# Encodage spatial variable basé sur les propriétés du système visuel humain.

Jérôme Schmaltz
École de Technologie Supérieure

***Résumé* – Cet article présente différentes approches de la compression vidéo basée sur les caractéristiques du système visuel humain (SVH). On y présente les différentes techniques d'acquisition des zones d'intérêt, les techniques de traitement des images en appliquant un filtrage fovéal et finalement les principes de codage de données en relation avec les caractéristiques du SVH. L'article propose une implémentation d'une approche multi résolution par pyramides gaussiennes pour filtrer des images selon les caractéristiques du SVH exposées. Nous pourrons voir des gains de l'ordre de 22% lors de compression des images.**

***Termes* – fovéal, filtrage.**

## I. Introduction

La compression vidéo est une des technologies clés pour les medias digitaux. Elle est utilisée afin d'accroître l'efficacité de la transmission et du stockage des trames vidéo. Le but ultime est de compresser les trames sans pour autant occasionner, de façon significative, une dégradation de la qualité visuelle. Cet objectif peut être accompli en utilisant plusieurs types d'informations redondantes issues des trames soit : spatiale, temporelle, spectrale ou psycho visuelle. Les trois premières étant dépendantes entièrement du contenu de l'image, la dernière quant à elle, directement influencée par les caractéristiques du système visuel humain. Plusieurs caractéristiques, tel que le masquage de texture, la fréquence et la luminescence ont été investigué et apportés de meilleurs ratios de compression et une qualité visuelle accrue [2].

L'œil humain est doté de deux classes de photorécepteurs : les cônes et bâtonnets. Les cônes répondent aux stimuli durant l'activité diurne tandis que les bâtonnets réagissent avec des lumières de basses intensités. L'acuité visuelle augmente au fur et à mesure que le stimulus se rapproche de la fovéa, zone qui compte une densité accrue des photorécepteurs.

La résolution spatiale est donc directement proportionnelle à la concentration des photorécepteurs en fonction de la distance du stimulus avec la fovéa. La résolution maximale peut donc être atteinte lorsqu'une luminescence élevée est projetée à un angle visuel nul avec la fovéa. Dans le cas contraire, lorsque la distance avec la fovéa augmente (plus de $5^o$), l'acuité diminue de façon logarithmique [1].

Plusieurs approches [9, 10, 11, 12, 13, 14, 15] de compression vidéo utilisent ce concept afin d'améliorer la taille des flux lors des transmissions. Une première étape consiste en la division des images en régions d'intérêt [1, 2, 3] en fonction du point de regard de l'utilisateur ce qui permet de prioriser l'encodage.

Cette priorisation repose sur la distribution non uniforme des photorécepteurs dans la région de l'œil. Une approche simple consiste en une compression accrue en tolérant une plus grande dégradation de la qualité visuelle dans les zones de moins grande priorité [2]. (*À titre indicatif, d'autres approches plus complexes, prennent en considération plusieurs autres paramètres environnementaux tels que la bande passante du réseau de transmission par exemple* [3]).

Une fois les zones d'intérêts ségrégées et priorisées, chacune d'elle devra être encodée avec diverses résolutions et ce, de manière efficace. Dans un contexte temps réel (imaginons une transmission vidéoconférence), des contraintes de temps rendent l'encodage spatial variable complexe et onéreux en termes de temps computationnel [2].

Cet article présente un survol des différentes méthodes contenues dans chacune des étapes servant à produire des images reposant sur les caractéristiques du SVH. Nous présenterons les techniques d'acquisition, de filtrage et d'encodage.

## II. Revue de Littérature

### A. Méthodes d'acquisitions

La sélection des régions de priorités demeurent un problème non-résolu. Plusieurs avancées [1,3] ont cependant été constatées dans des contextes distincts : dans un contexte temps réel avec transmission vidéo sur un réseau avec bande passante limitée puis, dans un second cas, dans un contexte plus générique non-interactif.

Le contexte temps réel (interactif) est réalisé par le biais d'un capteur indiquant les fixations du regard de l'observateur humain, ce qui permet d'appliquer un filtre répliquant les caractéristiques spatiales de la fovéa sur le contenu vidéo en temps réel. Cela permet donc d'allouer le maximum de la bande passante aux données correspondant à la zone fixée du regard (pour une représentation haute fidélité) puis de transmettre le reste de la vision périphérique sur la bande passante restante.

