



HAL
open science

Recherche d'images basée sur la saillance visuelle pour l'imagerie urbaine

Kamel Guissous, Valérie Gouet-Brunet

► **To cite this version:**

Kamel Guissous, Valérie Gouet-Brunet. Recherche d'images basée sur la saillance visuelle pour l'imagerie urbaine. ORASIS 2017, GREYC, Jun 2017, Colleville-sur-Mer, France. hal-01866664

HAL Id: hal-01866664

<https://hal.science/hal-01866664>

Submitted on 3 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Recherche d'images basée sur la saillance visuelle pour l'imagerie urbaine

Kamel Guissous¹

Valérie Gouet-Brunet¹

¹ Université Paris-Est, LASTIG MATIS, IGN, ENSG

73 Avenue de Paris, 94160 Saint-Mandé, France

Kamel.Guissous@ign.fr

Résumé

Avec l'augmentation de la taille des bases d'images et de la complexité des descripteurs dans les domaines de la recherche d'images par contenu visuel et de la vision par ordinateur, il est nécessaire de trouver un moyen pour limiter la quantité de données manipulées, tout en conservant leur représentativité. Au lieu d'analyser l'image entière, la sélection des régions qui détiennent l'essence de l'information est une option pertinente pour atteindre cet objectif. Comme la saillance visuelle a pour objectif de sélectionner les zones les plus importantes de l'image pour une tâche donnée, dans cet article, nous proposons d'exploiter des cartes de saillance visuelle pour filtrer les caractéristiques visuelles les plus saillantes de l'image. Une nouvelle approche de saillance visuelle basée sur l'analyse de la distribution locale de l'orientation des contours, en particulier dédiée aux contenus image structurés, comme les images de type street-view de l'environnement urbain, est proposée. Il est évalué pour la recherche d'images par contenu visuel à partir d'un exemple selon trois critères : la qualité de la recherche, le volume des caractéristiques manipulées et le temps de calcul. L'approche proposée peut être exploitée dans diverses applications qui manipulent de grands nombres de caractéristiques visuelles ; ici il est expérimenté dans deux applications : la recherche d'images cross-domain et la localisation de véhicule basée image.

Mots Clef

recherche d'images par contenu, saillance visuelle, descripteurs locaux, orientation des contours.

Abstract

With the increase of image datasets size and of descriptors complexity in Content-Based Image Retrieval (CBIR) and Computer Vision, it is essential to find a way to limit the amount of manipulated data, while keeping its quality. Instead of treating the entire image, the selection of regions which hold the essence of information is a relevant option to reach this goal. As the visual saliency aims at highlighting the areas of the image which are the most important for a given task, in this paper we propose to exploit visual

saliency maps to prune the most salient image features. A novel visual saliency approach based on the local distribution analysis of the edges orientation, particularly dedicated to structured contents, such as street view images of urban environments, is proposed. It is evaluated for CBIR according to three criteria : quality of retrieval, volume of manipulated features and computation time. The proposal can be exploited into various applications involving large sets of local visual features ; here it is experimented within two applications : cross-domain image retrieval and image-based vehicle localization.

Keywords

CBIR, visual saliency, local descriptors, edge orientation.

1 Introduction

La saillance visuelle et la recherche d'images par contenu visuel (en anglais : Content-Based Image Retrieval ou CBIR) sont deux sujets de recherche très actifs avec le besoin croissant de solutions pour interpréter et gérer les contenus visuels à large échelle. La saillance visuelle a pour objectif de sélectionner les régions les plus saillantes de l'image ou qui contiennent l'essentiel de l'information visuelle. A leur début, les approches de la saillance visuelle étaient bio-inspirées, avec l'objectif de la modélisation de l'attention visuelle humaine [1]. Aujourd'hui, ce terme couvre des significations différentes, principalement guidées par la tâche ; par exemple, la saillance basée sur la détection de symétries dans l'image permet de détecter des architectures particulières dans les images d'environnements urbains [2]). Plusieurs modèles computationnels ont été proposés pour extraire ces régions saillantes. Borji et Itti [3] ont proposé de classer les modèles de la saillance visuelle en plusieurs catégories, en fonction de leur mécanisme pour obtenir les régions saillantes : les modèles cognitifs [1], les modèles d'analyse spectrale [4], les modèles graphiques [5, 6], les modèles bayésiens [7], les modèles de la théorie de décision [8] et les modèles de la théorie de l'information [9].

L'objectif principal des systèmes de recherche d'images par contenu visuel est de décrire et d'indexer les images

dans une base par l'analyse de leurs contenus, dans le but de faciliter la gestion et la consultation de cette base de données à grande échelle. Avec l'augmentation de la taille des bases d'images et la complexité des descripteurs visuels d'aujourd'hui, trouver rapidement et avec précision les images de la base d'images qui sont similaires à une image requête donnée est encore un sujet de recherche aux nombreux verrous scientifiques non résolus.

