_i, p_j)}{2\varepsilon^2}\right) + \mu \quad (2.47)$$

avec  $\mu$  une constante d'une petite valeur ajustée empiriquement à 0.1.

Le premier terme de l'équation 2.46 encourage le superpixel  $p_i$  avec une grande connectivité avec l'arrière-plan  $w_i^{bg}$  à avoir une petite valeur de saillance (proche de 0). Le second terme encourage le superpixel  $p_i$  avec une grande valeur de contraste  $w_i^{fg}$  à avoir une grande valeur de saillance (proche de 1). Le troisième terme vise à former des surfaces homogènes et continues en saillance. La figure 2.5 montre les différentes étapes d'estimation de la saillance basée sur la connectivité avec l'arrière plan [18].

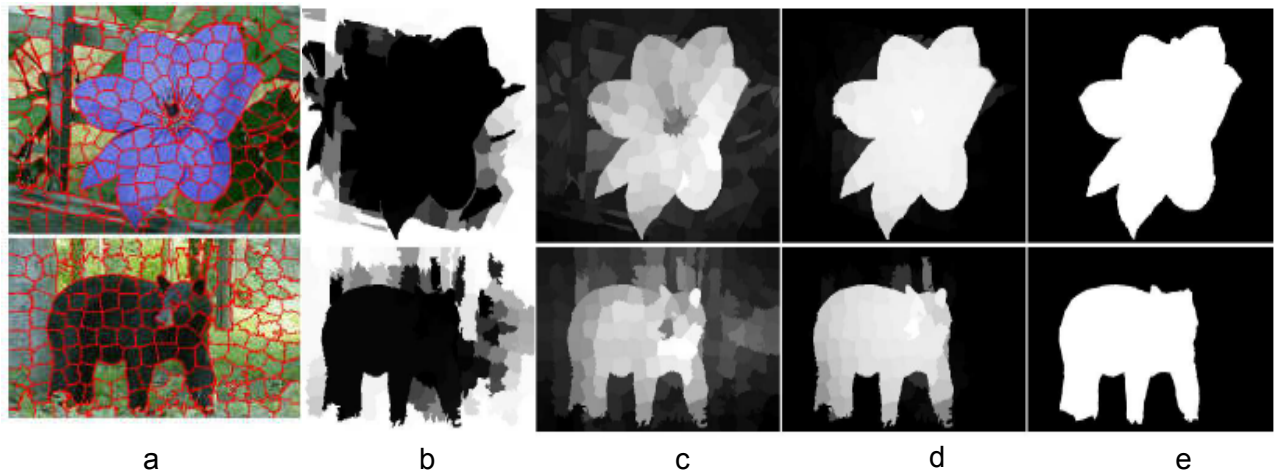


Figure 2.5 Différentes étapes de la méthode présentée dans [82]. (a) segmentation en superpixels, (b) Estimation de l'arrière-plan, (c) carte de contraste, (d) carte de saillance optimisée, (e) vérité terrain

#### 2.4.2.1.3 Ordonnancement sur un graphe mono-couche

Dans [9], les auteurs se basent sur une méthode de diffusion des valeurs de l'ordonnancement des nœuds d'un graphe construit sur l'image à traiter pour estimer la saillance des objets dans des images naturelles. Chaque nœud du graphe est connecté uniquement à ses nœuds voisins ainsi qu'aux voisins de ses voisins. La diffusion est alors limitée au maximum à deux pas dans le graphe (cf. figure 2.6).

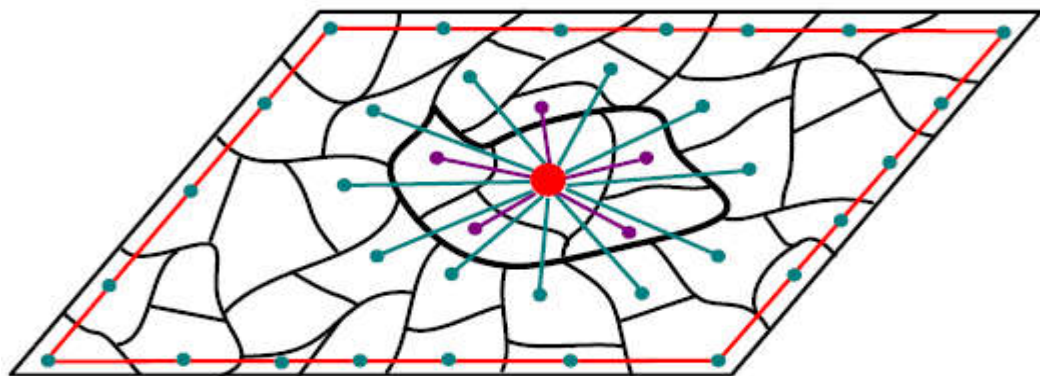


Figure 2.6 Structure du graphe utilisé dans [9].

Ils modélisent la détection de la saillance comme un problème d'ordonnancement [83] et proposent une démarche à deux étapes pour l'ordonnancement des nœuds sur le graphe.

Dans la première étape, les auteurs utilisent l'information à priori sur l'arrière-plan afin d'étiqueter les nœuds positionnés sur les quatre bords de l'image comme arrière-plan (parallélogramme rouge sur la figure 2.6). Une diffusion est alors effectuée à partir de chaque bord sur toute l'image. Ils obtiennent alors quatre cartes d'ordonnancement qu'ils fusionnent par multiplication pour obtenir une carte de saillance.

Dans la seconde étape, une segmentation binaire est effectuée sur la carte de saillance obtenue afin d'effectuer une seconde diffusion à partir des régions étiquetées objet saillant vers le reste de l'image. La figure 2.7 illustre les différentes étapes de l'algorithme proposée.

Plus précisément, ils segmentent d'abord l'image en un ensemble  $\Omega$  composé de  $n$  régions (superpixels) ( $\Omega = \{r_1, r_2, \dots, r_n\}$ ) en utilisant la méthode SLIC [61]. Ils construisent alors un graphe  $\langle V, E \rangle$  avec  $V$  l'ensemble des nœuds qui correspondent aux régions et  $E$  l'ensemble des arcs pondérés du graphe. Un poids  $w_{ij}$  associé à un arc reliant deux nœuds  $r_i$  et  $r_j$  voisins ou bien partageant un voisin est calculé sur la base de la similarité de couleur selon l'équation 2.35.

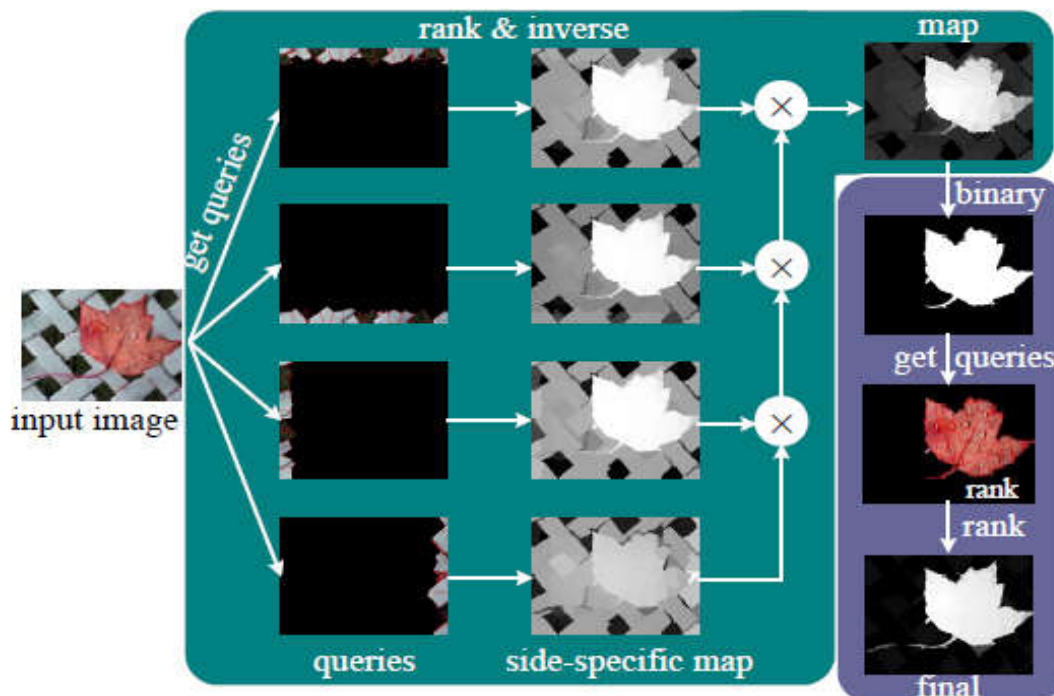


Figure 2.7 Différentes étape de la méthode proposée dans [9].

Soit  $\phi : \Omega \rightarrow \mathbb{R}^n$  la fonction d'ordonnement qui affecte une valeur de  $f_i$  pour chaque nœud  $r_i \in \Omega$ ,  $f = [f_1, f_2, \dots, f_n]^T$  l'ordonnement des nœuds du graphe et  $y = [y_1, y_2, \dots, y_n]^T$  le vecteur indiciel dans lequel  $y_i=1$  si la région  $r_i$  est une requête et  $y_i=0$  sinon. La propagation des valeurs de l'ordonnement initial se fait grâce à la minimisation de la fonction suivante:

$$f^* = \arg \min_f \left( \sum_{i,j=1}^n w_{ij} \left( \frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right)^2 + \lambda \sum_{i=1}^n (f_i - y_i)^2 \right) \quad (2.48)$$

avec  $d_i = \sum_{k=1}^n w_{ik}$  et  $\lambda$  un paramètre constant qui contrôle la contribution du terme de fidélité (le second terme). Ce terme force l'ordonnement à ne pas différer significativement de l'initialisation des valeurs de l'ordonnement. Par exemple les nœuds du bord initialisés comme arrière-plan doivent garder une valeur d'ordonnement similaire après le processus de diffusion. Le premier terme (terme d'uniformisation) force l'ordonnement à ne pas trop varier entre les nœuds voisins et entre les nœuds qui partagent au moins un voisin.

La figure 2.8 montre l'avantage de considérer la fusion de quatre diffusions où chacune est initialisée à partir d'un des quatre bords de l'image spécialement dans des cas où l'objet saillant touche l'un des bords de l'image. Nous pouvons constater également la nette amélioration apportée par la deuxième étape de l'algorithme. Même si dans certains cas, l'ordonnement retourné par la première étape est mal estimé dû au fait que l'objet touche au moins un des bords, la deuxième étape arrive à corriger efficacement l'estimation de la saillance.

Généralement, l'utilisation des superpixels comme unités de traitement permet d'améliorer sensiblement les résultats dans la détection d'objets saillants. Cependant, l'échelle de présentation des superpixels est tributaire de la surface occupée par les objets saillants dans l'image. Dans le cas des méthodes graphique, l'utilisation d'une forte granularité de superpixels augmente la similarité entre régions adjacentes. L'opération de diffusion retourne alors des régions saillantes plus étendues générant un taux de faux positifs important. Par contre, l'utilisation d'une basse granularité limite la similarité entre régions voisines et

atténue le processus de diffusion à des surfaces réduites. En outre, dans certains cas, l'utilisation d'une forte granularité de superpixels permet de retourner uniquement certaines parties de l'objet saillant avec certains détails de l'arrière-plan. Par contre, l'utilisation d'une granularité basse permet de retourner de grandes régions contenant les objets ainsi que des parties de l'arrière-plan. Nous allons présenter dans la section suivante les méthodes qui définissent plusieurs échelles de segmentation de l'image comme solution à ces limites.

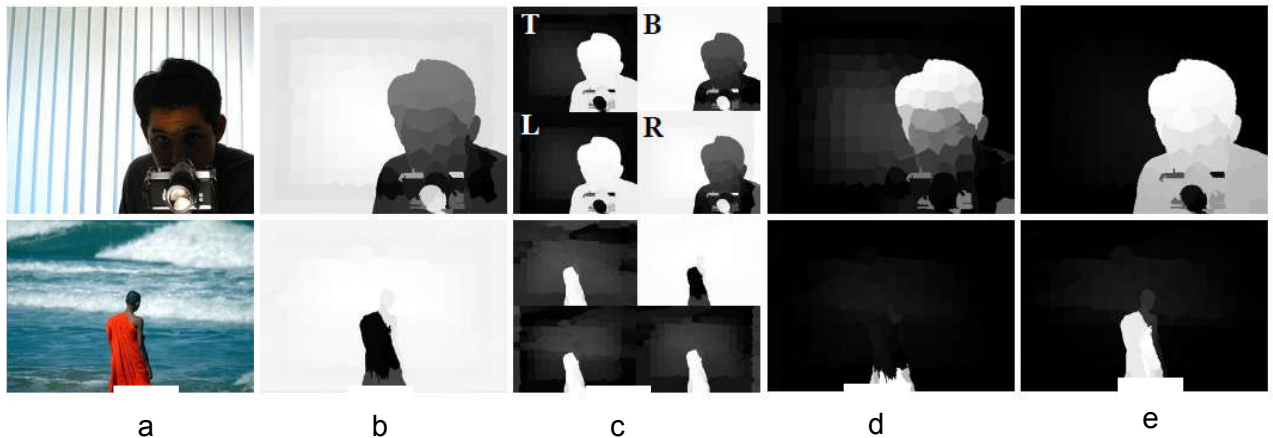


Figure 2.8 Exemples où l'objet saillant touche au moins un bord de l'image. (a) Image en couleur, (b) résultat d'une seule diffusion à partir de tous les bords, (c) Résultats de diffusion à partir de chaque bord, (d) fusion des quatre cartes de saillance, (e) Carte de saillance finale.

#### 2.4.2.2 Méthodes basées sur plusieurs résolutions

Ces méthodes se basent généralement sur un graphe multi-couches ayant une structure hiérarchique.

##### 2.4.2.2.1 Structure arborescente pour la détection de la saillance

Dans [10], les auteurs segmentent d'abord l'image à traiter en  $n$  régions primitives en se basant sur une méthode de détection de contours [84]. Ils calculent ensuite la similarité régionale entre chaque paire de segments  $R_i$  et  $R_j$  en utilisant la différence de chi-square des histogrammes de couleur normalisés notée  $Sim_c$  (premier terme de l'équation 2.49) et la distance spatiale notée  $Sim_d$  (second terme de l'équation 2.49).

$$Sim(R_i, R_j) = \exp\left(-\frac{1}{2} \sum_{k=1}^m \frac{[H_i(k) - H_j(k)]^2}{H_i(k) + H_j(k)}\right) \cdot \left(1 - \frac{\|\mu_i - \mu_j\|}{d}\right) \quad (2.49)$$

avec  $H_i$  l'histogramme de couleur de la région  $i$ ,  $m$  le nombre de bins considérés,  $d$  la longueur spatiale de la diagonale de l'image et  $\mu_i$  la position spatiale du centroïde de  $R_i$  définie comme suit:

$$\mu_i = \frac{\sum_{p \in R_i} x_p}{|R_i|} \quad (2.50)$$

avec  $x_p$  représente les coordonnées spatiales du pixel  $p$  et  $|R_i|$  représente le nombre de pixel dans  $R_i$ . Nous pouvons aisément constater que la similarité est élevée quand la la distance entre les régions est petite et leur distribution colorimétrique est similaire.

A partir de la mesure de similarité régionale, les auteurs estiment ensuite le niveau de saillance de chaque région en se basant sur trois mesures: le contraste global, la dispersion spatiale et la connectivité. Le contraste global est défini pour une région  $R_i$  comme la différence de couleur pondérée avec toutes les autres régions de l'image selon l'équation suivante:

$$GC(R_i) = \sum_{j=1}^n |R_j| \cdot Sim_d(R_i, R_j) \cdot \|mc_i - mc_j\| \quad (2.51)$$

avec  $mc_i$  dénote la moyenne de couleur de la région  $R_i$ . Le poids  $|R_j| \cdot Sim_d(R_i, R_j)$  indique que les régions larges et proches de  $R_i$  possèdent une meilleur contribution dans l'évaluation du contraste globale de  $R_i$ . La normalisation du contraste globale de  $R_i$  se fait comme suit:

$$NGC(R_i) = \frac{GC(R_i) - GC_{min}}{GC_{max} - GC_{min}} \quad (2.52)$$

avec  $GC_{min}$  et  $GC_{max}$  sont les valeurs minimums et maximums respectivement dans la mesure du contraste de toutes les régions de l'image. Comme les régions ayant des valeurs de similarités régionales similaires doivent avoir des valeurs de

contraste global similaires, les auteurs utilisent la similarité régionale pour raffiner le contraste global normalisé comme suit:

$$RGC(R_i) = \frac{\sum_{j=1}^n Sim(R_i, R_j) \cdot NGC(R_j)}{\sum_{j=1}^n Sim(R_i, R_j)} \quad (2.53)$$

La dispersion spatiale de la distribution colorimétrique d'une région  $R_i$  est définie comme suit:

$$SS(R_i) = \frac{\sum_{j=1}^n Sim(R_i, R_j) \cdot D(R_j)}{\sum_{j=1}^n Sim(R_i, R_j)} \quad (2.54)$$

avec  $D(R_j)$  la distance spatiale du centroïde de  $R_j$  au centre de l'image. De la même manière que précédemment, cette dispersion est normalisée puis raffinée et est notée  $RSS$ . Selon l'équation 2.54, les régions qui se trouvent éloignées du centre de l'image contribuent plus dans l'estimation de la dispersion spatiale.

La connectivité permet de mesurer le taux d'overlapping d'une région par rapport à l'arrière-plan (notée  $OP$ ). Pour se rapprocher au niveau objet, l'auteur utilise une segmentation plus grossière [84] comme indiqué sur la figure 2.9.f. Les auteurs considèrent alors les régions positionnés sur les bords de cette segmentation avec  $n_c$  régions comme template de l'arrière-plan notée  $B$ . La mesure  $OP$  est alors calculée par rapport à  $B$  pour un nœud  $R_j$ ,  $j \in \{1, 2, \dots, n_c\}$  par rapport à son périmètre  $\delta R_j$  et  $B$  comme suit:

$$OP(R_j) = \exp\left(-\lambda \frac{|R_j \cap B|}{\delta R_j}\right) \quad (2.55)$$

avec  $\lambda$  un coefficient ajusté à 2 pour une atténuation modérée de la connectivité des régions touchant les bords de la carte de segmentation. Ensuite, la connectivité de chaque région primitive  $R_i$ ,  $i \in \{1, 2, \dots, n\}$  est calculée comme suit:

$$OP(R_i) = OP(R_j), \forall R_i \in R_j \quad (2.56)$$

La valeur de saillance de chaque région est alors estimée en multipliant les trois mesures ( $S_I(R_i) = RGC(R_i).RSS(R_i).OP(R_i)$ ). La figure 2.9 montre les différentes étapes de génération de la carte de saillance initiale.

Afin de raffiner la carte de saillance initiale, les auteurs ont défini un arbre de partitions binaire pour fusionner les régions adjacentes de la carte de segmentation primitive selon leur similarité au niveau couleur et saillance initiale. Les feuilles de l'arbre sont les régions de la carte de segmentation primitive et chaque nœud de niveau supérieur est formé par la fusion de deux nœuds qui se trouvent à un niveau inférieur (cf. figure 2.10). Les auteurs définissent le critère de fusion entre les nœuds adjacents comme suit:

$$Mrg(R_i, R_j) = Sim_c(R_i, R_j). (1 - |S_I(R_i) - S_I(R_j)|) \quad (2.57)$$

Il est facile de constater que ce critère a une valeur élevée si  $R_i$  et  $R_j$  possèdent des distributions de couleur similaires et des valeurs de saillance régionales similaires. La fusion des nœuds de l'arbre est effectuée selon l'ordre décroissant de leurs critères de fusion avec un seuil d'arrêt qui limite le processus de fusion. Après chaque opération de fusion, l'histogramme de couleur régionale et la saillance régionale sont calculés pour le nouveau nœud généré.

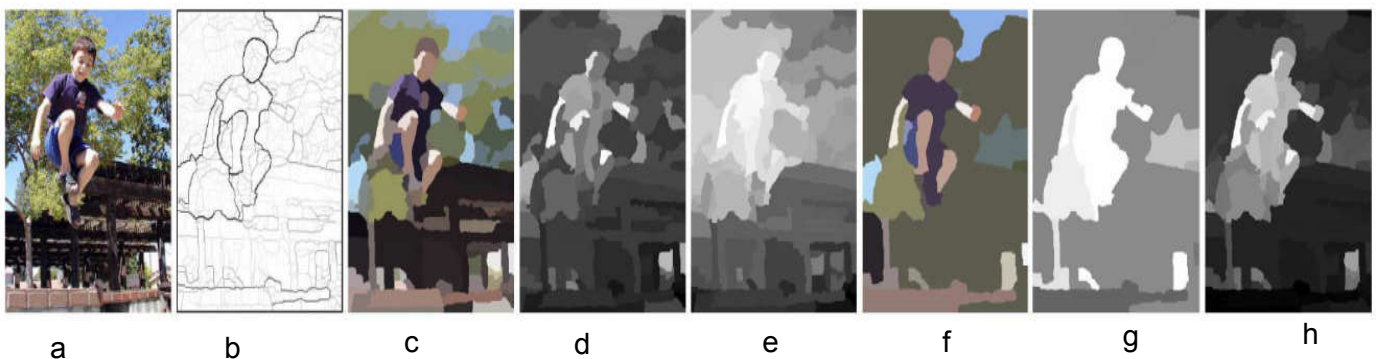


Figure 2.9 Différentes étapes de génération de la carte de saillance initiale [10] (a) l'image en entrée, (b) carte de contour, (c) Segmentation primitive (d) carte de contraste globale, (e) carte de dispersion spatiale de couleur, (f) segmentation grossière, (g) Carte de connectivité, (h) carte de saillance initiale.

Nous notons que l'estimation du contraste globale dans cette méthode peut nuire à la saillance finale dans le cas d'un objet saillant uniforme en couleur assez



large de surface entouré de petites régions homogènes avec un contraste évident avec l'objet (cf. équation 2.51). La proximité de ces régions à l'objet favorise leur contraste et génère une valeur de saillance à leurs niveaux qui accentue le taux de faux positifs. Ce phénomène est illustré sur la figure 2.11.

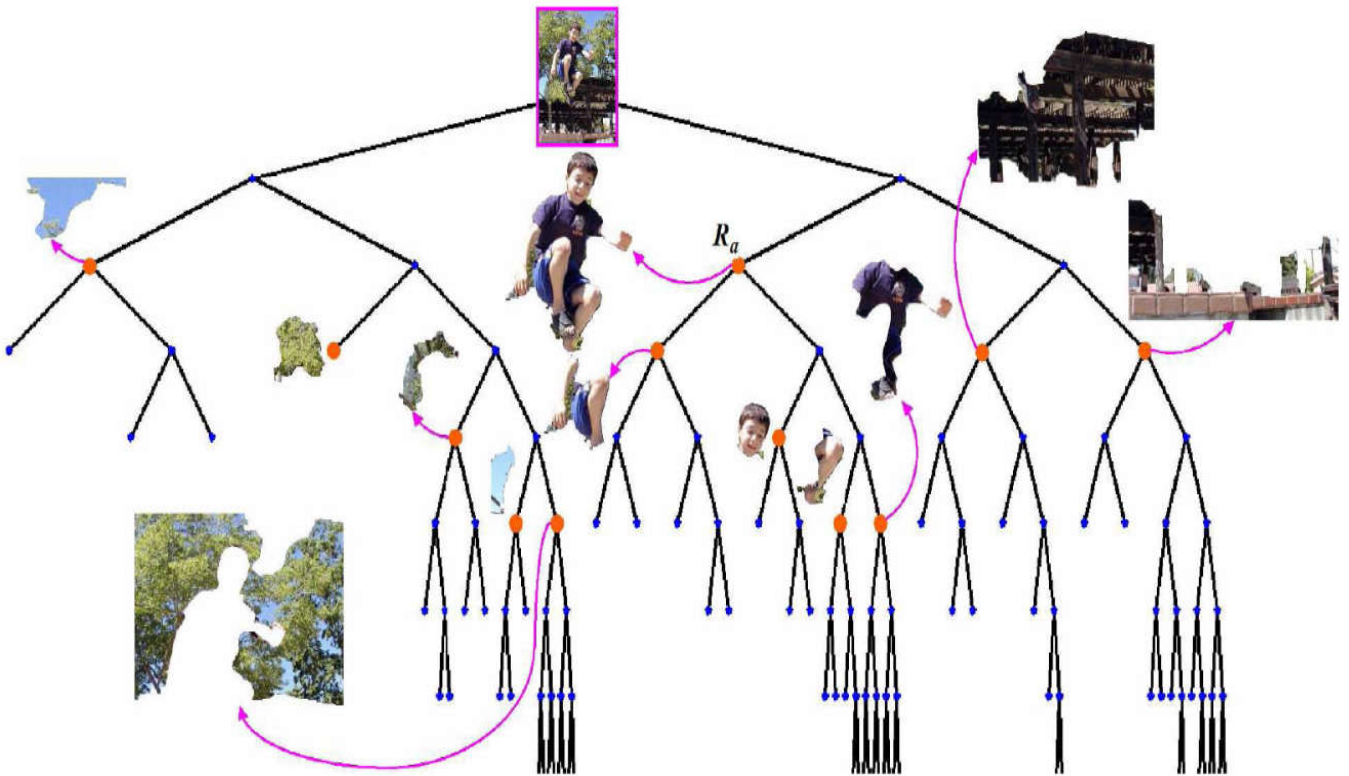


Figure 2.10 Exemple illustrant la fusion des nœuds dans l'arbre proposé dans [10].

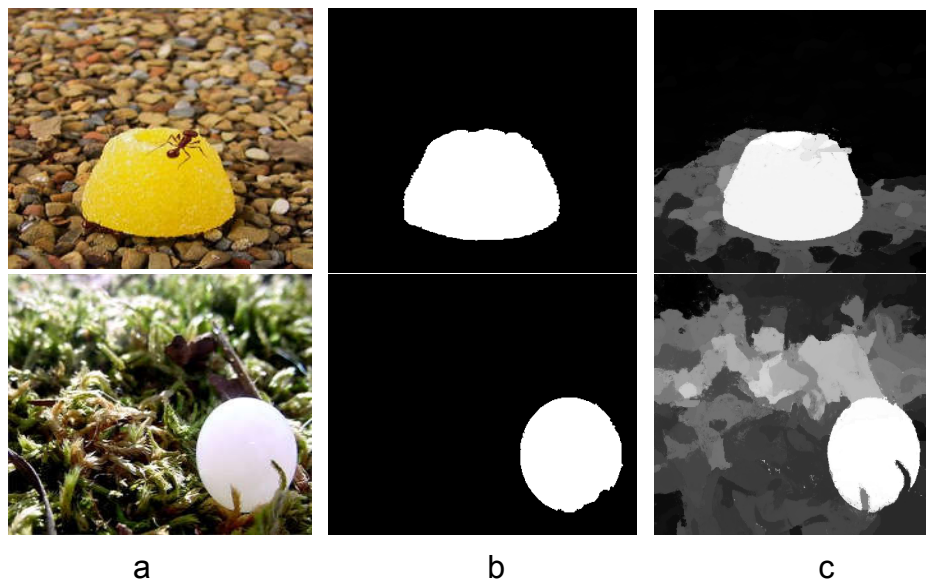


Figure 2.11 Exemple où la méthode présentée dans [10] génère un taux de faux positifs important.

#### 2.4.2.2.2 Structure hiérarchique pour la détection de la saillance

Dans [18], un modèle de détection d'objets saillants basé sur un graphe hiérarchique à plusieurs résolutions a été proposé. Les auteurs définissent une structure de trois résolutions superposées organisés de bas en haut selon l'ordre décroissant des granularités des différentes segmentations. La couche la plus basse est construite à partir d'un processus de fusion des régions adjacentes et similaires d'une carte de segmentation à forte granularité (Over-segmentation). Cette segmentation est obtenue en utilisant la méthode présentée dans [85]. Chacune des autres résolutions est obtenue en appliquant le même processus de fusion sur la résolution se trouvant au niveau inférieur (cf. Figure 2.12). Une carte de saillance intermédiaire est calculée ensuite pour chaque résolution.



Figure 2.12 Exemple illustrant les différentes résolutions du graphe proposé dans [18].

Le contraste local d'une région  $R_i$  est calculé comme une somme pondérée de son écart de couleur et de sa distance spatiale par rapport aux  $n$  autres régions voisines comme suit:

$$C_i = \sum_{j=1}^n \exp(-D(R_i, R_j)/\sigma^2) w(R_j) \|c_i - c_j\| \quad (2.58)$$

avec  $D(R_i, R_j)$  représente la distance euclidienne au carré des centres des régions  $R_i$  et  $R_j$  et  $\sigma$  représente une constante qui contrôle la contribution de la distance spatiale et  $c_i$  correspond à la valeur colorimétrique du centroïde de  $R_i$ . Le paramètre  $w(R_j)$  représente le nombre de pixels dans la région  $R_j$ . Nous remarquons que les régions voisines larges contribuent mieux au contraste local.

Les auteurs ont considéré ensuite l'information à priori sur le centre de l'image et définissent une nouvelle mesure  $H_i$  appliquée à une région  $R_i$  qui est calculée en fonction du nombre de pixels dans  $R_i$  et de leurs distances spatiales au centre de l'image selon l'équation suivante:

$$H_i = \frac{1}{w(R_i)} \sum_{x_i \in R_i} \exp(-\lambda \|x_i - x_c\|^2) \quad (2.59)$$

avec  $\{x_0, x_1 \dots\}$  est l'ensemble des coordonnées des pixels se trouvant dans la région  $R_i$ ,  $x_c$  les coordonnées du centre de l'image et  $\lambda$  un paramètre de régulation souvent ajusté à 9. Cette mesure permet d'attribuer aux régions proches du centre de l'image de plus grandes valeurs. La saillance intermédiaire est alors estimée comme la multiplication de cette mesure avec le contraste local ( $\bar{s}_i = C_i \cdot H_i$ ). Etant donné que chaque résolution est traitée à part, les auteurs proposent une fusion des résultats de saillance obtenus pour chaque résolution sur un graphe hiérarchique. Un nœud (région) se trouvant sur la résolution  $l$  ( $R_i^l$ ) est relié à un nœud de niveau supérieur ( $R_j^{l+1}$ ) si  $R_i^l$  est totalement contenu dans  $R_j^{l+1}$  sur une même surface ( $R_i^l \cap R_j^{l+1} = R_i^l$ ).

Le principe de cette méthode est que la saillance d'une région ne doit pas changer significativement par rapport à la saillance initiale et doit être la plus similaire possible à la saillance des régions de niveaux supérieurs auxquelles elle est liée. Cela est réalisé grâce à la minimisation de la fonction suivante:

$$S = \sum_l \sum_i \beta^l \|s_i^l - \bar{s}_i^l\|^2 + \sum_l \sum_{i, R_i^l \subseteq R_j^{l+1}} \lambda^l \|s_i^l - s_j^{l+1}\|^2 \quad (2.60)$$

avec  $\beta^l$  et  $\lambda^l$  des paramètres qui contrôlent la consistance de calcul dans la même résolution et entre les résolutions respectivement. La valeur de saillance initiale est symbolisée dans l'équation par  $\bar{s}_i^l$  et la saillance finale par  $s_i^l$ . La figure 2.13 illustre le principe générale de cette méthode. Nous notons que cette méthode possède la même limite que [10], vu qu'elle considère la superficie des régions dans l'estimation du contraste local (cf. équation 2.59).

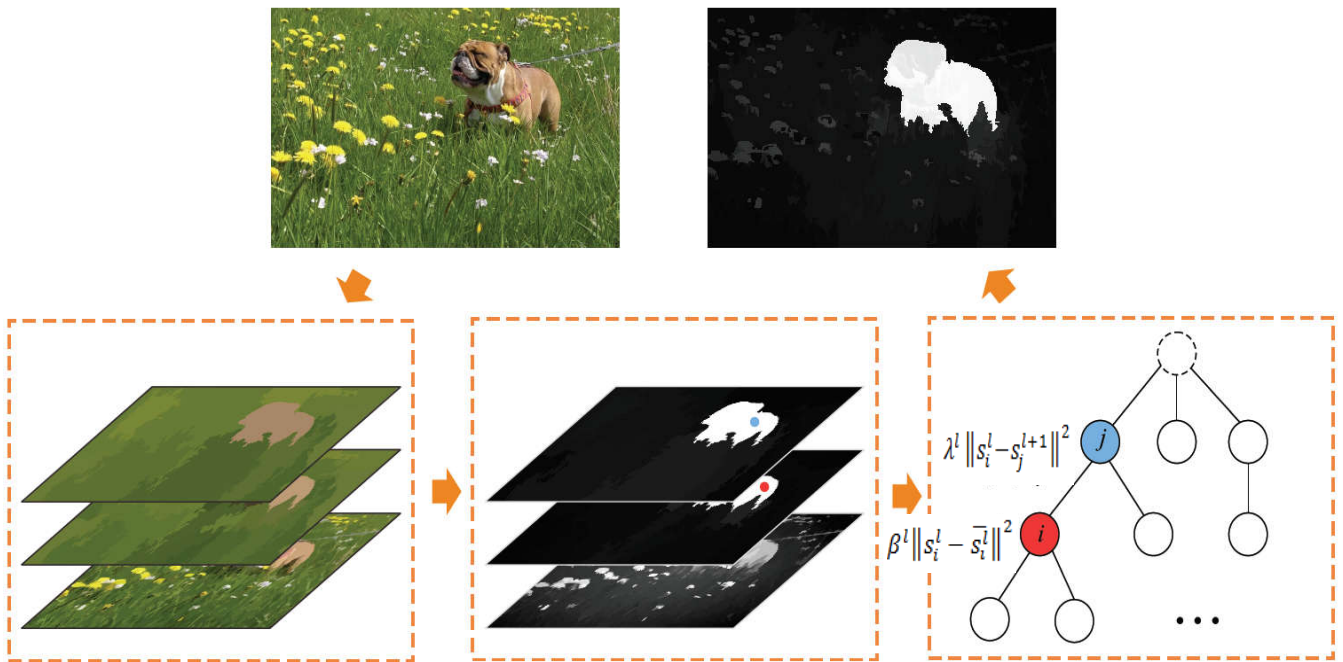


Figure 2.13 Aperçu général de la méthode présentée dans [18].

Nous constatons cependant une autre difficulté pour cette méthode dans le cas de similarité de distributions colorimétriques entre l'arrière-plan et l'objet saillant (première image de la figure 2.14) ou dans le cas d'arrière-plan hétérogène et complexe (deuxième image de la figure 2.14).