10 Décembre 2009. École de technologie supérieure.

Cette approche à démontrée son efficacité lorsque les conditions ambiantes et les systèmes visuels des observateurs étaient semblables [1] et elle à même trouvé sa place dans des contextes où des zones d'intérêts étaient fixes et pré déterminé ou lorsque le capteur était remplacé par un simple pointeur [4].

Cependant, dans un contexte où plusieurs observateurs sont présents dans la même scène, il est difficile de concevoir la réalisation technique d'un système de capteurs ainsi que la priorisation et filtrage des différentes zones d'intérêts en temps réel. De plus, la latence occasionnée par les réseaux lors de la transmission cause souvent plusieurs difficultés lors de l'application du filtrage car les modèles existants ne tiennent pas compte de la nature rapide de l'œil humain causant plusieurs erreurs [3].

À cet effet, l'utilisation d'une technique non-interactive est de mise. Grâce à plusieurs algorithmes de vision par ordinateur, maintes études [6, 7] ont démontrées qu'il était possible d'identifier les zones d'intérêts sans l'aide d'humains en utilisant des méthodes algorithmiques ayant le potentiel de rendre ce processus pratique et à moindres coûts.

L'identification des zones d'intérêts repose sur les propriétés du système visuel humain permettant de définir les régions de perception importantes (taille de l'objet, contraste, forme, couleur, etc.). Cette approche est fonctionnelle dans le cas où les propriétés sont bien définies.

Une des limitations connue à cet égard est que les propriétés du système visuel humain sont difficiles à implémenter puisque l'évaluation de chacune d'elle repose souvent sur un pré traitement coûteux. Par exemple, pour l'évaluation de la taille d'un objet, une segmentation de la forme est nécessaire.

*Technique Interactive*

De manière générale, afin d'appliquer le filtrage fovéal sur une séquence vidéo, un système de capteur doit être installé afin d'identifier le point de regard de l'utilisateur sur l'image (*Voir* Figure 1). Chaque image est ensuite encodée individuellement en prenant en compte la coordonnée du point de regard de l'observateur.

Par la suite, afin de permettre un rendu en temps réel, la compression des trames est effectuée selon une stratégie d'encodage par sous-blocs, en relation avec la priorisation des zones d'intérêts identifié par le point de regard. Ainsi, la quantification est implémentée au sein des sous-blocs et non sur la trame en entier, ce qui permet en plus d'atteindre des performances temps réel, une implémentation simple en termes de requis matériel pour des solutions matériels [2].

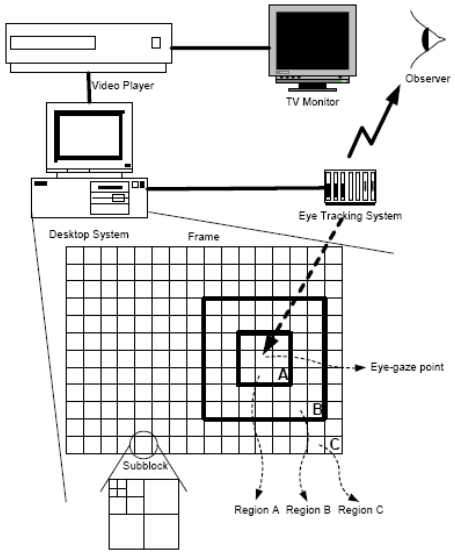


Figure 1. Technique Interactive. (Source : [2])

La ségrégation des zones d'intérêts selon le point de regard permet d'apparenté chacune d'elle à un type de codage de sous-bloc. Ainsi, si on prend en exemple la Figure 1, un codage sans perte sera effectué dans la zone d'intérêt A, la zone B comptera des pertes négligeables tandis que la zone C contiendra des blocs avec pertes.

Tel que mentionné, dans un contexte temps réel avec transmission sur réseaux, cette technique comporte malheureusement des limitations importantes : elle ne tient pas compte du mouvement saccadé de l'œil humain changeant de position rapidement. Ainsi, avec un réseau de télécommunication ayant un délai entre 20ms et quelques secondes, les saccades de l'œil humain variant de 10 à 100 degrés pendant ce laps de temps, réduisent potentiellement les avantages d'un design d'une fenêtre para fovéale de grande précision [3].