De nombreuses approches ont été proposées pour décrire le contenu de l'image [10, 11], ainsi que de nombreuses solutions pour indexer ces descriptions pour faire face aux grands volumes [12]. Pour résoudre le problème du passage à l'échelle, plusieurs options existent, soit en exploitant des structures d'index dédiées (centralisées ou distribuées) ou des plateformes Big-Data [13], soit en se concentrant sur des descripteurs produisant des signatures très compactes. Dans ce travail, nous suivons la seconde option en proposant une nouvelle approche de saillance visuelle basée sur l'analyse de la distribution locale de l'orientation des contours afin de filtrer les principales caractéristiques de l'image.

Les cartes de saillance produites sont intégrées dans un système de recherche d'images. Ensuite, seules les régions informatives sélectionnées par les cartes de saillance sont décrites par le contenu et exploitées dans le système de recherche d'images, dans le but de réduire la quantité d'informations traitées, ainsi de maintenir le passage à l'échelle du système. En raison de la variété des contenus des images, il est difficile de développer une approche de salience efficace permettant de sélectionner les régions saillantes dans tout type de contenu et pour tout type d'application. Ici, nous nous concentrons sur des contenus avec des structures, *i.e.* des paysages urbains ou des scènes d'intérieurs. Deuxièmement, nous comparons l'approche proposée avec huit méthodes de l'état de l'art de la saillance visuelle. Troisièmement, nous expérimentons l'approche proposée pour deux applications : la recherche d'images *cross-domain* utilisant des documents anciens et la localisation de véhicules basée image.

Le papier est organisé comme suit : les travaux antérieurs sur le sujet sont présentés dans la section 2. Dans la section 3, l'approche proposée de saillance visuelle est décrite. Les expériences réalisées et les résultats obtenus sont détaillés dans la section 4 ; enfin la section 5 conclut notre travail.

2 Travaux antérieurs

Dans cette section nous présentons l'état de l'art de la saillance visuelle et ses applications liés à la sélection de caractéristiques visuelles dans le domaine de la recherche d'images. Nous commençons avec certains modèles de saillance visuels, qui seront utilisés comme base de comparaison avec notre proposition. Nous supposons que ces modèles sont représentatifs de la littérature en raison de leurs mécanismes différents de détection des régions saillantes. Itti et al. [1] ont proposé le premier modèle computationnel de la saillance visuelle. Ce modèle est devenu la base des

modèles subséquents et une référence de comparaison. Il est bio-inspiré, basé sur la théorie de l'intégration de caractéristiques visuelles proposée par Treisman et Gelade [14]. Trois primitives de bas-niveau sont utilisées pour sélectionner les régions saillantes : intensité, couleur et orientation. Hou et Zhang [4] ont proposé un modèle simple et facile à implémenter, basé sur le filtrage du spectre de Fourier d'une image. Un autre modèle est proposé par Rosin [15] basé sur la transformation en distance [16] des contours de l'image. Harel et al. [6] ont proposé un modèle basé sur la normalisation des caractéristiques RVB à partir de la convergence des graphes. Zhang et al. [7] ont proposé un modèle basé sur des statistiques bayésiennes. Yang et al. dans [5] ont proposé une approche de détection de saillance basée sur le contraste et les centres des objets saillants et pour affiner la carte de saillance finale, ils ont utilisé le lissage.

Dans le domaine de la recherche d'image, la saillance visuelle est généralement utilisée pour filtrer les caractéristiques visuelles ou pour concevoir un nouveau descripteur : Lei et al. [17] ont proposé de segmenter les régions saillantes identifiées par [1], puis d'extraire les descripteurs à l'aide de l'entropie. La recherche d'images est alors effectuée en calculant la similarité entre ces descripteurs. Wan et al. [18] ont proposé une carte de saillance basée sur trois caractéristiques : couleur, intensité et texture, puis l'histogramme de la carte de saillance est utilisé comme nouveau descripteur pour la recherche d'image.

Awad et al. [19] ont proposé un nouveau descripteur texture-perceptuel qui analyse la fréquence du contenu des caractéristiques perceptuelles (couleur, intensité et orientation) dans une pyramide multi-résolution calculée à partir de l'approche de saillance visuelle proposée par Da Silva et al. [20].

Marques et al. [21] ont utilisé [1] pour calculer les cartes de saillance afin de définir une nouvelle métrique de similarité en utilisant les valeurs d'attention calculées à partir des régions d'intérêt. Ozyer et Vural [22] ont également utilisé [1] pour proposer une nouvelle métrique de similarité basée sur les régions saillantes.