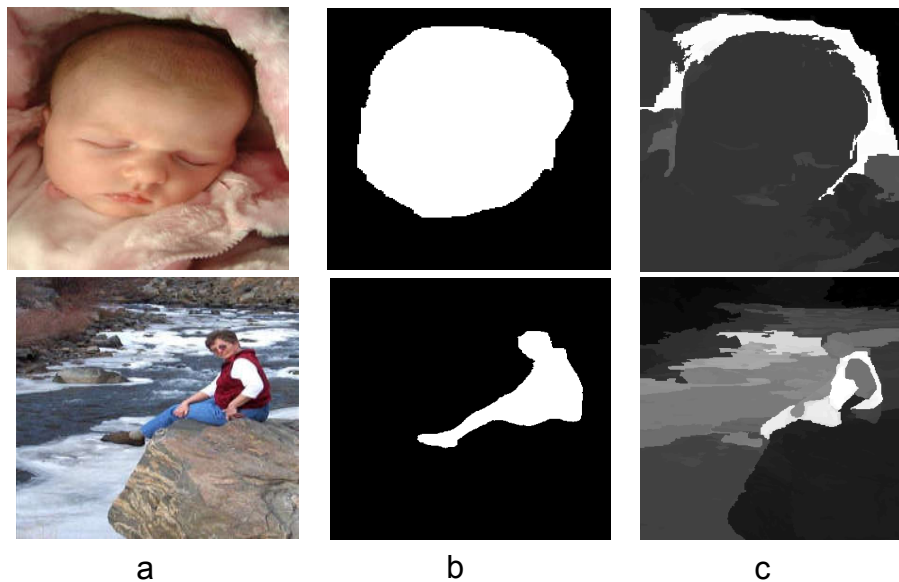


Figure 2.14 Exemple où la méthode présentée dans [18] estime mal la saillance.  
(a) Image en couleur, (b) Vérité terrain, (c) Carte de saillance retournée

### 2.4.2.2.3 Multi-graphes pour la détection d'objets saillants

Dans [15], les auteurs ont utilisé les principes de la physique quantiques sur une structure multi-graphes pour estimer la saillance des objets. Ils définissent trois résolutions pour l'image à traiter. Les différentes résolutions sont obtenues en segmentant l'image à traiter en superpixels avec des niveaux de granularité différents. Pour chaque résolution, les auteurs construisent un graphe où les superpixels représentent les nœuds. Ils définissent un arc pondéré non orienté entre chaque paire de nœuds adjacents. Les auteurs considèrent que deux nœuds sont  $m$ -voisins s'il est possible d'atteindre un nœud à partir de l'autre en parcourant  $m-1$  arcs. A partir d'un nœud du graphe, Le  $n^{\text{ième}}$  niveau de voisinage est alors défini comme l'ensemble des nœuds voisins appartenant à l'intervalle  $[1 + \text{floor}(2^{n-2}), 2^{n-1}]$ . Les auteurs enrichissent donc le graphe par des arcs reliant chaque nœud avec ses voisins du 1<sup>er</sup> jusqu'au 5<sup>ème</sup> niveau. Le poids  $w_{ij}$  relatif à un arc reliant deux nœuds  $i$  et  $j$  est défini en fonction de la distance colorimétrique et de l'information spatiale sur le voisinage comme suit:

$$w_{ij} = \left( \frac{1}{\varepsilon + d(\text{lab}_i, \text{lab}_j)^2} \right) \left( \frac{1}{|N_{i,C(i,j)}| \cdot |N_j|} \right)^2 \quad (2.61)$$

avec  $\text{lab}_i$  est le vecteur qui représente la moyenne de couleur dans l'espace CIE LAB,  $\varepsilon$  est une petite valeur qui évite la division par 0 et  $d(., .)$  correspond à la distance euclidienne. Le nombre de voisins du nœud  $j$  est représenté par  $|N_j|$  et le nombre de voisins du nœud  $i$  dans le niveau  $C(i,j)$  est représenté par  $|N_{i,C(i,j)}|$  avec  $C(i,j)$  la fonction qui retourne le niveau d'adjacence entre les nœuds  $i$  et  $j$ . Une fois les poids entre les nœuds sont définis, les auteurs construisent la matrice Hamiltonienne suivante:

$$H(i,j) = \begin{cases} V(i) + \frac{\hbar^2}{2} \sum_{k \in N_i} w_{ik}, & i = j \\ \frac{\hbar^2}{2} w_{ij}, & j \in N_i \\ 0 & \text{sinon} \end{cases} \quad (2.62)$$

avec  $h$  la constante de Dirac et  $V(i)$  correspond à la potentialité d'un nœud à appartenir à l'arrière-plan. Etant donné que la potentialité des nœuds positionnés sur les bords de l'image à appartenir à l'arrière-plan est élevée (information à priori sur l'arrière-plan),  $V(i)$  est ajusté à 1 sur tous les nœuds du bord et ajusté à 0 ailleurs. Une fois la matrice  $H$  est établie, les auteurs calculent le vecteur propre avec la valeur propre la plus basse  $\Psi$ . La carte de saillance  $L_i$  est alors estimée comme suit:

$$L_i = \Psi(i) \circ \Psi(i) \quad (2.63)$$

Ceci revient à trouver les régions de l'image les plus contrastées par rapport à leur voisinage en optimisant le critère suivant:  $\arg \min_A (cut(A, \bar{A}) / area(A))$  avec  $cut(A, \bar{A})$  une mesure de similarité entre la région  $A$  avec son voisinage  $\bar{A}$  calculé comme suit  $cut(A, \bar{A}) = \sum_{u \in A, v \in \bar{A}} w_{u,v}$ . La fonction  $area(A)$  retourne la surface de  $A$ . Une carte de saillance intermédiaire est ainsi estimée pour chaque résolution. La carte de saillance finale est ensuite obtenue en effectuant la moyenne de toutes ces cartes de saillance. Comme cette méthode utilise l'information à priori sur l'arrière-plan, toute partie de l'objet saillant touchant un des bords de l'image est fortement considérée comme arrière-plan et toute région similaire à cette partie est également considérée comme arrière-plan (cf. figure 2.15).

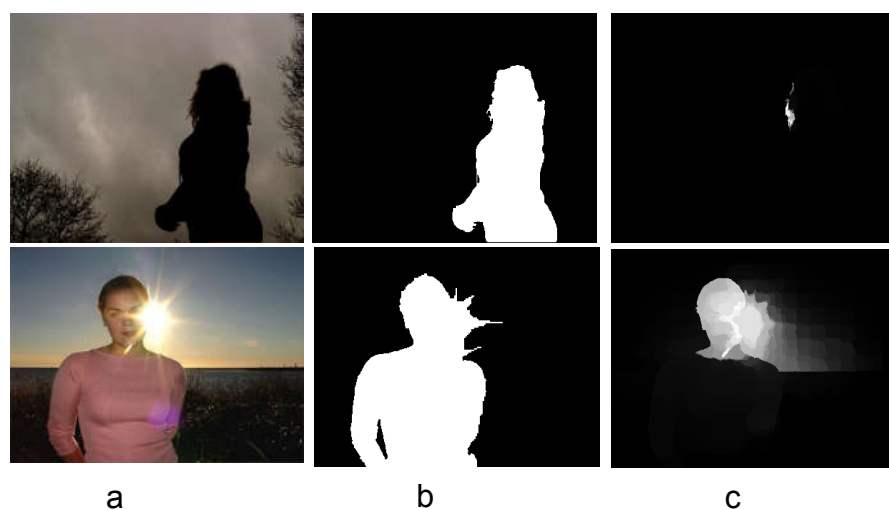


Figure 2.15 Exemple où l'objet saillant est connecté à un des bords de l'image. (a) Image en couleur, (b) Vérité terrain, (c) Carte de saillance retournée

### 2.4.3 Autres méthodes

#### 2.4.3.1 Reconstruction dense et éparses d'erreur

Une méthode basée sur la reconstruction dense et éparses d'erreur est proposée dans [86]. Les auteurs segmentent l'image en superpixels en utilisant l'algorithme SLIC. Chaque région est caractérisée par sa moyenne de couleur dans les espaces  $RGB$  et  $Lab$  ainsi que les coordonnées spatiales des pixels lui appartenant ( $x = (L, a, b, r, g, b, x, y)$ ). L'image est alors représentée par le vecteur  $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{D \times N}$ , avec  $N$  le nombre de régions et  $D$  la dimension des caractéristiques. Les auteurs définissent les régions qui se situent sur les bords de l'image comme information à priori sur l'arrière-plan et les représentent par leurs vecteurs de caractéristiques ( $B = [b_1, b_2, \dots, b_M]$  avec  $M$  le nombre de régions qui se situent sur les bords). Ils définissent par la suite pour chaque région, deux types de mesures de reconstruction de l'erreur: éparses et dense.

Les auteurs considèrent qu'une région avec une importante construction d'erreur basée sur le template de l'arrière-plan  $B$  est plus susceptible d'appartenir au fond de l'image. Ils définissent une représentation dense de l'arrière-plan (les contours de l'image) en appliquant l'ACP sur  $B$ . Les vecteurs propres de  $B$ ,  $U_B = [u_1, u_2, \dots, u_{D'}]$  correspondant aux plus grandes valeurs propres  $D'$  sont calculées pour former l'ACP de  $B$ . Le coefficient de reconstruction pour une région  $i$  ( $i \in [1, N]$ ) est calculé comme suit:

$$\beta_i = U_B^T (x_i - \bar{x}) \quad (2.64)$$

avec  $\bar{x}$  est le vecteur moyenne de  $X$ . La reconstruction dense de l'erreur est représentée comme suit:

$$\varepsilon_i^d = \|x_i - (U_B \beta_i + \bar{x})\|_2^2 \quad (2.65)$$

La représentation dense du template de l'arrière-plan peut particulièrement affecter l'estimation du fond de l'image à cause de la perte de précision de la représentation (cf. figure 2.16 (b)). Pour remédier à cela, les auteurs proposent une représentation éparses du template de l'arrière-plan dans un espace plus large de taille  $M$ . Ils associent à une région  $i$  le vecteur des caractéristiques  $\alpha_i$  qui

représente la corrélation de  $i$  avec chaque élément du template de l'arrière-plan. Ce vecteur est estimé comme suit:

$$\alpha_i = \arg \min_{\alpha_i} \|x_i - B\alpha_i\|_2^2 + \lambda \|\alpha_i\| \quad (2.66)$$

Notons que le produit  $B\alpha_i$  permet une représentation dans l'espace d'origine (de taille  $D$ ) à partir de l'espace des caractéristiques (taille  $M$ ). L'erreur éparsée est donc calculée comme suit:

$$\varepsilon_i^s = \|x_i - B.\alpha_i\|_2^2 \quad (2.67)$$

Comme tous les éléments du template de l'arrière-plan sont considérés comme fonctions de bases, l'estimation de l'erreur est plus sévère et le fond de l'image est écarté plus efficacement. Cependant une connexion d'une partie de l'objet saillant avec l'un des bords de l'image peut affecter l'estimation de la saillance (cf. Figure 2.16 (c)). La reconstruction éparsée est plus robuste au traitement de fonds d'image complexes et la reconstruction dense est plus adaptée au cas où l'objet est connectée aux éléments du template de l'arrière-plan. Les deux mesures sont donc complémentaires. Les auteurs proposent un algorithme de propagation contextuelle pour chaque type de reconstruction pour améliorer l'estimation de la saillance. Ils segmentent l'image en  $K$  ( $k= 1, 2, \dots, K$ ) clusters et initialisent la reconstruction d'erreur pour chaque segment selon les équations 2.65 et 2.67. La reconstruction de l'erreur propagée d'une région  $i$  appartenant à un cluster  $k$  est définie comme suit:

$$\tilde{\varepsilon}_i = \tau \sum_{j=1}^{N_c} w_{ikj} \tilde{\varepsilon}_{kj} + (1 - \tau)\varepsilon_i \quad (2.68)$$

$$w_{ikj} = \frac{\exp\left(-\frac{\|x_i - x_{kj}\|^2}{2\sigma_x^2}\right) (1 - \delta(k_j, i))}{\sum_{j=1}^{N_c} \exp\left(-\frac{\|x_i - x_{kj}\|^2}{2\sigma_x^2}\right)} \quad (2.69)$$



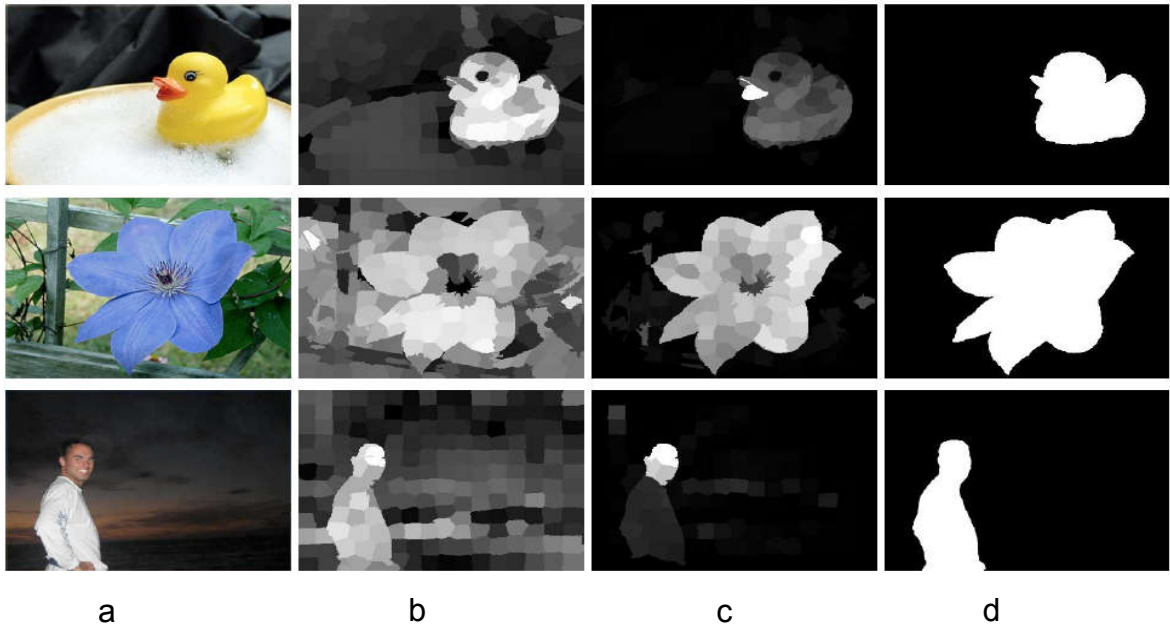


Figure 2.16 Cartes de saillance basées sur la reconstruction dense et éparses de l'erreur [86]. (a) l'image en couleur, (b) carte de saillance obtenue à partir de la reconstruction dense, (c) carte de saillance obtenue à partir de la reconstruction éparses et (d) vérité terrain.

avec  $\{k_1, k_2, \dots, k_{N_c}\}$  dénote les  $N_c$  segments dans le cluster  $k$ ,  $\tau$  est un paramètre de régularisation,  $\sigma_x^2$  la somme des variances de  $X$  pour chacune des  $D$  dimensions et  $\delta(k_j, i)$  une fonction indicatrice renseignant sur l'appartenance de  $i$  à  $k_j$ . Le premier terme de l'équation représente la somme pondérée de tous les autres segments du même cluster et le second terme correspond à la valeur de reconstruction initiale. Les auteurs réitèrent le même processus pour  $N_s$  différentes résolutions définies sur différentes échelles de segmentation, puis intègrent les différents résultats au niveau pixel comme suit:

$$E(z) = \frac{\sum_{s=1}^{N_s} w_{zn^{(s)}} \tilde{\epsilon}_{n^{(s)}}}{\sum_{s=1}^{N_s} w_{zn^{(s)}}}, \quad w_{zn^{(s)}} = \frac{1}{\|f_z - x_{n^{(s)}}\|_2} \quad (2.70)$$

avec  $f_z$  la présentation d'un pixel  $z$  dans l'espace  $D$ -dimensionnel et  $n^{(s)}$  le segment qui contient  $z$  définit sur l'échelle  $s$ . Les auteurs proposent un processus de raffinement basé sur un formalisme gaussien qui favorise les valeurs proches du centre de l'objet saillant (estimé par l'équation 2.68) à avoir une plus grande valeur de saillance. Comme la reconstruction de l'erreur éparses et la reconstruction de

l'erreur dense sont deux processus complémentaires exécutés séparément et qui retournent deux cartes de saillances différentes, un processus de fusion bayésien des deux cartes est alors effectué pour former une carte de saillance finale.

#### 2.4.3.2 Décomposition matricielle de l'image: partie redondante et partie saillante

Les objets saillants, en terme de leur unicité dans l'image, peuvent également être définis comme les bruits épars dans un certain espace de caractéristiques dans lequel l'image à traiter (cf. figure 2.17) est représentée comme une matrice de rang faible [87,88,17,89].

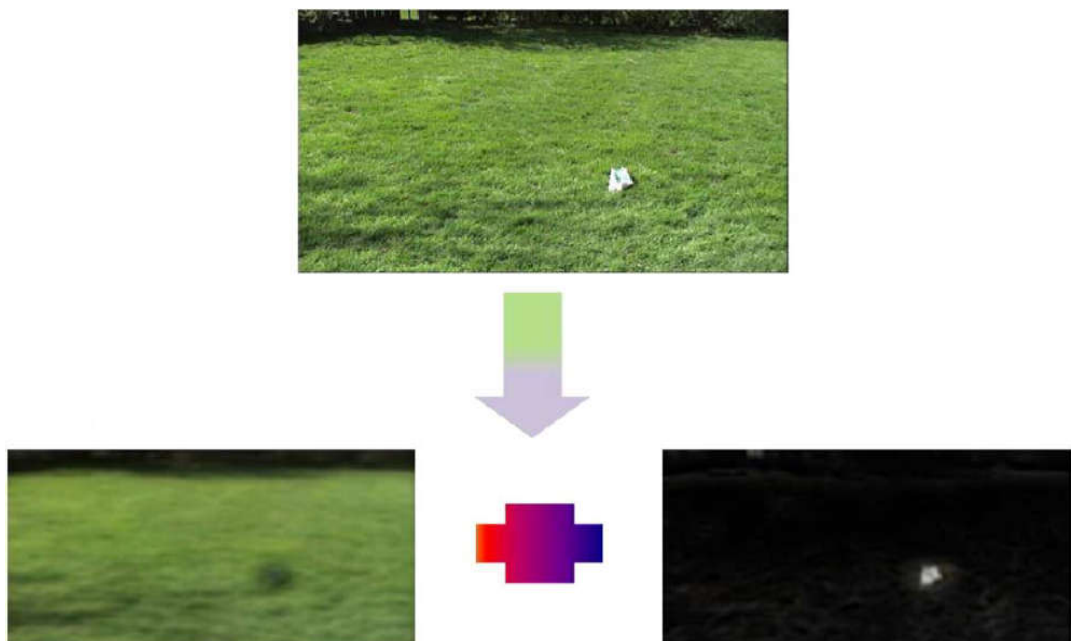


Figure 2.17 L'image en couleur peut être décomposée en deux parties: partie redondante et partie saillante [89].

L'hypothèse de base est que les régions non-saillantes (arrière-plan) peuvent être représentées par la matrice de rang faible alors que les régions saillantes sont représentées par les bruits épars sur l'image. Formellement, chaque région  $r_i$  est représentée par son vecteur de caractéristiques  $f_i$ . L'image est alors représentée par la matrice des caractéristiques  $F = [f_1, f_2, \dots, f_N] \in \mathbb{R}^{D \times N}$  avec  $D$  la dimension de chaque vecteur de caractéristiques.  $F$  peut être décomposé en deux parties, la matrice de redondance  $L = [l_1, l_2, \dots, l_N] \in \mathbb{R}^{D \times N}$  et la matrice éparsée  $S[s_1, s_2, \dots, s_N] \in \mathbb{R}^{D \times N}$  en optimisant la fonction objective suivante:

$$\arg \min_{L,S} \|L\| + \lambda \|S\|, \quad F = L + S \quad (2.71)$$

avec  $\lambda$  un coefficient de régularisation.  $L$  correspond à la présentation de l'arrière-plan et  $S$  à l'objet saillant. La saillance de  $r_i$  est alors définie comme suit:

$$s(r_i) = \|s_i\|_2, \quad \text{ou } s(r_i) = \|s_i\|_1 \quad (2.72)$$

Dans [17], les auteurs intègrent un processus de régularisation Laplacien qui tient en compte de la structure géométrique et encourage les patches voisins à avoir une représentation similaire.

#### 2.4.3.3 Apprentissage de régression de type forêts aléatoire

Dans [16], les auteurs segmentent l'image en plusieurs résolutions, puis se basent sur l'apprentissage d'une régression de type forêts aléatoire pour estimer la saillance d'une région à partir de ses caractéristiques régionales. Les auteurs définissent un espace de 92 caractéristiques répertoriées dans trois types de descripteurs. Le descripteur de contraste local utilise les caractéristiques photométrique (espaces de couleur, différents types de texture) pour estimer le contraste d'une région  $R_i$ . Pour chaque caractéristique  $k$ , il est calculé comme suit:

$$x_k(R_i) = \sum_{j=1}^N \alpha_j \cdot \exp\left(\frac{\|p_i - p_j\|^2}{2\sigma^2}\right) \cdot D(v_{R_i}, v_{R_j}) \quad (2.73)$$

avec :  $D(v_{R_i}, v_{R_j})$  la différence entre les régions  $R_i$  et  $R_j$  en considérant  $k$  ;

$N$  représente le nombre de régions de l'image ;

$p_i$  la position spatiale du centre de la région  $R_i$  ;

$\sigma^2$  un paramètre qui contrôle la contribution de la distance spatiale ;

$v_{R_i}$  la valeur de la caractéristique  $k$  dans  $R_i$  ;

et  $\alpha_j$  correspond à la surface normalisée de la région  $R_j$ .

Le descripteur de l'arrière-plan considère une bande de 5 pixels de largeur tout au long du contour de l'image comme template de l'arrière-plan. Ce descripteur est calculé pour une région et une caractéristique  $k$  comme sa différence avec le

template prédéfini. Le descripteur de caractéristiques régionales englobe les caractéristiques géométriques de l'image comme la forme, la taille et la position spatiale ainsi que les caractéristiques d'apparence comme la distribution de couleur et de texture dans une région. Par la suite, les auteurs font une sélection des 60 meilleurs caractéristiques en effectuant des tests sur l'apprentissage de la régression basée sur les arbres aléatoires. L'application de la régression retourne une carte de saillance relative à chaque résolution. La fusion de ces cartes se fait par une somme pondérée dont les poids sont estimés par apprentissage.

Comme cette méthode est basée sur l'apprentissage, elle peut souffrir du problème de la généralisation. Sa principale faiblesse réside dans son instabilité à estimer la saillance dans des images avec des arrière-plans et des objets saillants dotés de nouvelles caractéristiques différentes de celles présentes dans l'ensemble d'apprentissage. La figure 2.18 montre deux exemples où l'objet saillant est mal détecté.

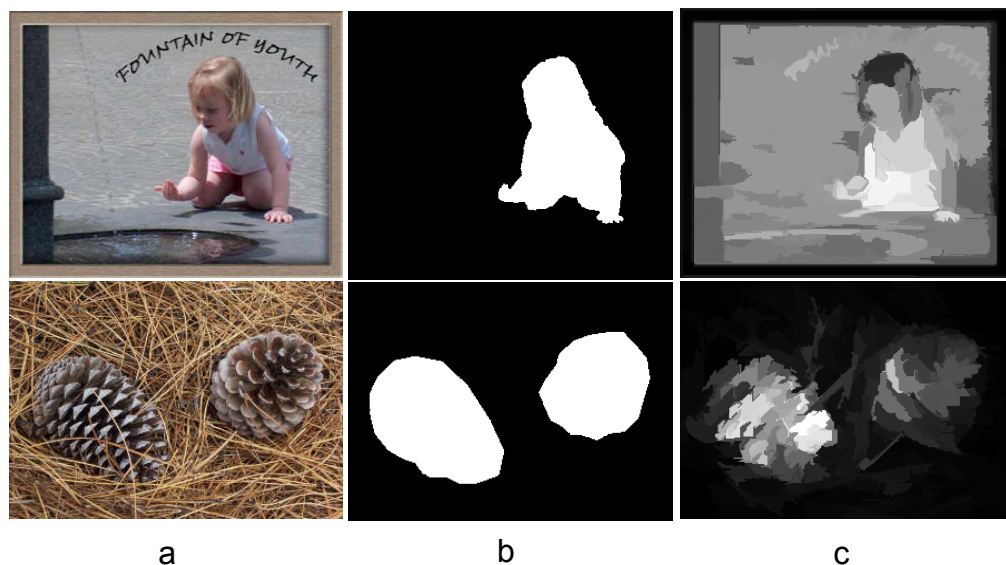


Figure 2.18 Exemple où l'objet saillant est mal détecté dans [16]. (a) image en couleur, (b) Vérité terrain, (c) Carte de saillance retournée

#### 2.4.3.4 Espace de haute dimension pour la détection de la saillance

J. Kim et al. [90] ont proposé une méthode de détection d'objets saillants dans un espace à haute dimension. Ils segmentent d'abord l'image à traiter en superpixels. Puis, Chaque région est représentée par un vecteur qui définit ses caractéristiques de couleur (moyenne et histogrammes), spatiales, de texture et

de forme (gradient, surface,...) ainsi que les caractéristiques de contraste. La caractéristique d'histogramme relative à une composante de couleur pour une région est calculée comme la somme de ses distances de chi-square par rapport à tous les bins considérés et toutes les autres régions de l'image.

La caractéristique de contraste globale relative à une composante de couleur pour une région est calculée comme la somme des distances euclidiennes par rapport à toutes les autres régions de l'image. La caractéristique de contraste locale relative à une composante de couleur pour une région est calculée comme la somme de ses distances euclidiennes pondérées par rapport à toutes les régions de l'image. Le poids entre deux régions est calculé en fonction de leur distance spatiale. Plus deux régions sont proches, plus le poids est important. Chaque composante de couleur lui est ensuite appliquée trois transformations Gamma avec des valeurs 0.5, 1.0 et 2.0 pour enrichir la représentation des caractéristiques. Les auteurs se basent ensuite sur une carte de saillance préalable obtenue par l'application d'un classifieur de type forêts aléatoire sur l'image est pour initialiser les régions candidates pour l'arrière-plan et celles candidates pour le pseudo-objet. Ils définissent ensuite des échantillons de régions candidates pour les deux classes considérées et les représentent dans une matrice  $\tilde{K}$  où les lignes correspondent aux régions et les colonnes aux différentes caractéristiques. Ils définissent également un vecteur d'indices qui renseigne sur l'appartenance de chaque région notée  $U$ . La carte de saillance globale est obtenue par une combinaison linéaire de coefficients de couleur. Ces coefficients représentés dans un vecteur  $\alpha$  sont alors estimés comme suit:

$$\arg \min_{\alpha} \|(U - \tilde{K}\alpha)\|_2^2 + \lambda \|\alpha\|_2^2 \quad (2.74)$$

avec  $\lambda$  un paramètre de régularisation.

Pour l'estimation de la saillance locale, les auteurs utilisent la distance spatiale et colorimétrique d'une région par rapport aux  $k$  plus proches régions du pseudo-objet et aux  $k$  plus proches régions de l'arrière-plan (cf. figure 2.19). A partir de ces caractéristiques, ils entraînent et appliquent une régression de type forêts aléatoires et génèrent une carte de saillance dite locale. Les deux cartes sont ensuite fusionnées d'une manière linéaire pour générer la carte de saillance

finale. Nous notons que la limite de toutes les méthodes basées sur une seule résolution est le problème de détermination de granularité adéquate pour la segmentation d'image.

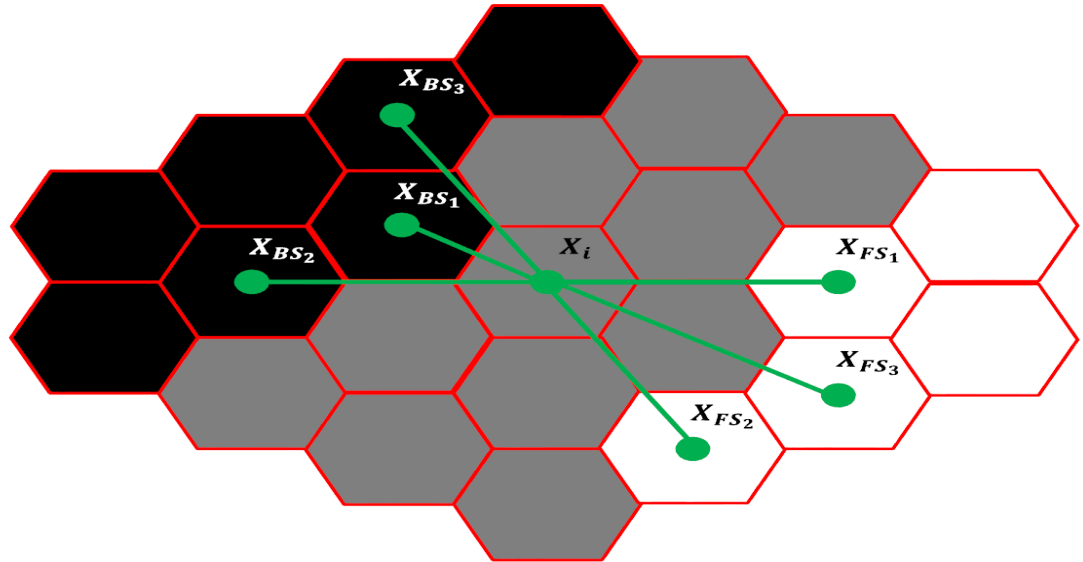


Figure 2.19 Illustration de la saillance locale. Les régions noires, blanches et grises dénotent respectivement, les superpixels de l'arrière-plan, du pseudo-objet et inconnus. La saillance est calculée en fonction de la distance spatiale et colorimétrique aux  $K$  plus proches régions de l'arrière-plan et aux  $K$  plus proches régions du pseudo-objet [90].

#### 2.4.3.5 Estimation de la saillance basée sur le calcul de variance

J. Liu et al. [91] proposent une méthode basée sur les statistiques de couleur pour estimer la saillance dans une image. Ils subdivisent d'abord l'image  $I_{M \times N}$  en blocs de taille  $l \times l$  avec un recouvrement de 50%. Pour chaque canal de couleur  $L$ ,  $a$  et  $b$ , ils définissent le contraste local comme une mesure de variance au sein de chaque bloc. Un bloc est défini par les coordonnées spatiales  $i=1 \dots [2(N-l)/l+1]$  et  $j=1 \dots [2(M-l)/l+1]$ . Sa variance est calculée comme suit:

$$V_{ij}^z = \frac{1}{l \times l} \sum_{k=1}^{|x|} (x_k - \mu)^2 \quad (2.75)$$

avec  $x_k$  la valeur colorimétrique du pixel  $k$ ,  $\mu$  la valeur colorimétrique moyenne dans tout le bloc et  $z \in \{L, a, b\}$ . Plus la distribution de couleur à l'intérieur du bloc est homogène, plus la valeur de variance est petite. La variance finale est alors estimée comme suit:

$$V_{ij} = (V_{ij}^L)^2 + (V_{ij}^a)^2 + (V_{ij}^b)^2 \quad (2.76)$$

Après normalisation dans l'intervalle  $[0, 255]$ , l'auteur obtient une carte de variance dans les blocs notée ( $BVM$ ). Afin d'améliorer l'estimation de la saillance, les auteurs ont intégré la distance au centre de l'attention qu'ils ont définis comme le centre de gravité de la carte  $BVM$ . Ils mettent à jour comme suit:

$$V'_{ij} = V_{ij} \exp \left( - \sqrt{\left( \frac{i - x_0}{k_w \times N} \right)^2 + \left( \frac{j - y_0}{k_w \times M} \right)^2} \right) \quad (2.77)$$

avec  $k_w$  un paramètre qui contrôle la sensibilité de la distance et  $(x_0, y_0)$  les coordonnées du centre de gravité de  $BVM$  calculées en fonction des moments d'ordre 1 ( $m_{10}$  et  $m_{01}$ ) et du moment d'ordre 0 ( $m_{00}$ ) comme suit :

$$x_0 = m_{10}/m_{00} ; y_0 = m_{01}/m_{00} \quad (2.78)$$

Une opération de binarisation est ensuite effectuée sur la carte  $BVM$  en appliquant un seuil de segmentation  $T$ . Ils définissent par la suite une fenêtre ajustée de manière à contenir toutes les parties saillantes (cf. figure 2.20 (d)). La partie à l'intérieur de la fenêtre (notée  $F$ ) est considérée comme potentiellement appartenant à l'objet saillant et la partie extérieure (notée  $B$ ) est considérée comme appartenant à l'arrière-plan. Les auteurs effectuent ensuite une quantification des couleurs utilisées dans un ensemble  $U = \{1, 2, \dots, K\}$ . Pour une couleur donnée, le rapport entre son histogramme dans  $F$  et son histogramme dans  $B$  est calculé comme suit:

$$R_c = \frac{n_{c \in F}}{n_{c \in B}} \quad (2.79)$$

avec  $n_{c \in X}$  le nombre de pixels de la couleur  $c$  dans l'ensemble  $X$ . La saillance finale d'une couleur est alors calculée comme suit:

$$S_c = \begin{cases} 1 & \text{si } c \in U - B \\ \frac{k}{K_c + 1} & \text{si } R_c \geq R_T \text{ et } c \in B \cap F \\ 0 & \text{sinon} \end{cases} \quad (2.80)$$

avec  $K_c$  l'ensemble total de couleurs dont  $R_c$  est supérieur à  $R_T$  et  $R_T$  un seuil qui contrôle  $K_c$ . L'ensemble des  $K_c$  couleurs est trié dans un ordre ascendant en fonction des valeurs de  $R_c$  et  $k \in \{1, 2, \dots, K_c\}$  un index des couleurs considérées correspondant à la couleur  $c$ . L'inconvénient de cette méthode est qu'elle retourne de larges régions avec des valeurs de saillance élevées mais générant un taux assez élevé de faux positifs.

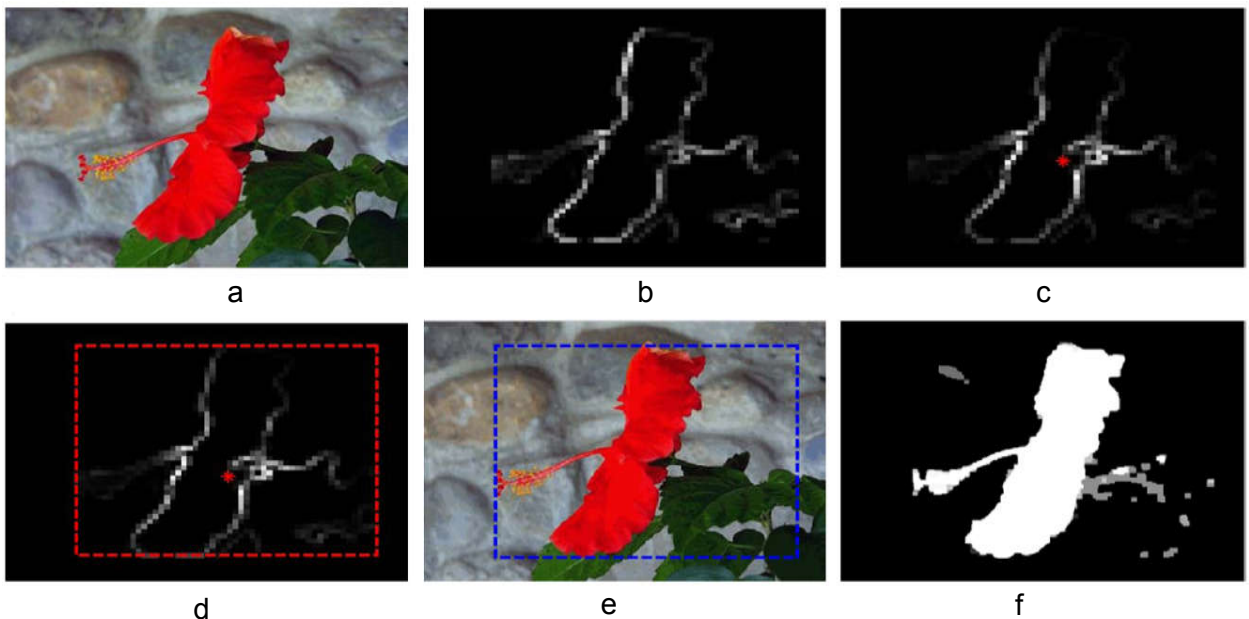


Figure 2.20 Exemple de déroulement de l'algorithme de [91]. (a) image en couleur, (b) carte VBM, (c) calcul du centre de gravité, (d-e) ajustement de la fenêtre, (f) estimation de saillance globale.

Le tableau 2.1 résume les différentes méthodes de l'état de l'art étudiées dans ce chapitre.



Tableau 2.1

Différentes méthodes de l'état de l'art sur la détection d'objets saillants

	Type De méthode	Niveau d'estimation de la saillance	Méthode	
Méthodes basées sur l'estimation du contraste local	Basée sur le modèle centre/périphérie	Bloc	Différence de gaussiennes sous différentes échelles de représentation [4].	
			différence de texture [71].	
			Information mutuelle [72].	
			modèle multi échelles basée sur l'entropie locale [74].	
		Modèle basé sur les phénomènes psycho-visuels [73].		
		Segment/Bloc	distance de $\chi^2$ entre les histogrammes de couleur et de texture [75,5,76].	
Méthodes basées sur l'estimation du contraste global	Probabiliste	Pixel	Mixture de gaussiennes [5,6].	
			Formalisme bayésien [77,78].	
	Graphique	Bloc	marches aléatoires markoviennes [79].	
		Segment	distance géodésique sur un graphe [57].	
	distance colorimétrique	Segment	distance spatiale et la distance entre histogrammes de couleur [62].	
Estimation de la saillance dans le domaine fréquentiel		Pixel	calcul du résidu spectral [7].	
			Différence entre Chaque pixel avec la moyenne de l'image [8].	
Méthodes hybrides	Probabiliste	Pixel	Analyse de la transformée en ondelettes [85].	
		Bloc	Apprentissage de codes de caractéristiques [70].	
		Superpixel	Approche basée sur les multi résolutions et les distributions normales [11].	
	Graphiques	Superpixel au niveau d'une seule résolution		Marches aléatoires markoviennes [54].
				Connectivité avec l'arrière plan [82].
				Ordonnancement sur un graphe mono-couche [9].
		Superpixel au niveau de plusieurs résolutions		Structure arborescente pour la détection de la saillance [10].
				Structure hiérarchique pour la détection de la saillance [18].
				Multi-graphes pour la détection

			d'objets saillants [15].
	Autres méthodes	Bloc	Décomposition matricielle de l'image: partie redondante et partie saillante [17].
			Estimation de la saillance basée sur le calcul de variance [91].
		Superpixel	Apprentissage de régression de type forêts aléatoire [16].
			Espace de haute dimension pour la détection de la saillance [90].
			Reconstruction dense et éparse d'erreur [86].