*Technique non-interactive*

Les techniques non-interactives reposent sur un modèle computationnel implémentant les caractéristiques du système visuel humain dans le but d'identifier les possibles zones d'intérêts typiquement déterminé par un observateur.

À ce fait, Osberger et Maeder [8] ont proposé une approche reposant sur le modèle d'attention visuel. L'algorithme développé utilise une carte d'importance (*Importance Map*) recueillant des données de segmentation reposant sur les facteurs influençant l'attention et le mouvement de l'œil. Parmi les facteurs retenus, notons le contraste, la taille, la forme, la location et le fond d'écran. Une combinaison des facteurs permet de prédire les régions d'intérêts.

La Figure 2 illustre les étapes de construction d'une carte d'importance.

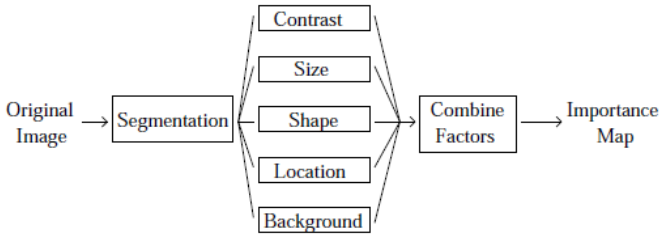


Figure 2. Schéma bloc de la construction d'une carte d'importance (Source : [8]).

Ce type d'approche pose un problème puisque des propriétés de la vision humaine restes difficile à implémentée. À cet égard, Itti [1] à présenté un modèle computationnel répliquant les caractéristiques de la vision humaine, plus particulièrement sur le processus engendré par neurones. Ainsi, les propriétés des réponses neuronales sont employées dans la construction d'une carte de saillence mettant en évidence les zones de l'image les plus frappantes pour l'œil humain.

### *B. Processus de traitement fovéal*

Une fois les zones d'intérêts déterminées, le processus de traitement fovéal permet de filtrer les images en considérant le point de regard de l'observateur en répliquant les caractéristiques spatiales de la fovéa.

#### *1) Méthodes géométriques*

L'application d'une méthode géométrique consiste à utiliser les propriétés géométriques décrites par le comportement de la fovéa lors de l'échantillonnage d'une image. Il y aura donc transformation des coordonnées dans le référentiel de la fovéa.

Cette transformation de référentiels est opérée associant à la fois l'aspect non-uniforme de l'échantillonnage géométrique et la transformation adaptive de coordonnées spatiales. Lorsque la transformation est appliquée aux points échantillonnés, une densité d'échantillonnage uniforme est obtenue dans le nouveau référentiel. Cette transformation de référentiel peut être décrite par la formule suivante [9] :

$$W = \log(z + a) \qquad (1)$$

Où ***a*** est une constante, ***z*** et ***w*** sont des nombres complexes représentant des position dans le système de coordonnées original et transformé respectivement. Cette approche peut être utilisée de plusieurs manières. Les deux sections suivantes en font état.

*Géométrie d'échantillonnage basé sur les propriétés de la fovéa.*

Une première utilisation consiste à appliquer la transformation de référentiel directement sur une image de résolution uniforme, l'espace de l'image étant directement portée dans le nouveau système de coordonnées. Dans le domaine de transformation, une image est traitée comme une image à résolution constante ce qui permet le l'application de filtres linéaires et non-linéaires, analogue au traitement d'une image à résolution uniforme. Finalement, une dernière étape consiste à reconvertir le référentiel afin d'obtenir une image traitée.

La difficulté principale à cette approche est l'indexation des pixels de l'image, provenant originalement d'une grille à position décrites par des entiers vers des positions non-entières lorsque le changement de référentiel est appliqué. Les différentes transformations, telle l'interpolation et l'échantillonnage doivent être appliqué dans les domaines de la transformation ainsi que dans le domaine inverse. Ces procédures ne sont pas complexes mais peuvent engendrer des distorsions [9].

*Approche par Super Pixels*

Une seconde approche consiste à grouper et moyenner un ensemble de pixels en un Super Pixel dont la forme et les dimensions sont déterminées par les propriétés de densité d'échantillonnage de la rétine. Wallace *et al* [10], ont tenté de développer une structure géométrique, présenté à la Figure 3, adhérant avec le concept de transformation de référentiels présenté à l'équation 1.