Le modèle d'Itti et al. [1] a été exploité dans différents travaux dans le cadre de la recherche d'images, afin de filtrer les descripteurs après avoir identifié des régions saillantes telles que [23, 24]. Dans [25, 26] la saillance visuelle est utilisée pour réduire le nombre de descripteurs dans le but de réduire le temps de calcul de la recherche d'images.

Le modèle de saillance visuelle proposé par Da Silva et al. [20] a été utilisé par Awad et al. [27] pour filtrer les descripteurs SIFT [28]. Avec 60% des points d'intérêt filtrés, la performance a été réduite dans certaines classes de la base d'images VOC2005. Dans [29], Awad et al. ont proposé une évaluation de cinq méthodes de l'état de l'art de la saillance visuelle en utilisant la base d'images VOC2007 et un système de recherche d'images. Leur évaluation est basée sur la capacité d'un modèle de saillance visuelle à maintenir la performance du système de re-

cherche d'images de référence lorsqu'il agit comme un filtre des points d'intérêt.

3 L'approche proposée

Notre approche de saillance visuelle, basée sur l'analyse de la distribution locale de l'orientation des contours, est présentée dans cette section. La détection des segments de droite et l'analyse de la distribution locale de leurs orientations sont présentées respectivement dans les sections 3.1 et 3.2. La section 3.3 décrit enfin comment plusieurs types de cartes de saillance sont extraites en fonction de ces caractéristiques.

3.1 Détection de segments

Les orientations des gradients contiennent des informations importantes, notamment sur les structures principales de l'image et sur la forme des objets. Cette orientation a été exploitée dans plusieurs approches et applications, afin de décrire le contenu de l'image (*i.e.*, l'histogramme des gradients orientés (HOG) [30], les descripteurs SIFT [28] et SURF [31]).

Comme l'orientation du gradient peut être sensible au bruit, et aussi dans le but de limiter la description aux principales structures de l'image, nous choisissons de caractériser l'orientation des gradients de pixels appartenant uniquement à des segments de droite extraits de l'image. Dans le domaine de la vision par ordinateur, un segment peut être défini par un ensemble de pixels connectés qui partagent la même orientation de gradient avec une tolérance donnée. Les segments représentent des structures locales discriminantes, ils sont aussi robustes aux occultations, aux petits changements et moins sensibles au bruit. C'est l'algorithme LSD (Line Segment Detector), proposé par Goi et al. [32], qui est choisi pour détecter les segments dans l'image. Cet algorithme est rapide et efficace.

3.2 Calcul des orientations locales dominantes

L'utilisation d'un histogramme est classique pour décrire la distribution des orientations dans l'image [30, 28]. Les histogrammes de l'orientation du gradient des pixels, avec une quantification grossière (*i.e.* 8 ou 9 bins) comme dans [30, 28, 31], présentent des inconvénients liés à la quantification uniforme et à la périodicité des angles à $0+$ et $\pi-$. La quantification uniforme divise les orientations relatives entre deux bins différents, alors qu'ils constituent ensemble une direction dominante. Les orientations des deux extrêmes de l'histogramme peuvent être reliées mais sont divisées en deux bins différents en utilisant la quantification uniforme. Afin de résoudre ces problèmes, nous utilisons une méthode adaptative basée sur la convolution circulaire pour étudier la distribution locale des orientations, comme suit.

Calcul de l'histogramme des orientations. Nous considérons que l'orientation du segment correspond à l'orientation de tous les pixels qui le constitue. Dans une fenêtre

circulaire de rayon r pixels, les pixels des segments sont détectés, puis leurs orientations sont quantifiées avec un pas de quantification q rad et l'histogramme des orientations est calculé avec $N = \pi \times \frac{1}{q}$ bins, couvrant chacun un angle de q rad. Nous avons choisi la valeur du pas de quantification $q = 0,01$ afin de minimiser l'erreur de quantification qui sera incluse dans un intervalle de $] +0,005, -0,005]$ rad.

Nous considérons également que la distribution est nulle si le nombre de pixels des segments détectés dans la fenêtre locale est inférieur à N_p . Après plusieurs évaluations, nous avons fixé $N_p = 60$ pixels.

Calcul de la convolution de l'histogramme. Dans cette étape, en utilisant l'équation 1, nous calculons la convolution de l'histogramme des orientations avec une fenêtre rectangulaire d'une largeur de $T = \frac{N}{8}$ bins centrée sur le pixel cible. Nous avons choisi cette largeur non seulement pour respecter l'angle de tolérance ($\frac{\pi}{8}$) utilisé dans l'algorithme LSD pour détecter les segments mais aussi pour respecter les 8 bins utilisés dans l'histogramme classique pour décrire la distribution des orientations du gradient.

$$c(n) = h(n) \times H(n) = \sum_{k=0}^{N-1} h(k) H(\|n - k\|) \quad (1)$$

où $n \in [0, N - 1]$ et $c(n)$ est l'histogramme convolué, $h(n)$ est la fenêtre de convolution, $H(n)$ est l'histogramme des orientations. Note, $\|n - k\|$ veut dire $(n - k)$ modulo N .