## 2.5 Conclusion

Nous avons présenté dans ce chapitre une taxonomie des différentes méthodes de détection d'objets saillants selon le contexte spatial pris en compte pour estimer le contraste dans l'image à traiter. Nous pouvons distinguer les méthodes basées sur le contraste local, les méthodes basées sur le contraste global et les méthodes hybrides basées sur l'estimation du contraste local et global à la fois.

Les méthodes basées sur l'estimation du contraste locale évaluent la saillance de chaque pixel/région par rapport à son voisinage dans un contexte spatial bien limité. Elles s'inspirent du modèle de fixation du regard de type centre/périphérie. Cependant ces méthodes tendent à retourner les contours comme les parties les plus saillantes de l'image et ne sont pas adaptés à retourner l'objet saillant dans sa globalité.

Les méthodes basées sur l'estimation du contraste global sont plus orientées pour une détection entière de l'objet saillant. Cependant, l'inconvénient de ces méthodes est qu'elles sont instables dans le cas de similarité colorimétrique entre l'arrière-plan et l'objet saillant et dans le cas de complexité et hétérogénéité de l'arrière-plan. En outre, ces méthodes sont dépendantes de la surface occupée par l'objet saillant et ne peuvent détecter les régions saillantes occupant une surface majoritaire dans l'image.

Comme solution aux inconvénients des méthodes basées sur l'estimation du contraste locale, plusieurs méthodes se sont intéressées à combiner l'information local et global pour estimer la saillance de chaque région/pixel de l'image. Nous distinguons dans ces méthodes les méthodes probabilistes, les

méthodes graphiques et les autres méthodes. Les méthodes probabilistes utilisent une théorie probabiliste ou statistique pour détecter la rareté d'un élément dans l'image. Les méthodes graphiques se basent sur la définition d'une ou plusieurs résolutions pour la segmentation de l'image en régions sur lesquelles est construit le graphe. L'utilisation de plusieurs résolutions offrent un meilleur avantage car l'échelle de la surface occupée par l'objet saillant est inconnue à priori. Les autres méthodes se basent sur différentes théorie mathématiques comme la variance, l'ACP, la décomposition matricielle, apprentissage de régression, projection dans des espaces de caractéristiques, etc.

D'une manière générale, les méthodes basées sur la segmentation en superpixels conservent mieux les contours des objets saillants par rapport aux méthodes basées sur les blocs et les méthodes basées sur les multi-résolutions sont plus stable à la variation d'échelle de la surface des objets sur l'image.

## CHAPITRE 3

### ORDONNANCEMENT SUR UN GRAPHE MULTI-COUCHES POUR LA DETECTION D'OBJETS SAILLANTS

#### 3.1 Introduction

Nous présentons dans ce chapitre notre méthode de détection d'objets saillants basée sur un ordonnancement sur un graphe multi-couches que nous notons *MLR* (multi layered graph ranking) [123,124]. Comme mentionné dans la section 2.4.1, le passage vers les approches de détection d'objets saillants basées sur les multi-résolutions permet d'améliorer considérablement les résultats d'estimation de la saillance et corrige les défauts liés au traitement d'une seule résolution. Dans ce contexte, plusieurs méthodes basées sur les graphes multi-couches ont été proposés dans le passé. Cependant, la façon dont notre modèle est construit et la manière dont les différentes résolutions sont combinées dans notre travail est très différente sur plusieurs aspects.

Nous rappelons que dans la méthode EQCUT [15], une carte de saillance est estimée à partir de chaque résolution, puis une fusion basée sur la moyenne de ces cartes est effectuée pour générer une carte de saillance finale. Dans les méthodes hiérarchiques [10,18], un graphe est construit dont les liens non pondérés sont définis exclusivement sur la direction verticale. Un nœud n'est connecté qu'aux nœuds de la couche supérieure ayant une surface qui l'englobe. Dans [18], les nœuds sont non orientés et le processus de diffusion se fait dans les directions haut-bas et bas-haut. Dans [10], les poids sont orientés du bas vers le haut. Une résolution n'est couplée alors qu'avec la résolution de niveau supérieur ou inférieure uniquement et aucun lien n'existe au sein d'une même résolution.

Dans le cadre de notre approche, nous définissons un graphe multi-couches compacte et pleinement connecté utilisé à la fois dans une seule fonction objective pour estimer la saillance. Nous définissons un couplage fort dans notre

structure graphique en connectant chaque nœud avec ses voisins situés sur la même résolution ainsi que ses voisins sur toutes les autres résolutions. Nous avons utilisé dans le processus de diffusion de l'ordonnancement, l'information régionale, spatiale et de gradient pour renforcer l'efficacité d'estimation de la saillance. Cette démarche permet d'obtenir des cartes de saillance plus cohérentes et lisses à travers les différentes résolutions. Nous détaillerons dans les sections suivantes les différentes étapes de notre démarche.

### 3.2 Segmentation en superpixels

Un superpixel correspond à un regroupement de pixels homogène en une région compacte. La segmentation en superpixels capture la redondance dans l'image et réduit considérablement la complexité des tâches de traitement d'image. Elle est caractérisée par une granularité adaptée au traitement effectué et à la taille de l'image considérée et possède la particularité de s'ajuster d'une manière adéquate sur les contours des objets.

#### 3.2.1 Interêt de la segmentation en superpixels

Dans la vision par ordinateur, la segmentation d'image se réfère au processus de partitionnement d'une image en groupes ou clusters de pixels spatialement connectés qui partagent des caractéristiques similaires, comme l'intensité, la couleur et la texture. Chacun des clusters est appelé un segment de l'image. Pour cette raison, le clustering et la segmentation sont souvent utilisés comme synonymes. L'objectif de la segmentation est généralement de simplifier la représentation d'une image sous une forme plus significative et facile à analyser. La plupart des problèmes de vision par ordinateur, comme la reconnaissance d'objets [92] et la détection de la saillance visuelle [3,14,50] nécessitent souvent une abstraction de l'image sous forme de segmentation puisque les pixels sont à un niveau de granularité trop fin difficile à utiliser comme unité de traitement. Les algorithmes de segmentation d'image populaires tentent de définir des regroupements de pixels sous formes de segments sémantiquement significatifs. Cependant, le problème majeur est la difficulté de maîtriser la taille des segments conformément à la sémantique souhaitée. Souvent, ces algorithmes génèrent soit un nombre trop restreint de segments (désigné par le terme under-segmentation)

soit un nombre très important (désignée par le terme *over-segmentation*) dépourvus de sémantique. Une manière pratique d'éviter les limitations de la segmentation liée à la sémantique est d'effectuer d'abord une *over-segmentation* à partir de laquelle les segments similaires sont regroupés afin de former des unités sémantiquement significatives. Les segments issus de l'*over-segmentation* sont désignés sous le terme de *superpixels*. Comme la segmentation en *superpixels* représente une étape préliminaire importante pour beaucoup d'autres disciplines liées à la vision par ordinateur, il est convenu que, pour une bonne segmentation, un modèle devrait satisfaire les critères suivants: 1) l'adhérence stricte aux frontières des objets, 2) contrôle sur la granularité des *superpixels*, 3) contrôle sur la taille de chaque *superpixel*, 4) Uniformité de chaque région délimitée par un *superpixel*, 5) complexité temporelle de calcul raisonnable et 6) régularité de la forme de chaque *superpixel*. La majorité des modèles de segmentation de *superpixels* ne répondent pas pleinement à ces critères [61]. Ils souffrent souvent d'une complexité de calcul élevée, qualité de segmentation médiocre et taille des segments difficile à contrôler.

Dans le cadre de notre travail, nous avons opté pour l'algorithme SLIC (*simple linear iterative clustering*) [61] qui répond favorablement à tous les critères requis pour une bonne segmentation d'une manière simple, efficace et produisant des segments homogènes et globalement de forme et de taille uniformes, compacts avec un contrôle sur la granularité des *superpixels* et caractérisé par une basse complexité temporelle et une adhérence remarquable des frontières des segments sur les contours des objets (cf. figure 3.1). En outre, Il a été montré dans [61], que cet algorithme surpasse la majorité des autres algorithmes de l'état de l'art en efficacité et qualité de segmentation. Nous notons que la totalité des méthodes de détection d'objets qui utilisent la segmentation en *superpixels* optent pour la méthode SLIC pour sa simplicité, sa rapidité et son efficacité. Nous allons nous intéresser dans la section suivante à la description de cette méthode.

### 3.2.2 La méthode SLIC

L'approche SLIC [61] génère une segmentation en *superpixels* en effectuant un clustering basé sur la similarité colorimétrique et la proximité des pixels de l'image calculées comme une mesure de distance. Pour cela, Les

auteurs utilisent un vecteur de caractéristiques  $[l \ a \ b \ x \ y]^T$ , où  $[l \ a \ b]^T$  correspond aux valeurs de couleur dans l'espace colorimétrique CIE LAB et  $[x \ y]^T$  à la position spatiale dans l'image. Bien que la distance maximale possible entre deux couleurs soit limitée, la distance spatiale est tributaire de la taille de l'image. Il convient alors d'effectuer une normalisation de chaque distance avant d'être utilisées conjointement.



Figure 3.1 Images segmentées en superpixels de taille approximative de 64, 256 et 1014 pixels en utilisant l'algorithme SLIC [61].

Afin de regrouper les pixels dans cet espace 5D, les auteurs introduisent une nouvelle mesure de distance entre deux pixels  $i$  et  $j$  qui tient compte de la taille des superpixels calculée comme suit:

$$D_s(i, j) = d_{lab}(i, j) + \frac{m}{S} d_{xy}(i, j) \quad (3.1)$$

avec  $d_{lab}$  la distance euclidienne colorimétrique dans l'espace CIE LAB,  $d_{xy}$  la distance euclidienne spatiale,  $S$  la distance spatiale approximative entre chaque paire de superpixels voisins et  $m$  une constante qui contrôle la compacité des segments. Plus la valeur de  $m$  est grande, plus la distance spatiale est privilégiée et plus le cluster est compact. Cette valeur est ajustée d'une manière empirique à la valeur 10. L'algorithme SLIC prend en entrée le nombre de superpixels fixé préalablement à une constante  $K$ . Pour une image avec  $N$  pixels, la taille approximative d'un superpixel est donc  $N/K$ . Pour une segmentation approximativement uniforme en taille, la distance entre chaque paire de centroides représentant des superpixels adjacents est de l'ordre de  $S = \sqrt{N/K}$ . Chaque cluster  $k \in \{1, 2, \dots, K\}$  est représenté par son centroïde  $C_k = [l_k, a_k, b_k, x_k, y_k]^T$ . Les

centroïdes sont initialisés sur une grille régulière d'intervalles  $S$ . Étant donné que la surface approximative d'un superpixel est de l'ordre de  $S^2$ , les pixels susceptibles d'être associés à un cluster sont pris dans un intervalle spatial de  $Y = 2S \times 2S$  autour du centroïde. Chaque pixel est associé au cluster ayant le centroïde qui lui est le plus proche en terme de distance  $D_S$ . Afin d'éviter que les centroïdes soient positionnés sur les contours, un réajustement itératif de ces positions est effectué entraînant une mise à jour du clustering adopté. Chaque centroïde est déplacé vers une position ayant la plus faible valeur du gradient dans l'espace de voisinage  $3 \times 3$ . Le gradient de l'image est calculé comme suit:

$$G(x, y) = \|I(x + 1, y) - I(x - 1, y)\|_2^2 + \|I(x, y + 1) - I(x, y - 1)\|_2^2 \quad (3.2)$$

avec  $I(x, y)$  le vecteur de couleur *lab* du pixel positionné sur  $(x, y)$  et  $\|\cdot\|_2$  la norme euclidienne  $L_2$ . Après que chaque pixel soit associé au centroïde qui lui est le plus proche dans l'intervalle  $Y$ , un nouveau centroïde est calculé dans chaque cluster. Il est considéré comme le vecteur moyen  $[l \ a \ b \ x \ y]$  de tous les pixels contenus dans le cluster. Le processus d'association de pixels aux centroïdes et la réestimation des centroïdes au sein des nouveaux clusters produits est réitéré jusqu'à l'atteinte d'une convergence. Comme cet algorithme n'utilise pas de mesure de connectivité entre clusters, des petits segments ou pixels disjoints peuvent se trouver au voisinage d'un segment plus grands et qui leurs sont similaires. Afin de renforcer la connectivité dans le modèle, les segments disjoints sont incorporés dans le segment voisin le plus important en surface.

### 3.3 Extraction de caractéristiques

Les caractéristiques visuelles représentent les propriétés mesurables d'un phénomène physique observé au sein d'une scène visuelle. Cela consiste à définir des transformations mathématiques calculées sur les pixels d'une image numérique. Les caractéristiques visuelles permettent généralement de mieux rendre compte de certaines propriétés visuelles de l'image. Nous distinguons usuellement<sup>1</sup> les caractéristiques globales qui sont calculées sur toute l'image (comme la couleur) et les caractéristiques locales qui sont calculées au niveau du voisinage de chaque pixels (comme la texture). Nous allons présenter présenter



dans cette section les différentes caractéristiques utilisées au sein de notre approche.

### 3.3.1 La couleur

La couleur est une caractéristique importante associée aux objets. Elle fournit une information utile pour ces objets afin de faciliter leur localisation, identification et leur interprétation. On identifie aisément son équipe favorite lors d'un match de football en se basant sur la couleur de la tenue et on reconnaît facilement un fruit mur apte à la consommation en examinant sa couleur.

La couleur est un phénomène psycho-physiologique provoqué par l'excitation de photorécepteurs situés sur la rétine de l'œil par une onde électromagnétique [93]. Nous avons choisis d'utiliser l'espace de couleur *RGB* pour caractériser les différentes régions d'une image. Cependant, La couleur n'est pas suffisante à elle seule pour détecter les objets efficacement. En effet, si les différentes parties de l'objet sont soumises à différentes conditions d'éclairage, la couleur serait affectée considérablement. Cependant, la variation d'illumination n'a aucun effet sur la structure spatiale de couleur et d'intensité des différentes régions. Pour cette raison, nous nous sommes intéressés à utiliser la texture comme caractéristique complémentaire à la couleur.

### 3.3.2 La texture

La texture est une caractéristique importante en vision par ordinateur. Elle décrit la structure inhérente à un objet ou région d'une image sous forme d'un arrangement spatial de couleurs ou d'intensités. Elle peut correspondre par exemple à un relief qui se répète dans l'image. Il existe plusieurs mesures de texture. Parmi les plus utilisées, en vision artificielle, nous notons l'entropie [85] et *LBP (local binary pattern)* [94]. Nous utiliserons la caractéristique *LBP* afin de mesurer les différences de luminance entre un pixel et ses voisins dans chaque canal de couleur utilisé. Nous utiliserons la caractéristique d'entropie afin de mesurer le niveau de désordre au voisinage d'un pixel. Plus l'intensité de luminance est diversifiée autour d'un pixel, plus ce pixel possède une entropie élevée et plus son voisinage est uniforme en intensité, plus l'entropie converge vers 0. Nous extrayons ensuite les histogrammes issus de l'entropie et *LBP* pour chaque superpixel et l'utiliserons pour estimer son contraste.

### 3.3.3 Les contours

La détection de contours est une caractéristique importante et largement étudiée dans la vision par ordinateur [95,96,97]. Un contour se matérialise par une rupture et un changement brutal d'intensité lumineuse dans l'image suivant une direction donnée. Elle permet de réduire de manière significative la quantité de données et élimine les informations jugées moins pertinentes, tout en préservant les propriétés structurelles importantes de l'image. Elle est utilisée comme une étape préliminaire dans plusieurs domaines comme la reconnaissance d'objets [98,99], la segmentation d'image [100,101] et les contours actifs [102], etc. Une bonne détection de contour doit alors identifier efficacement les bords des objets et des segments de l'image avec un temps de calcul réduit. La méthode qui respecte le mieux ces deux critères est celle de [97] où les auteurs proposent une méthode basée sur l'entraînement des forêts aléatoires décisionnelles pour prédire la position de contours. Les résultats sont retournés en temps réel et les contours estimés se positionnent efficacement sur les contours des objets.

## 3.4 Graphe multi-couches pour la détection d'objets saillants

Nous proposons dans cette section une méthode d'ordonnement graphique pour la détection des objets saillants dont le principe est le suivant: étant donné un nœud requête sur le graphe, les autres nœuds sont évalués par une mesure d'ordonnement selon leur pertinence à cette requête. L'objectif est d'effectuer l'apprentissage d'une fonction objective qui définit la pertinence entre les nœuds non étiquetés et les requêtes.

### 3.4.1 Construction du graphe

#### 3.4.1.1 Structure du graphe

Nous utilisons la segmentation en superpixels présentée dans la section 3.2.2 pour définir les nœuds de notre graphe. Utiliser une seule granularité ne peut s'accommoder à la diversité des tailles des objets saillants. Dans le cas de l'ordonnement sur un graphe, une forte granularité tend à retourner des régions plus étendues générant un taux de faux positifs important et inversement une basse granularité tend à retourner des régions saillantes réduites générant un taux de faux négatifs important. Un exemple d'utilisation d'une seule granularité

est montré sur les figures 3.6 et 3.7. Nous avons défini un modèle basé sur une représentation compacte d'un graphe multi couches  $G = (V, E)$  où chaque couche correspond à une résolution.

Les couches sont organisées de haut en bas de la résolution avec la granularité la plus basse à la résolution ayant la granularité la plus haute. Chaque nœud est relié par un poids à ses voisins *directs* se trouvant sur la même résolution aussi bien qu'avec ses voisins indirects se trouvant sur les autres résolutions. Deux nœuds sont voisins indirects s'ils sont situés sur deux résolutions différentes et s'ils partagent une frontière commune ou une surface commune dans l'image. Toute cette structure est utilisée dans l'optimisation d'une fonction objective pour l'estimation de la saillance. Contrairement à certaines méthodes basées sur plusieurs résolutions, toutes les couches contribuent à la fois dans la détection d'objets saillants.

Nous proposons d'utiliser  $L$  décompositions de l'image en différentes résolutions  $\{\Omega_1, \Omega_2, \dots, \Omega_L\}$ . Chaque décomposition  $\Omega_\ell$ ,  $\ell \in \{1, 2, \dots, L\}$  est générée en utilisant l'algorithme SLIC qui produit  $n_\ell$  superpixels  $\{r_1^{(\ell)}, r_2^{(\ell)}, \dots, r_{n_\ell}^{(\ell)}\}$ . Afin d'obtenir plusieurs résolutions, nous avons utilisé différents nombres de superpixels dans la segmentation  $n_1 < n_2 < \dots < n_L$ . Nous agencions ces résolutions sur une nouvelle structure graphique multi-couches organisée de haut en bas selon un ordre croissant du nombre de superpixels correspondant à chaque couche comme illustré sur la figure 3.2. Chaque nœud du graphe est relié avec tous ses voisins directs et indirects avec des liens pondérés. Nous allons présenter dans la section suivante l'estimation du poids d'un lien du graphe en fonction des distances spatiale, de couleur et de texture entre les nœuds qu'il raccorde.

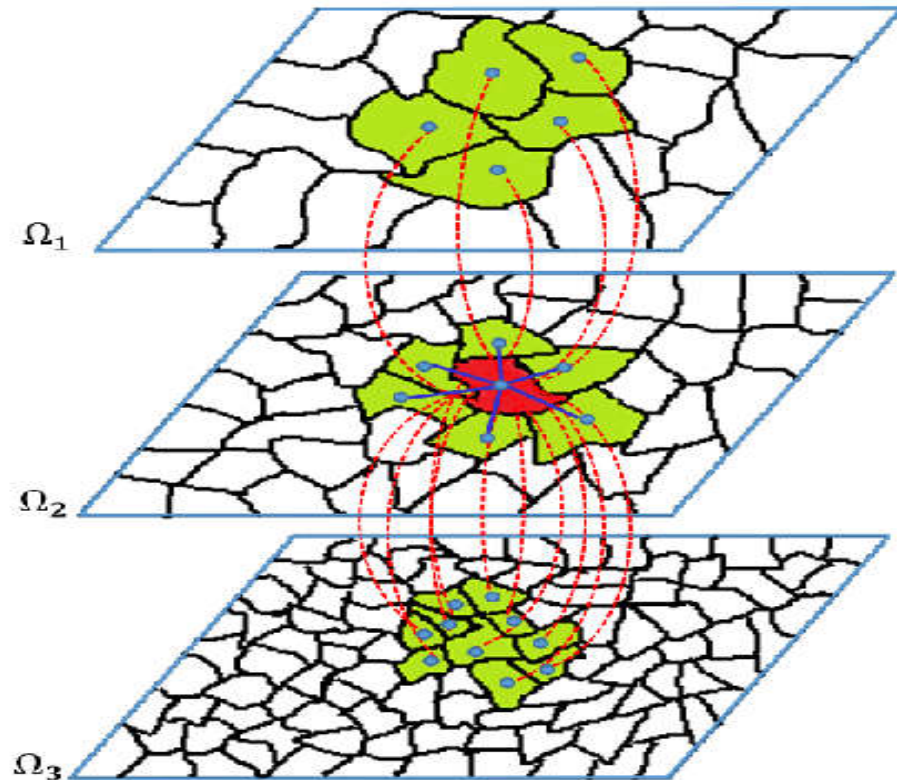


Figure 3.2 Illustration de notre graphe multi-couches avec trois résolutions superposées pour la détection d'objets saillants.

#### 3.4.1.2 Ajustement des poids du graphe

L'estimation des poids de notre graphe multi-couches est essentielle pour l'établissement de l'ordonnement de tous les nœuds comme il sera montré dans la section 3.4.2. Une bonne estimation de ces poids est donc nécessaire pour assurer une bonne diffusion de la saillance depuis un nœud à son voisin. La majorité des méthodes de l'état de l'art ont utilisé uniquement l'information de couleur pour la détection d'objets saillants [50]. Cependant, la texture peut contribuer favorablement là où la couleur subit une altération au changement d'illumination ou l'effet de l'ombrage. En outre, les objets (ou arrière-plan) à motifs fortement récurrents sont identifiables efficacement grâce à leurs texture (cf. figure 3.3).

Une autre caractéristique peu étudiée dans la détection d'objets saillants est le contour. En effet, les objets possèdent une forte densité de gradient au niveau de leurs contours. L'idée d'utiliser le gradient est d'atténuer le processus de diffusion de la saillance voir l'arrêter à proximité des bords des objets. Cela permet de mieux ressortir l'objet saillant du fond et de mieux mettre en évidence ses

contours. Dans le cas de notre étude, afin d'obtenir une meilleure description des objets saillants, nous combinons l'information régionale, spatiale et de gradient pour définir les poids relatifs aux liens du graphe. Pour l'information régionale, nous utilisons pour chaque canal de couleur  $k$ ,  $k \in \{R, G, B\}$  les caractéristiques de couleur et de texture. Pour la couleur, nous considérons la valeur moyenne de chaque superpixel  $i$  au niveau de la résolution  $\ell$ ,  $C_{i,k}^{(\ell)}$  après normalisation des valeurs de couleur dans l'intervalle  $[0,1]$ . Pour la texture, nous considérons les caractéristiques dérivées des histogrammes du modèle à motifs binaires locaux LBP (*Local Binary Pattern*) [94] et l'entropie [103] de couleur notés par  $l_{i,k}^{(\ell)}$  et  $e_{i,k}^{(\ell)}$  respectivement. La figure 3.3 montre l'avantage de combiner la couleur et la texture dans notre méthode pour la détection d'objets saillants.

Pour l'information du gradient, nous estimons la discontinuité de couleur pour chaque superpixel  $i$  en utilisant la méthode [97]. Nous incluons l'information spatiale dans notre méthode afin de renforcer la contiguïté entre les nœuds. Plus les centres de gravité de deux superpixels sont proches, plus les nœuds leur correspondant sont similaires et plus le poids les reliant est faible.

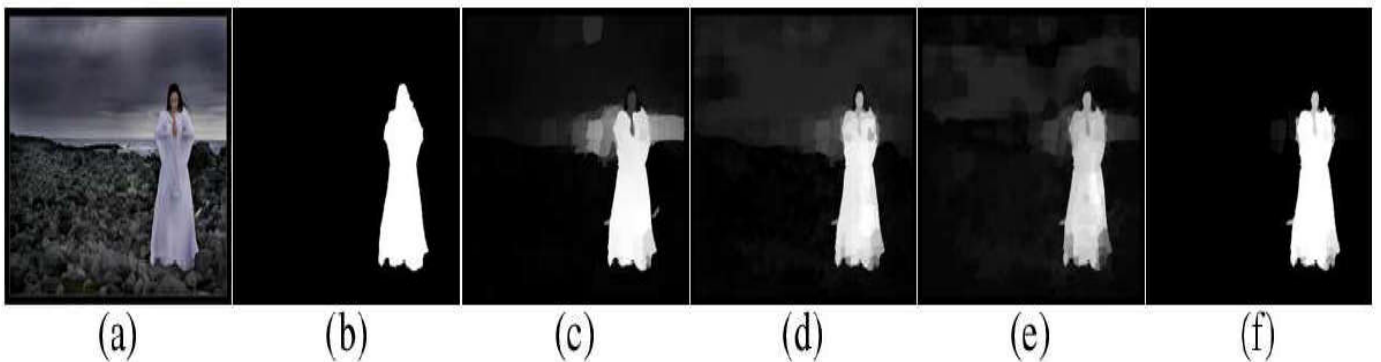


Figure 3.3 Comparaison des cartes de saillance retournées par notre méthode d'ordonnancement sur un graphe multi-couches en considérant (c) la couleur, (d) LBP, (e) l'entropie et (f) Leur combinaison respectivement. (a) et (b) représentent l'image en couleur et la vérité terrain respectivement.

Nous pouvons observer ce phénomène aisément dans le cas d'une seule résolution. Plus la granularité de la segmentation en superpixels est grande, plus les distributions de texture et de couleur sont similaires d'un nœud à son voisin, plus la distance qui sépare leurs centres de gravité est petite et plus la valeur de saillance est similaire. Pour un superpixel  $i$ , l'information spatiale correspond aux

coordonnées géométriques du centre de gravité noté  $g_i^{(\ell)}$ . Nous notons que cette valeur est aussi normalisée dans l'intervalle  $[0, 1]$ .

En combinant toutes ces caractéristiques, nous proposons la formule suivante pour le calcul de poids entre les superpixels  $r_i^{(\ell)}$  et  $r_j^{(m)}$  dans le graphe multi-couches:

$$w_{ij}^{(\ell,m)} = \exp \left( -\gamma \sum_{k=1}^3 \left( \varepsilon_{c,k} \cdot |c_{i,k}^{(\ell)} - c_{j,k}^{(m)}| + \varepsilon_{l,k} \cdot d_{\mathbb{B}}(l_{i,k}^{(\ell)} - l_{j,k}^{(m)}) + \varepsilon_{e,k} \cdot d_{\mathbb{B}}(e_{i,k}^{(\ell)} - e_{j,k}^{(m)}) + \varepsilon_g \cdot \|g_i^{(\ell)} - g_j^{(m)}\| \right) \right) \quad (3.3)$$

avec  $\varepsilon_{c,k}$ ,  $\varepsilon_{l,k}$ ,  $\varepsilon_{e,k}$  et  $\varepsilon_g$  sont des poids qui contrôlent la contribution des différentes caractéristiques. Nous notons que contrairement à l'entropie et *LBP*, nous n'avons pas utilisé la différence entre les histogrammes pour les caractéristiques de couleur. Ceci est dû au fait que l'utilisation de la différence d'histogramme s'avère bien plus précise que la distance entre les moyennes des superpixels. Ainsi la distance entre les histogrammes renvoie une valeur plus importante que la distance entre les moyennes des superpixels. Dans le cas de notre étude, une telle précision peut considérer deux couleurs perceptuellement similaires comme assez distantes et le résultat de saillance est altéré. Les poids correspondant aux caractéristiques de couleur et de texture sont estimés en analysant leur pouvoir discriminatif uniquement durant la phase de raffinement de la saillance où l'extraction du noyau de l'objet est nécessaire. Nous allons présenter ce processus dans le chapitre suivant. Les poids des différentes caractéristiques sont fixés à des constantes dans cette première étape de notre algorithme. Nous avons pondéré la couleur avec des valeurs plus importantes que pour la texture du moment que la couleur est généralement plus fiable que la texture dans la détection d'objets saillants. Nous avons donc choisi de pondérer empiriquement la couleur avec une valeur de 0.8 et chaque composante de la texture à 0.2. Nous avons accordé un poids de 0.02 pour la distance spatiale entre nœuds. Plus deux nœuds sont proches et similaire en couleur et en texture, plus

le poids les raccordant est élevé et moins la valeur de saillance se propage de l'un à l'autre.

Le paramètre  $\gamma$  quant à lui contrôle la contribution de l'information de contour dans la phase du processus de raffinement de la saillance comme cela va être montré dans la section 4.3. Nous notons qu'il peut exister une forte densité de contours en dehors de l'objet saillant dans des parties de l'arrière-plan qui peut affecter considérablement l'estimation de la saillance (cf. figure 3.4). Nous choisissons alors d'inhiber la contribution de l'information du contour en ajustant la valeur de  $\gamma$  à 1 dans cette première étape de notre algorithme. La fonction  $d_{\mathbb{B}}$  dénote la distance de Bhattacharyya entre histogrammes [104]. Elle est définie pour les caractéristiques de *LBP* et d'entropie respectivement par les formules suivantes:

$$d_{\mathbb{B}}(l_{i,k}^{(\ell)}, l_{j,k}^{(m)}) = 1 - \sum_{u=0}^{p-1} \sqrt{l_{i,k}^{(\ell)}(u) - l_{j,k}^{(m)}(u)} \quad (3.4)$$

$$d_{\mathbb{B}}(e_{i,k}^{(\ell)}, e_{j,k}^{(m)}) = 1 - \sum_{u=0}^{p-1} \sqrt{e_{i,k}^{(\ell)}(u) - e_{j,k}^{(m)}(u)} \quad (3.5)$$

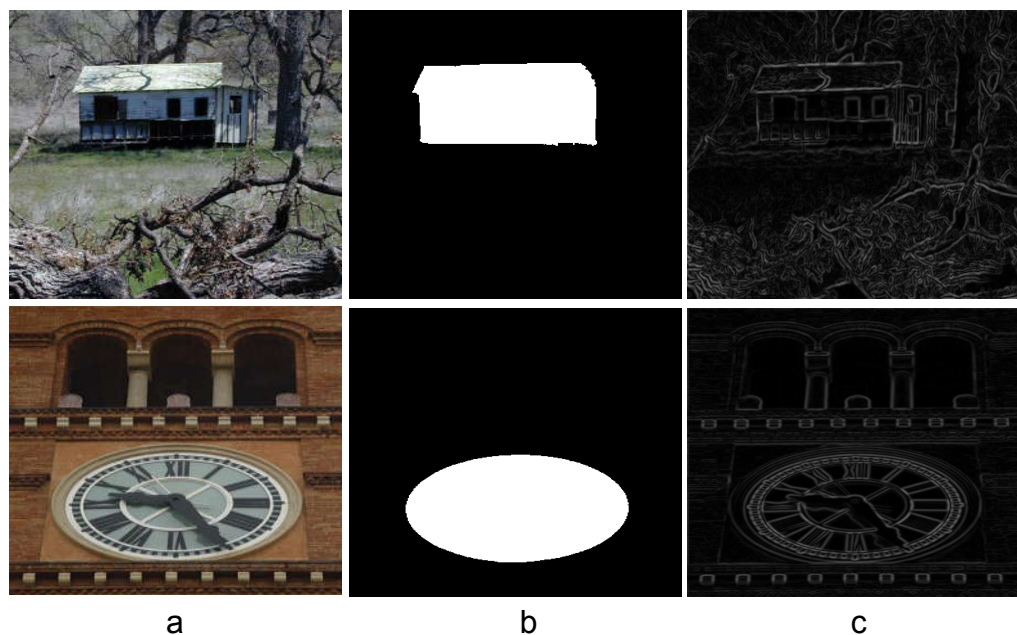


Figure 3.4 Exemples d'estimation de contour. (a) l'image en couleur, (b) la vérité terrain et (c) la carte de contour correspondante.

avec  $p$  le nombre de bins de l'histogramme et  $u, (u \in \{0, 1, \dots, p-1\})$  un indice représentant un bin de l'histogramme. L'entropie au sein de la résolution ( $\ell$ ) est calculée au voisinage de chaque pixel  $\varrho(x, y)$  appartenant au superpixel  $i$  comme suit:

$$e_{i,k}^{(\ell)}(\varrho) = - \sum_{u=0}^{p-1} H_k(u) \cdot \log_2(H_k(u)) \quad (3.6)$$

avec  $H_k$  l'histogramme local normalisé correspondant au canal de couleur  $k$  dans le voisinage  $v \times v$  autour du pixel  $\varrho$  et  $p$  le nombre de bins considérés. Dans notre cas, nous fixons  $v$  à 9 et  $p$  à 32. La caractéristique *LBP* au sein de la résolution ( $\ell$ ) est calculée au voisinage de chaque pixel  $\varrho(x, y)$  appartenant au superpixel  $i$  comme suit:

$$l_{i,k}^{(\ell)}(\varrho) = \sum_{\tau=1}^{\Gamma} s(g_{\tau} - g_{\varrho}) \cdot 2^{\tau-1} \quad (3.7)$$

où  $\Gamma$  le nombre de voisins entourant  $\varrho$  fixé empiriquement à 8,  $g_{\tau}$  la valeur du pixel  $\tau$  dans le canal de couleur  $k$  et  $s$  une fonction indicielle définie comme suit:

$$s(g_{\tau} - g_{\varrho}) = \begin{cases} 1 & \text{si } g_{\tau} - g_{\varrho} > 0 \\ 0 & \text{sinon} \end{cases} \quad (3.8)$$

### 3.4.2 Ordonnement sur notre graphe multi-couches

Soit  $f^{\ell} = [f_1^{(\ell)}, f_2^{(\ell)}, \dots, f_{n_{\ell}}^{(\ell)}]$  l'ordonnement des nœuds au niveau de la couche  $\ell$  et  $y = [y_1^{(\ell)}, y_2^{(\ell)}, \dots, y_{n_{\ell}}^{(\ell)}]$  un vecteur indiciel dans lequel  $y_1^{(\ell)} = 1$  si la région  $r_i^{(\ell)}$  est une requête sur la couche  $\Omega_{\ell}$  et  $y_1^{(\ell)} = 0$  sinon. Nous proposons la fonction objective suivante qui généralise l'équation 2.48 à un modèle d'ordonnement sur un graphe multi-couches:



$$\begin{aligned}
f^* = \arg \min_f & \left( \sum_{\ell=1}^L \left( \sum_{i,j=1}^{n_\ell} w_{ij}^{(\ell)} \left( f_i^{(\ell)} / \sqrt{d_i^{(\ell)}} - f_j^{(\ell)} / \sqrt{d_j^{(\ell)}} \right)^2 \right. \right. \\
& + \sum_{m=1, m \neq \ell}^L \left( \sum_{i=1}^{n_\ell} \sum_{j=1}^{n_m} w_{ij}^{(\ell,m)} \left( f_i^{(\ell)} / \sqrt{\tilde{d}_i^{(\ell)}} - f_j^{(m)} / \sqrt{\tilde{d}_j^{(m)}} \right)^2 \right) \\
& \left. \left. + \lambda \sum_{i=1}^{n_\ell} \left( f_i^{(\ell)} - y_i^{(\ell)} \right) \right) \right) \quad (3.9)
\end{aligned}$$

avec  $f = [f^1, f^2, \dots, f^L]^T$ ,  $w_{ik}^\ell$  le poids reliant les nœuds  $r_i^\ell$  et  $r_k^\ell$  appartenant à la même couche et  $w_{ik}^{\ell,m}$  le poids reliant les nœuds  $r_i^\ell$  et  $r_k^m$  appartenant aux couches  $\Omega_\ell$  et  $\Omega_m$  respectivement. Le terme  $d_i^{(\ell)}$  correspond à la somme des poids reliant le nœud  $r_i^\ell$  à tous ses voisins directs situés sur la même résolution et le terme  $\tilde{d}_i^{(\ell)}$  correspond à la somme des poids reliant le même nœud à ses voisins indirects situés sur toutes les autres résolutions. Ces deux termes sont calculés comme suit:

$$d_i^{(\ell)} = \sum_{k=1}^{n_\ell} w_{ik}^{(\ell)} \quad (3.10)$$

$$\tilde{d}_i^{(\ell)} = \sum_{m=1, m \neq \ell}^L \sum_{k=1}^{n_m} w_{ik}^{(\ell,m)} \quad (3.11)$$

Le paramètre  $\lambda$  est une constante de régularisation qui contrôle l'équilibre entre la contrainte de lissage qui correspond au premier et second terme et la contrainte d'ajustement (fitting) de la diffusion qui correspond au troisième terme. Le premier terme encode les interactions entre les nœuds appartenant à une même couche de notre structure et le second terme encode les interactions entre les différentes couches (nœuds connectés par les liens verticaux dans la figure 3.2). Ces deux termes assurent une dégradation lisse de la valeur de saillance entre les nœuds voisins (directs et indirects) lors du processus de diffusion dans le graphe. Autrement dit, la valeur de l'ordonnancement ne doit pas

changer drastiquement d'un nœud à son voisin. Le troisième terme renforce la préservation des valeurs des nœuds requêtes. La minimisation de la fonction 3.9 peut être résolue de la même manière que pour l'équation 2.48. D'abord, nous définissons les quantités suivantes:

$$\left\{ \begin{array}{l} D_i^{(\ell)} = d_i^{(\ell)} + \tilde{d}_i^{(\ell)} \\ W^\ell = [w_{ij}^{(\ell)}]; \quad i, j \in \{1, \dots, n_\ell\} \\ W^{(\ell, m)} = [w_{ij}^{(\ell, m)}]; \quad i \in \{1, \dots, n_\ell\}, \quad j \in \{1, \dots, n_m\} \end{array} \right. \quad (3.12)$$

avec  $W^\ell$  une matrice de dimension  $n_\ell \times n_\ell$  contenant les poids reliant les nœuds dans une même couche ( $\ell$ ) du graphe et  $W^{(\ell, m)}$  une matrice de dimension  $n_\ell \times n_m$  contenant les poids reliant les différentes couches du graphe (liens verticaux). Ces deux matrices sont définies dans une matrice plus générale  $W$  comme suit:

$$W = \begin{bmatrix} W^{(1)} & W^{(1,2)} & \dots & W^{(1,L)} \\ W^{(2,1)} & W^{(2)} & \dots & W^{(2,L)} \\ \vdots & \vdots & \ddots & \vdots \\ W^{(L,1)} & W^{(L,2)} & \dots & W^{(L)} \end{bmatrix} \quad (3.13)$$

$D_i^{(\ell)}$  correspond à la somme des poids du nœud  $i$  avec tous ces voisins directs et indirects. Nous définissons la matrice diagonale (appelée matrice des degrés) suivante qui contient la somme des poids de chaque nœud avec ses voisins:

$$D = \text{diag}[D_1^{(1)}, \dots, D_{n_1}^{(1)}, \dots, \dots, D_1^{(L)}, \dots, D_{n_L}^{(L)}] \quad (3.14)$$

Nous notons que les éléments se trouvant sur la diagonale de chaque sous-matrice  $W^{(\ell)}$ ,  $\ell \in \{1, 2, \dots, L\}$  sont ajustés à 0 du moment qu'un nœud ne contribue jamais à son propre ordonnancement. En outre, les positions qui correspondent aux liens entre les nœuds qui ne sont pas voisins sont tous mis à 0 du moment qu'un nœud ne contribue jamais à l'ordonnancement des nœuds qui

ne lui sont pas voisins. Comme proposée dans [83], une fonction d'ordonnement peut être résolue par la fonction suivante:

$$f^* = (I - \alpha L)^{-1}Y \quad (3.15)$$

avec  $\alpha = 1/(1 + \lambda)$ ,  $I$  la matrice d'identité et  $L$  la matrice de Laplace définie comme suit:

$$L = D^{-1/2}WD^{1/2} \quad (3.16)$$

Nous notons que le vecteur  $Y = [Y^{(1)}, Y^{(2)}, \dots, Y^{(L)}]^T$  retourne les valeurs de saillance initiales dans toutes les résolutions (désignation des nœuds requêtes).