Comme on peut l'observer, la structure proposée rend son application difficile étant donné les formes complexes des Super Pixels.

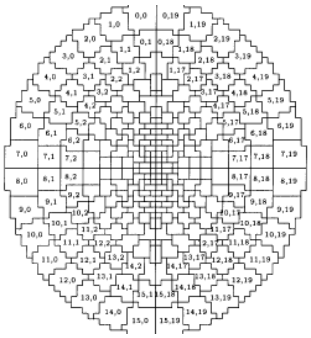

Figure 3. Structure de Super Pixels de Wallace *et al*. (Source : [10])

Kortum et Geisler [12] ont proposés une structure plus simpliste proposant une division uniforme et rectangulaire des Super Pixels (*Voir* Figure 4).

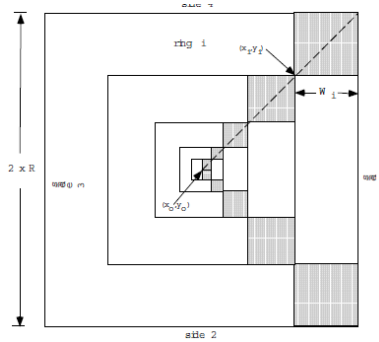


Figure 4. Structure de Super Pixels proposée par Krotum et Geisler. (Source : [12])

L'arrangement de Kortum et Geisler [12] est appelé Grille de Résolution où la taille de chaque Super Pixel est défini par rapport à sa distance au centre de l'écran et les pixels qui les composent sont tous dotés du même niveau de gris.

Il y a deux inconvénients à l'application des approches Super Pixels [9]:

- Premièrement, les discontinuités observées peuvent causer des artefacts importants. Un moyen palliatif consiste à appliquer des méthodes de fondu afin de réduire l'effet de frontière, mais causant du coup une augmentation du temps computationnel.
- Deuxièmement, le masque de Super Pixel doit être recalculé lorsque le point de fixation de l'observateur change.

*Approche de la Géométrie Rétinal*

En terminant, mentionnons qu'une autre approche consiste à employer les propriétés géométriques de la rétine afin de guider le design d'une structure non-uniforme pour le sous-échantillonnage d'une image à résolution uniforme. Un exemple de structure est donné à la Figure 5. Kuyel *et al* [11] propose de ré échantillonné une image à résolution uniforme en une image à résolution reposant sur les caractéristiques de densité d'échantillonnage de la rétine humaine. Finalement, une interpolation B-Spline est utilisée afin de reconstruire les images filtrées.

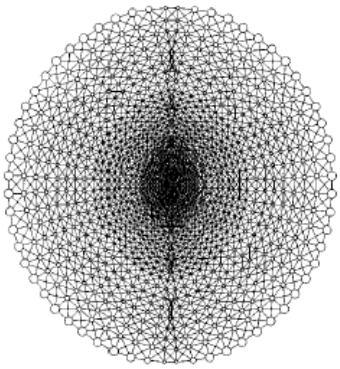

Figure 5. Structure fovéale proposée par Kuyel *et al*. (Source:[11]).

*2) Méthodes par application de filtres*

La théorie de l'échantillonnage stipule que la fréquence la plus élevée d'un signal peut être représentée sans aliasing par la moitié de la fréquence d'échantillonnage. De manière analogue, on peut exprimer que la bande passante du signal d'une image perçue est limitée par la densité d'échantillonnage de la rétine. Cette idée est reprise à travers les approches présentées dans cette section puisque le processus de traitement fovéal est implémenté en utilisant des filtres passes bas, variant au décalage, où les fréquences de coupure sont déterminées par la densité d'échantillonnage de la rétine.

Puisque l'échantillonnage spatial de la rétine varie graduellement, une implémentation du filtrage consiste à utiliser différents filtres passes bas à différentes location de l'image. Cette méthode optimale permet ainsi de produire des images de grandes qualités mais en étant extrêmement gourmande en temps computationnel et difficile à réaliser dans le contexte d'une faible bande passante locale.