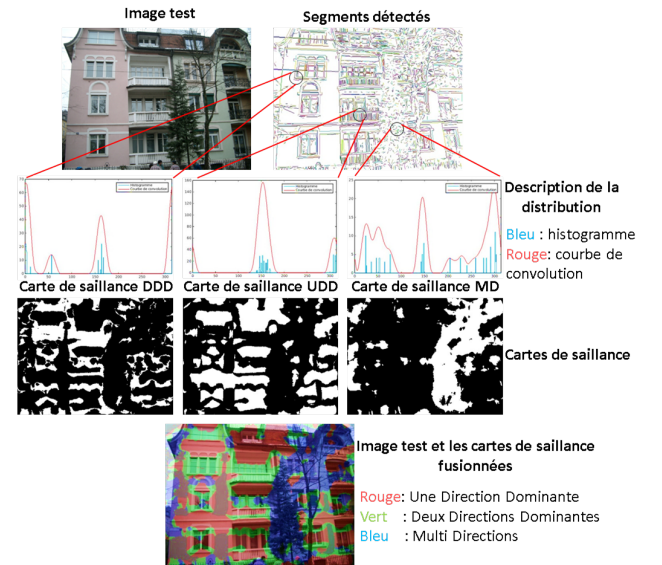


FIGURE 1 – Illustration des différentes étapes de la génération des cartes de saillance.

3.3 Génération des cartes de saillance

La dernière phase de l'approche proposée consiste en la génération de cartes de saillance basées sur le type de la distribution locale des orientations des pixels des segments.

Nous nous concentrons sur trois types de distributions : la distribution unimodale, bimodale et multimodale. Nous pourrions considérer plus de modes, mais les expériences effectuées ont démontré que ces trois modes encapsulent les informations les plus pertinentes sur les principales structures dans l'image. Chaque type correspond respectivement à une carte de saillance, nommée "Une Direction Dominante" (UDD), "Deux Directions Dominantes" (DDD) et "Multi Directions" (MD). Pour chaque pixel de l'image, si la distribution des orientations des pixels dans son voisinage (à l'intérieur de la fenêtre circulaire) est unimodale, bimodale ou multimodale, un pixel blanc est placé sur la carte de saillance correspondante UDD ou DDD ou MD. Avec cette proposition, chaque catégorie de carte de saillance binaire porte une information différente sur les structures saillantes de l'image. Les principales étapes de notre approche sont illustrées dans la figure 1. Dans cet exemple, il est intéressant de constater que ces trois cartes de saillance, basées sur des caractéristiques de bas niveau, sont corrélées aux principaux objets sémantiques dans l'imagerie urbaine (*i.e.* les principales structures de bâtiment avec UDD, les fenêtres avec DDD et la végétation avec MD).

Les cartes de saillance doivent être calculées pour toutes les images ; pour celles de la base d'images ainsi que pour celles de la requête. En pratique, lorsque les cartes sont exploitées pour filtrer des primitives d'image telles que des points d'intérêt, nous ne générons que des indices de saillance correspondant à un voisinage local des points d'intérêt, afin d'accélérer le traitement.

4 Tests et évaluation

Nous avons mis en œuvre l'approche proposée en utilisant C++ sur une plate-forme de processeur Intel Core(TM) i5-2500 3.30GHz CPU et 8Go de mémoire.

Afin d'évaluer les cartes de saillance proposées, nous effectuons plusieurs expériences. Tout d'abord dans la section 4.1, notre proposition est appliquée au problème de la recherche d'image par l'exemple. Nous utilisons un système de recherche d'images classique où les images sont représentées et indexées par des sacs de mots visuels, comme celui proposé par Bhowmik et al. [33] qui permet de combiner plusieurs catégories de descripteurs locaux pour effectuer la recherche. Deuxièmement, nous comparons la performance de recherche de notre approche avec huit approches de l'état de l'art dans la section 4.2. Enfin, dans la section 4.3 nous expérimentons notre méthode pour deux applications particulières : la recherche d'images *cross-domain* à l'aide de peintures et photographies anciennes de monuments et de bâtiments, et la localisation de véhicule basée image.

4.1 Évaluation pour la recherche d'images à partir d'un exemple

Notre objectif est d'exploiter les cartes de saillance proposées pour réduire le volume des descriptions (mots visuels)

en ne retenant que les plus pertinents dans chaque image de la base et des requêtes. Pour ce faire, nous sélectionnons les points d'intérêt situés sur des zones non nulles des cartes de saillance binaires. En pratique, les points sont maintenus si la région autour d'eux est considérée comme saillante (des indices de saillance sont calculés localement autour des points détectés, et non sur toute l'image).