Pour de meilleures performances, nous considérons deux autres versions d'ordonnement basées sur la matrice de Laplace. La première est non normalisée. Elle est représentée comme suit:

$$f^* = (D - \alpha W)^{-1}Y \quad (3.17)$$

La seconde est normalisée par  $D^{-1}$  et est représentée comme suit:

$$f^* = (I - \alpha D^{-1}W)^{-1}Y \quad (3.18)$$

Nous avons comparé les résultats d'estimation de la saillance retournées en utilisant les équations (3.15), (3.17) et (3.18). Nous avons constaté que les résultats retournés par les équations (3.17) et (3.18) sont similaires et que les résultats retournés par l'équation (3.15) sont les moins performants. Nous avons choisi d'utiliser la fonction (3.18) pour sa formulation plus adaptée à contrôler le processus de diffusion à partir des nœuds requêtes grâce au paramètre  $\alpha$  avec  $\alpha \in [0, 1]$ . Plus la valeur de  $\alpha$  converge vers 1, plus la contribution de la diffusion est importante et plus sa valeur converge vers 0, plus le résultat de la saillance conserve les valeurs des nœuds requêtes, moins donc la contribution de la diffusion est importante dans l'estimation de la saillance. Nous allons montrer dans la section 4.5 la manière dont nous pouvons varier les valeurs de  $\alpha$  dans un

processus de raffinement itératif de l'estimation de la saillance. Dans ce chapitre, nous nous contentons de fixer d'une manière empirique la valeur de  $\alpha$  à la constante 0.98.

### 3.4.3 Etapes de détection de la saillance

Nous détaillons dans cette section le processus de détection de la saillance basée sur l'ordonnancement sur graphe en deux étapes. D'abord, nous effectuons un ordonnancement en utilisant des requêtes extraites de l'arrière-plan sur lequel nous estimons le noyau de l'objet saillant. Ensuite, nous effectuons un deuxième ordonnancement en utilisant des requêtes extraites du noyau de l'objet.

#### 3.4.3.1 Ordonnancement avec des requêtes extraites de l'arrière-plan

Comme la majorité des méthodes de détection d'objets saillants basées sur les graphes, nous utilisons l'information à priori sur l'arrière-plan dans notre méthode. Nous initialisons les nœuds du graphe qui se situent sur les bords de l'image comme requête et nous calculons l'ordonnancement des autres nœuds selon leur pertinence à cette requête. Nous adoptons une approche séparation/combinaison (SC) pour propager les valeurs d'ordonnancement à partir des requêtes vers tous les autres nœuds du graphe.

Plus précisément, pour chaque bord de l'image, nous effectuons une diffusion de l'ordonnancement à partir des nœuds (appartenant à toutes les résolutions) qui le constituent vers tous les autres nœuds du graphe. Nous obtenons alors quatre cartes de saillance que nous combinons pour constituer une carte de saillance finale pour cette étape.

Comme exemple, nous considérons le bord de la partie supérieure de l'image. Nous utilisons les nœuds de cette partie comme requête et nous considérons les autres nœuds du graphe comme non étiquetés. Ainsi, le vecteur indiciel  $Y$  de taille  $N$ , (avec  $N = n_1 + n_2 + \dots + n_L$ ) est défini en effectuant l'initialisation appropriée (en mettant des valeur 1 sur les nœuds requêtes et 0 ailleurs), puis l'ordonnancement de tous les autres nœuds non étiquetés est calculé dans le vecteur  $f^*$  selon l'équation 3.18. Chaque élément de ce vecteur correspond à la pertinence d'un nœud par rapport aux requêtes de l'arrière-plan dont le complément correspond à la valeur de saillance. Nous normalisons les

valeurs de  $f^*$  dans l'intervalle  $[0,1]$  et la carte de saillance relative à la diffusion à partir du bord supérieur notée  $S_T$  est calculée comme suit:

$$S_T(i) = 1 - \overline{f^*}(i), i = 1, 2, \dots, N \quad (3.19)$$

avec  $i$  un indice du superpixel dans le graphe et  $\overline{f^*}$  représente le vecteur d'ordonnement normalisé. D'une manière similaire, nous générons les trois autres cartes de saillance  $S_B$ ,  $S_L$  et  $S_R$  en utilisant les nœuds des bords de la partie inférieure, de gauche et de droite respectivement comme requêtes. Ces cartes sont ensuite combinées pour générer une carte de saillance  $S_{BK}$  comme suit:

$$S_{BK} = S_T \circ S_D \circ S_L \circ S_R \quad (3.20)$$

avec  $\circ$  correspond au produit de Hadamard entre matrices.

Il y a deux raisons majeures qui nous ont conduit à utiliser l'approche SC pour la génération des cartes de saillance. D'abord les bords de l'image possèdent généralement des distributions de couleur et de texture différentes. L'utilisation de tous les nœuds des bords de l'image à la fois comme requêtes, conduit à des résultats moins optimaux dus à la différence significative entre les nœuds requêtes (cf. Figure 3.5 (g)). Ensuite comme les objets saillants sont de structure compacte (véhicules, animaux,...,etc.) contrairement à certaines composantes de l'image (ciel, herbe, etc.), ils occupent très rarement trois ou tous les bords de l'image. Dans le cas où des parties de l'objet saillant se trouvent sur un bord de l'image, l'approche SC permet de faire des rectifications sur la carte de saillance finale en assurant au moins deux cartes de saillances qui ne considèrent pas ces parties comme requêtes (cf. figure 3.5 ). Grâce à l'intégration de quatre cartes de saillance, certaines parties de l'objet saillant omises par au moins une carte de saillance peuvent être identifiées. Nous pouvons constater sur la figure 3.5 la supériorité de l'approche SC par rapport à l'estimation de la saillance issue d'un ordonnancement à partir de tous les bords à la fois.

La majorité des parties saillantes sont identifiées dans cette première étape. Cependant, nous pouvons remarquer que sur certaines images, la valeur

de saillance de certains nœuds de l'arrière-plan ne sont pas efficacement estimés (cf. figure 3.5 (h)). Afin d'obtenir un arrière-plan plus net, nous considérons un deuxième ordonnancement en exécutant l'équation (3.9) une seconde fois avec une initialisation de requêtes à partir des parties de l'objet saillant. Nous effectuons alors une diffusion de l'ordonnancement à partir de l'objet saillant vers toute l'image. Cela permet d'améliorer l'estimation de l'objet saillant spécialement dans le cas où il y a des parties omises durant la première étape à cause de l'intersection de l'objet saillant avec au moins l'un des bords de l'image.

#### 3.4.3.2 Ordonnancement avec des requêtes extraites du noyau de l'objet saillant

Nous effectuons d'abord une segmentation binaire de  $S_{BK}$  en utilisant un seuil adaptatif. Nous sommes intéressés dans ce processus par une segmentation qui couvre le plus possible de l'objet saillant. Pour cela, nous définissons la valeur du seuil comme la valeur moyenne de  $S_{BK}$ . À partir de cette segmentation, nous identifions les nœuds de l'objet saillant que nous utiliserons comme requêtes. Nous effectuons alors une initialisation appropriée sur  $Y$  et nous estimons l'ordonnancement de tous les autres nœuds de l'image par rapport aux nœuds requêtes. Comme pour la première étape, le vecteur d'ordonnancement  $f^*$  qui en résulte est normalisé dans l'intervalle  $[0, 1]$  pour former la carte de saillance finale selon l'équation suivante:

$$S_1 = \overline{f^*}(i), i = 1, 2, \dots, N \quad (3.21)$$

avec  $i$  un indice des nœuds sur le graphe et  $\overline{f^*}$  représente le vecteur d'ordonnancement normalisé. La figure 3.5 montre l'amélioration que peut apporter la deuxième phase de l'ordonnancement et la figure 3.6 montre l'impact de la granularité de la segmentation en superpixels sur la carte de saillance générée. En effet, l'application de l'ordonnancement sur un graphe mono-couche basé sur une résolution à faible granularité retourne généralement des portions d'objets réduites. Par contre, l'application de l'ordonnancement sur un graphe mono-couche basé sur une résolution à forte granularité retourne les objets saillants avec certaines parties de l'arrière-plan voisines favorisant un taux de faux positifs important.

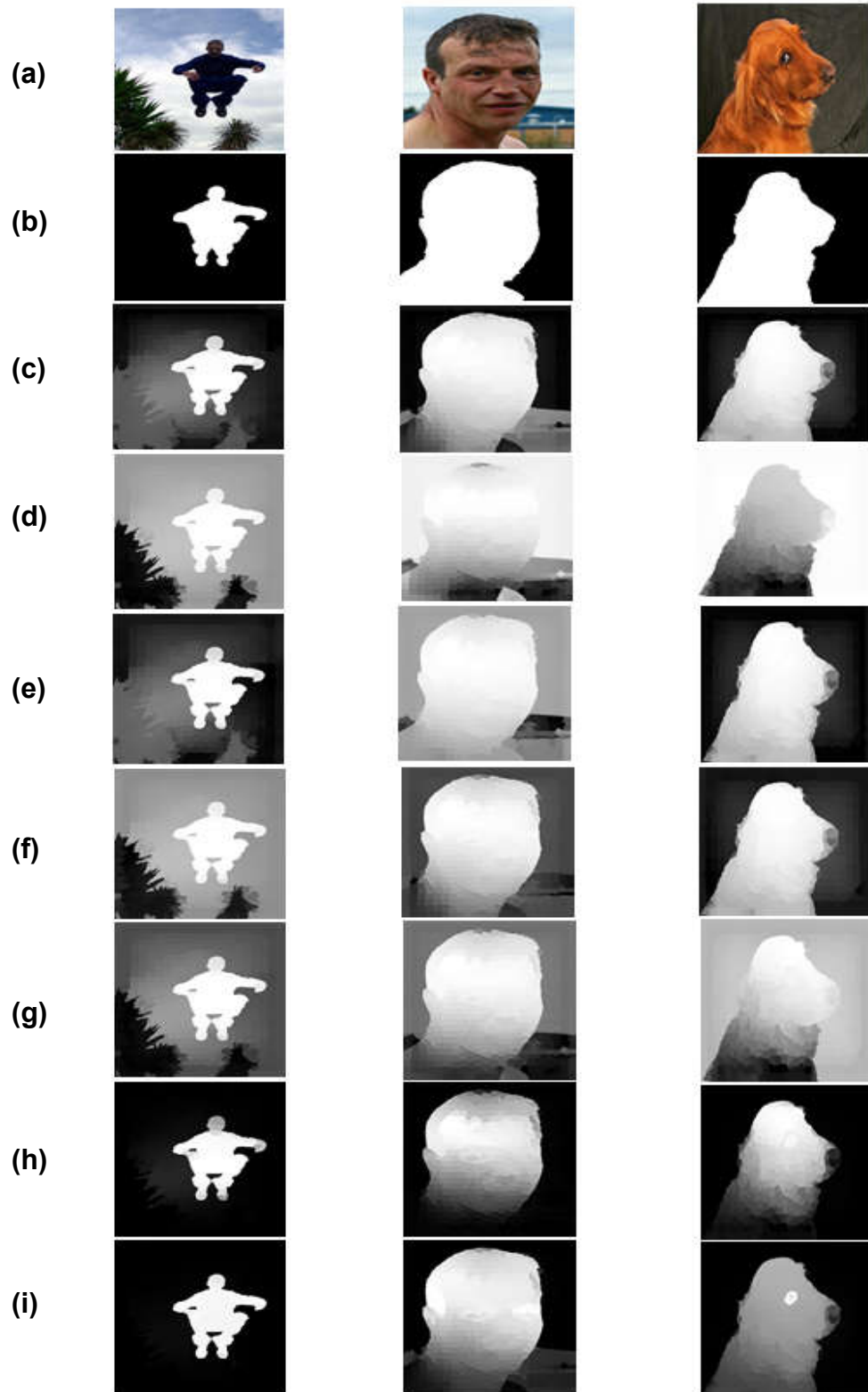


Figure 3.5 Différentes étapes de l'ordonnement effectué au sein de notre méthode. (a) image en couleur, (b) vérité terrain, (c) ordonnancement à partir du bord: (c) du haut, (d) du bas, (e) de gauche et (f) de droite respectivement, (g) ordonnancement à partir de tous les bords à la fois, (h) résultat de l'approche *SR* et enfin (i) ordonnancement à partir du noyau de l'objet saillant.

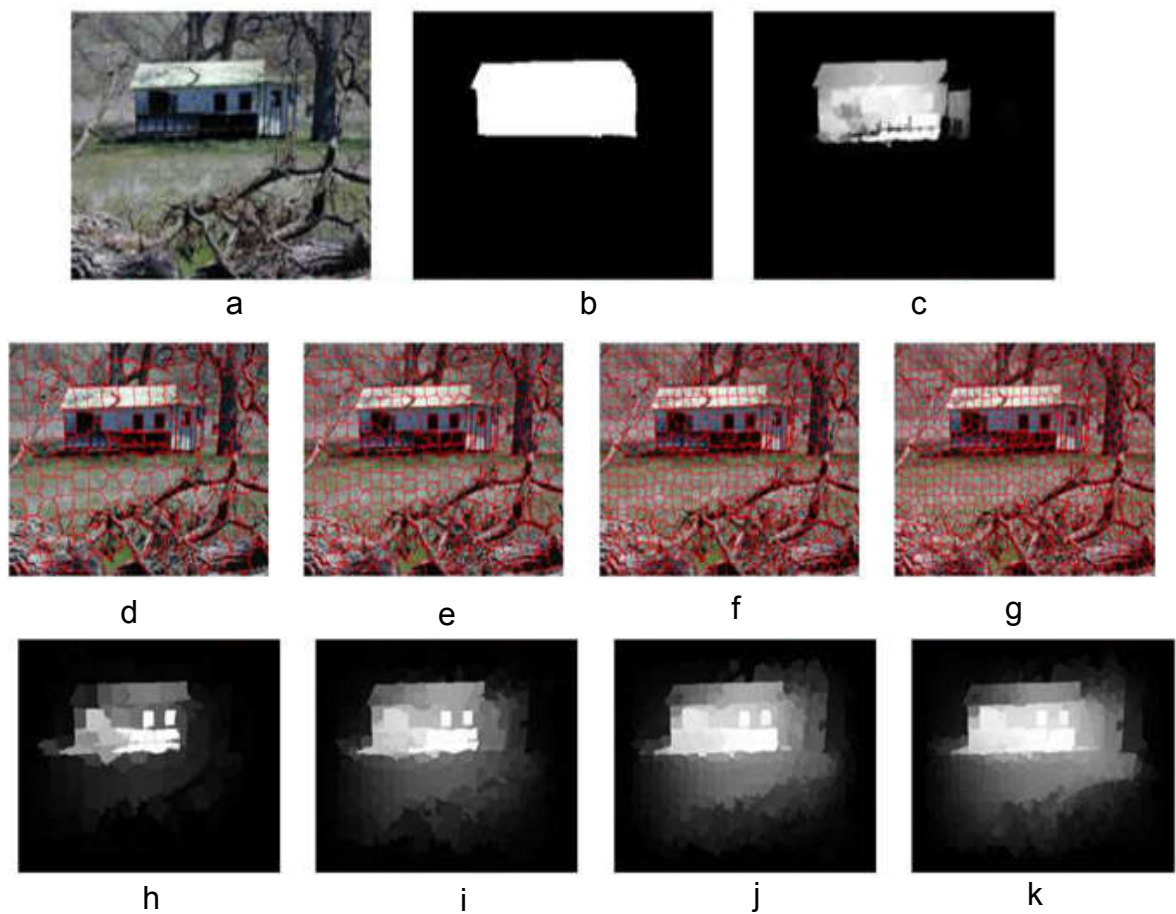


Figure 3.6 Comparaison entre les cartes de saillance retournées par l'ordonnement sur des graphes mono-couches (h ,i ,j , k) construits sur l'image segmentée en 200 (d), 400 (e), 600 (f) et 800 (g) superpixels respectivement et l'ordonnement sur notre graphe multi-couches (c) construit sur (d, e f, g). (a) image en couleur et (b) vérité terrain.

Bien que les régions à proximité de l'objet saillant renseignent sur des informations importantes sur le contexte de l'objet [105,106], elles sont indésirables dans certaines applications comme la reconnaissance d'objets [92], estimation des positions des objets [107, 108] et retouche d'image [109]. L'avantage d'utiliser un graphe multi-couches où chaque couche correspond à une résolution est montré sur la figure 3.6. L'objet saillant est mieux identifié et l'arrière-plan plus net. La figure 3.7 illustre quelques exemples qui comparent notre méthode de saillance basée sur le graphe multi-couches et la méthode [9] basée sur un graphe mono-couche. Nous notons que l'objet saillant est mieux ressorti et l'arrière-plan mieux écarté.



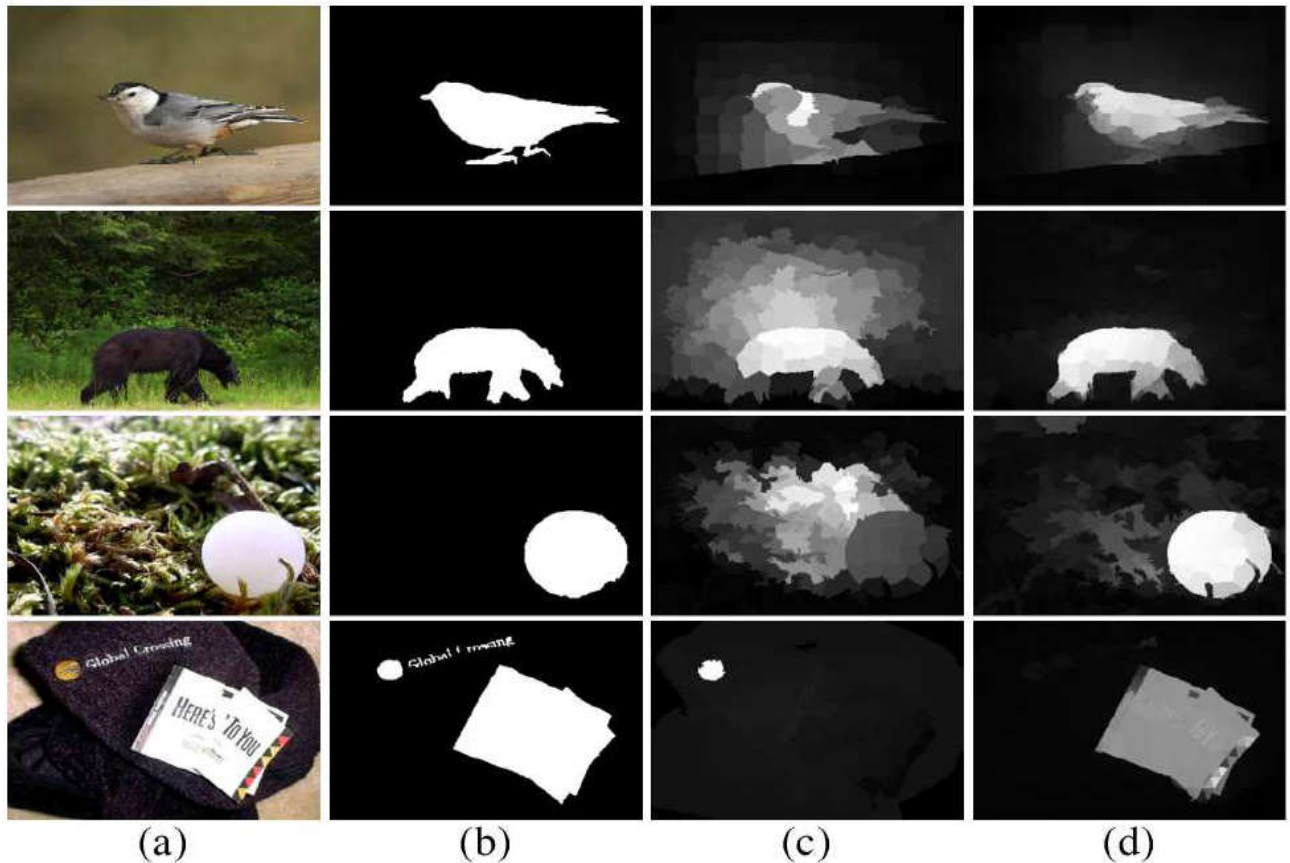


Figure 3.7 Exemples qui comparent le résultat de notre ordonnancement (d) et le résultat de l'ordonnancement proposé dans [9] (c) ; (a) correspond à l'image en couleur, et (b) à la vérité terrain.

### 3.5 Conclusion

Nous avons présenté dans ce chapitre notre détecteur d'objets saillants basé sur une approche d'ordonnancement sur un graphe multi-couches. La structure de notre graphe est construite à base de la superposition de plusieurs résolutions obtenues grâce à la segmentation en superpixels avec des granularités différentes. Les résolutions sont agencées de bas en haut, de la granularité la plus forte à la granularité la plus faible. Nous avons définis un couplage fort entre les nœuds du graphe avec des liens reliant chaque nœud avec ses voisins situés sur la même couche ainsi qu'avec ses voisins se trouvant sur toutes les autres résolutions. Toute cette structure est utilisée à la fois dans l'optimisation d'une fonction objective pour estimer la saillance dans l'image à traiter.

D'abord, nous avons utilisé l'information à priori sur l'arrière-plan pour diffuser les valeurs de l'ordonnement des nœuds situés sur les bords de l'image vers toute notre structure. Les caractéristiques que nous avons utilisées dans le processus de diffusion sont l'information régionale de couleur et de texture ainsi que l'information spatiale.

Ensuite, une fois la carte de saillance estimée, une autre diffusion est effectuée à partir des nœuds les plus saillants vers le reste de notre structure pour constituer la carte de saillance finale. Les résultats retournés par notre approche basée sur un graphe multi-couches surpassent significativement les résultats retournés par l'ordonnement basé sur un graphe mono-couche.

Nous allons détailler dans le chapitre suivant comment améliorer les performances de notre détecteur d'objets saillants dans un processus de raffinement spatiale.

## **CHAPITRE 4**

### **RAFFINEMENT SPATIAL DE LA SAILLANCE**

#### 4.1 Introduction

Dans le but de consolider notre approche de détection d'objets saillants, nous avons proposé un processus de raffinement spatial de la saillance [123,124]. L'objectif est d'augmenter le contraste de l'objet saillant retournée par rapport à l'arrière-plan et mieux ressortir ses contours tout en assurant une présentation plus nette de l'arrière-plan. Ce processus assure trois améliorations majeures effectuées sur la carte de saillance générée durant la première étape de notre approche présentée au chapitre 3.

La première amélioration consiste en la localisation spatiale de l'objet saillant. Celle-ci est effectuée grâce à la génération d'une fenêtre ajustée aux abords des contours de l'objet saillant. Cette fenêtre est le résultat de la combinaison d'un processus de localisation spatiale guidée par la saillance avec la mesure de localisation d'objets présentée dans [66]. Nous notons qu'il est intéressant d'inclure dans notre calcul une mesure de localisation d'objets. En effet, celle-ci génère une multitude de fenêtres susceptibles de contenir des objets. Orienter cette estimation vers les parties saillantes de l'image peut contribuer à la localisation de l'objet saillant. Une telle opération permet de réduire l'espace où s'effectue l'estimation de la saillance et de mieux écarter les parties de l'arrière-plan.

La seconde amélioration consiste à introduire conjointement l'information de contour et une procédure d'évaluation de la pertinence des caractéristiques utilisées pour l'estimation de la saillance dans la zone délimitée par notre fenêtre. En effet, les caractéristiques utilisées ne peuvent pas contribuer d'une manière similaire dans la détection des objets saillants. Par exemple, un objet saillant partageant une distribution de couleur similaire avec l'arrière-plan est mieux

discriminé par la texture. Dans ce contexte, les caractéristiques sont pondérées dans l'équation 3.3 selon leur pouvoir discriminatif. L'inclusion de l'information de contour permet quant à elle d'améliorer la localisation des contours des objets saillants.

La troisième amélioration consiste en l'introduction d'une approche globale de détection d'objets saillants afin d'assurer que le résultat d'estimation de la saillance soit conforme avec les statistique globales de l'image (objets saillants /arrière-plan). Cette opération est effectuée en définissant une approche basée sur les forêts aléatoires combinées avec les mixtures de gaussiennes. Ces améliorations sont détaillées dans ce chapitre en trois étapes fondamentales.

#### 4.2 Localisation spatiale de l'objet saillant

Pour une meilleure précision de détection d'objets saillants, nous définissons une procédure de raffinement qui vise à localiser l'objet saillant et réduire l'espace d'estimation de la saillance. Ceci est effectué en ajustant une fenêtre aux abords des contours de l'objet saillant. Soit  $S_1$  la carte de saillance générée durant la première étape de notre algorithme en minimisant l'équation 3.9. Pour un ajustement approprié de la fenêtre-objet, nous procédons d'abord à l'extraction du noyau de l'objet saillant à partir des parties ayant les valeurs de saillance les plus importantes dans  $S_1$ . Ceci est réalisé en effectuant une segmentation binaire sur  $S_1$  en utilisant le modèle de mixtures de gaussiennes notée *GMM (Gaussians Mixture Model)* [110].

Soient  $C_1, C_2, \dots, C_K$  les  $K$  groupes générées en appliquant sur  $S_1$  un clustering basé sur les GMM. Ces classes sont triées selon un ordre croissant des valeurs du paramètre moyenne  $m_1, m_2, \dots, m_k$ . Le paramètre  $K$  représente le nombre de composantes qui constituent la mixture de gaussiennes. Celui ci est estimé en utilisant la méthode *Minimum message length (MML)* [110,111]. Les paramètres de chaque gaussienne (moyenne et matrice de covariance) sont estimés grâce à la méthode *Expectation maximisation algorithm* [110]. Soient  $w_1, w_2, \dots, w_k$  les poids reliés à ces classes où  $w_i$  représente la probabilité à priori qu'un pixel appartienne à la classe  $C_i$ . Comme les zones les plus claires dans la carte de saillance  $S_1$  sont plus susceptibles d'appartenir au premier plan, nous extrayons le noyau de l'objet saillant en effectuant une segmentation binaire de  $S_1$

à base des valeurs de saillance. Cette segmentation est effectuée en séparant les différentes composantes en deux groupes distincts:  $C_b = \{C_1, C_2, \dots, C_{k^*}\}$  pour l'arrière-plan et  $C_f = \{C_{k^*+1}, C_{k^*+2}, \dots, C_K\}$  pour le noyau de l'objet saillant avec  $k^*$  défini comme suit:

$$k^* = \arg \min_h \left( \sum_{i=1}^h w_i > \tau \right) \quad (4.1)$$

avec  $\tau$  un seuil défini dans l'intervalle  $[0, 1]$  qui permet de déterminer le niveau de confiance utilisé pour extraire le noyau de l'objet saillant. La *valeur* de ce seuil est empiriquement ajustée à 0.94.

Soit  $Seg$  le mask binaire obtenu après la classification des pixels de l'image dans les deux classes  $C_f$  et  $C_b$  respectivement définie comme suit:

$$Seg(x, y) = \begin{cases} 1 & \text{si } S_1(x, y) \in C_f \\ 0 & \text{si } S_1(x, y) \in C_b \end{cases} \quad (4.2)$$

avec  $S_1(x, y)$  correspond à la valeur de saillance d'un pixel défini par les coordonnées spatiales  $x$  sur l'axe des abscisses et  $y$  sur l'axe des ordonnées dans  $S_1$ .

A Partir de  $Seg$  et  $S_1$ , nous pouvons calculer le centroïde de notre fenêtre-objet (appelé point d'ancrage) caractérisé par les coordonnées  $(x_0, y_0)$  en utilisant la formule suivante:

$$(x_0, y_0) = \frac{1}{Z} \sum_{x=1}^{m_1} \sum_{y=1}^{m_2} (x, y) \times \log(1 + Seg(x, y) \times S_1(x, y)) \quad (4.3)$$

avec  $Z$  défini comme suit:

$$Z = \sum_{x=1}^{m_1} \sum_{y=1}^{m_2} \log(1 + Seg(x, y) \times S_1(x, y)) \quad (4.4)$$

où  $(m_1, m_2)$  représente les dimensions de l'image et  $S_1(x, y) \in [0, 255]$ .

A partir du point d'ancrage, nous définissons une mesure de localisation d'objets saillants guidée par la saillance. Il s'agit de définir une fenêtre rectangulaire notée  $W$  avec une forte plausibilité de contenir l'objet saillant. Nous notons que  $W$  est défini par les intervalles spatiaux  $\mathcal{X}_{\delta_1}$  sur l'axe des abscisses et  $\mathcal{Y}_{\delta_2}$  sur l'axe des ordonnées comme suit:

$$\begin{cases} \mathcal{X}_{\delta_1} = [x_0 - \delta_1, x_0 + \delta_1] \\ \mathcal{Y}_{\delta_2} = [y_0 - \delta_2, y_0 + \delta_2] \end{cases} \quad (4.5)$$

avec  $\delta_1$  et  $\delta_2$  des paramètres indiquant les dimensions de  $W$  définis comme suit:

$$(\delta_1, \delta_2) = \arg \max_{\delta_1, \delta_2} \left\{ \frac{1}{Z'} \sum_{x \in \mathcal{X}_{\delta_1}} \sum_{y \in \mathcal{Y}_{\delta_2}} \log(1 + S_1(x, y)) \right\} \leq 1 - \eta \quad (4.6)$$

avec  $\eta$  un seuil ajusté empiriquement à 0.01 et  $Z'$  défini comme suit:

$$Z' = \sum_{x=1}^{m_1} \sum_{y=1}^{m_2} \log(1 + S_1(x, y)) \quad (4.7)$$

Fondamentalement, l'équation 4.6 permet de générer une fenêtre entourant les parties les plus saillantes de l'image. L'estimation de la saillance dans la fenêtre  $W$  peut améliorer de manière significative les résultats. Cependant, cette estimation est entièrement dépendante de la carte de saillance initiale  $S_I$ .

- D'une part, si certaines parties de l'objet saillant ne sont pas convenablement détectées dans  $S_I$ , la position spatiale de  $W$  ne permettrait pas de contenir tout l'objet saillant.
- D'autre part, si  $S_I$  contient un taux important de faux positifs (certaines parties de l'arrière-plan sont considérées comme saillantes), la position

spatiale de  $w$  ne serait pas convenablement ajustée aux abords de l'objet et inclurait des parties de l'arrière-plan.

Par exemple, dans la figure 4.1, sur les rangées 5, 6 et 7, la localisation spatiale de l'objet saillant par la génération de notre fenêtre (présentée avec la couleur jaune) rate certains détails saillants importants. Dans la même figure, sur la rangée 4, notre fenêtre est positionnée sur une surface large contenant un taux élevé de faux positifs. Elle n'est donc pas convenablement ajustée pour détecter la position spatiale de l'objet saillant.

Afin d'éviter ces deux problèmes de détection, l'utilisation d'une mesure de localisation spatiale d'objets indépendante de  $S_I$  peut améliorer les résultats. Il existe une multitude de travaux effectués sur la localisation d'objets [66,21,112]. Cependant, comme il s'agit d'une étape préliminaire, et que la complexité temporelle d'un bon détecteur d'objets saillant est un critère important à respecter (cf. section 1.5.3), nous avons choisi d'utiliser la méthode *Edge Boxes* [66]. Nous notons que cette méthode est particulièrement efficace en comparaison avec les autres méthodes de l'état de l'art sur la localisation d'objets. En outre, la complexité temporelle de cette méthode est l'une des meilleures de l'état de l'art. La réponse est presque instantanée et les résultats satisfaisants. Il s'agit d'une méthode qui se base sur les contours et utilise la méthode de [97] à cet effet. Pour chaque image, cette méthode retourne approximativement un ordonnancement de 1000 fenêtres en une fraction de seconde. Nous considérons  $n$  fenêtres dotées des meilleurs scores parmi lesquelles nous choisissons les  $m \ll n$  fenêtres ayant le taux d'overlapping le plus important avec notre fenêtre  $W$ . nous avons considéré d'une manière empirique  $n = 20$  et  $m = 3$ .