*Approche par Banque de filtres*

L'approche par banque de filtres permet un compromis entre l'acuité et le coût du processus de filtrage. Tel qu'illustré à la Figure 6, une banque finie de filtres avec des réponses fréquentielles variantes sont appliquées de manière uniforme à l'image en entrée, produisant plusieurs images filtrées.

L'image finale est produite par la combinaison des images filtrée. Le processus combinatoire se fait de manière variable où la représentation spatiale est fonction de la densité d'échantillonnage de la rétine.

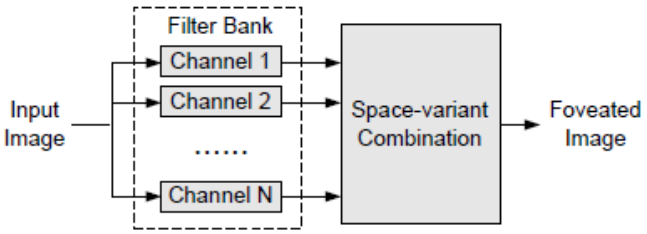


Figure 6. Approche par Banque de Filtres (Source : [9)].

Plusieurs contraintes limitent l'utilisation d'une telle méthode. Premièrement, le design des filtres qui composent la banque de filtres peut être orienté en utilisant des filtres passes bas ou même passe bande ce qui implique l'ajustement du processus combinatoire. Deuxièmement, des problèmes de design de filtres peuvent survenir lors du design d'un filtre à réponse finie, par exemple. Dans ce cas particulier il faudra alors considérer les compromis à faire entre l'effet de vagues, la largeur de bande de transition et la complexité de l'implémentation.

Toutefois, il est à noter que l'intégration de cette méthode à des systèmes de codage vidéo se fait de manière complètement transparente puisqu'aucune modification n'est nécessaire à l'encodeur ou au décodeur. Effectivement, il ne s'agît que d'ajouter le module de filtrage en amont de l'encodeur.

*Approche quantificateur DCT*

Plusieurs auteurs [13, 14] ont proposé d'appliquer le traitement d'un filtrage fovéal dans le domaine DCT combiné à un processus de quantification, que l'on retrouve dans les standards H.26x et MPEG. L'implémentation de tels systèmes est cependant difficile à réaliser car le codage par blocs de la trame courante doit se faire en utilisant des régions de la trame précédente pouvant couvrir plusieurs blocs DCT ayant des résolutions variables.

Point à noter est que le l'intégration du processus du traitement fovéal doit nécessiter des modifications à

l'encodeur; le décodeur quant à lui, agira de manière transparente.

*3) Méthodes multi résolutions*

Les méthodes multi résolutions peuvent être considérées comme étant une combinaison des méthodes géométriques et de filtrages de part le fait que l'image originale, de résolution uniforme, est transformée en différentes résolutions sur lesquels différents filtres sont appliqués. Les méthodes multi résolutions possèdent les avantages des méthodes géométriques et de filtrage :

Premièrement, les transformations géométriques ne sont pas complexes et ne reposent pas sur les Super Pixels puisque la transformation de résolution peut se faire en utilisant un sous-échantillonnage simple. Cela permet donc d'économiser de l'espace de stockage et permet une indexation directe des pixels.

Deuxièmement, suivant le sous-échantillonnage, le nombre de coefficients transformés pour chaque résolution produite est grandement diminué, ce qui en résulte un coût computationnel moindre occasionné le processus de filtrage.

*Approche par pyramides multi résolutions*

Burt [15] à proposé l'utilisation de pyramides multi résolutions afin de sélectionner de l'information sujette à produire une image fovéale. Afin d'empêcher de nombreuses discontinuités, des filtres de fondu sont appliqués. Cette structure efficiente de l'information permet de produire des images dans des contextes temps réels[1].

La dernière méthode présentée pour le processus de traitement fovéal d'une image concerne les méthodes d'encodage.

*4) Méthodes d'encodage*

Une tendance récente dans le domaine des communications vidéo est le développement d'algorithmes à débit variable permettant l'extraction et l'encodage d'information visuelle en taux de bits variables continus à partir d'un flux de bits compressés.

Un exemple de ce principe est illustré à la Figure 7 montrant une séquence vidéo encodée avec un taux variable de bits, stockés trame par trame.