Le filtrage des points d'intérêt peut avoir des conséquences sur le vocabulaire visuel calculé au préalable. Pour cela, nous reconstruisons un nouveau dictionnaire à partir du sous-ensemble de points sélectionnés pour chaque type de carte de saillance. La taille du vocabulaire est déterminée comme une proportion optimale du volume des points d'intérêt retenus.

Dans cette expérience, nous utilisons le détecteur Hessian-affine [34] et les descripteurs SIFT [28] et SURF [31] car cette combinaison a donné la meilleure performance parmi plusieurs autres testées dans [33].

Les évaluations sont effectuées sur trois bases d'images publiques. Deux d'entre eux, la base de Paris (DS_{Paris}) [35] et la base des immeubles de Zurich (DS_{ZuBuD}) [36], sont des paysages urbains - donc des contenus plutôt structurés - alors que la base INRIA-Holidays (DS_{InrHol}) [37] est très variée et beaucoup moins structurée. La figure 2 montre quelques exemples des trois bases et le tableau 1 résume leurs principales caractéristiques.



FIGURE 2 – Exemples des images des trois base d'images.

	DS_{Paris}	DS_{ZuBuD}	DS_{InrHol}
Nb. d'images	6412	1005	1491
Nb. de classes	12	201	500
Nb. de requêtes	55	201	500

Tableau 1 – Caractéristiques principales des trois bases d'images.

Comme les cartes de saillance ne sont calculées que localement autour des points d'intérêt détectés, leur temps de calcul dépend principalement du nombre de points détectés dans l'image ; celui-ci est presque identique pour les trois cartes de saillance ou leurs unions. Le tableau 2 montre les temps d'exécution moyens de notre méthode de saillance visuelle sur deux des bases d'images.

L'approche proposée est évaluée en termes de qualité de la recherche, en utilisant la moyenne de la précision moyenne

	DS _{Paris}	DS _{ZuBuD}
Nb. points d'intérêt (moy)	1571	1011
Temps CPU (moy) (sec)	1.705	0.762
Temps réel (moy) (sec)	1.923	0.976

Tableau 2 – Statistiques de temps de calcul de notre approche de saillance visuelle.

(mAP), et de performance à partir du taux moyen de points d'intérêt sélectionnés (équation 2) et du taux moyen du temps de recherche (équation 3). K est le nombre des images requêtes employées.

$$\tau_{kp} = \frac{\sum_{k=0}^K \text{nombre_de_point_selectionnes}(k)}{\sum_{k=0}^K \text{tous_les_points_detectes}(k)} \quad (2)$$

$$\tau_t = \frac{\sum_{k=0}^K \text{temps_reel_avec_saillance}(k)}{\sum_{k=0}^K \text{temps_reel_sans_saillance}(k)} \quad (3)$$

Les résultats de la méthode de référence [33] sont comparés aux résultats obtenus en exploitant les cartes de saillance et les différentes combinaisons de leurs unions; voir le tableau 3.

Dans les deux bases d'images DS_{Paris} et DS_{ZuBuD}, la meilleure performance, d'après les critères mAP et τ_{kp} , est obtenue en utilisant conjointement les cartes de saillance UDD et DDD (UDD + DDD) : les mAPs sont préservées tout en réduisant la quantité de descriptions notamment, en sélectionnant respectivement 40% à 73% des points d'intérêt. Le temps réel de la recherche dans DS_{Paris} et DS_{ZuBuD} est respectivement réduit de 42,9% et de 26,5%. En outre, la carte UDD+DDD a amélioré légèrement la performance de 1,9% dans DS_{Paris}; Il pourrait être justifié par l'élimination des points d'intérêt situés dans des zones mal structurées, généralement associés à des cartes MD (*i.e.* arbres, personnes). Cependant, les performances ne sont pas améliorées en utilisant nos cartes de saillance dans la base DS_{InrHol}; cela peut être expliqué par la diversité des contenus qui ne présentent pas de structures régulières (c'est principalement des contenu naturels, personnes, eau, effets du feu, etc.) comme dans les deux autres bases, ce qui rend les cartes de saillance proposées moins adaptées à ces genre de contenu.

4.2 Comparaison avec l'état de l'art

Nous considérons huit méthodes de saillance visuelle de l'état de l'art : GBVS [6], SUN [7], ITTI [1], ROSIN [15], SR [4], COV [38], MC [39], GR [5]. La Figure 3 illustre un exemple des cartes de saillance obtenues de ces méthodes. Ces méthodes ont été choisies parce que leurs mécanismes pour obtenir les régions saillantes sont différents. De plus, nous utilisons également une 9^{me} approche qui consiste à sélectionner dans chaque image aléatoirement des mots visuels (Rand).