Soient  $B_1, B_2, \dots, B_n$  les  $n$  fenêtres avec les meilleures valeurs d'ordonnancement. Pour chaque fenêtre  $B_i, i \in \{1, 2, \dots, n\}$ , nous calculons son score d'overlapping spatial avec  $W$  comme suit:

$$O_i = \frac{B_i \cap W}{B_i \cup W} \quad (4.8)$$

En choisissant les  $m$  fenêtres ayant les meilleurs scores d'overlapping,  $B_{(1)}, B_{(2)}, \dots, B_{(m)}$ , nous augmentons la probabilité d'obtenir une nouvelle fenêtre

bien ajustée sur la position spatiale de l'objet saillant. Nous nous intéressons donc à combiner l'ensemble des fenêtres  $\mathcal{W} = \{W, B_{(1)}, B_{(2)}, \dots, B_{(m)}\}$ . Nous supposons que chaque fenêtre  $B_i, i \in \{1, 2, \dots, n\}$  est définie par les intervalles spatiaux  $[x_{i,1}, x_{i,2}]$  sur l'axe des abscisses et  $[y_{i,1}, y_{i,2}]$  sur l'axe des ordonnées.

Soient:

$X_b = (x_{1,1}, x_{1,2}, \dots, x_{1,m}, x_{1,m+1})$  la liste triée selon un ordre croissant des premières coordonnées spatiales sur l'axe des abscisses des  $m$  fenêtres et de  $W$  conjointement.

$X_e = (x_{2,1}, x_{2,2}, \dots, x_{2,m}, x_{2,m+1})$  la liste triée selon un ordre croissant des secondes coordonnées spatiales sur l'axe des abscisses des  $m$  fenêtres et de  $W$  conjointement.

$Y_b = (y_{1,1}, y_{1,2}, \dots, y_{1,m}, y_{1,m+1})$  la liste triée selon un ordre croissant des premières coordonnées spatiales sur l'axe des ordonnées des  $m$  fenêtres et de  $W$  conjointement.

$Y_e = (y_{2,1}, y_{2,2}, \dots, y_{2,m}, y_{2,m+1})$  la liste triée selon un ordre croissant des secondes coordonnées spatiales sur l'axe des ordonnées des  $m$  fenêtres et de  $W$  conjointement.

Nous combinons les  $m+1$  fenêtres pour former la fenêtre finale  $W'$  en inhibant le premier élément de la liste des premières coordonnées sur chaque axe et le second élément de la liste des secondes coordonnées sur chaque axe. Cette fenêtre est donc définie par les intervalles spatiaux suivant:

$$W' \begin{cases} \theta = [X_b(2), X_e(m)] \\ \vartheta = [Y_b(2), Y_e(m)] \end{cases} \quad (4.9)$$

avec  $\theta$  correspond à l'intervalle de définition de  $W'$  sur l'axe des abscisses et  $\vartheta$  à l'intervalle de définition de  $W'$  sur l'axe des ordonnées. La figure 4.1 montre quelques exemples d'estimation de la fenêtre  $W'$ . Nous constatons clairement que l'emplacement spatial de l'objet saillant est estimé efficacement en utilisant notre approche.



### 4.3 Raffinement de saillance en utilisant les informations de contour et de pertinence des caractéristiques

Une fois la fenêtre  $W'$  est générée, nous appliquons une autre fois notre approche d'ordonnancement sur notre graphe multi-couches en minimisant la fonction 3.9 sur la surface délimitée par  $W'$ . Cette fois-ci, l'initialisation des requêtes est effectuée directement à base des valeurs de saillance extraites de  $S_I$  au sein de la surface de  $W'$ . Pour une meilleure localisation des contours des objets saillants et un meilleur contraste entre le premier plan et l'arrière-plan, nous introduisons une mesure d'évaluation de la pertinence des caractéristiques utilisées à travers les paramètres  $\varepsilon_{c,k}$ ,  $\varepsilon_{l,k}$  et  $\varepsilon_{e,k}$ ,  $k \in \{1,2,3\}$  ainsi que l'information de contour à travers le paramètre  $\gamma$  dans l'équation 3.3.

#### 4.3.1 Information de contour

Nous notons que contrairement à la première étape de notre algorithme, cette fois-ci, nous pouvons utiliser l'information de contour du moment qu'il n'existe pas une forte densité de contours en dehors de l'objet saillant. En effet, comme  $W'$  est ajusté sur l'objet saillant, la plupart des régions de l'arrière-plan sont écartées et les contours qui se trouvent à l'intérieur de  $S_I$  appartiennent majoritairement à l'objet saillant. Il est possible donc d'utiliser cette information pour arrêter le processus de diffusion des valeurs de l'ordonnancement à partir des bords de  $W'$  vers les régions contenant des contours. Nous utilisons la méthode [66] pour estimer le gradient naturel des images comme indiqué dans la section 3.3.3. Nous supposons que la somme des gradients d'une région  $r_i^{(\ell)}$  est notée par  $G_i^\ell$  avec  $\ell \in \{1,2, \dots, L\}$  l'indice de la résolution utilisée et  $i$  l'indice du superpixel dans la résolution  $\ell$  (cf. section 3.4.1). Nous proposons donc d'ajuster la valeur de  $\gamma$  comme suit:

$$\gamma = \frac{(G_i^{(\ell)} + G_j^{(m)})}{2} \quad (4.10)$$

Comme les objets saillants possèdent une forte densité de gradient au niveau de leurs bords, le poids entre deux nœuds adjacents situés au voisinage de ces bords tend vers 0 et la diffusion tend à être annulée. Nous notons que la

considération du gradient est uniquement prise en compte dans le processus de diffusion à partir des parties arrière-plan de  $W'$  vers l'objet saillant. En effet, étant donné que l'objet saillant peut avoir une forte densité de gradient non seulement au niveau de ses bords mais aussi sur toute sa surface, la diffusion à partir de celui-ci vers l'arrière-plan peut endommager considérablement le processus d'estimation de la saillance.

#### 4.3.2 Évaluation de la pertinence des caractéristiques

L'évaluation de la pertinence des caractéristiques a comme objectif d'attribuer plus d'importance pour les caractéristiques qui discriminent mieux l'objet saillant de l'arrière-plan. Du moment que nous avons déjà estimé le noyau de l'objet obtenu en effectuant la segmentation menée grâce à l'équation 4.1, nous pouvons évaluer la pertinence de chaque caractéristique d'une manière supervisée.

Pour une caractéristique donnée, nous considérons ses histogrammes normalisés dans l'intervalle  $[0, 1]$  notés  $p$  et  $q$  au sein du noyau de l'objet et l'arrière-plan respectivement. Nous supposons que chaque caractéristique est échantillonnée sur 256 bins et soient  $p(i)$  et  $q(i)$  les fréquences normalisées du  $i^{\text{ème}}$  bin. Le pouvoir discriminatif d'une caractéristique est représenté par le degré d'overlapping entre  $p$  et  $q$ . Ceci peut être formalisé comme suit [113]:

$$V = \frac{\text{var}\left(F; \frac{(p+q)}{2}\right)}{\text{var}(F; p) + \text{var}(F; q)} \quad (4.11)$$

avec  $\text{var}(F; p)$ ,  $\text{var}(F; q)$  et  $\text{var}\left(F; \frac{(p+q)}{2}\right)$  représentent la variance de la fonction  $F$  obtenue par rapport aux distributions  $p$ ,  $q$  et  $\frac{(p+q)}{2}$ . La variance de  $F$  par rapport à une distribution  $R$  est calculée comme suit:

$$\text{var}(F, R) = \sum_{i=1}^{256} R(i) \cdot F^2(i) - \left[ \sum_{i=1}^{256} R(i) \cdot F(i) \right]^2 \quad (4.12)$$

La valeur de  $F$  pour le  $i^{\text{ème}}$  bin est calculée comme suit:

$$F(i) = \log \left( \frac{\max(p(i), \epsilon)}{\max(q(i), \epsilon)} \right) \quad (4.13)$$

avec  $\epsilon$  une petite valeur qui évite les fréquences nulles. Cette valeur est ajustée empiriquement à 0.001. Nous notons que par un tel processus, il est facile de constater que  $V$  soit défini dans l'intervalle  $[0.5, +\infty]$ . Afin de normaliser l'évaluation de la pertinence dans l'intervalle  $[0, 1]$ , nous utilisons la fonction suivante:

$$\mathcal{E} = 1 - \exp(-\delta|V - 0.5|) \quad (4.14)$$

avec  $\delta$  un facteur qui contrôle la sensibilité du calcul de la pertinence. Augmenter la valeur de ce facteur, augmenterait le nombre de caractéristiques considérées pertinentes, et vice versa. Ce vecteur est ajusté empiriquement à 0.2. Une fois que tous les poids sont estimés par l'évaluation de la pertinence des caractéristiques, nous appliquons l'équation 3.9 sur la sous image délimitée par  $\mathcal{W}'$  pour la génération d'une carte de saillance notée  $S_2$ .

La figure 4.1 montre quelques exemples qui montrent l'apport de chacune des étapes de raffinement citées précédemment. Notre procédure de localisation spatiale de l'objet saillant améliore à elle toute seule les résultats (colonne f). L'inclusion de l'information de contour et d'évaluation de la pertinence des caractéristiques permet d'écarter l'arrière-plan plus efficacement tout en assurant un contraste Objet saillant /arrière-plan plus important et une meilleur localisation des contours des objets saillants (colonnes g et h). Ces deux informations agissent au niveau local de l'image. Afin que les résultats soient conformes avec les statistiques globales de l'image, nous proposons un processus de raffinement supplémentaire basé sur les forêts aléatoire et les mixtures de gaussiennes.

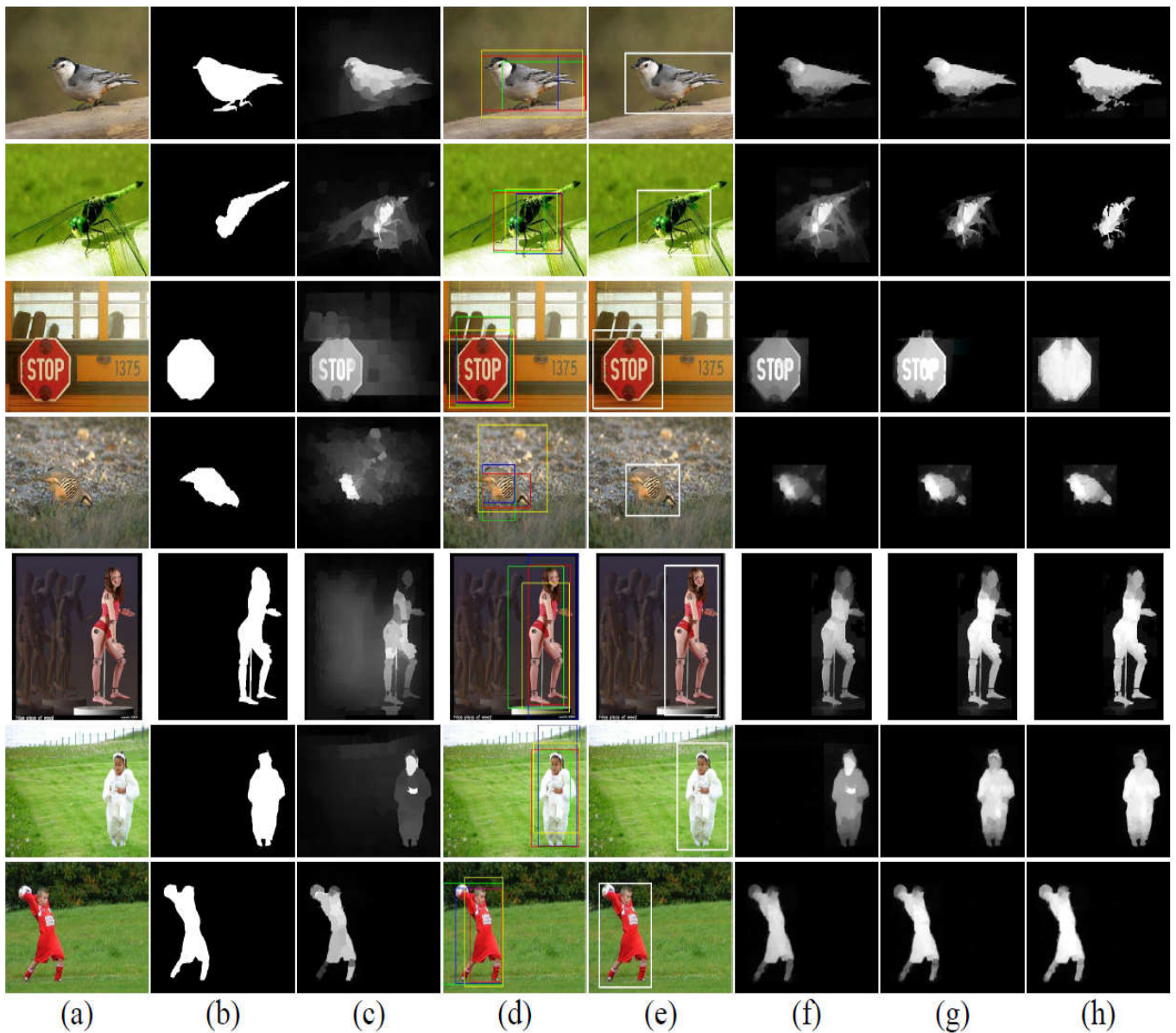


Figure 4.1 Exemples de localisation spatiale d'objets saillants: (a) image en couleur, (b) vérité terrain, (c) carte de saillance  $S_I$ , (d) position spatiales des fenêtres ( $\mathcal{W}$ ): la fenêtre jaune correspond à  $W$  et les autres à  $B_{(i)}, i \in \{1, 2, \dots, m\}$ , (e) position de la fenêtre finale  $W'$ , (f), (g) et (h) sont les cartes de saillance obtenu après la localisation spatiale: (f) sans inclure les informations de contours et de pertinence des caractéristiques, (g) avec l'information de pertinence des caractéristiques et sans l'information de contour, (h) en incluant les informations de pertinence des caractéristiques et de contour conjointement.

#### 4.4 Raffinement global de la saillance

Nous proposons dans cette section une procédure de détection d'objets saillants basée sur les forêts aléatoires notées *RF* (*random forest*) [114] combinés aux mixtures de gaussiennes notées *GMM* (*gaussians mixture model*) pour inclure l'information global de l'image dans l'estimation de la saillance. Dans le but de réduire le taux de faux positifs notamment au voisinage des contours des objets saillants ainsi que pour augmenter les valeurs de saillance pour les objets détectés, nous proposons d'appliquer un classifieur basé sur les forêts aléatoires aux superpixels tout en tenant en compte l'information sur le niveau de contraste entre l'objet saillant et l'arrière-plan. Pour cela, nous reprenons les mêmes résolutions présentées à la section 3.4. Pour chaque résolution, nous effectuons l'apprentissage d'un classifieur à partir des parties premier-plan et arrière-plan extraites du noyau de l'objet saillant et des régions de l'un des quatre bords de l'image respectivement. Nous entraînons alors quatre classifieurs pour chaque résolution. Nous utilisons les mêmes caractéristiques de couleur et de texture présentés dans la section 3.3 (neufs caractéristiques en tout) pour l'entraînement de chaque classifieur. Chaque classifieur est constitué de trois arbres et chaque arbre est entraîné en sélectionnant d'une manière aléatoire 50 % des données disponibles et trois caractéristiques parmi les neufs utilisées.

Nous utilisons les mixtures de gaussiennes pour évaluer le niveau de contraste entre le noyau de l'objet et les bords de l'image qui va servir par la suite à raffiner le résultat de classification retourné. D'abord, Nous modélisons le noyau de l'objet et les parties de l'arrière-plan en utilisant deux mixtures de gaussiennes  $M_{obj}$  et  $M_{bck}$  respectivement. Soient  $K_{obj}$  et  $K_{bck}$  le nombre de composantes constituant les deux modèles  $M_{obj}$  et  $M_{bck}$  respectivement. L'estimation de ces nombres se fait en utilisant l'algorithme *MML*. Ensuite, nous estimons le niveau de contraste noté  $d$  entre l'objet saillant et l'arrière-plan en utilisant la divergence de Kullback Leibler entre les composantes  $M_{obj}$  et  $M_{bck}$  comme suit:

$$d = \min_{i,j} \left( tr(\Sigma_j^{-1}\Sigma_i) + (\mu_j - \mu_i)^T \Sigma_j^{-1}(\mu_j - \mu_i) + \ln \left( \frac{|\Sigma_i|}{|\Sigma_j|} \right) \right) \quad (4.15)$$

avec  $i \in \{1, \dots, K_{obj}\}$  et  $j \in \{1, \dots, K_{bck}\}$ . Les paramètres  $(\mu_i, \Sigma_i)$  et  $(\mu_j, \Sigma_j)$  représentent le vecteur moyenne et la matrice de covariance de la  $i^{\text{ème}}$  et la  $j^{\text{ème}}$  gaussienne appartenant aux modèles  $M_{obj}$  et  $M_{bck}$  respectivement. En utilisant le niveau de contraste  $d$ , nous calculons un seuil dynamique qui permet d'affecter une valeur unifiée (étiquette) pour chaque superpixel  $r_i^{(\ell)}$ . D'abord nous effectuons la prédiction de la saillance pour chaque pixel au niveau de chaque résolution  $\Omega_\ell$ ,  $\ell \in \{1, 2, \dots, L\}$  et de chaque bord de l'image. Soit  $P_i^{(\ell)}$  le pourcentage de pixels dans  $r_i^{(\ell)}$  classifié comme saillants par notre classifieur  $RF$ . Le processus d'unification des valeurs de saillance au sein du superpixel génère une étiquette notée  $s_i^{(\ell)}$  dont la valeur est calculée comme suit:

$$s_i^{(\ell)} = \begin{cases} 1 & \text{si } p_i^{(\ell)} \geq \exp(-\beta \cdot d) \\ 0 & \text{sinon} \end{cases} \quad (4.16)$$

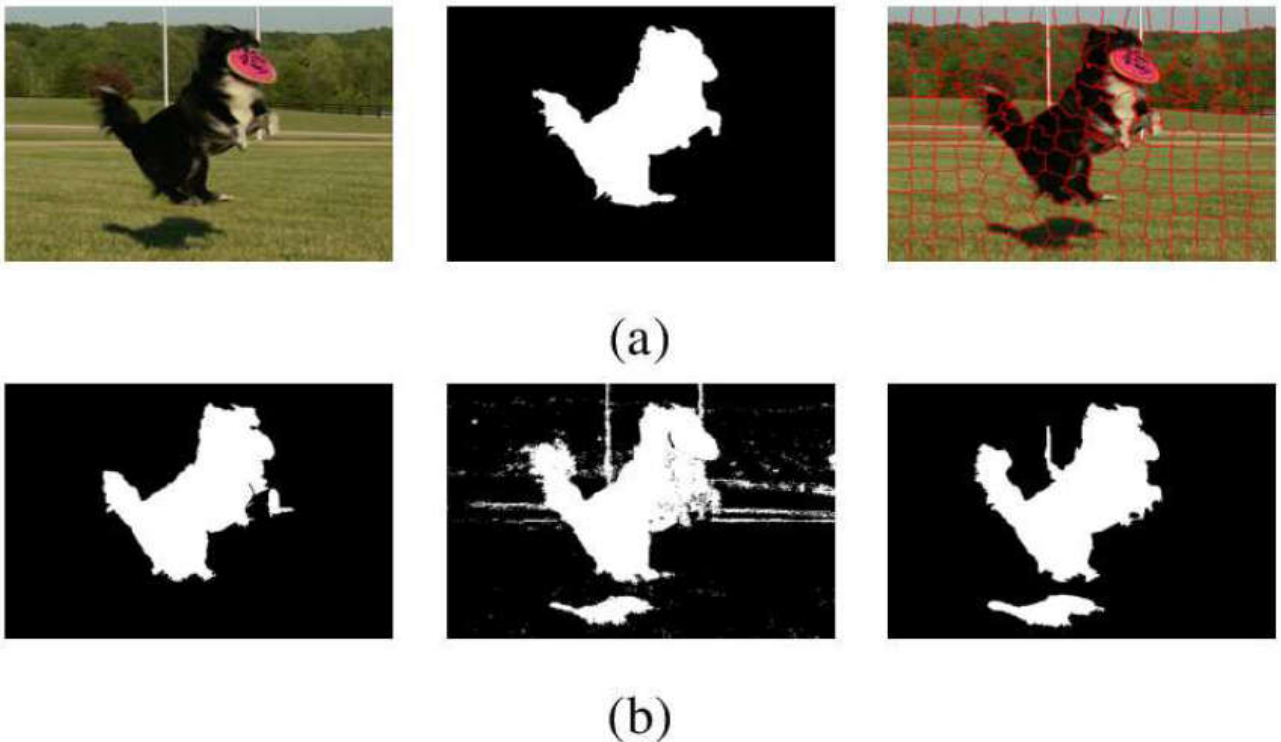


Figure 4.2 Exemple montrant le résultat du classifieur RF en considérant le bord de gauche de l'image template d'arrière-plan. De gauche à droite: (a) image en couleur, vérité terrain et segmentation de l'image en 200 superpixel; (b) noyau de l'objet saillant, résultat du classifieur RF au niveau pixel et résultat d'unification de la saillance au niveau superpixels.

Le paramètre  $\beta > 0$  contrôle la sensibilité du processus d'unification de la saillance au niveau des superpixels (génération d'étiquettes). D'une manière empirique, nous ajustons  $\beta$  à 0.1. La génération de l'étiquette au sein de chaque superpixel dépend du niveau la similarité entre les distributions de l'objet saillant et l'arrière-plan. Plus la valeur du contraste  $d$  est importante, plus le niveau de confiance accordé par notre algorithme à la prédiction effectuée par le classifieur *RF* est important.

La figure 4.2 illustre un exemple d'application du classifieur basé sur les forêts aléatoires entraîné sur le bord gauche de l'image comme information à priori sur l'arrière-plan et le noyau de l'objet comme information à priori sur l'objet saillant. Nous pouvons aisément constater que notre processus d'unification de la saillance améliore la qualité de la carte de saillance retournée. En effet, les parties de l'arrière-plan sont plus efficacement écartées dans le processus d'estimation de la saillance. Le taux de faux positifs est donc considérablement réduit au sein de l'objet saillant.

Le même processus est effectué en considérant chaque bord de l'image comme information à priori sur l'arrière-plan au sein d'une même résolution. La carte de saillance  $T_\ell$  correspondant à la résolution  $\ell \in \{1, 2, \dots, L\}$  est alors calculée en effectuant la moyenne de toutes les cartes de saillances générées. Finalement, nous obtenons  $L$  cartes de saillances chacune relative à une résolution. Nous combinons alors toutes ces cartes pour générer une carte de saillance notée  $T_f$  et calculée par l'opérateur de factorisation selon l'équation suivante (cf. figure 4.3):

$$T_f = T_1 \times T_2 \times \dots \times T_L \quad (4.17)$$

La figure 4.3 montre un exemple de notre démarche d'estimation de la saillance globale basée sur les RF et les GMM. Nous notons que les valeurs de saillance affectées aux parties saillantes sont élevées. L'objet saillant est mieux contrasté avec un taux de rappel accru. Cependant, nous notons que les régions saillantes retournées contiennent généralement certains faux positifs qui réduisent significativement la précision de détection. D'une autre part, notre approche basée sur notre graphe multi-échelles raffiné grâce à la localisation spatiale, l'information de contour et l'évaluation de la pertinence des caractéristiques retourne des

régions saillantes plus restreintes. Dans ce cas, la précision de détection est mieux favorisée. Cependant, les régions saillantes sont moins contrastées et la valeur de rappel peut connaître une certaine diminution dans certains cas. Afin d'atteindre des taux de précision et de rappel conjointement élevés, nous fusionnons les deux cartes de saillance selon la formule suivante:

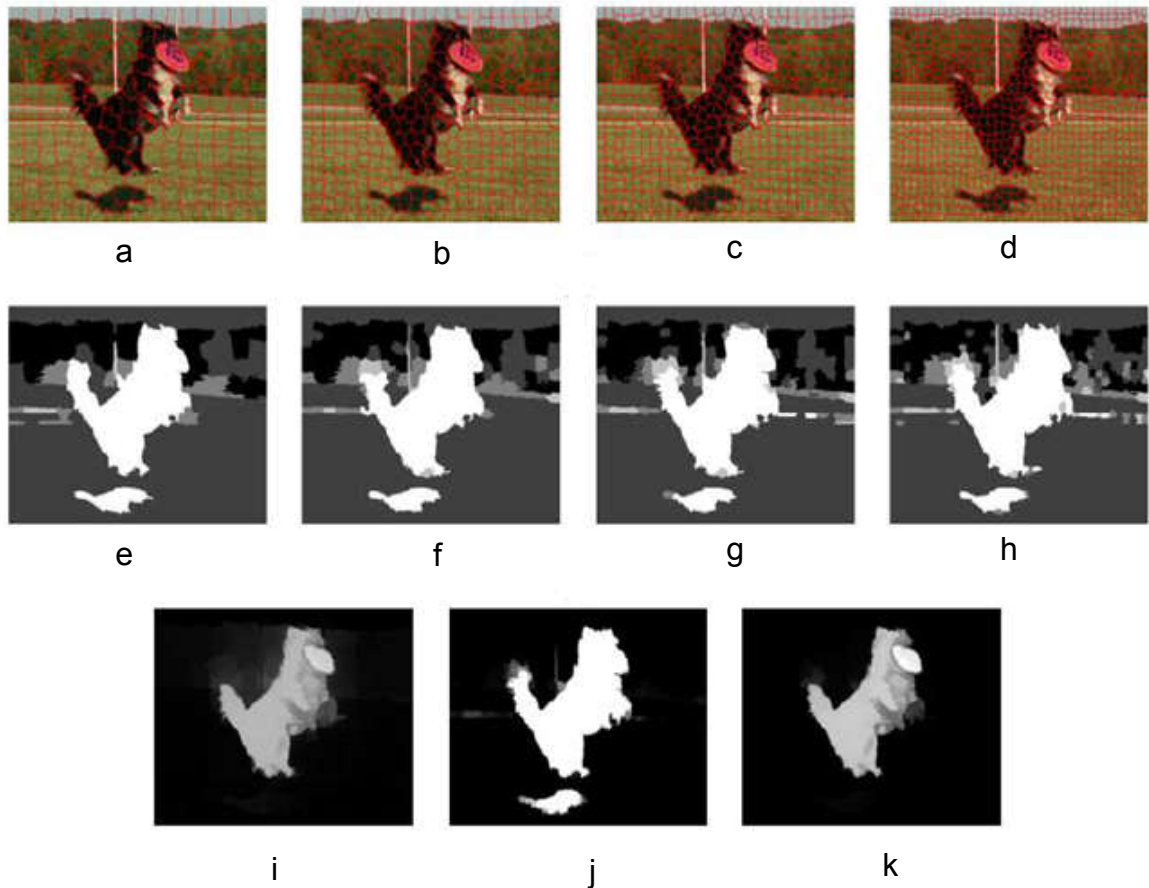


Figure 4.3 Fusion des cartes de saillance de l'exemple de la figure 4.2. (e,f,g,h) cartes de saillance obtenues par notre démarche de classification basée sur les RF et les GMM appliquées aux résolutions à granularité de (a)200, (b) 400, (c) 600 et (d) 800 superpixels, (i) le résultat de raffinement basé sur l'information de contour et la pertinence des caractéristiques, (j) résultat de fusion des cartes (e,f,g,h) et (k) carte de saillance finale correspondant à la combinaison des cartes (i) et (j).

$$S_3 = [\omega \times T_f + (1 - \omega) \times S_2] \times S_2, \quad \omega \in [0,1] \quad (4.18)$$



avec  $\omega$  un paramètre qui équilibre la contribution de  $T_f$  et  $S_2$  pour constituer la carte de saillance finale. Généralement, nous accordons plus d'importance à  $S_2$  en fixant  $\omega < 0.5$  (nous ajustons empiriquement  $\omega$  à 0.33). Cela permet de mieux contraster l'objet avec l'arrière-plan par rapport à  $S_2$ . L'opérateur de factorisation permet de mieux écarter les faux positifs générés dans  $T_f$  tout en gardant des valeurs de saillance élevées pour l'objet saillant.

#### 4.5 Notre algorithme

Nous retraçons dans ces sections les différentes étapes de notre approche de détection d'objets saillants. L'algorithme 1 présente d'une manière concise notre démarche. Le paramètre  $iter_{max}$  fixe le nombre maximal d'itérations impliquées dans le processus de raffinement. Nous avons ajusté cette valeur à 2. Notons que plus cette valeur est élevée, plus le résultat de la saillance se dégrade. En effet, la localisation spatiale de l'objet saillant génère une fenêtre qui se rétrécit progressivement autour de l'objet à chaque itération. Afin d'éviter des situations de dégradation, nous avons mis au point un système plus stable qui s'accommode à l'augmentation du nombre d'itérations. Pour cela, nous avons défini un processus de mise à jour des paramètres  $\alpha$  et  $\eta$  qui permettent de contrôler les équations 3.18 et 4.6 respectivement. Cette mise à jour itérative est définie comme suit:

$$\alpha = 9 \times 10^{-iter} \quad (4.19)$$

$$\eta = 100^{-iter} \quad (4.20)$$

Plus concrètement, plus la valeur de  $\alpha$  est basse, plus l'algorithme préserve les valeurs de saillance de l'itération précédente et plus la valeur de  $\eta$  est basse, plus les dimensions de notre fenêtre se stabilisent. Bien que cela fournisse une stabilité globale de l'algorithme, l'amélioration de saillance pour des itérations plus élevées est très faible. La figure 4.4 montre des exemples de cartes de saillance générées en utilisant l'équation 4.18. Nous pouvons constater que le raffinement global améliore considérablement les résultats. L'objet saillant est mieux contrasté par rapport aux cartes obtenues après le raffinement local et la localisation

spatiale et l'arrière-plan est mieux écarté par rapport aux cartes de saillances générées par notre classifieur basé sur les RF et les GMM. Globalement, nos cartes de saillance finale se rapprochent mieux de la vérité terrain.

---

Algorithme 1: Graphe multi-échelles pour la détection d'objets saillants

---

**Données:** Image en couleur, nombre de résolutions  $L$ , nombre de superpixels

$$n_1, n_2, \dots, n_L$$

**Résultat:** Carte de saillance finale.

- Segmentation multiple en superpixels en utilisant l'algorithme SLIC pour la génération des différentes résolutions.
- Génération de la carte de saillance initiale  $S_1$  en utilisant les équations 3.3 et 3.9.

**For**  $iter=1 : iter_{max}$  **do**

- Appliquer les équations 4.3 et 4.5 pour la génération de notre fenêtre  $W$  pour la localisation de l'objet saillant;
- Construction de  $W'$  en combinant  $W$  avec les fenêtres ayant les meilleurs scores retournées par [82] (équations 4.8 et 4.9);
- Calculer la pertinence des caractéristiques utilisées en appliquant les équations 4.11 et 4.14
- Calculer l'information du contour en appliquant l'équation 4.10.
- Mise à jour de l'équation 3.3 pour la prise en compte de la pondération des caractéristiques et l'information de contour;
- Génération de la carte de saillance  $S_2$  en appliquant les équations 3.3 et 3.9;
- Raffinement global de la saillance en appliquant les équations 4.16 et 4.17;
- Génération de la carte de saillance finale  $S_3$  en appliquant l'équation 4.18.

**End**

---

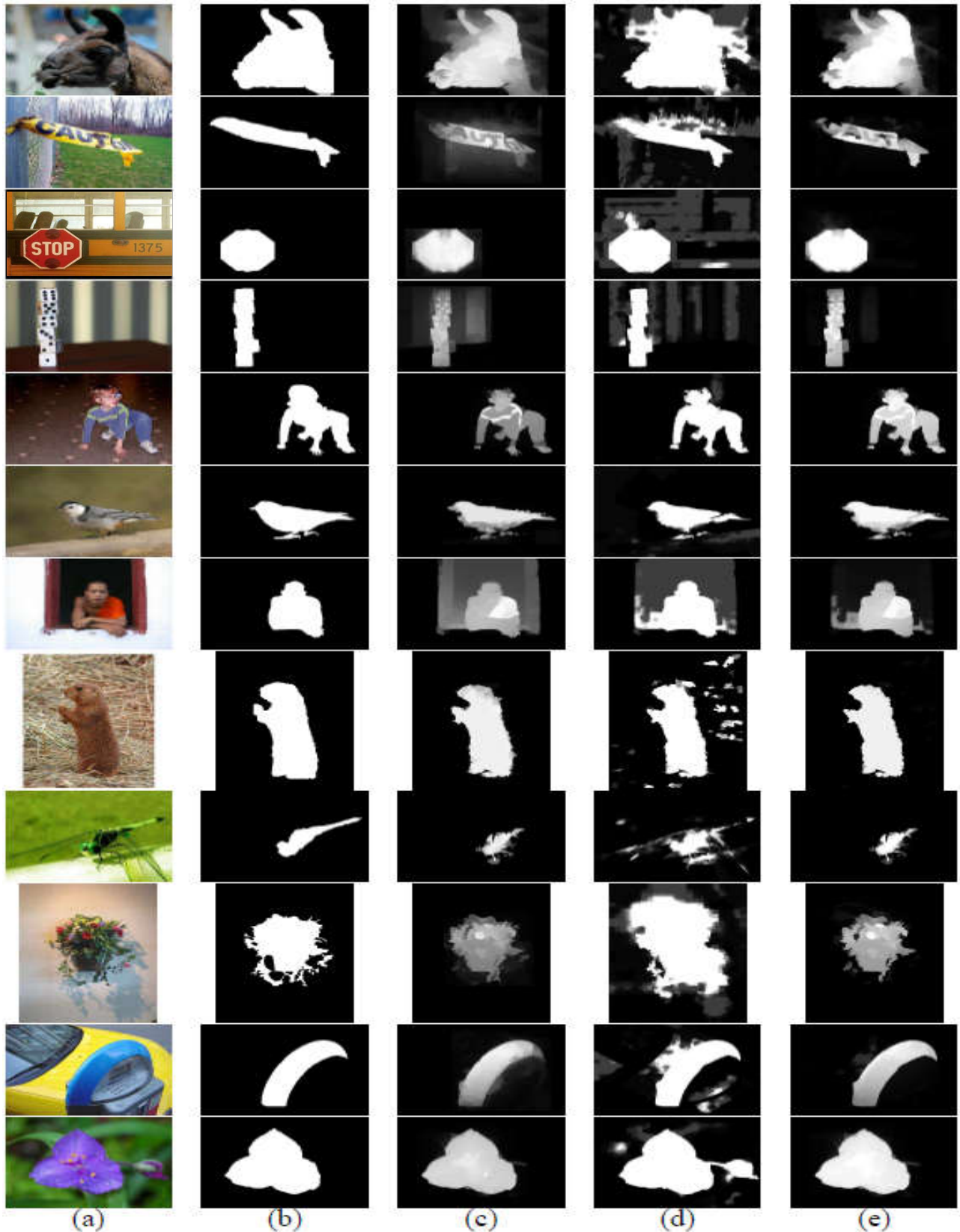


Figure 4.4 Exemple illustrant notre processus de raffinement global: (a) image en couleur, (b) vérité terrain, (c) carte de saillance obtenue en appliquant le raffinement local et la localisation spatiale, (d) résultat de prédiction de notre classifieur basé sur les *RF* et les *GMM*, (e) notre carte de saillance finale.

## 4.6 Conclusion

Nous avons présenté dans ce chapitre notre démarche de raffinement de la saillance qui permet d'améliorer considérablement les résultats. Le contraste entre l'objet saillant et l'arrière plan est augmenté et les contours de l'objet saillant mieux conservés. Nous avons défini notre processus de raffinement en trois étapes majeures.

D'abord, nous avons développé un processus de localisation spatiale des objets saillants par une fenêtre ajustée aux bords de l'objet saillant. Une telle disposition permet de restreindre l'espace d'estimation de la saillance qui se fera exclusivement dans la partie délimitée par les bords de la fenêtre. Appliquer notre détecteur d'objets saillant sur cet espace permet d'améliorer les résultats et de mettre à l'écart les parties de l'arrière-plan plus efficacement.

Ensuite, nous avons défini un processus de raffinement en introduisant conjointement l'information de contour et une procédure d'évaluation de la pertinence des caractéristiques utilisées. L'évaluation de la pertinence des caractéristiques permet de contrôler la contribution de chaque caractéristique selon son pouvoir discriminatif. L'information de contour quant à elle permet d'atténuer la diffusion des valeurs de l'ordonnancement à partir des parties du fond de l'image vers l'objet saillant. Ce processus permet de mieux ressortir l'objet saillant avec une meilleure localisation de ses contours.