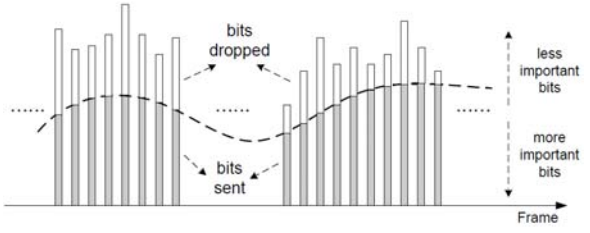


Figure 7. Flux de données à débit variable. Chacun des bits est ordonné selon son importance, chaque barre représente le flux de bits pour une trame d'une séquence vidéo. (Source : [16]).

[1] Il est à noter que la section III se base sur cette approche.

Durant la transmission des données codées sur le réseau, il est donc possible de tronquer les données et de n'envoyer que les bits les plus importants du flux. Cette capacité à faire varier la quantité d'information permet d'envoyer une même séquence selon un niveau de qualité et quantité d'information variables. Cette caractéristique convient parfaitement à la transmission sur des canaux hétérogènes, multi utilisateurs tel qu'Internet.

Des solutions traditionnelles, tel le transcodage vidéo [17] ou le codage à répétition requièrent plus de ressources en termes computationnels et d'espace de stockage. De plus, elles ne sont permettent pas de s'adapter rapidement aux conditions changeantes des canaux puisqu'il n'est pas possible de faire varier, de façon arbitraire, le taux des données dans les séquences vidéos compressées.

À l'inverse, les codecs à taux variable continu permettent l'encodage variable des séquences en fonctions des paramètres environnement tel la bande passante attribuable sur le canal de transmission.

L'application du processus lors de l'encodage vidéo, consiste à organiser les flux afin de fournir des informations visuelles, encodées selon le processus de traitement fovéal, à un taux de données variables.

## III. Filtrage Multi-Résolution à l'aide de Pyramides Gaussiennes

Une solution prometteuse retenue présentée dans cet article est issue des méthodes multi résolutions présentée à la section II.3. Basé sur les concepts avancés par Kuyel *et al*. [11], l'emploi d'une pyramide multi résolutions basée sur un filtre passe-bas Gaussien à été implémentée afin de générer des images à résolution spatiale variable reflétant les caractéristiques fovéales [18].

Pour ce faire, une première étape consiste à générer de multiples images à résolution variables suivant une pyramide Gaussienne de plusieurs niveaux, tel qu'illustré à la Figure 8.

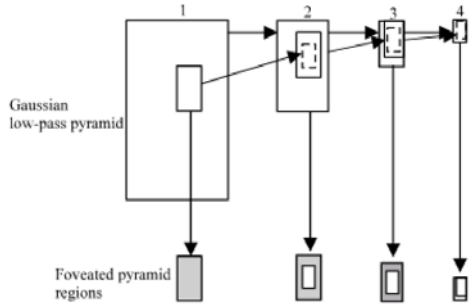


Figure 8. Pyramide Gaussienne à multiples niveaux (Source : [18]).

Chaque image générée se trouve à être deux fois plus petite que le niveau inférieur, le premier niveau étant constitué de l'image originale. Lors du processus fovéal, le filtre Gaussien appliqué à chaque image de la pyramide, permettra de récupérer certaines portions floues lors de la construction de l'image finale. Le filtre Gaussien (2) utilisé se décrit de la façon suivante :

$$H(x, y) = \exp\left[-\frac{l^2(x,y)}{2\sigma_0^2}\right] \quad (2)$$

Où, σ représente la fréquence de coupure dans l'image filtrée. Dans notre cas, cette fréquence est déterminée par la un modèle psychométrique proposé par Perry et Geisler [4] :

$$CT(f, e) = CT_0 \exp\left(\alpha . f . \frac{e + e_2}{e_2}\right) \quad (3)$$

Où ***CT*** représente la tolérance de contraste minimale déterminée en fonction de ***f***, la fréquence spatiale (en cycle par degré) et ***e*** l'excentricité de la rétine. $CT_0$ est le contraste minimal (1/64), $e_2$ est la constante de la moitié de la résolution de l'excentricité (2.3) et α, la constante fréquence spatiale (0.106).