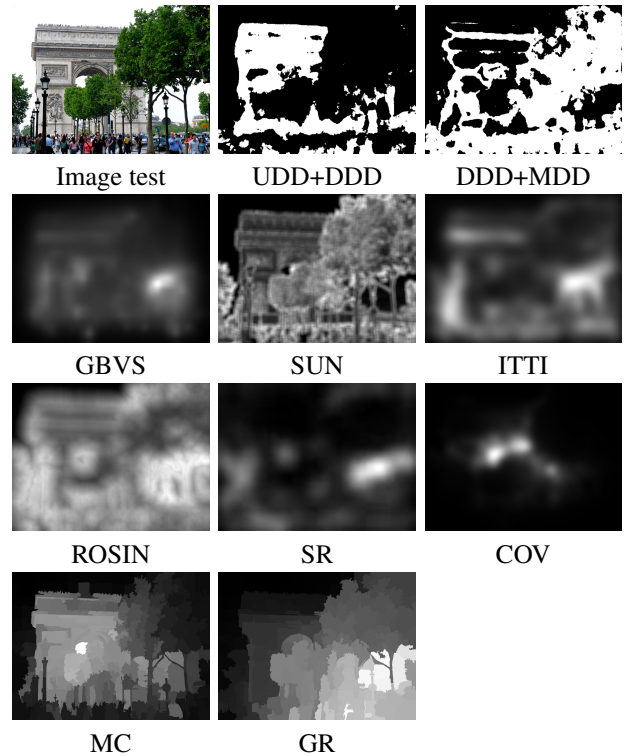


FIGURE 3 – Exemple des cartes de saillance des méthodes de l'état de l'art et nos cartes de saillance UDD+DDD et DDD+MD (l'image test de DS_{Paris}).

Pour chaque image de DS_{Paris}, chaque approche de saillance visuelle est accordée pour obtenir une quantité donnée de mots visuels. Afin de fournir une comparaison équitable, nous utilisons deux taux particuliers (40 % et 79 %) qui correspondent aux volumes de points d'intérêt sélectionnés respectivement UDD+DDD et DDD+MD (*i.e.* sont choisis parce qu'ils ont obtenu les meilleures performances, voir Tableau 3). Le Tableau 4 montre les mAPs obtenues utilisant DS_{Paris}.

En réduisant la quantité de descriptions en sélectionnant 79% et 40% de points d'intérêt, les mAPs sont réduites pour toutes les méthodes de l'état de l'art, à l'exception GBVS, COV et MC qui obtiennent la meilleure mAP avec le taux 79%. Néanmoins, nous pouvons voir que avec le taux de sélection de 40%, notre carte ODD+TDD a donné les meilleurs résultats.

4.3 Applications

La saillance visuelle a été exploitée dans plusieurs domaines, telles que la détection et la reconnaissance d'objets, la classification des images, l'évaluation de la qualité de l'image, la localisation, etc. Dans ce papier, nous nous concentrons sur deux applications particulières : la recherche d'images cross-domain; Elle peut être utilisé pour localiser des documents anciens comme a été fait dans [40], et la deuxième application est la localisation de véhicule dans la cartographie mobile (mobile mapping).

	DS _{Paris}			DS _{ZuBuD}			DS _{InrHol}	
	mAP	τ_{kp}	τ_t	mAP	τ_{kp}	τ_t	mAP	τ_{kp}
Référence	0.544	1	1	0.941	1	1	0.645	1
UDD	0.359	0.181	0.185	0.875	30.41	0.289	0.350	0.089
DDD	0.428	0.218	0.265	0.875	0.425	0.432	0.544	0.094
MD	0.448	0.571	0.591	0.654	0.253	0.303	0.340	0.808
UDD+DDD	0.563	0.400	0.571	0.937	0.730	0.735	0.416	0.183
DDD+MD	0.546	0.789	1.018	0.895	0.679	0.699	0.600	0.902
UDD+MD	0.488	0.752	0.677	0.905	0.557	0.590	0.602	0.897
UDD+DDD+MD	0.549	0.970	0.950	0.941	0.983	1.003	0.631	0.991

Tableau 3 – Évaluation de l’impact du filtrage des mots visuels sur différent cartes de saillance, et comparaison avec la référence.

Taux de sélection (%)	79	40
GBVS [6]	0.554	0.501
SUN [7]	0.537	0.477
ITTI [1]	0.515	0.416
ROSIN [15]	0.488	0.378
SR [4]	0.489	0.360
COV [38]	0.556	0.532
MC [39]	0.555	0.530
GR [5]	0.536	0.477
Selection aléatoire	0.533	0.476
Notre méthode	DDD+MD	UDD+DDD
	0.546	0.563
Référence	0.544	

Tableau 4 – mAPs obtenues dans DS_{Paris} avec filtrage (utilisant les cartes de saillance et la sélection aléatoire) et sans filtrage (référence).