Enfin, nous avons défini un processus de raffinement global afin de rester conforme aux statistiques globales de l'image. Nous nous sommes basés sur l'entraînement de classifieurs de type forêts aléatoires à travers plusieurs résolutions de l'image pour estimer la saillance au niveau globale. Ensuite, nous avons appliqué une démarche d'uniformisation de la saillance au niveau local. Cette démarche est basée sur les mixtures de gaussiennes. L'objet saillant est alors mieux contrasté du fond de l'image et l'arrière plan mieux ordonnancé. Nous allons présenter dans la prochaine section l'évaluation expérimentale de notre approche suivi d'une étude comparative avec les meilleures méthodes de l'état de l'art sur la détection d'objets saillants.

## CHAPITRE 5 EVALUATION EXPERIMENTALE

### 5.1 Introduction

Nous allons présenter dans ce chapitre l'évaluation expérimentale de notre approche en comparaison avec les meilleures méthodes de détection d'objets saillants que nous trouvons dans l'état de l'art. Pour une évaluation pertinente nous avons utilisé plusieurs (cinq) bases de tests parmi les plus utilisées dans l'état de l'art et une multitude de mesures d'évaluations. Nous allons commencer par présenter les bases de tests et les mesures d'évaluations utilisées puis nous nous attarderons à évaluer notre approche en comparaison avec 22 méthodes de l'état de l'art.

### 5.2 Bases de tests

Dans le cadre de l'évaluation de notre système de détection d'objets saillant, nous avons choisi les bases de tests suivantes :

**Microsoft MSRA-B** (Microsoft Research Area dataset) [75]: contient 5000 images en couleur incluant une large variété de scénarios comme les scènes naturelles (animaux, végétations, reliefs naturels), les scènes d'intérieur et d'extérieur, etc. La vérité terrain de cet ensemble d'images est fournie dans [54].

**ASD** (Accurate human labeled images for Saliency detection Dataset) [8]: contient 1000 images fournis dans [8] avec des masks binaires représentant la vérité terrain. Il s'agit de la base la plus utilisée dans l'évaluation des systèmes de détection d'objets saillant et un sous ensemble de la base MSRA-B.

**SOD** (Salient Object Dataset) [115] : contient 300 images extraites de la base *Berkeley segmentation dataset (BSD)* [116] pour lesquelles les bords des objets saillants sont annotés par sept utilisateurs. A partir de ces annotations, les masks binaires correspondant à la vérité terrain sont construits.

**SED1 et SED2** [117] : SED1 (Segmentation Evaluation Database on one salient object) est caractérisée par un seul objet saillant présentée dans chaque image tandis que SED2 (Segmentation Evaluation Database on two salient objects) est caractérisé par la présence de deux objets dans chaque image. Chacune de ces deux bases contient 100 images avec leur vérité terrain établi grâce à l’annotation manuelle.

### 5.3 Mesures d’évaluation

L’évaluation de la performance d’un système de détection d’objets saillants peut être effectuée en mesurant la cohérence des cartes de saillances générées avec la vérité terrain issue de l’annotation humaine. Pour cela, nous avons choisi d’utiliser les métriques suivantes: la courbe “*précision-rappel*” (*PR*), la mesure  $F_{\mu}$ , la courbe “*sensibilité/spécificité*” ou *receiver operating characteristic (ROC)*, la mesure “*aire sous la courbe ROC*” ou *area under the ROC curve (AUC)* et “*l’erreur moyenne absolue*” ou *mean absolute error (MAE)*. Pour une carte de saillance donnée, chaque pixel est défini dans l’intervalle  $[0,255]$ . Nous définissons un seuil  $T$  pour générer un mask binaire qui affecte la valeur 1 à l’objet saillant et la valeur 0 à l’arrière-plan. En faisant varier  $T$  de 1 à 254, nous obtenons 254 masks binaires qui sont comparés à la vérité terrain.

Pour chaque mask binaire, la précision ( $P$ ) représente le pourcentage de pixels saillants correctement classifiés. Soit  $M$  le mask binaire généré et  $G$  la vérité terrain. Cette mesure est définie comme suit:

$$P = \frac{|M \cap G|}{|M|} \quad (5.1)$$

Pour chaque mask binaire, le rappel ( $R$ ) correspond au pourcentage de pixels saillants retournés au sein de l’objet saillant. Cette mesure est calculée comme suit:

$$R = \frac{|M \cap G|}{|G|} \quad (5.2)$$

La mesure  $F_{\mu}$  (*F-measure*) correspond à une moyenne harmonique pondérée entre la précision et de rappel calculée comme suit:

$$F_{\mu} = \frac{(1 + \mu^2) \times P \times R}{\mu^2 \times P + R} \quad (5.3)$$

avec  $\mu$  un paramètre positif qui contrôle l'importance attribuée à  $P$  et  $R$ . Habituellement comme suggéré dans plusieurs travaux sur la détection d'objets saillants (par exemple, voir [8]),  $\mu^2$  est souvent ajusté à 0.3 afin d'accorder plus d'importance à la précision. La raison de pondérer la précision plus que le rappel est dû au fait que le rappel n'est pas une mesure aussi fiable que la précision (voir [75]). Par exemple, 100% de rappel peut aisément être atteint en considérant toute l'image entière comme objet saillant. Pour chaque seuil,  $T \in \{1, 2, \dots, 254\}$ , nous calculons la valeur moyenne de la précision et la valeur moyenne du rappel à travers toutes les images à évaluer. A partir de ces valeurs, nous construisons La courbe  $P$ - $R$ .

En plus de la précision, du rappel et de la mesure  $F_{\mu}$ , nous pouvons aussi évaluer nos masks binaires avec les mesures *false positive rate* ( $FPR$ ) et *true positive rate* ( $TPR$ ).  $TPR$  correspond au rappel déjà mentionné en haut et  $FPR$  correspond au taux de pixels non saillants non correctement classifiés parmi tous les pixels de l'arrière-plan. Cette mesure est définie comme suit:

$$FPR = \frac{|M| - |M \cap G|}{|\bar{G}|} \quad (5.4)$$

De la même manière que pour la précision et le rappel, nous calculons pour chaque seuil  $T$ , la valeur moyenne de  $TPR$  et la valeur moyenne de  $FPR$  à travers toutes les images à évaluer. A partir de ces valeurs, nous construisons La courbe  $ROC$ .

Alors que la courbe  $ROC$  est représenté en un espace à deux dimensions, la mesure  $AUC$  permet quant à elle d'interpréter la courbe par un seul scalaire.  $AUC$  permet de mesurer la surface sous la courbe  $ROC$ . Plus cette surface est grande, plus la segmentation est meilleure. Un modèle parfait retournerait une valeur de  $AUC$  égale à 1.

Toutes les mesures d'évaluation que nous avons définies jusqu'ici ne tiennent pas en compte le taux de vrais négatifs dans les objets saillants retournés

(les pixels correctement classifié comme non saillants). Omettre une telle information dans l'évaluation, permet de favoriser les modèles qui affectent des valeurs élevées aux pixels saillants mais échouent dans la détection des régions non saillantes. Pour une évaluation plus précise et complète, nous utilisons la mesure *mean absolute error (MAE)* pour une comparaison de la carte de saillance  $M$  avec la vérité terrain normalisées conjointement dans l'intervalle  $[0, 1]$ . La mesure  $MAE$  est définie comme suit:

$$MAE = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H \|M(x, y) - G(x, y)\| \quad (5.5)$$

Avec  $W$  et  $H$  correspondent à la largeur et la longueur de l'image respectivement.

Nous notons que le coin supérieur droit dans la courbe  $P-R$  représenté par les coordonnées  $(1,1)$  correspond à une détection parfaite. Plus la courbe est proche de ce point, plus le système de détection de la saillance est meilleur. D'une manière similaire, plus la courbe ROC est proche du coin supérieur gauche représenté par les coordonnées  $(0,1)$ , plus le système de détection de saillance est meilleur. Afin de visualiser la meilleure valeur moyenne de  $F_\mu$  ainsi que les valeurs de  $AUC$  et  $MAE$ , nous utilisons les diagrammes à bar.

#### 5.4 Évaluation expérimentale de notre algorithme

Nous allons d'abord justifier le choix sur lequel nous nous sommes portés sur le nombre de résolutions utilisés dans notre modèle. Généralement, plus de résolutions sont prises en compte, plus nous obtenons de meilleurs résultats. Cependant, plus la structure de notre graphe possède de couche, plus la complexité de calcul est élevée. Nous avons effectué des tests sur 500 images de la base MSRA-B choisis aléatoirement en variant le nombre de résolutions utilisées (cf. figure 5.1). Nous constatons que la prise en compte de résolutions ayant une granularité plus fine que 800 superpixels n'améliore pas significativement les résultats. Pour un bon compromis entre l'efficacité de notre système et une complexité de calcul raisonnable, nous avons choisi d'utiliser quatre résolutions avec les granularités suivantes: 200, 400, 600 et 800 superpixels respectivement.



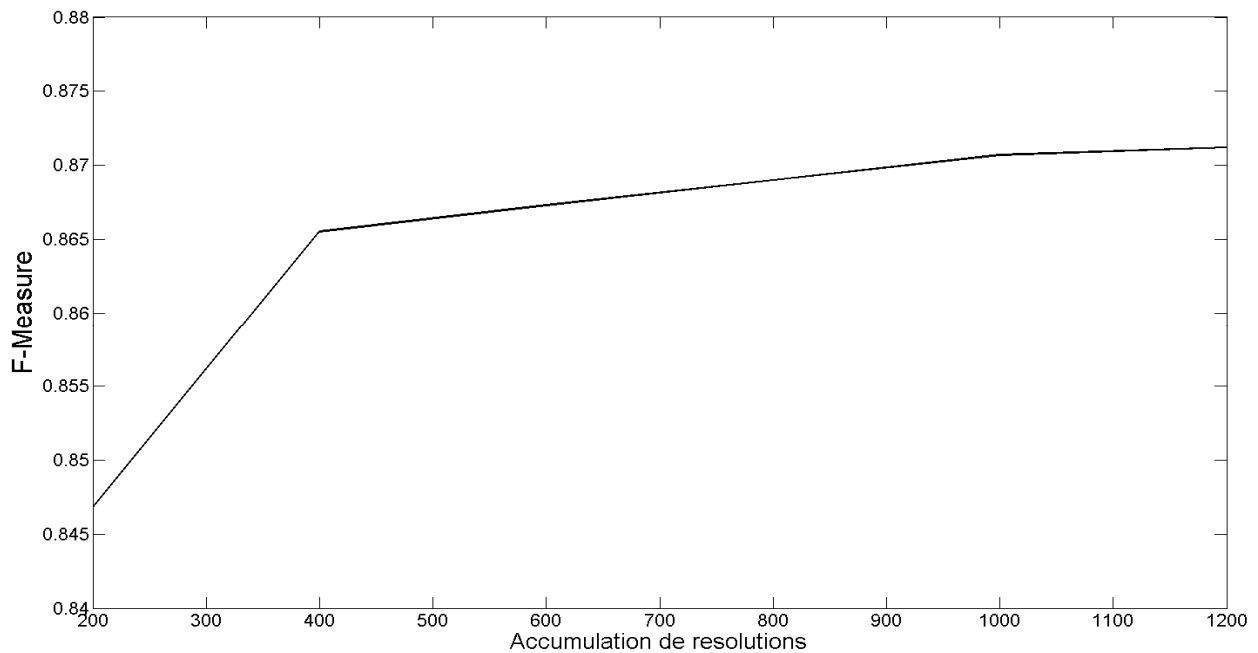


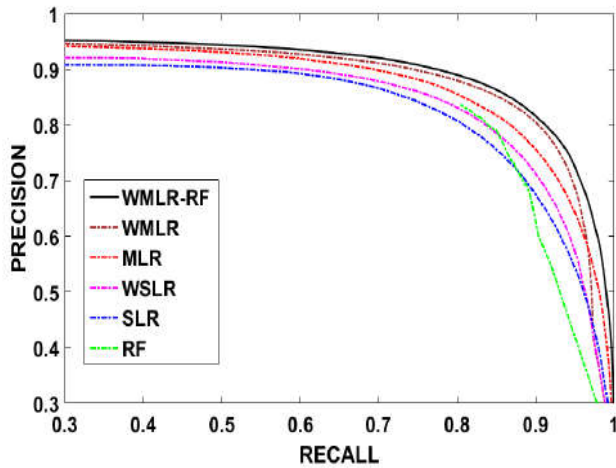
Figure 5.1 Évolution de la mesure  $F_{\mu}$  en fonction du nombre croissant des résolutions sur 500 images de la base MSRA-B

Nous avons évalué notre algorithme et ses différentes variantes sur les bases *MSRA-B* et *ASD* [123]:

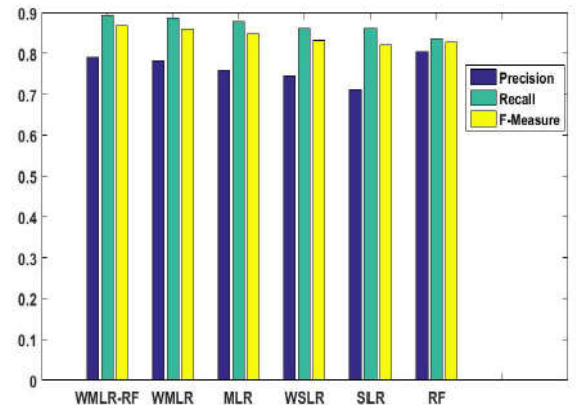
- 1) saillance obtenue par un graphe mono-couche en appliquant la fonction 2.48 sur l'image notée *SLR*,
- 2) saillance obtenu après avoir effectué notre localisation spatiale et application de la fonction 2.48 noté *WSLR*,
- 3) saillance obtenue par notre graphe multi-couche en appliquant la fonction 3.9 notée(*MLR*),
- 4) saillance obtenue par le raffinement local et la localisation spatiale de notre graphe multi-couches noté *WMLR*,
- 5) saillance obtenue par notre classifieur global basé sur les RF et GMM noté *RF*,
- 6) notre saillance finale obtenu par notre méthode après l'application de tous les raffinements considérés notée *WMLR-RF*.

Les résultats sont montrés dans les figures 5.2 et 5.3. Nous pouvons observer que *MLR* surpasse clairement *SLR* et *WSLR* pour les deux bases de tests, tandis que *WMLR* et *WSLR* améliorent les résultats de *MLR* et *SLR* respectivement. Nous notons également que l'application de *RF* à elle toute seule ne retourne pas des résultats compétitifs par rapport aux autres approches.

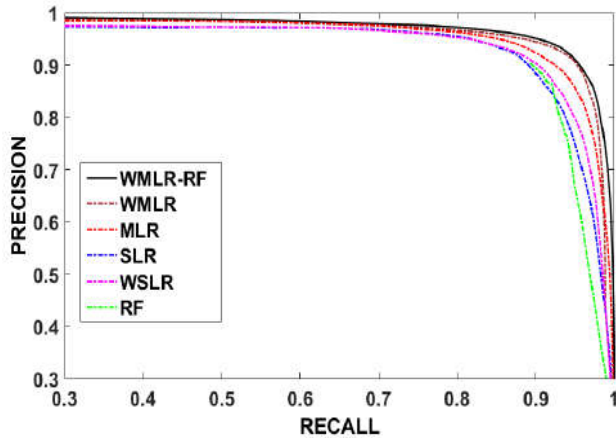
Cependant, la combinaison de *WMLR* et *RF* (*WMLR-RF*) atteint une meilleure performance et les résultats retournés pour les mesures (*F-measure*, *AUC*, *MAE*) sont (85.78%, 85.40%, 10.76%) dans MSRA-B et (92.92%, 87.95%, 6.67%) dans ASD respectivement. La seconde meilleure performance est obtenue par *WMLR* avec les résultats suivants (85.06%, 84.11%, 11.74%) dans MSRA-B et (92.43%, 87.64%, 7.74%) dans ASD respectivement.



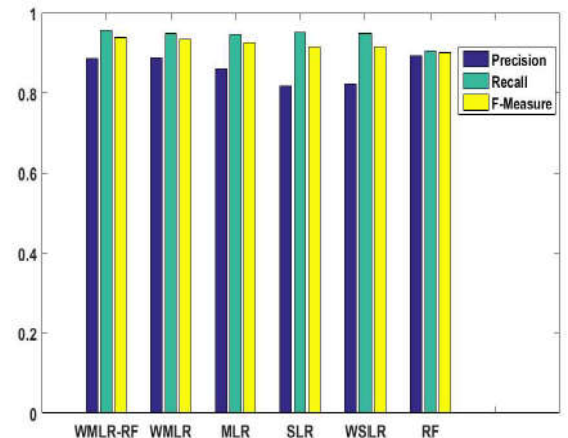
(a)



(b)

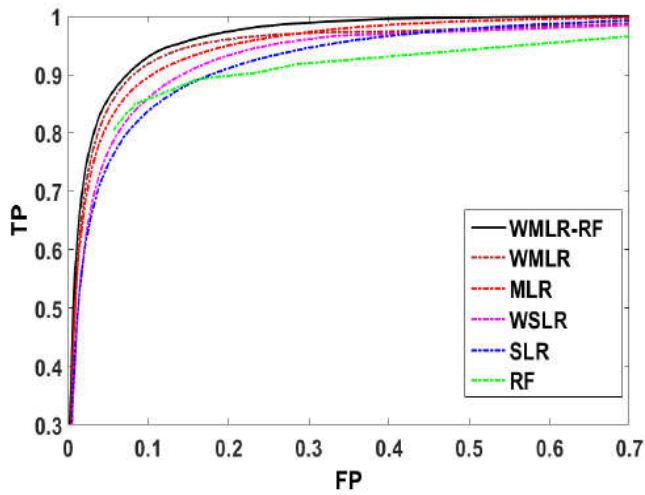


(c)

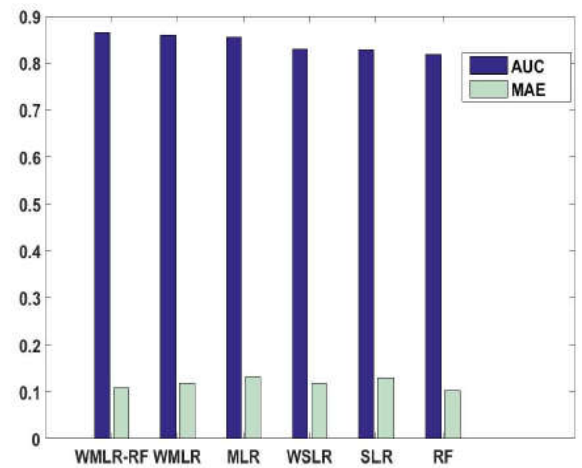


(d)

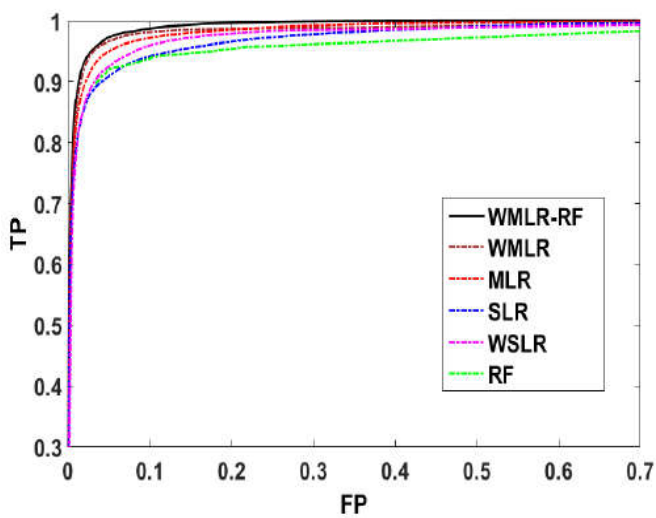
Figure 5.2 Résultats comparatifs des différentes méthodes *SLR*, *WSLR*, *RF*, *MLR*, *WMLR* et *WMLR-RF* en utilisant: (a,b) la base MSRA-B et (c,d) la base ASD. Pour chaque base de tests, le graphe de gauche correspond à courbe *P-R* et le graphe de droite aux meilleures valeurs de  $F_{\mu}$  et les valeurs de *P*, *R* correspondantes.



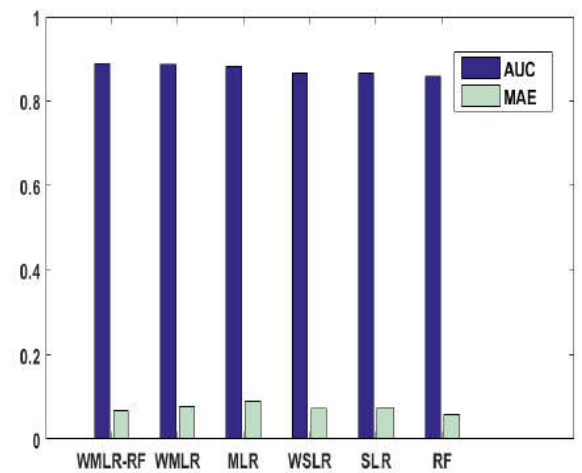
(a)



(b)



(c)



(d)

Figure 5.3 Résultats comparatifs des différentes méthodes *SLR*, *WSLR*, *RF*, *MLR*, *WMLR* et *WMLR-RF* en utilisant: (a,b) la base MSRA-B et (c,d) la base ASD. Pour chaque base de tests, le graphe de gauche correspond à courbe *ROC* et le graphe de droite aux valeurs de *AUC* et *MAE*.

### 5.5 Étude comparative sur les bases MSRA et ASD

Nous comparons notre algorithme à 22 méthodes de l'état de l'art sur la détection d'objets saillants: Context and shape prior (*CS*) [55], Human visual sensitivity and amplitude spectrum (*VSAS*) [118], High dimensional color transform and local spatial support (*HDCT*) [90], Saliency optimization from robust Background detection (*RBD*) [82], Structured matrix decomposition (*SMD*) [17], Segmentation driven low-rank matrix recovery (*SLMR*) [88], Hierarchical saliency

detection (*HSL*) [18], Joint spatial-color constraint and multi-scale segmentation (*SCC*) [119], Saliency Tree (*ST*) [10], Unified approach to salient object detection via Low rank matrix recovery (*LMR*) [87], Bayesian Saliency via Low and Mid Level Cues (*BS*) [77], single-Layer Graph-Based Manifold ranking (*SLR*) [9], PCA based method (*PCA*) [120], Combining simple priors method (*CSP*) [121], Low level features based on wavelet transform (*WT*) [81], multi-size superpixel approach based on multivariate normal distribution estimation (*MSN*) [11], Salient region detection via simple local and global contrast (*LGC*) [91], Salient Region Detection with Soft Image Abstraction (*SIA*) [6], Saliency Detection Via Dense and Sparse Reconstruction (*DSR*) [86], Saliency detection via absorb Markov chain (*MC*) [54], Extended Quantum Cuts (*EQCUT*) [15] et Discriminative Regional Feature Integration Approach (*DRFI*) [16]. Nous utilisons les codes sources avec les paramètres par défaut fournis par les auteurs de toutes ces méthodes. Nous notons que *DRFI* est la seule de ces méthodes qui est basée sur une phase d'apprentissage. Les auteurs utilisent 2500 images de la base de test MSRA-B pour la phase d'apprentissage. A cet effet, exceptionnellement pour la comparaison de notre méthode avec *DRFI*, nous utiliseront 2500 images de MSRA-B qui n'ont pas servi dans la phase d'apprentissage.

Nous pouvons observer dans les figures 5.4, 5.6.a et 5.7 que la courbe générée par notre méthode *WMLR-RF* est la plus haute et la plus proche du point optimale (1,1) que toutes les autres courbes générées par les autres méthodes dans les deux bases MSRA-B et ASD. Pour chaque base de test, les résultats des mesures *P*, *R* et *F-measure* sont illustrés sur les figures 5.5 et 5.6.b pour MSRA-B et sur la figure 5.8 pour ASD. Nous remarquons que la valeur de *F-measure* la plus importante est toujours retournée par notre méthode. Cela démontre clairement l'efficacité de notre système de détection d'objets saillants. Nous notons également que même notre approche *MLR* démontre une performance meilleure que la majorité des autres méthodes. Après notre approche *WMLR-RF*, les meilleurs résultats sont ceux retournés par *DRFI*, *EQCUT*, *SMD*, *ST*, *MSN*, *HDCT*, *RBD*, *MR*, *HSL*, *MC*, *SLR* et *DSR* respectivement.

Nous effectuons alors une seconde évaluation comparative de notre approche *WMLR-RF* et sa variante *MLR* avec ces 12 méthodes en utilisant la courbe *ROC* et les mesures *AUC* et *MAE*. Nous notons que d'une manière

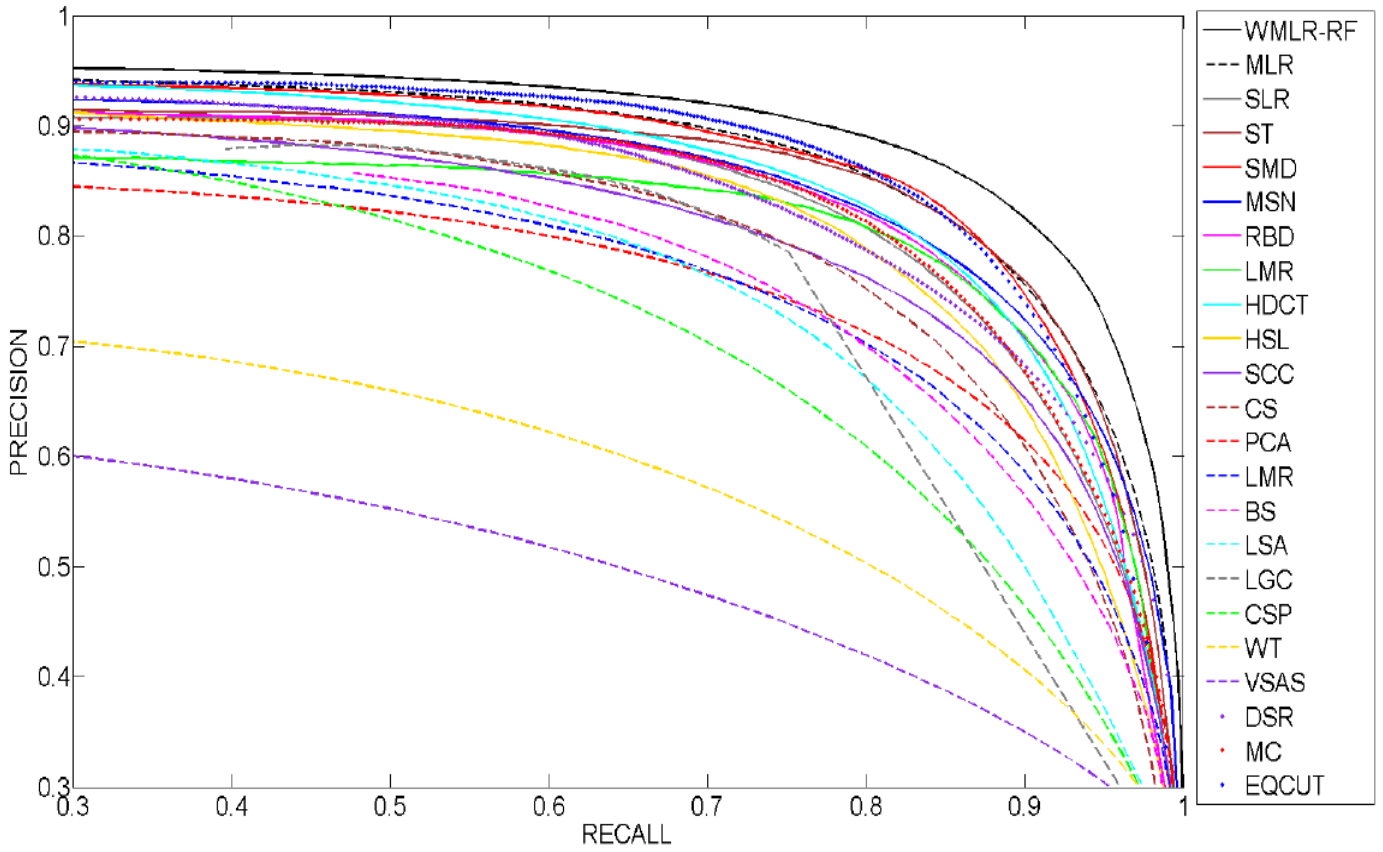


Figure 5.4 Courbes P-R obtenus pour 21 méthodes de l'état de l'art en comparaison avec notre méthode en utilisant la base MSRA-B

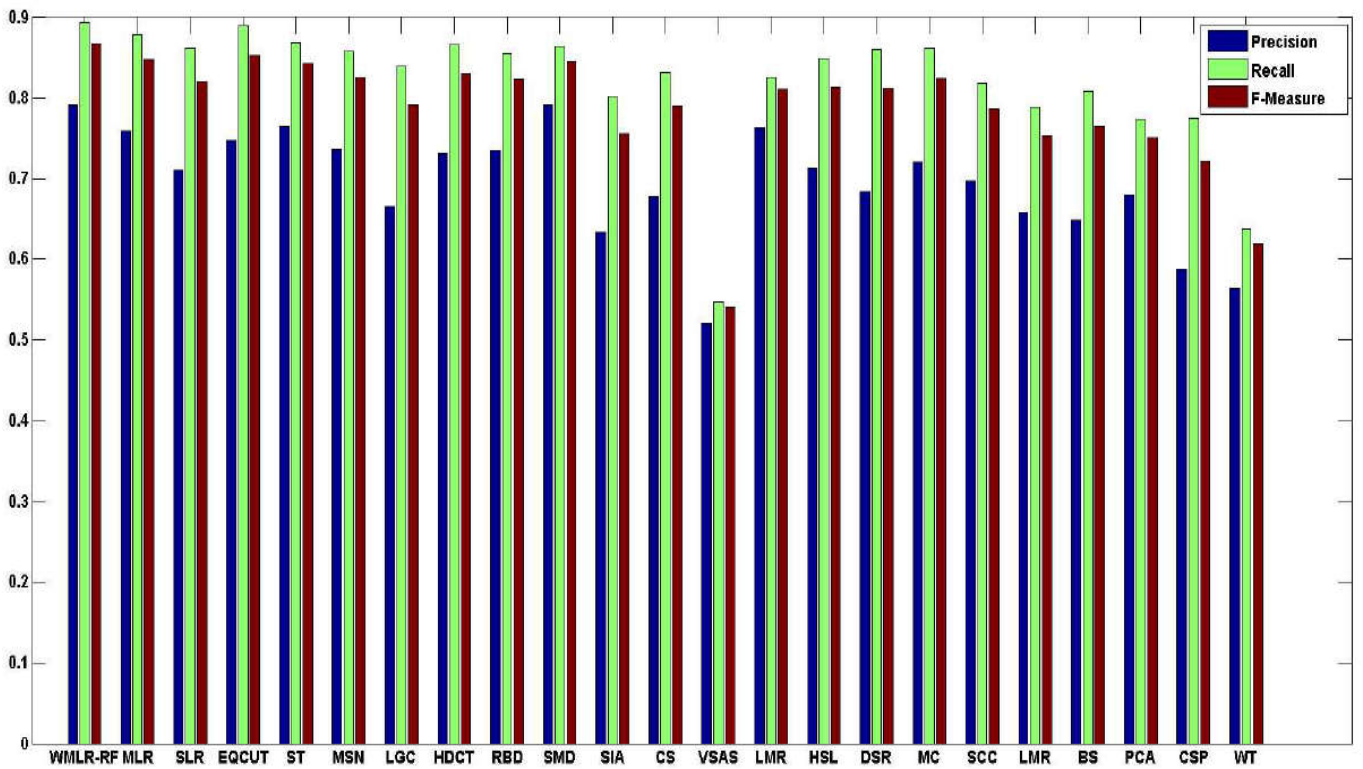


Figure 5.5 Valeurs obtenues pour les mesures  $P$ ,  $R$  et  $F$ -measure pour 21 méthodes de l'état de l'art en comparaison avec notre méthode sur MSRA-B

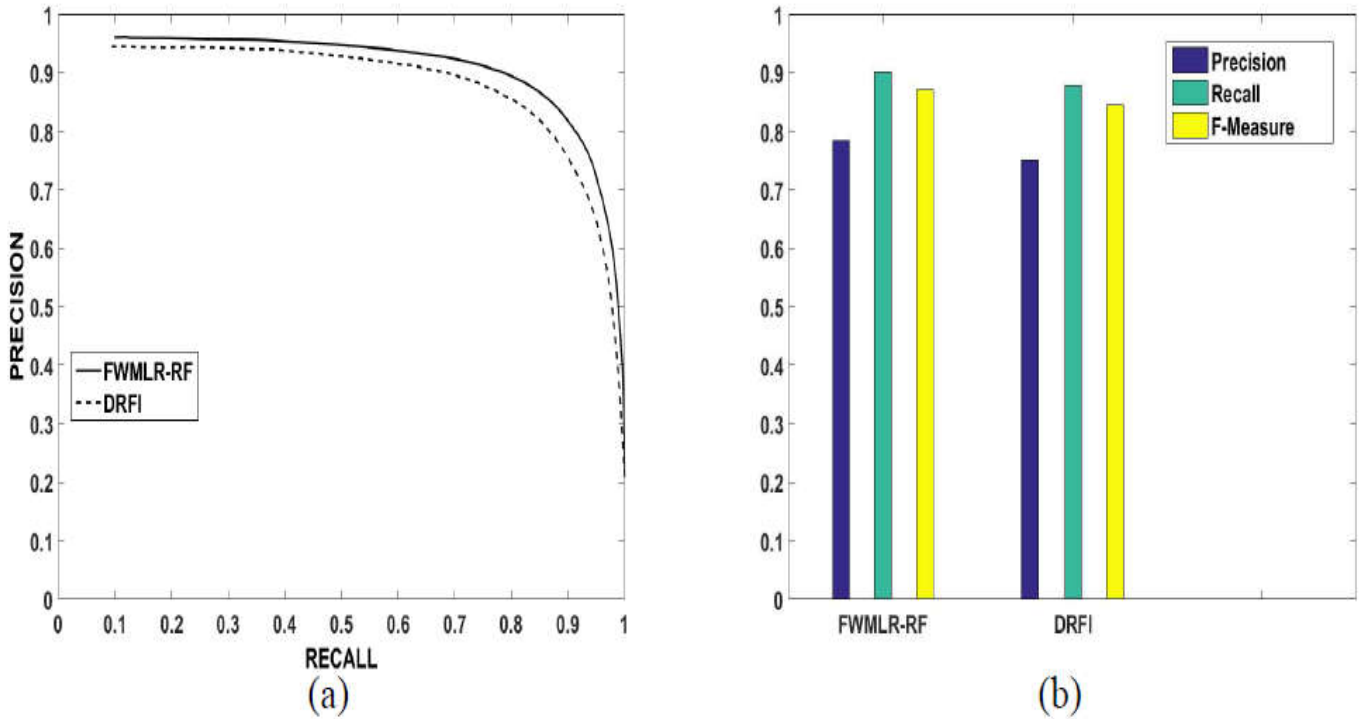


Figure 5.6 Comparaison de notre méthode avec *DRFI* sur 2500 images de MSRA-B: (a) courbes P-R, (b) meilleure valeur de F-measure et les valeurs de *PR* correspondantes.

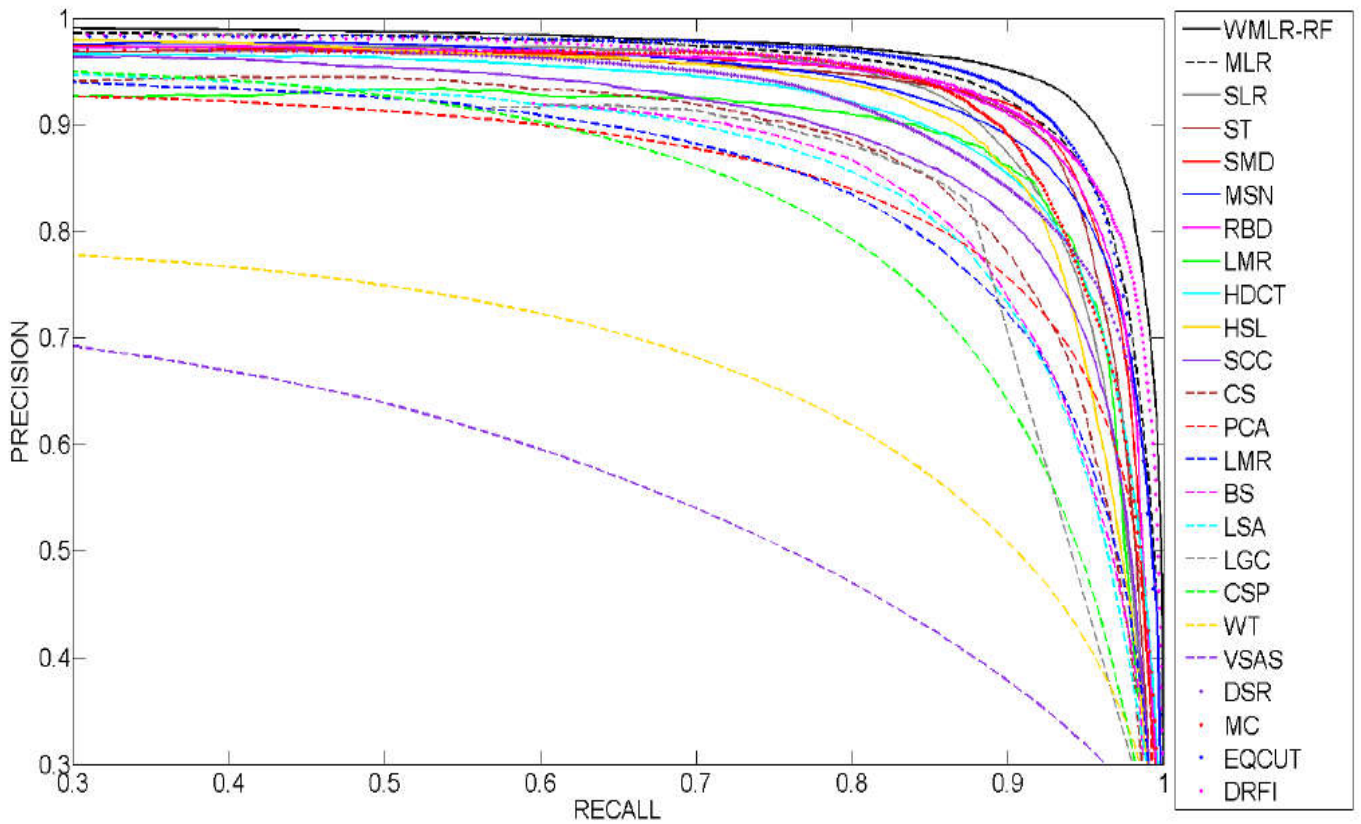


Figure 5.7 Courbes P-R obtenus pour 22 méthodes de l'état de l'art en comparaison avec notre méthode en utilisant la base ASD

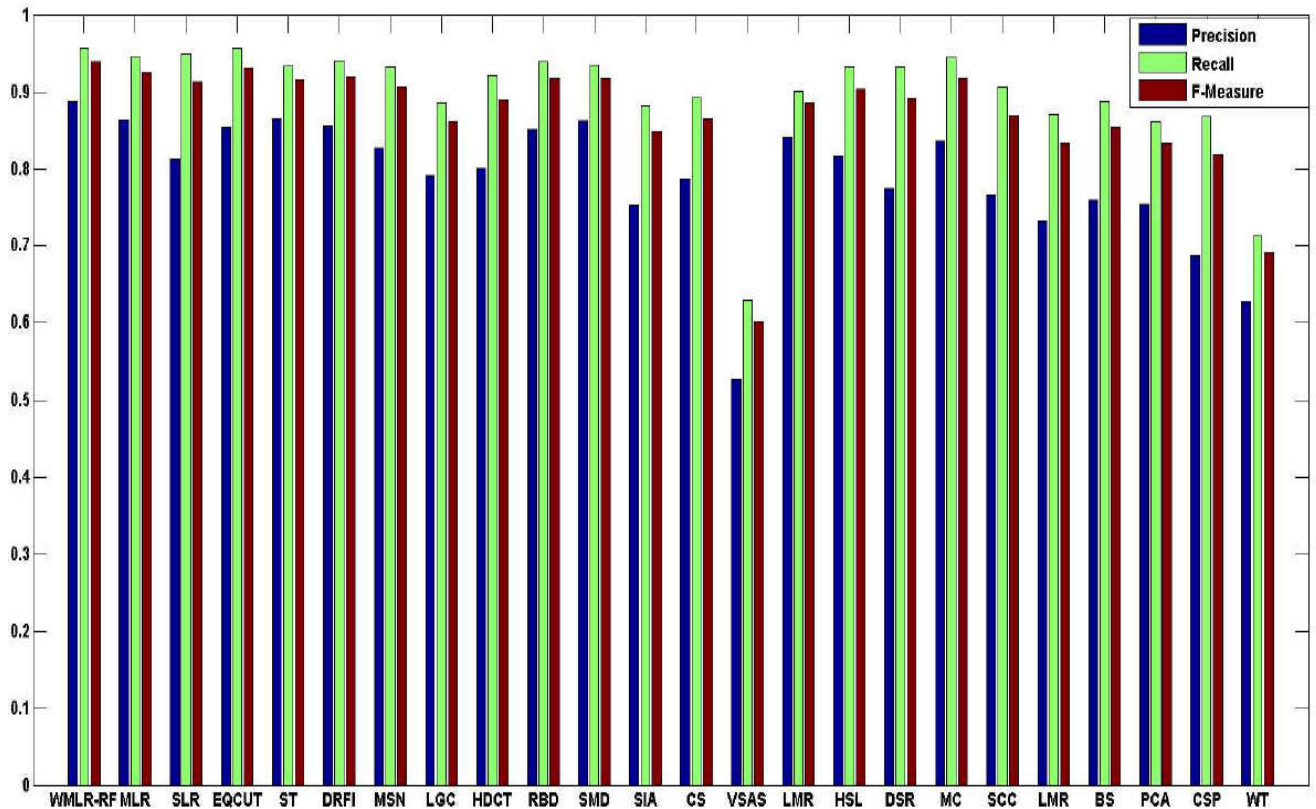


Figure 5.8 Valeurs obtenue pour les mesures  $P$ ,  $R$  et  $F$ -measure pour 21 méthodes de l'état de l'art en comparaison avec notre méthode sur ASD

similaire que précédemment, nous effectueront une comparaison de notre méthode avec *DRFI* non pas sur toute la base de test MSRA-B mais uniquement sur 2500 images qui n'ont pas servi dans la phase d'apprentissage de *DRFI*. Les figures 5.9, 5.10 et 5.11 montrent les performances obtenues pour chaque méthode respectivement. Nous pouvons noter que la courbe *ROC* générée par notre approche *WMLR-RF* est la plus haute que celles générées par les autres méthodes de l'état de l'art aussi bien dans MSRA-B que dans ASD. La valeur la plus importante de la mesure *AUC* et la valeur la plus petite de la mesure *MAE* sont aussi retournées par *WMLR-RF*.

Nous pouvons alors conclure que notre approche *WMLR-RF* retourne les valeurs les plus élevées pour les mesures *F-measure* et *AUC* la valeur la plus petite pour la mesure *MAE* dans les deux bases de tests MSRA-B et ASD. Ceci démontre la qualité de notre approche et la supériorité de notre méthode par rapport aux 22 autres méthodes de l'état de l'art.

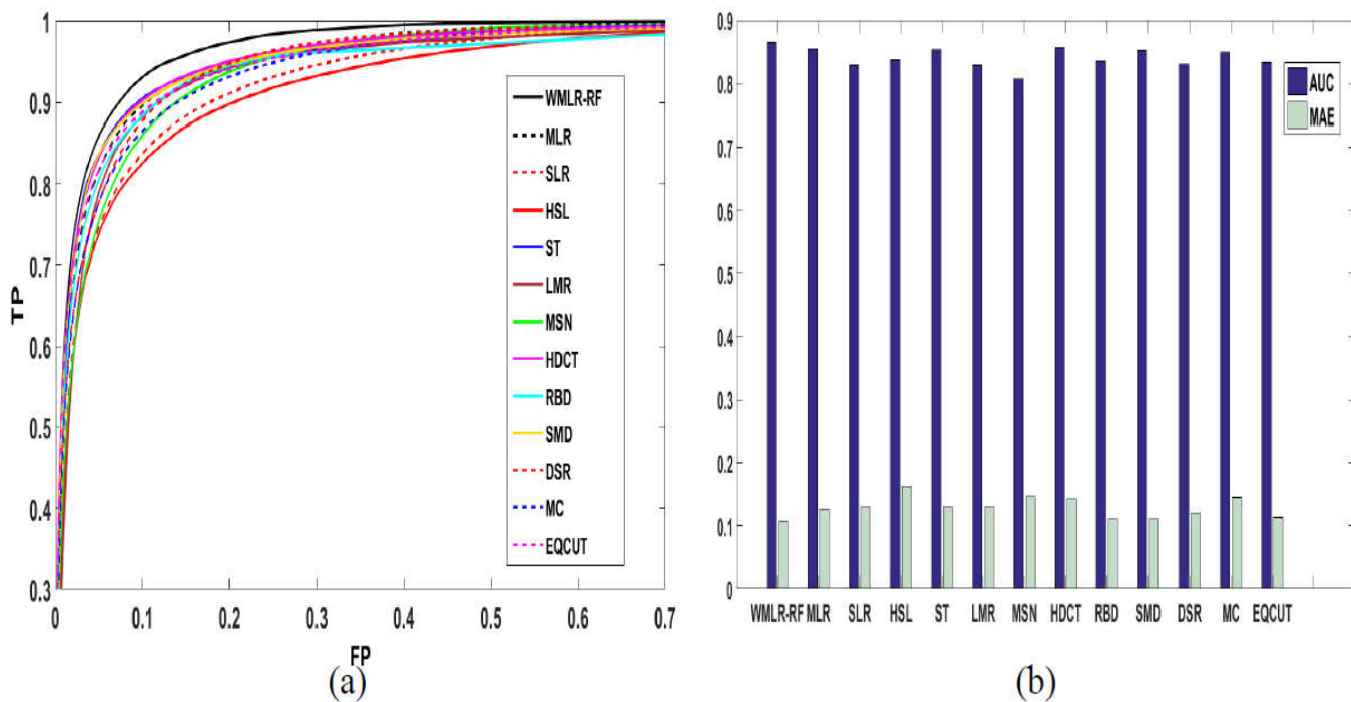


Figure 5.9 Résultats comparatifs entre notre méthode et les onze meilleures méthodes de l'état de l'art sur la base MSRA-B: (a) courbes ROC obtenues, (b) valeur des mesure AUC et MAE.

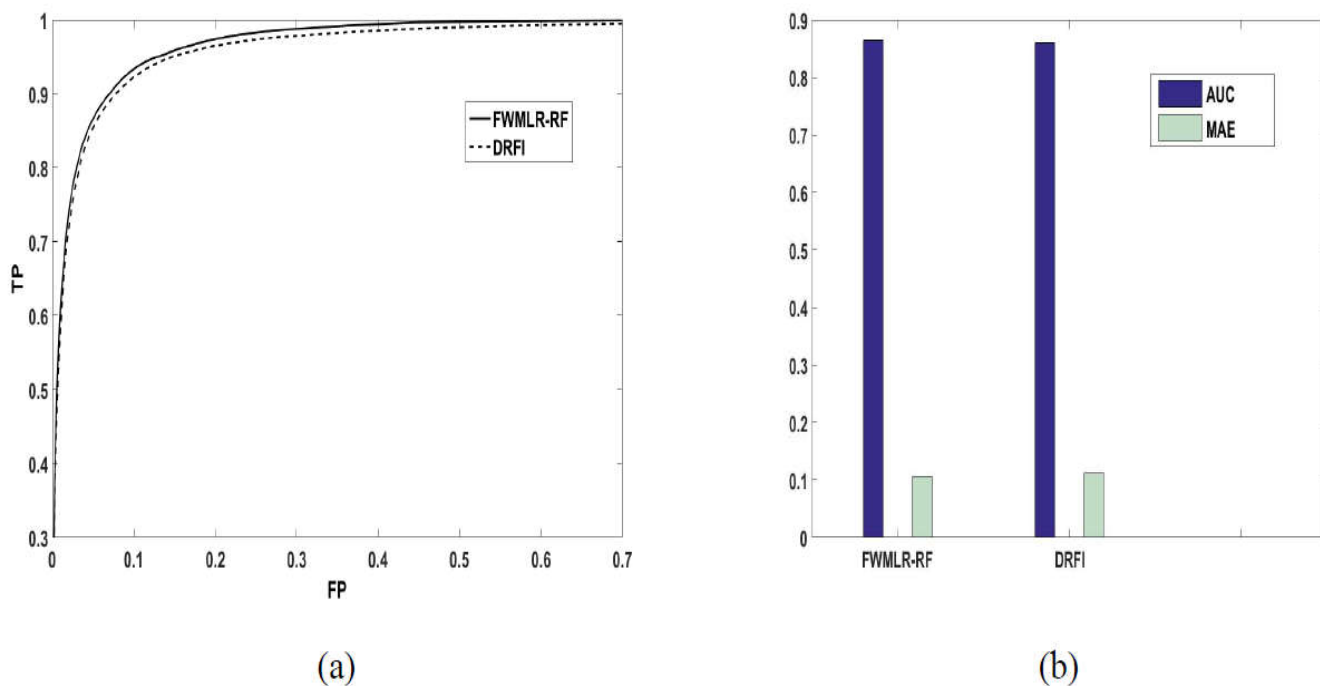


Figure 5.10 Résultats comparatifs entre notre méthode et *DRFI* sur 2500 image de la base MSRA-B: (a) courbes ROC obtenues, (b) valeur des mesure AUC et MAE.



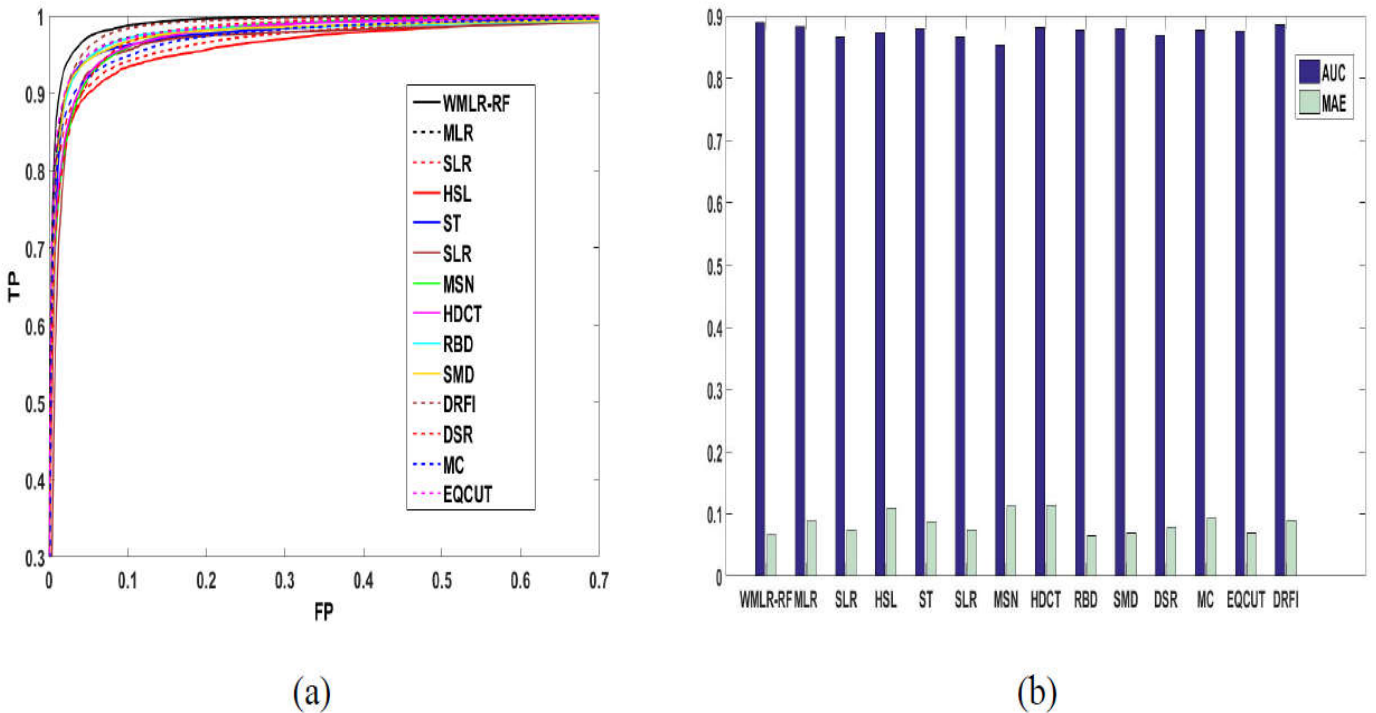


Figure 5.11 Résultats comparatifs entre notre méthode et les douze meilleures méthodes de l'état de l'art sur la base ASD: (a) courbes ROC obtenues, (b) valeur des mesure AUC et MAE.

### 5.6 Étude comparative sur les bases SOD, SED1 et SED2

Pour une évaluation expérimentale plus complète, nous étendons notre étude comparative sur les bases de test: SOD, SED1 et SED2. Nous effectuons dans ces bases une comparaison entre notre méthode avec les 12 meilleures méthodes citées dans la section 5.5 à savoir *DRFI*, *EQCUT*, *SMD*, *ST*, *MSN*, *HDCT*, *RBD*, *MR*, *HSL*, *MC*, *SLR* et *DSR*. Les figures 5.12, 5.13 et 5.14 montrent les résultats obtenus pour les 3 bases de test en utilisant toutes les mesures mentionnées dans la section 5.3.

Nous pouvons constater que les performances des douze méthodes de l'état de l'art varient selon la base de test utilisée. Les courbes *P-R* (partie (a) de chacune des trois figures) et ROC (partie (c) de chacune des trois figures) générées par notre méthode sont plus hautes que toutes les autres courbes générées par les autres méthodes. Notre courbe *P-R* est la plus proche du point optimal (11) et notre courbe ROC est plus proche du point optimal (0,1). Les diagrammes à barres quant à eux montrent clairement la supériorité de notre méthode. Notre méthode retourne le taux de *F-measure* et le taux de *AUC* les plus élevés ainsi

que le taux de *MAE* le plus bas (voir les parties (b) pour les taux de *F-measure* et les parties (d) pour les taux de *MAE* et *AUC* sur les figures 5.12, 5.13 et 5.14).

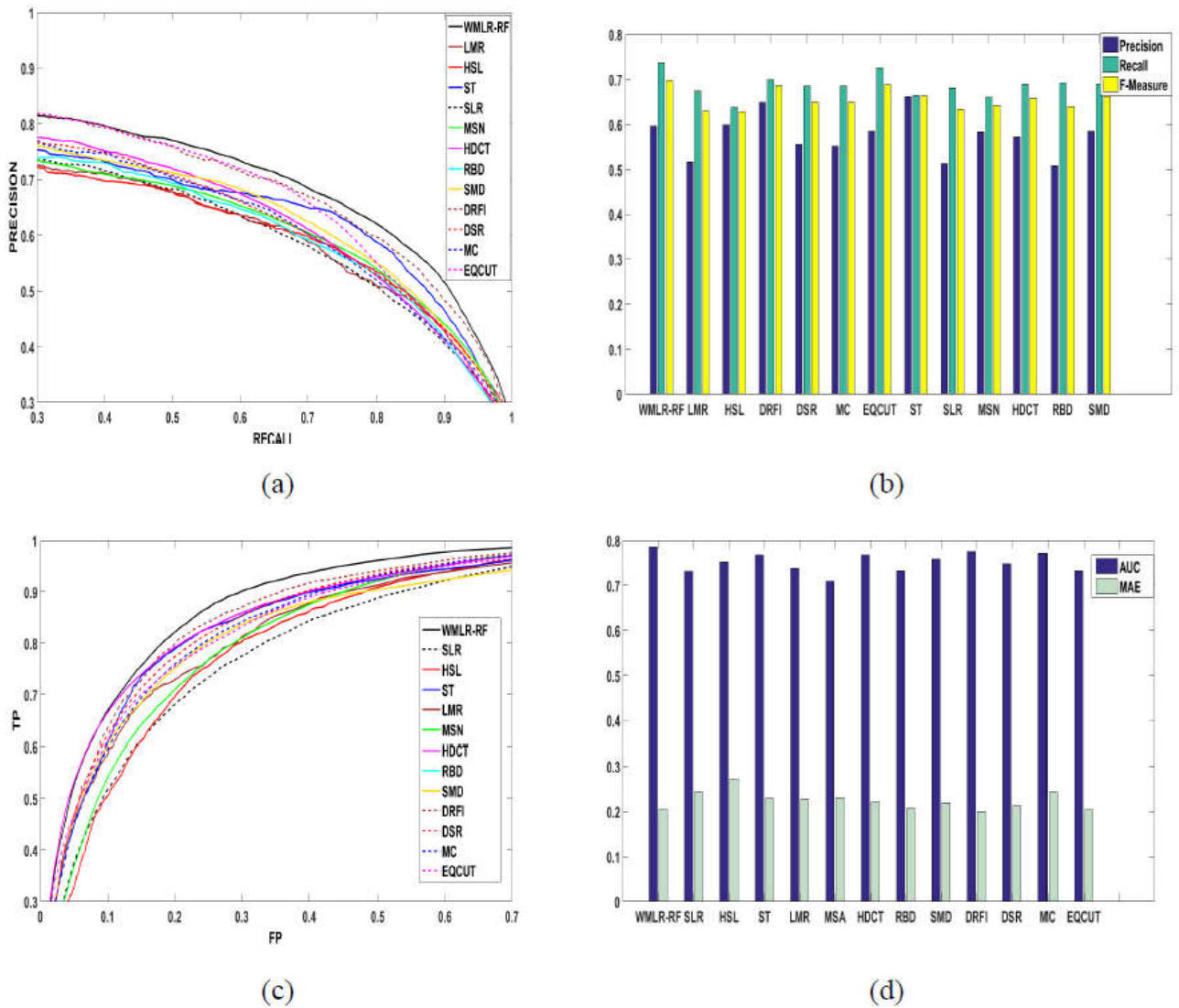


Figure 5.12 Résultats comparatifs entre notre méthode et les 12 meilleures méthodes de l'état de l'art sur la base de test SOD: (a) courbe P-R, (b) taux de *P*, *R* et *F-measure*, (c) courbe ROC, (d) taux de *AUC* et *MAE*.

Dans la base *SOD*, après notre méthode, *DRFI* montre une meilleure performance par rapport aux autres méthodes restantes, suivi de *SMD* et *EQCUT* en considérant la mesure *F-measure* (cf. figure 5.12). Concernant les mesures *AUC* et *MAE*, les meilleurs résultats sont retournées par *DRFI*, *HDCT* et *ST* respectivement.

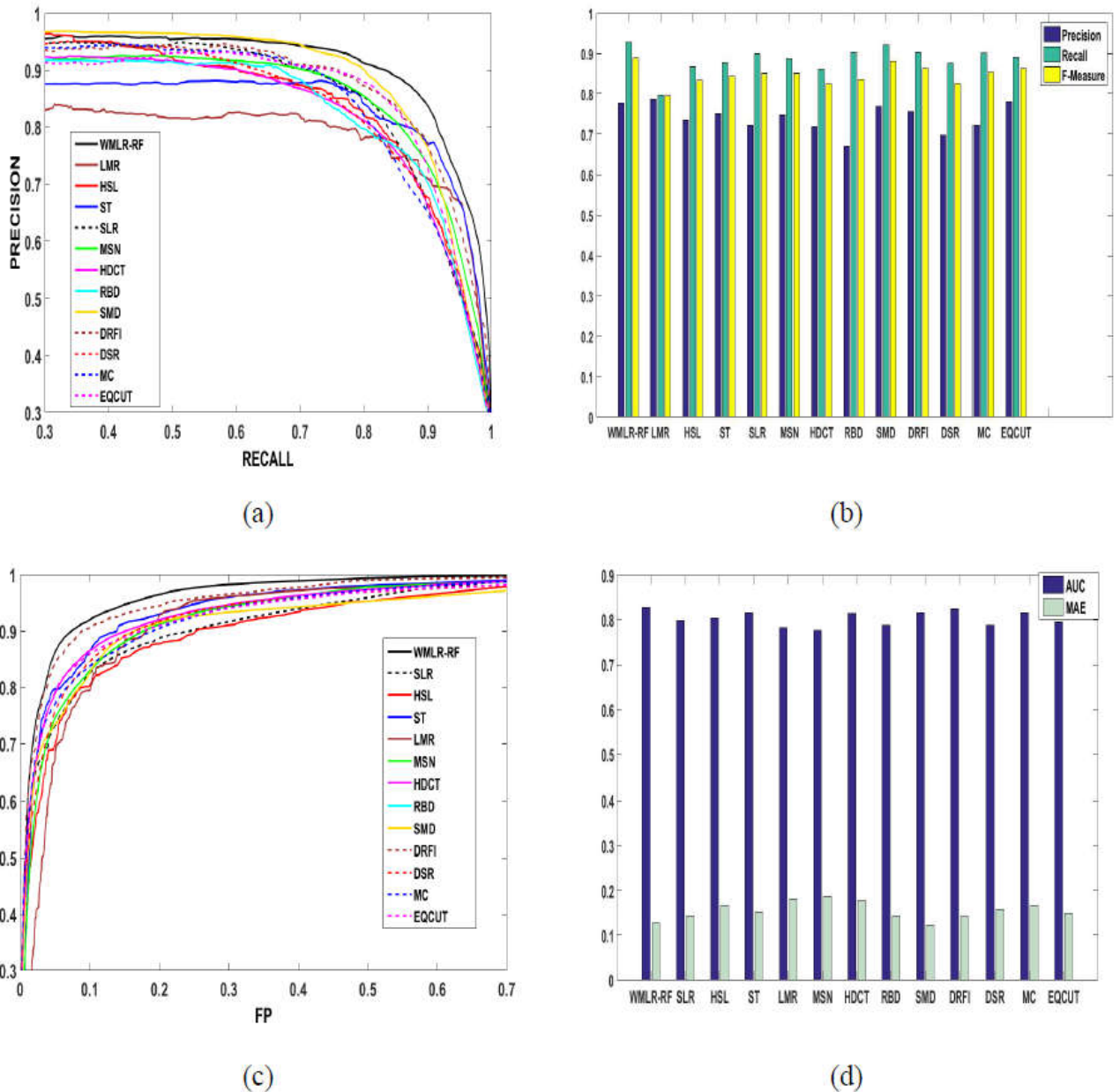


Figure 5.13 Résultats comparatifs entre notre méthode et les 12 meilleures méthodes de l'état de l'art sur la base de test SED1 : (a) courbe P-R, (b) taux de  $P$ ,  $R$  et  $F$ -measure, (c) courbe ROC, (d) taux de AUC et MAE.

Dans la base *SED1*, les meilleurs taux de  $F$ -measure après celui retourné par notre méthode sont ceux des méthodes *SMD*, *EQCUT* et *DRFI* (cf. figure 5.13). En ce qui concerne les mesures AUC et MAE, les meilleurs résultats après ceux retournés par notre approche sont ceux retournés par les méthodes *DRFI*, *SMD*, *MC* et *HDCT* respectivement.

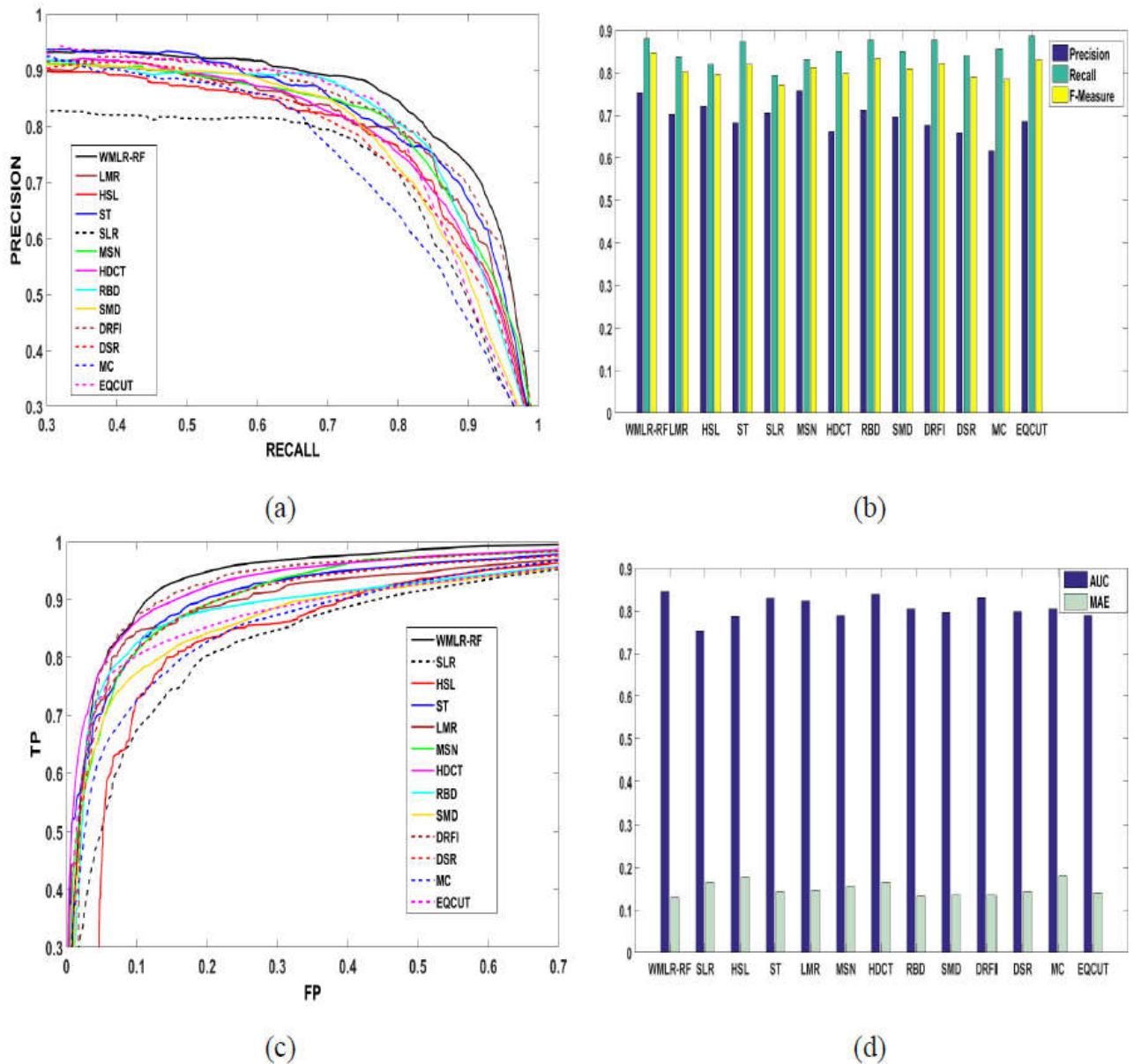


Figure 5.14 Résultats comparatifs entre notre méthode et les 12 meilleures méthodes de l'état de l'art sur la base de test SED2 : (a) courbe P-R, (b) taux de  $P$ ,  $R$  et  $F$ -measure, (c) courbe ROC, (d) taux de AUC et MAE.

Dans la base SED2, les meilleurs taux de  $F$ -measure après celui retourné par notre méthode sont ceux des méthodes *EQCUT*, *RBD* et *RDFI* (cf. figure 5.14). Concernant les mesures AUC et MAE, les meilleurs résultats après ceux retournés par notre approche sont ceux retournés par les méthodes *DRFI*, *HDCT* et *ST*.

D'une manière générale la méthode qui possède des performances qui se rapprochent le plus de notre méthode est *DRFI*. Comme cette méthode est basée

sur une phase d'apprentissage, elle serait incapable de se généraliser sur des images qui contiennent des objets saillants dont la structure et la composition est différente de ceux qui se trouvent dans la base d'entraînement. La figure 5.15 montre des cas où DRFI n'arrive pas à bien se généraliser en comparaison avec notre méthode *WMLR-RF*.



Figure 5.15 Exemple montrant des cas où DRFI n'arrive pas à bien se généraliser: (a) l'image en couleur, (b) la vérité terrain, (c) carte de saillance retournés par notre approche (*WMLR-RF*), (d) carte de saillance retournés par *DRFI*.

La figure 5.16 illustre des exemples de cartes de saillances générées par notre méthode et les meilleures méthodes de l'état de l'art dans la base MSRA-B. Nous pouvons observer que notre méthode retourne des objets saillants avec un meilleur contraste par rapport à l'arrière-plan en comparaison avec les autres méthodes. Le fond de l'image est nettement mieux écarté également. La figure 5.17 illustre quant à elle des exemples de cartes de saillances générées par notre méthode et les meilleures méthodes de l'état de l'art dans les bases SOD, SED1 et SED2. Nous pouvons observer que notre méthode permet de mieux ressortir l'objet saillant qui se présente sous différentes échelles avec plus d'efficacité (voir par exemple les rangées 1, 5, 6, 7, 10 et 12 où la taille de l'objet varie). Nous notons aussi que contrairement à notre méthode, certaines approches n'arrivent pas à s'accommoder avec la taille des objets saillants. Dans certains cas, certaines méthodes omettent de détecter certains objets présentés sous une petite échelle (cf. figure 5.17, rangées 1, 8, 10, 12 et 13). Dans d'autres cas, les objets présentés sur une grande échelle ne sont pas bien détectés (cf. figure 5.17, rangées 2, 3, 6, 9 et figure 5.16, rangée 3).

### 5.7 Analyse du temps de calcul

Le tableau 5.1 résume le temps de calcul moyen pris par chaque tâche de notre algorithme dans la base de test MSRA-B dans laquelle chaque image est généralement présentée avec la résolution 400 x300 ou bien 300 x400. Nous avons utilisé une machine avec un processeur *I7* et une fréquence d'horloge 2.92 GHZ. Nous exécutons notre méthode sur un environnement Matlab 64 bits. Le tableau 5.2 résume le temps de calcul moyen de notre méthode exécutée dans un environnement parallèle en comparaison avec différentes méthodes de l'état de l'art. Nous pouvons noter que le temps d'exécution de la méthode proposée est comparable avec celles des autres méthodes.

Soient  $n_1, n_2, \dots, n_L$  le nombre de superpixels dans les résolutions  $\Omega_1, \Omega_2, \dots, \Omega_L$  respectivement. Soit  $N_1$  le nombre total des superpixels et  $N_2$  le nombre de pixels dans l'image. La complexité algorithmique de l'algorithme SLIC est linéaire et en rapport avec  $N_2$ . Du moment que nous pouvons effectuer les  $L$  segmentations en parallèle, la complexité totale est  $\sim O(N_2)$  pour cette tâche.

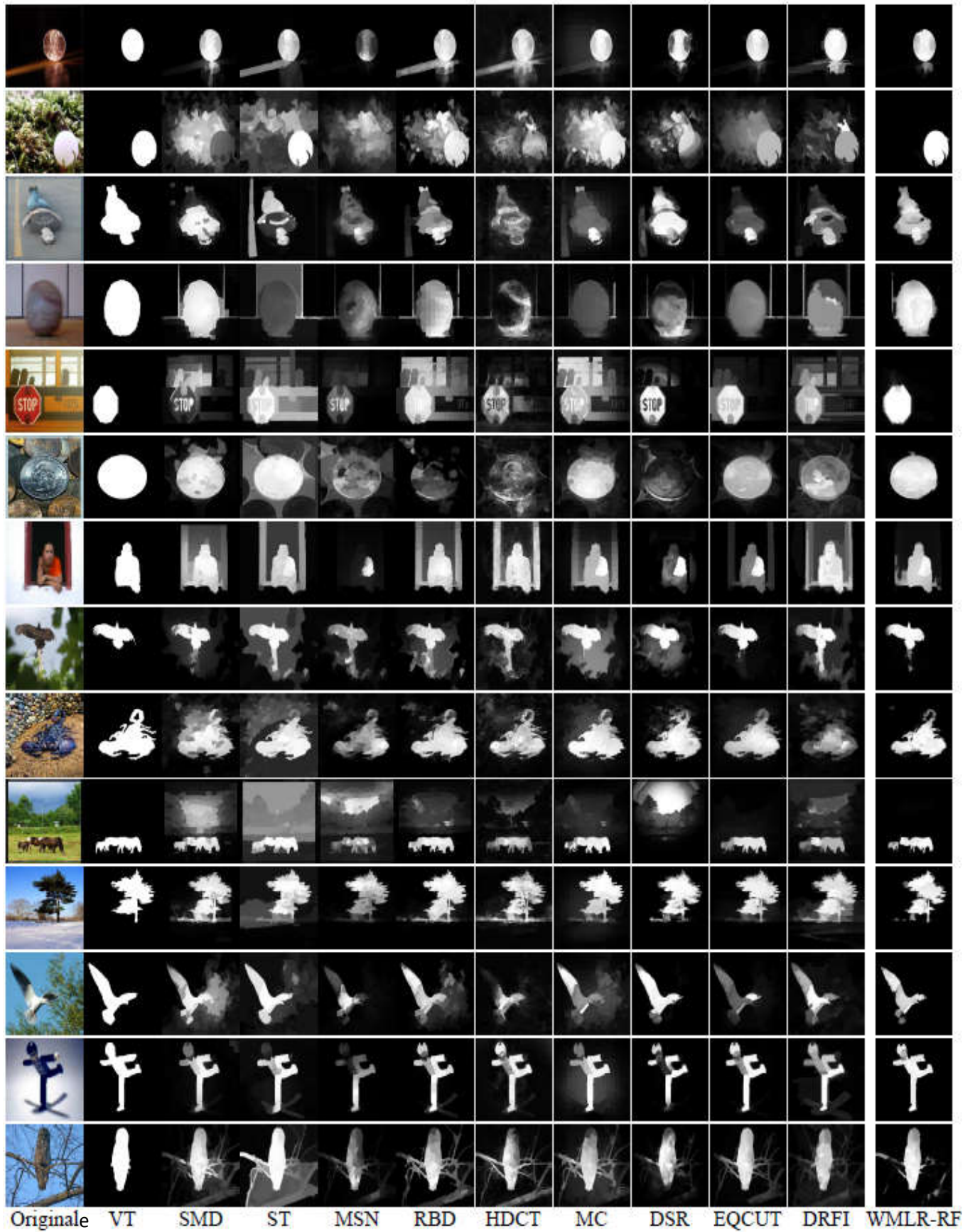


Figure 5.16 Comparaison visuelle entre notre méthode et les meilleures méthodes de l'état de l'art sur la détection d'objets saillants sur la base MSRA-B.