Recherche d’image en Cross-domain. Dans cette expérience, nous utilisons DS_{Paris} comme base d’images et comme images requêtes nous utilisons 32 peintures et images anciennes de différentes tailles trouvées dans flickr et google. Les images requêtes contiennent quatre cibles de DS_{Paris} (Arc de Triomphe, Tour Eiffel, Sacre Coeur, Pantheon and Notre Dame). La figure 4 montre deux échantillons d’images utilisées et le Tableau 5 résume les résultats obtenus.

	mAP	Nb. kp	τ_t
Sans saillance	0.319	986	1
Avec saillance	0.359	381	0.479

Tableau 5 – Résultats de la recherche d’images obtenus utilisant peintures et images anciennes comme requêtes

Notre carte de saillance UDD+DDD a amélioré la qualité de la recherche de 4% tandis que le temps de calcul est amélioré de 52.1% avec 62.1% de points d’intérêt filtrés ; peut être justifié par le fait que notre approche de saillance visuelle basée sur des segments de droite qui préservent l’information dans le cas de photos prises à des moments

différents et avec des technologies différentes (cas de la peinture).

Localisation de véhicule basée image. Nous avons intégré notre carte de saillance UDD+DDD dans le système de localisation de véhicules proposé par Qu et al. [41]. Ce système effectue une estimation de pose 6D basée sur les signalisations routières et des points d’intérêt extraits dans des images de type *street-view* acquises avec un système de cartographie mobile (mobile mapping).

Trois séquences d’images de type *street-view* (voir Figure 5) ont été utilisées pour évaluer notre méthode de saillance visuelle . Ces séquences ont des longueurs de 520, 721 et 976 mètres et elles contiennent respectivement 206, 465 et 331 images. Le Tableau 6 montre la moyenne des erreurs de translation et de rotation obtenues lors de l’estimation de pose 6D des trois séquences, la moyenne du temps de calcul par séquence et la moyenne des points d’intérêt détectés et retenus par image.



FIGURE 4 – Exemples des images *street-view*.

	avec saillance	sans saillance
Erreur translation (%)	0.7160	0.8973
Erreur rotation (deg/m)	0.000025	0.000031
Temps (sec/seq)	155.92	212.67
Points d’intérêt (kp/img)	2063.73	3056.97

Tableau 6 – Résultats de la localisation utilisant trois séquences d’images de type *street-view*.

On peut conclure que notre méthode de saillance améliore considérablement les résultats de localisation (translation et rotation) de 20,21% en réduisant le nombre de points d’intérêt de 32,41% et le temps de calcul de 26,69%.

5 CONCLUSION

Nous avons présenté dans ce papier une nouvelle approche de saillance visuelle basée sur l'analyse de la distribution locale de l'orientation des contours dans l'image. Cette approche a été expérimentée comme une solution intéressante pour filtrer efficacement les caractéristiques locales lorsque le contenu de l'images est structuré. cette approche a été évaluée dans le cadre de la recherche d'image par contenu visuel. Nous avons montré que la qualité de la recherche est maintenue, tout en filtrant les points d'intérêt notamment (jusqu'à 60 %) en ne gardant que ceux correspondant à des zones particulièrement structurées (cartes UDD+DDD). La méthode proposée a prouvé son efficacité lorsqu'elle était exploitée dans le cadre de deux applications : la recherche d'images cross-domain et la localisation basée image dans les paysages urbains. Dans les travaux futurs, nous prévoyons d'améliorer notre approche en adaptant la sélection de la meilleure configuration de cartes de saillance à chaque image de requête et non seulement globalement à toutes la base d'image.