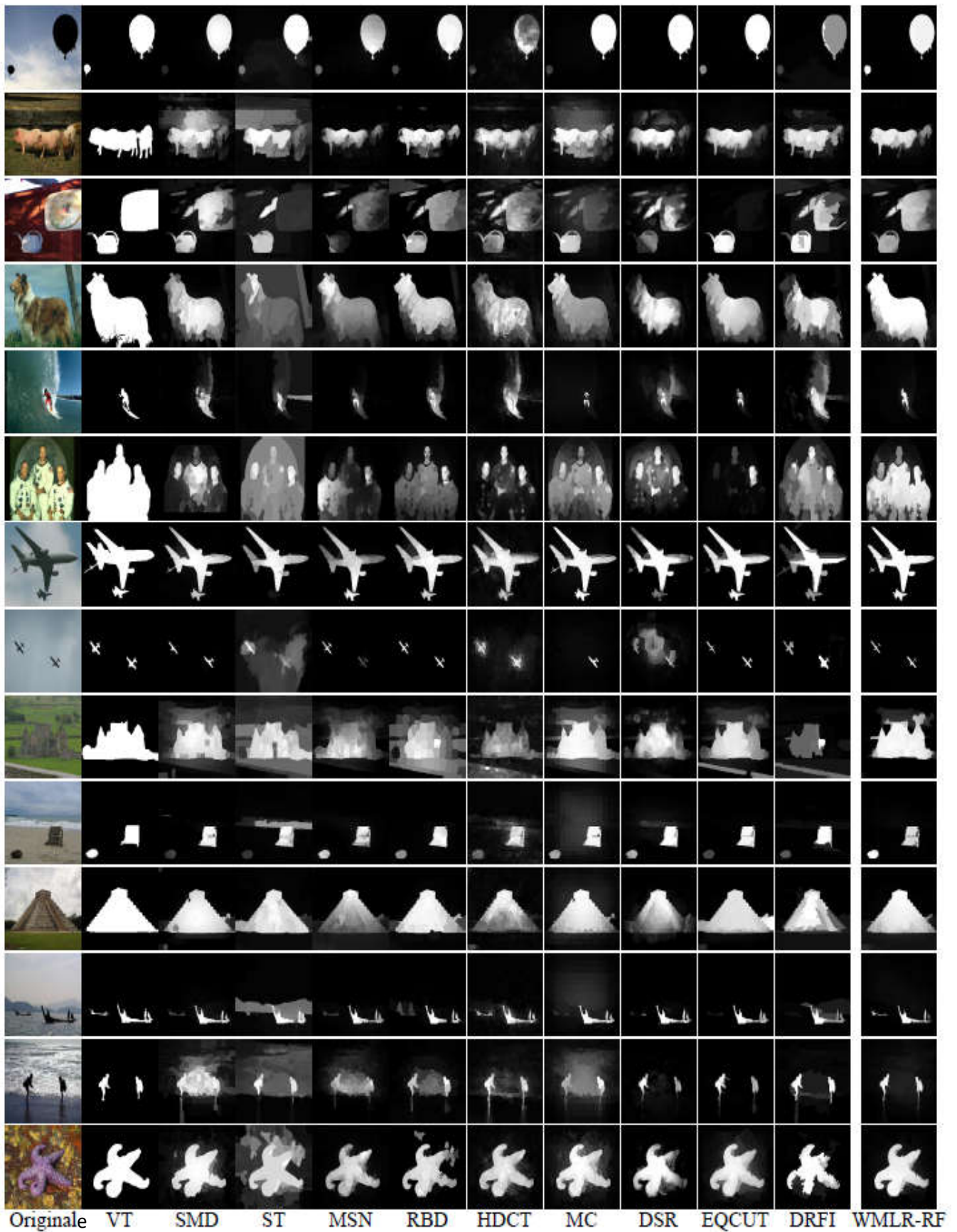


Figure 5.17 Comparaison visuelle entre notre méthode et les meilleures méthodes de l'état de l'art sur la détection d'objets saillants sur les bases SOD, SED1 et SED2.



La construction de notre graphe requiert la constitution de la matrice des poids  $W$  (cf. équation 3.13) de taille  $N_I \times N_I$ . Un seul balayage de l'image est effectué pour construire la matrice d'adjacence. Les poids sont alors calculés en utilisant la fonction 3.3. Cette étape est relativement rapide du moment que seuls les nœuds adjacents sont considérés dans le calcul des poids. La complexité algorithmique de cette tâche est approximativement linéaire avec la borne supérieure  $\sim O((N_1(N_1 - 1))/2)$ . Cette borne correspond au pire cas où un nœud est relié à tous les autres nœuds. Notons également que le calcul de poids peut également se faire en parallèle entre les différents nœuds et résolutions.

La diffusion de la saillance en utilisant l'équation 3.9 requiert une opération d'inversion de la matrice  $I - \alpha D^{-1}W$ . Celle-ci est cependant éparsée et symétrique, l'opération d'inversion est relativement rapide ( $\approx 0.83 s$  dans la manière conventionnelle). Nous utilisons la décomposition de Cholesky qui ramène ce temps d'exécution à  $\approx 0.346 s$ . La complexité algorithmique peut ensuite être réduite de  $\sim O(N_1^{2.8})$  à  $O(N_1)$  en utilisant l'algorithme parallèle proposé dans [122].

Nous notons que l'évaluation de la pertinence des caractéristiques et la localisation spatiale ont conjointement une complexité linéaire de  $\sim O(N_2)$ .

Tableau 5.1

Temps de calcul moyen pour chaque tâche de notre modèle WMLR-RF

	Tâche	Temps (secondes)
1	Segmentation en superpixels	0.135
2	Construction du graphe	1.447
3	Diffusion de la saillance	0.666
4	Localisation spatiale	0.154
5	Évaluation de pertinence des caractéristiques	0.008
6	Entraînement et classification du RF et génération de la carte finale	0.805

L'entraînement du modèle  $RF$  possède une complexité de  $\sim O(N_2 \log N_2)$  (du moment que tous les arbres peuvent être entraîné en parallèle). La complexité algorithmique des mixtures de gaussiennes est de l'ordre  $\sim O((d + K)n)$  avec  $d$  la dimension des données,  $K$  le nombre de composantes et  $n$  le nombre de données. Nous notons également que cette opération est effectuée pour chaque bord de l'image d'une manière parallèle et le tout peut être exécuté en parallèle avec l'entraînement de  $RF$ .

Tableau 5.2

Temps de calcul moyen de notre méthode dans un environnement parallèle en comparaison avec certaines méthodes de l'état de l'art

Méthode	WMLR-RF	HDCT	RBD	DRFI	DSR	MC	EQCUT	SLR	SMD
Temps (secondes)	2.74	5.53	0.27	19.34	6.14	0.11	0.842	1.373	1.217
Code	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab +C++

### 5.8 Conclusion

Nous avons présenté dans cette section l'évaluation expérimentale de notre méthode avec ses différentes variantes en comparaison avec 22 des meilleures méthodes de l'état de l'art sur la détection d'objets saillants. Nous avons utilisé 5 bases de test qui sont parmi les plus utilisées dans la littérature: MSRA-B, ASD, SOD, SED1 et SED2. Pour une étude comparative pertinente, nous avons utilisé plusieurs mesures d'évaluation: la précision, le rappel,  $F$ -measure, ROC, AUC et MAE. Les résultats aussi bien qualitatifs que quantitatifs montrent clairement la supériorité de notre approche sur toutes les méthodes considérées. L'objet saillant est mieux ressorti du fond de l'image et l'arrière-plan mieux écarté. Notre modèle s'adapte mieux également aux changements d'échelle de présentation des objets saillants.

## CONCLUSION GENERALE

La détection d'objets saillants est un problème de vision par ordinateur qui consiste à analyser des scènes visuelles afin de ressortir les objets d'intérêt qui attire l'attention humaine dans les zones spatiales les plus informatives. Une partie de l'image est dite saillante si elle se distingue de son entourage par sa rareté, son contraste ou son unicité dans un contexte spatiale local (entourage direct) ou bien global (l'image dans sa globalité). Il existe deux motivations majeures derrière ce domaine de recherche. La première est la diversité de ses applications comme la reconnaissance d'objets, analyse de mouvement, résumés d'images, etc. La seconde est la difficulté rencontrée à extraire les parties saillantes spécialement dans des cas de scènes complexes avec des fonds d'image chargés, de présence de distracteurs dans l'image et de manque de contraste de l'objet saillant par rapport à l'arrière-plan.

Nous avons d'abord présenté dans ce manuscrit le fonctionnement biologique de l'œil humain dont les travaux sur l'estimation de la saillance puisent leur inspiration. Les travaux précurseurs sur la détection de la saillance se basent sur les fixations du regard humain qui considèrent que le regard est attiré par les points spatiaux qui se distinguent de leur entourage. Par la suite avec l'avènement de la théorie de Gestalt qui définit certaines lois géométrique qui permettent un groupement perceptuel, plusieurs travaux se sont penchés à détecter les objets entiers dans l'image. Ils considèrent que le regard humain est non pas attiré uniquement par des points saillants mais également par des régions entières qui correspondent à des objets.

Nous avons ensuite présenté l'état de l'art sur les méthodes de détection d'objets saillants. Nous avons organisé ces méthodes en trois classes majeures. Les méthodes basées sur le contraste local sont généralement reliées aux fixations du regard humain. Elles mesurent le niveau de contraste d'une région par

rapport à son entourage. L'inconvénient majeur de ces méthodes est que la taille de l'objet à détecter est inconnue a priori. Les cartes de saillance retournées tendent à ressortir les contours des objets plutôt que leur contenu. Les méthodes basées sur le contraste globale mesurent l'unicité d'une partie de l'image (région ou bloc) par rapport à l'image entière en termes de statistiques globales. L'inconvénient de ces méthodes est qu'elles sont fortement dépendantes de la taille des objets saillants. Si celui-ci occupe une part majoritaire de l'image, il ne serait pas considéré comme saillant. En outre, la performance de ces méthodes est limitée dans le cas de fond d'image chargé.

Comme solution aux inconvénients des approches basées sur le contraste local et celles basées sur le contraste global, Les méthode hybrides considèrent le contraste local et global conjointement dans l'estimation de la saillance d'une région ou bloc de l'image. Au sein de ces méthodes, certaines approches se basent sur les blocs comme unité de traitement. Cependant l'inconvénient de ces méthodes est leur imprécision aux abords des contours des objets saillants. Contrairement aux blocs, les superpixels s'accommodent parfaitement aux contours et délimitent des régions homogènes. A cet effet, les méthodes basées sur les superpixels ont considérablement amélioré les performances des systèmes de détection d'objets saillants. Parmi ces méthodes, on dénombre les méthodes probabilistes qui utilisent une théorie probabiliste ou statistique pour détecter la rareté d'un élément dans l'image. Les méthodes graphiques se basent sur la définition d'un graphe pondéré sur l'image. Ces méthodes ont comme avantage d'être adaptées à l'encodage de la contiguïté spatiale qui renforce la cohérence des résultats retournés. Dans ce contexte, l'utilisation de plusieurs résolutions dans la construction du graphe offre un avantage important par rapport à l'utilisation d'une seule résolution en s'adaptant mieux à la variation de la taille des objets saillants. Cependant, aucune de ces méthodes ne considèrent un fort couplage entre un nœud et ses voisins à travers les différentes résolutions. Les autres méthodes se basent sur différentes théorie mathématiques comme la variance, l'ACP, la décomposition matricielle, etc.

Nous avons par la suite présenté notre approche basée sur une structure graphique multi-couches compacte et fortement couplée. Chaque nœud du graphe est connecté non seulement à ses voisins direct se trouvant sur la même

résolution mais également à ses voisins indirects situés sur chacune de toutes les autres résolutions. Deux nœuds sont indirectement voisins si chacun se trouve sur une résolution différente et l'intersection spatiale des surfaces qu'ils occupent n'est pas nulle. Nous avons alors proposé d'optimiser une fonction objective qui établit un ordonnancement des nœuds sur cette structure. Cette fonction diffuse les valeurs de l'ordonnancement depuis les bords de chaque résolution vers toute la structure du graphe, puis effectue une seconde diffusion depuis le noyau de l'objet vers toutes les autres parties du graphe.

Afin de consolider notre approche de détection d'objets saillants, nous avons proposé trois étapes de raffinement. D'abord, nous avons définis une procédure de localisation spatiale de l'objet saillant guidée par la saillance. Une fenêtre est alors ajustée aux dimensions de l'objet saillant qui permet de mieux estimer la saillance dans cet espace restreint et d'écarter les parties appartenant à l'arrière-plan. Ensuite, nous avons introduit conjointement l'information de contour et l'évaluation de la pertinence des caractéristiques utilisées. Cela permet de contrôler la contribution de chaque caractéristique selon son pouvoir discriminatif et d'améliorer la détection des bords de l'objet saillant. L'objet retourné est mieux contrasté avec des bords nettement visibles. Finalement, afin de rester conforme aux statistiques globales de l'image, nous avons définis une approche de détection de la saillance se basant sur l'apprentissage des forêts aléatoires sur les bords des différentes résolutions comme information à priori sur l'arrière-plan et le noyau de l'objet saillant comme information à priori sur le premier-plan. Un processus d'unification de la saillance est ensuite appliqué au niveau superpixel. Ce processus est principalement guidé par le niveau de contraste entre l'arrière-plan et le noyau de l'objet mesuré en utilisant les mixtures de gaussiennes.

Les résultats expérimentaux sur de larges bases de test, ont montré que les performances de notre approche surpassent celles des meilleures méthodes de l'état de l'art sur la détection d'objets saillants. En outre, Notre approche s'adapte mieux à la variation de la taille des objets et retourne des cartes de saillance où l'objet saillant est mieux contrasté avec l'arrière-plan.

## Perspectives

D'une part, nous planifions dans nos travaux futurs d'exploiter notre système de détection d'objets saillants dans plusieurs applications. Comme notre méthode retourne des objets saillants bien contrastés avec le fond de l'image, elle peut être exploitée comme une étape préliminaire pour l'estimation des points saillant correspondant aux fixations du regard. En effet, ces points sont forcément contenus dans les objets saillant retournés. Ensuite, comme notre processus de localisation spatiale de l'objet saillant délimite bien une fenêtre autour de l'objet saillant, notre approche peut être intégrée dans un processus de localisation d'objets comme information préliminaire sur les objets contenus dans la scène visuelle. Elle permettrait ainsi de réduire l'espace de recherche des autres objets non saillants et améliorer les résultats en offrant au moins une position d'objet fiable. En outre, notre système de détection de la saillance peut également servir comme mesure d'analyse de saillance au niveau des fenêtres retournées par un processus de localisation d'objets. Enfin, notre approche peut servir comme information à priori pour la reconnaissance d'objets. En effet, une recherche des objets guidée par l'information de saillance peut réduire considérablement la complexité de recherche à travers toute l'image et offre une zone de recherche fiable d'au moins un objet dans l'image. Cela réduit considérablement l'espace de recherche et améliore la complexité de calcul.

D'une autre part, nous notons que tout comme la majorité des méthodes de détection de la saillance basée sur les graphes, nous avons utilisé une information à priori sur l'arrière-plan qui ne représente pas toujours efficacement le fond de l'image. Un arrière plan complexe avec des parties bien contrasté avec les bords de l'image pourrait affecter la qualité de nos résultats. Il conviendrait alors de définir une approche qui estime l'arrière-plan d'une meilleure façon.

## REFERENCES

1. J.K. Tsotsos. Analysing vision at the complexity level. *Behavioral and Brain Sciences*, 13: 423-469, 1990.
2. L. Paletta et J.K. Tsotsos. *Preface of Attention in Cognitive Systems*. In R. Goebel, J. Siekmann & W. Wahlster, editors, *Attention in Cognitive Systems*, page Preface. Springer Berlin /Heidelberg, 2008.
3. A. Borji and L. Itti. State-of-the-art in Visual Attention Modeling. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35(1):185-207, 2013.
4. L. Itti, C. Koch and E. Niebur. A Model of Saliency Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 20(11):1254-1259, 1998.
5. T. Liu, J. Sun, N. Zheng, X. Tang and H.Y. Shum. Learning to detect a salient object. *in CVPR*, 1–8, 2007.
6. M.M. Cheng, J. Warrell, W.Y. Lin, S. Zheng, V. Vineet, and N. Crook. Efficient salient region detection with soft image abstraction. *in ICCV*, 1529–1536, 2013.
7. X. Hou and L. Zhang. Saliency Detection: A Spectral Residual Approach. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
8. R. Achanta, S.S. Hemami, F.J. Estrada and S. Susstrunk. Frequency-Tuned Salient Region Detection. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1597 – 1604, 2009.
9. C. Yang, L. Zhang, H. Lu and X. Ruan. Saliency Detection via Graph Based Manifold Ranking. *IEEE Conf. on Computer Vision and Pattern Recognition*, 3166-3173, 2013.
10. Z. Liu, W. Zou and O. Le Meur. Saliency Tree: A Novel Saliency Detection

- Framework. *IEEE Trans. on Image Processing*, 23(5):1937-1952, 2014.
11. L. Zhu, D.A. Klein, S. Frintrop, Z. Cao and A.B. Cremers. A Multi-size Superpixel Approach for Salient Object Detection based on Multivariate Normal Distribution Estimation. *IEEE Trans. on image processing*, 23(12):5094-5107, 2014.
  12. V. Gopalakrishnan, Y. Hu and D. Rajan. Random Walks on Graphs for Salient Object Detection in Images. *IEEE Trans. on Image Processing*, 19(12):3232-3242, 2010.
  13. V. Vilaplana. SaliencyMaps on Image Hierarchies. *Signal Processing: Image Communication*, 38:84-99, 2015.
  14. A. Borji, M-M. Cheng, H. Jiang and J. Li: Salient Object Detection: A Benchmark. *IEEE Trans. on Image Processing*, 24(12): 5706-5722, 2015.
  15. C. Aytekin, E. Ozan, S. Kiranyaz and M. Gabbouj. Visual Saliency by Extended Quantum Cuts. *IEEE Int'l Conf. on Image Processing*, 112-117, 2014.
  16. H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng and S. Li. Salient Object Detection: A Discriminative Regional Feature Integration Approach. *IEEE Conf. on Computer Vision and Pattern Recognition*, 2083-2090, 2014.
  17. H. Peng, B. Li, R. Ji, W. Hu, W. Xiong, and C. Lang. Salient Object Detection via Low-Rank and Structured Sparse Matrix Decomposition. *AAAI Conference on Artificial Intelligence*, 796-802, 2013.
  18. Q. Yan, L. Xu, J. Shi and J. Jia. Hierarchical Saliency Detection. *IEEE Conf. on Computer Vision and Pattern Recognition*, 1155-1162, 2013.
  19. M.S. Allili and D. Ziou. Object of Interest segmentation and Tracking by Using Feature Selection and Active Contours. *IEEE Conf. on Computer Vision and Pattern Recognition*, 1-8, 2007.
  20. M.S. Allili and D. Ziou. Active Contours for Video Object Tracking Using Region, Boundary and Shape Information. *Signal, Image and Video Processing*, 1(2):101-117, 2007.
  21. B. Alexe, T. Deselaers and V. Ferrari. Measuring the Objectness of Image



- Windows. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(11):2189-2202, 2012.
22. M.E. Raichle, M.E.: The brain's dark energy. *Scientific American Magazine*, 44-49, 2010.
  23. C.Koch. Biophysics of Computation: Information Processing in Single Neurons, 1<sup>st</sup>edn. *Oxford University Press, New York*, 2004.
  24. J. Li, W. Gao. Visual Saliency Computation, a Machine Learning Perspective, *Lecture Notes in Computer Science, Springer*, 8408 2014, 2014.
  25. W. James. The principles of psychology. *Dover Publications, volume 1 edition*, 1890.
  26. A.R. Luria. The Working Brain. Basic Books (1973)
  27. R.L. Solso. Cognitive Psychology, 2<sup>nd</sup> edn. *Allyn and Bacon, Boston*, 1988
  28. M.M. Sohlberg, and C.A. Mateer. Introduction to cognitive rehabilitation: Theory and practice. Guilford Press, New York, 1989.
  29. Anne Treisman and Garry Gelade. *A Feature-Integration Theory of Attention*. *Cognitive Psychology*, 136(2): 97-136, 1980.
  30. O. Le Meur, P. Le Callet, D. Barba, D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(5), 802-817, 2006.
  31. R. A. Rensink. The dynamic representation of scenes. *Visual Cognition*, 7: 17-42, 2000.
  32. R. VanRullen. Visual saliency and spike timing in the ventral visual pathway. *Journal of physiology, Paris*, 97(2-3): 365-77, 2003.
  33. M.P. Da Silva. Modèle computationnel d'attention pour la vision adaptative. *Thèse de doctorat, Université de la Rochelle*, 2010.
  34. R.J. Sternberg, K. Sternberg, J. Mio. *Cognitive Psychology, 6th edn. Cengage Learning*, 2008.
  35. J.L Wilkinson-Berka. Diabetes and retinal vascular disorders: role of the

- reninangiotensin system. *Expert Reviews in Molecular Medicine*, 6: 1-18, 2004.
36. L. Itti, G. Rees et J.K. Tsotsos, editeurs. *Neurobiology of attention*. *Academic Press*, 1st edn, 2005.
  37. M.I. Posner. Orienting of attention. *The Quarterly Journal of Experimental Psychology*, vol. 32, no. 1, :3–25, 1980.
  38. M.I. Posner et S.E. Petersen. The attention system of the human brain. *Annual review of neuroscience*, 13: 25-42, 1990.
  39. J. Duncan. Selective attention and the organization of visual information. *Journal of Experimental Psychology:General*, 113(4): 501–517, 1984.
  40. P.Guillaume. *La Psychologie de la forme*. *Flammarion, Paris*, 1937.
  41. S.E Palmer. Les théories contemporaines de la perception de Gestalt. *Intellectica*, 28(1): 53–91,1999.
  42. Yuka Sasaki. Processing local signals into global patterns. *Current opinion in neurobiology*, vol. 17, no. 2, pages 132–9, 2007.
  43. P.R. Roelfsema. Cortical algorithms for perceptual grouping. *Annual review of neuroscience*, 29: 203–27,2006.
  44. S. Sabatini and F. Solari. An Early Cognitive Approach to Visual Motion Analysis. *Advances in ArtificialIntelligence*, 385-397, 2003.
  45. F. Wörgötter, N. Krüger, N. Pugeault, D. Calow, M. Lappe, K. Pauwels, M.V. Hulle, S. Tan and A. Johnston. Early cognitive vision: Using Gestalt-laws for task dependent,active image-processing. *Natural computing*, 3(3): 293–321, 2004.
  46. A. Desolneux, L. Moisan & J.M. Morel. Gestalt theory and computer vision, chapitre 3, 71-101. *Springer, Berlin, Heidelberg*, 2004.
  47. A. Desolneux, L. Moisan and J.M. Morel. From Gestalt Theory to Image Analysis, volume 34 of *InterdisciplinaryApplied Mathematics*. *Springer Verlag, New York, NY, interdisci edition*, 2008.
  48. Jean Petitot. *Neurogéométrie de la vision*. Paris, France, les éditions de

- l'école polytechnique, 2008.
49. Q. Zou, S. Luo and J. Li. Selective attention guided perceptual grouping model. *Advances in Natural Computation*, 867–876, 2005.
  50. A. Borji, M.-M. Cheng, H. Jiang and J. Li. Salient Object Detection: A Survey. [online] Available: <http://arxiv.org/abs/1411.5878>, 2014.
  51. M. Wang, J. Konrad, P. Ishwar, K. Jing, and H. A. Rowley. Image Saliency: From Intrinsic to Extrinsic Context. *IEEE Conf. Computer Vision and Pattern Recognition*, 417424, 2011.
  52. F. Huazhu, C. Xiaochun and T. Zhuowen. Cluster-based Co-saliency Detection. *IEEE Trans. on Image Processing*, 22(10):3766-3778, 2013.
  53. P. Wang, J. Wang, G. Zeng, J. Feng, H. Zha, and S. Li, Salient Object Detection for Searched Web Images via Global Saliency. *IEEE Conf. on Computer Vision and Pattern Recognition*, 3194-3201, 2012.
  54. B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang. Saliency Detection via Absorbing Markov Chain. *IEEE Int'l Conf. on Computer Vision*, 1665-1672, 2013.
  55. H. Jiang, J. Wang, Z. Yuan, T Liu and N. Zheng. Automatic Salient object Segmentation Based on Context and Shape Prior. *British Machine Vision Conference*, 1-12, 2011.
  56. Q. Fan and C. Qi. Two-stage Salient Region Detection by Exploiting Multiple Priors. *J. of Visual Communication and Image Representation*, 25(8):1823-1834, 2014.
  57. Y. Wei, F. Wen, W. Zhu and J. Sun. Geodesic saliency using background priors. *European Conf. on Computer Vision*, 29-42, 2012.
  58. Y. Li, X. Hou, C. Koch, J-M. Rehg and A-L. Yuille. The Secrets of Salient Object Segmentation. *IEEE Conf. on Computer Vision and Pattern Recognition*, 280-287, 2014.
  59. F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. *in CVPR*, 2012, 733-740, 2012.

60. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE TPAMI*, 24, 2002.
61. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE TPAMI*, 34(11), 2012.
62. M-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S-M. Hu. Global Contrast Based Salient Region Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 37(3): 569-582, 2015.
63. G. Larivière and M.S. Allili. A Learning Probabilistic Approach for Object Segmentation. *IEEE Canadian Conf. On Computer and Robot Vision*, 86-93, 2012.
64. Y-F. Ma, X-S. Hua, L. Lu and H-J. Zhang. A Generic Framework of User Attention Model and its Application in Video Summarization. *IEEE Trans. on Multimedia*, 7(5):907-919, 2005.
65. C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *in Matters of Intelligence*, 115–141, 1987.
66. C.L. Zitnick and P. Dollar. Edge Boxes: Locating Object Proposals from Edges. *EuropeanConf. on Computer Vision*, 391-405, 2014.
67. P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE TPAMI*, 1627-1645, 2010.
68. R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 1995.
69. A. Borji, D. N. Sihite, and L. Itti. What stands out in a scene? a study of human explicit saliency judgment. *Vision Research*, 91(0), 62-77, 2013.
70. A. Borji and L. Itti. Exploiting Local and Global Patch Rarities for Saliency Detection. *IEEE Conf. on Computer Vision and Pattern Recognition*, 478-485, 2012.
71. Y. Hu, D. Rajan, L.T. Chia. Adaptive local context suppression of multiple cues for salient visual attention detection. *In: Proceedings of the IEEE*

*International Conference on Multimedia and Expo, ICME, 2005.*

72. D. Gao, V. Mahadevan, and N. Vasconcelos. The discriminant center surround hypothesis for bottom-up saliency. *In: Advances in Neural Information Processing Systems, 2009.*
73. O. Le Meur, P. Le Callet, D. Barba, D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5): 802–817, 2006.*
74. T. Kadir, M. Brady. Saliency, scale and image description. *International Journal of Computer Vision, 45(2): 83-105, 2001.*
75. T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.Y. Shum, Learning to detect a salient object. *IEEE TPAMI, 33(2): 353–367, 2011.*
76. D.R. Martin, C.C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE TPAMI, 26(5): 530-549, 2004.*
77. Y. Xie, H. Lu and M.H. Yang. Bayesian saliency via low and mid level cues. *IEEE TIP, 22(5): 2013.*
78. E. Rahtu, J. Kannala, M. Salo, and J. Heikkila. Segmenting salient objects from images and videos. *in ECCV, 2010.*
79. J. Harel, C. Koch and P. Perona. Graph-based visual saliency. *In: Advances in Neural Information Processing Systems, 545-552, 2007.*
80. C. Guo, Q. Ma, and L. Zhang. Spatio-Temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform. *Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008.*
81. N. Imamoglu, W. Lin and Y. Fang. A saliency detection model using low-level features based on Wavelet Transform. *IEEE Trans. on Multimedia, 15(1):96-105, 2013.*
82. W. Zhu, S. Liang, Y. Wei and J. Sun. Saliency Optimization from Robust Background Detection. *IEEE Conf. Computer Vision and Pattern Recognition, 2814-2821, 2014.*

83. D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Scholkopf. Ranking on data manifolds. In NIPS, 2004.
84. P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach.Intell.*, 33(5): 898–916, 2011.
85. R. C. Gonzalez and R. E. Woods, Digital Image Processing (3rd Edition). *Upper Saddle River, NJ, USA: Prentice-Hall, Inc.*, 2006.
86. L. Xiaohui, L. Huchuan, Y. Ming-Hsuan, Z. Lihe and R. Xiang. Saliency Detection Via Dense and Sparse Reconstruction. *International Conference on Computer Vision (ICCV)*, 2013.
87. X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. *in CVPR*, 2012.
88. W. Zou, K. Kpalma, Z. Liu, J. Ronsin et al. Segmentation driven low-rank matrix recovery for saliency detection. *in BMVC*, 1-13, 2013.
89. J. Yan, M. Zhu, H. Liu, and Y. Liu. Visual saliency detection via sparsity pursuit. *IEEE SPL*, 17(8): 739-742, 2010.
90. J. Kim, D. Han, Y.W Tai and J. Kim. Salient Region Detection via High-Dimensional Color Transform and Local Spatial Support. *IEEE Trans. on Image Processing*, 25(1): 9-23, 2016.
91. J. Liu and S. Wang. Salient Region Detection via Simple Local and Global Contrast Representation. *Neurocomputing*, 147(5):435-443, 2015.
92. D. Walther, L. Itti, M. Riesenhuber, T. Poggio, C. Koch. Attentional Selection for Object Recognition-A Gentle Way. *Biologically Motivated Computer Vision*, LNCS 2525, 472-479, 2002.
93. M. Craven and J.W. Shavlik. Extracting tree-structured representations of trained networks. *Advances in Neural Information Processing Systems*, pp. 24-30, 1996.
94. T. Ojala, M. Pietikainen and T. Maenpaa. Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *European Conf. on Computer Vision*, 404-420, 2000.

95. J. Lim, C. L. Zitnick, and P. Dollar. Sketch tokens: A learned mid-level representation for contour and object detection. *In CVPR*, 3158-3165, 2013.
96. S. Zheng, Z. Tu, and A. Yuille. Detecting object boundaries using low, mid and high-level information. *In CVPR*, 2007.
97. P. Dollar and C.L. Zitnick. Structured Forests for Fast Edge Detection. *IEEE Int'l Conf. on Computer Vision*, 1841-1848, 2013.
98. V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contour segments for object detection. *PAMI*, 30(1):36–51, 2008.
99. S. Ullman and R. Basri. Recognition by linear combinations of models. *PAMI*, 13(10): 992-1006, 1991.
100. J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *IJCV*, 43(1):7-27, 2001.
101. P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *PAMI*, 33, 2011.
102. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1988.
103. Gonzalez, R.C., R.E. Woods, S.L. Eddins, *Digital Image Processing Using MATLAB*, New Jersey, Prentice Hall, 2003, Chapter 11.
104. T. Kailath. The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Trans. on Communication Technology*, 15(1):5260, 1967.
105. S. Goferman, L. Zelnik-Manor and A. Tal. Context-aware Saliency Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 34(10):1915-1926, 2011.
106. X. Wang, X. Bai, W. Liu and L.J. Latecki. Feature Context for Image Classification and Object Detection. *IEEE Conf. On Computer Vision and Pattern Recognition*, 961-968, 2011.
107. P. Liu, M. Reale, X. Zhang and L. Yin. Saliency-Guided 3D Head Pose Estimation on 3D Expression Models. *ACM Int'l Conf. on Multimodal*

*Interaction*, 75-78, 2013.

108. B. Pepik, M. Stark, P. Gehler, and B. Schiele. Teaching 3D Geometry to Deformable Part Models. *IEEE Conf. on Computer Vision and Pattern Recognition*, 3362-3369, 2012.
109. T. Chen, M-M. Cheng, P. Tan, A. Shamir and S-M. Hu. Sketch2Photo: Internet Image Montage. *ACM Trans. on Graphics*, 28(5), Article 124, 2009.
110. G. McLachlan and D. Peel. Finite Mixture Models. *John Wiley and Sons*, 2000.
111. Wavelet Modeling Using Finite Mixtures of Generalized Gaussian Distributions: Application to Texture Discrimination and Retrieval. *IEEE Trans. Image Processing*, 21(4): 1452-1464, 2012.
112. M.M. Cheng, Z. Zhang, W.Y. Lin, P. Torr. BING: Binarized normed gradients for objectness estimation at 300fps. *In: CVPR.*,3286-3293, 2014.
113. R.T. Collins, Y. Liu, M. Leordeanu, Online selection of discriminative tracking features, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (10) (2005) 1631–1643.
114. L. Breiman. Random forests. *Machine Learning*, 45(1):5-32, 2001.
115. V. Movahedi and J.H. Elder. Design and Perceptual Validation of Performance Measures for Salient Object Segmentation. *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, 4956, 2010.
116. D. Martin, C. Fowlkes, D. Tal, and J. Malik. A Database of Human segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. *IEEE Int'l Conf. on Computer Vision*, 416-423, 2001.
117. S. Alpert, M. Galun, A. Brandt and R. Basri. Image Segmentation by Probabilistic Bottom-Up Aggregation and Cue Integration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(2):315-327, 2012.
118. Y. Fang, W. Lin, B. Lee, C. Lau, Z. Chen and C. Lin. Bottom-up Saliency Detection Model Based on Human Visual Sensitivity and Amplitude



- Spectrum. *IEEE Trans. on Multimedia*, 14(1):187-198, 2012.
119. L. Xu, H. Li, L. Zeng and K.N. Ngan. Saliency Detection Using Joint Spatial-Color constraint and multi-scale segmentation. *Visual Commun. Image Represent*, 465-476, 2013.
  120. R. Margolin, A. Tal, L.Z. Manor. What Makes a Patch Distinct?. *IEEE Conf. on Computer Vision and Pattern Recognition*, 1139-1146, 2013.
  121. L. Zhang, Z. Gu and H. Li. SDSP: A Novel Saliency Detection Method by Combining Simple Priors. *IEEE Int'l Conf. on Image Processing*, 171-175, 2013.
  122. G. Sharma, A. Agarwala, B. Bhattacharya. A fast Parallel Gauss Jordan Algorithm for Matrix Inversion Using CUDA. *Computers & Structures*, 128:31-37, 2013.
  123. I. Filali, M.S Allili and N. Benblidia. Multi-scale salient object detection using graph ranking and global-local saliency refinement. *Signal processing: Image Communication*, 47:380-401, 2016
  124. I. Filali, M.S Allili and N. Benblidia. Multi-graph based salient object detection. *ICIAR*, 318-324, 2016.