Références

- [1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis & Machine Intelligence*, , no. 11, pp. 1254–1259, 1998.
- [2] R. Achanta and S. Ssstrunk, "Saliency detection using maximum symmetric surround," in *International Conference on Image Processing (ICIP)*, 2010, pp. 2653–2656.
- [3] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.
- [4] X. Hou and L. Zhang, "Saliency detection : A spectral residual approach," in *Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.
- [5] C. Yang, L. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 637–640, 2013.
- [6] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2006, pp. 545–552.
- [7] L. Zhang, M.H. Tong, T.K. Marks, H. Shan, and G.W. Cottrell, "Sun : A bayesian framework for saliency using natural statistics," *Journal of vision*, vol. 8, no. 7, pp. 32–32, 2008.
- [8] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 171–177, 2010.
- [9] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Advances in neural information processing systems*, 2005, pp. 155–162.
- [10] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image retrieval : Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 55 :1–5 :60, 2008.
- [11] T. Tuytelaars and k. Mikolajczyk, "Local invariant feature detectors : a survey," *Foundations and trends® in computer graphics and vision*, vol. 3, no. 3, pp. 177–280, 2008.
- [12] H. Samet, *Foundations of Multidimensional And Metric Data Structures*, 2006.
- [13] M. Chen, S. Mao, and Y. Liu, "Big data : A survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.
- [14] A.M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive psychology*, vol. 12, pp. 97–136, 1980.
- [15] P.L. Rosin, "A simple method for detecting salient regions," *Pattern Recognition*, vol. 42, no. 11, pp. 2363–2371, 2009.
- [16] G. Borgefors, "Distance transformations in arbitrary dimensions," *Computer vision, graphics, and image processing*, vol. 27, no. 3, pp. 321–345, 1984.
- [17] Y. Lei, X. Gui, and Z. Shi, "Feature description and image retrieval based on visual attention model," *Journal of Multimedia*, vol. 6, no. 1, pp. 56–65, 2011.
- [18] S. Wan, P. Jin, and L. Yue, "An approach for image retrieval based on visual saliency," in *Image Analysis and Signal Processing*, 2009, pp. 172–175.
- [19] D. Awad, V. Courboulay, and A. Revel, "A new hybrid texture-perceptual descriptor : application cibir," in *International Conference on Pattern Recognition (ICPR)*, 2014, pp. 1150–1155.
- [20] M.P. Da Silva, V. Courboulay, and P. Estrailier, "Objective validation of a dynamical and plausible computational model of visual attention," in *Visual Information Processing (EUVIP)*, 2011, pp. 223–228.
- [21] O. Marques, L.M. Mayron, G.B. Borba, and H.R. Gamba, "Using visual attention to extract regions of interest in the context of image retrieval," in *ACM-SE*, 2006, pp. 638–643.
- [22] G.T. Ozyer and F.Y. Vural, "An attention-based image retrieval system," in *Machine Learning and Applications and Workshops (ICMLA)*, 2011, vol. 1, pp. 96–99.
- [23] Z. Zdziarski and R. Dahyot, "Feature selection using visual saliency for content-based image retrieval," in *Signals and Systems Conference (ISSC)*, 2012, pp. 1–6.
- [24] H. Gao and Z. Yang, "Integrated visual saliency based local feature selection for image retrieval," in *Intelligence Information Processing and Trusted Computing*, 2011, pp. 47–50.

- [25] Z. Wen, J. Gao, R. Luo, and H. Wu, "Image retrieval based on saliency attention," in *Foundations of Intelligent Systems*, pp. 177–188. Springer, 2014.
- [26] J. Liu, F. Meng, F. Mu, and Y. Zhang, "An improved image retrieval method based on sift algorithm and saliency map," in *Fuzzy Systems and Knowledge Discovery*, 2014, pp. 766–770.
- [27] D. Awad, V. Courboulay, and A. Revel, "Saliency filtering of sift detectors : application to cbir," in *Advanced Concepts for Intelligent Vision Systems*, 2012, pp. 290–300.
- [28] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [29] D. Awad, M. Mancas, N. Riche, V. Courboulay, and A. Revel, "A cbir-based evaluation framework for visual attention models," in *European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 1526–1530.
- [30] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Computer Vision and Pattern Recognition*, vol. 1, pp. 886–893, 2005.
- [31] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf : Speeded up robust features," in *European conference on computer vision*, 2006, pp. 404–417.
- [32] R.G. Von Gioi, J. Jakubowicz, J.M. Morel, and G. Randall, "Lsd : A fast line segment detector with a false detection control," *Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 722–732, 2010.
- [33] N. Bhowmik, R. González, V. Gouet-Brunet, H. Pedrini, and G. Bloch, "Efficient fusion of multidimensional descriptors for image retrieval," *International Conference on Image Processing (ICIP)*, 2014.
- [34] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [35] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization : Improving particular object retrieval in large scale image databases," in *Conference on Computer Vision and Pattern Recognition*, 2008.
- [36] H. Shao, T. Svoboda, and L. Van Gool, "Zubud-zurich buildings database for image based recognition," *Computer Vision Lab, Swiss Federal Institute of Technology*, vol. 260, 2003.
- [37] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," *European conference on computer vision*, pp. 304–317, 2008.
- [38] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances," *Journal of vision*, vol. 13, no. 4, pp. 11–11, 2013.
- [39] B. Jiang, L. Zhang, H. Lu, C. Yang, and M. H. Yang, "Saliency detection via absorbing markov chain," in *Computer Vision*, 2013, pp. 1665–1672.
- [40] N. Bhowmik, L. Weng, V. Gouet-Brunet, and B. Soheilian, "Cross-domain image localization by adaptive feature fusion," *JURSE*, 2017.
- [41] B. Qu, X. and Soheilian and N. Paparoditis, "Vehicle localization using mono-camera and geo-referenced traffic signs," in *Intelligent Vehicles Symposium (IV)*, June 2015, pp. 605–610.