Afin de déterminer la fréquence de coupure $f_c$ pour une excentricité ***e*** donnée, on peut présumer un contraste maximal ***CT*** = 1.0. L'équation (3) se traduit de la manière suivante :

$$f_c = \frac{e_2 \ln(1/CT_0)}{\alpha(e + e_2)} \quad (4)$$

L'équation (4) décrit la relation entre la fréquence de coupure et une excentricité ***e*** pour chaque pixel dans l'image perçue. La Figure 9 présente les éléments géométriques des relations exprimées.

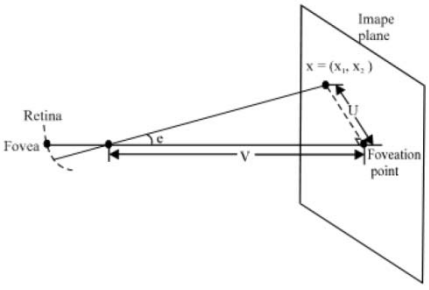


Figure 9. Relation entre le point fovéal et un pixel x (Source : [18]).

Afin de calculer l'excentricité ***e*** pour chaque point, nous devons utiliser la relation suivante :

$$e_c = 180 \cdot tg^{-1}(D \cdot l / V)/\pi \quad (5)$$

Où ***D*** est la précision du moniteur, habituellement ***D***=0.25x10E-3 m [18], ***V*** est la distance écran-rétine (en m) et ***l***, la distance entre un point et le point fovéal est calculé par l'équation (6).

$$l = \left(x_c^2 + y_c^2\right)^{\wedge}(1/2) \quad (6)$$

La résolution perçue par l'œil est cependant limitée par la résolution de l'écran, ***fm***. Cette résolution maximale peut être exprimée de la façon suivante :

$$f_m = \pi/((tg^{-1}((D \cdot (l+1)/V - tg^{-1}((D \cdot (l-1)/V) \times 180) \quad (7)$$

L'équation (7) nous permet de comprendre que la distance est utilisée pour déterminer la fréquence spatiale maximale que le moniteur peut reproduire à chaque pixel. Par la suite, la fréquence maximale du moniteur est divisée par la fréquence maximale attribuable par le système visuel pour chaque pixel :

$$Pyrlevel = f_m / f_c \quad (8)$$

Le résultat ***pyrlevel*** correspond à un niveau fractionnaire pyramidal permettant d'attribuer une valeur pour chaque pixel de l'image finale.

## IV. DESCRIPTION DE LA SOLUTION ET IMPLÉMENTATION

Chaque valeur de ***pyrlevel*** est stockée dans une matrice pour chaque point de l'image. Lorsque des valeurs plus petites que 1 surviennent, cela peut correspond au fait que l'œil peut résoudre des fréquences plus élevées que le moniteur. Rappelons à cet effet que la fréquence maximale du moniteur est un facteur limitant donc, il faudra s'assurer de tronquer correctement la valeur à 1.0, ce qui correspond au niveau le plus élevé de la pyramide, donc ayant une résolution maximale. Inversement il se pourrait que des valeurs soient supérieures au niveau le plus élevé de la pyramide, il faut aussi dans ce cas s'assurer de rester dans les bornes 1 à ***N***, où ***N*** est le nombre de niveaux.

Afin d'utiliser la matrice de ***pyrlevel***, on doit s'assurer que les images de chaque niveaux de la pyramide soient de dimensions identiques à l'image originale. Ainsi, on devra sur échantillonner et appliquer un effet de flou aux images composant la pyramide.

Toutes ces images composeront un tableau tri dimensionnel. Chaque valeur de l'image finale sera déterminée en utilisant une interpolation des valeurs provenant de la matrice ***pyrlevel*** ainsi que des images de chaque niveau associé.

Afin d'appliquer l'effet fovéal à des images en couleur, on doit s'assurer de maintenir la fidélité des couleurs à travers les transformations. Pour ce faire, il faut convertir le système de couleur RGB à YCbCr afin de séparer les composantes de luminescence des composantes de chrominance.

### *Limitations*

Afin d'appliquer le modèle fovéal décrit précédemment à un contexte de transmission de données avec compression, la solution serait de permettre au transmetteur d'envoyer les zones nécessaires de la pyramide multi résolution, la distance ***V*** ainsi que la taille de l'image. L'implémentation

courante ne tient pas compte de ce contexte et ne s'est concentré qu'en la génération de l'image fovéale.

De plus, l'implémentation ne tient pas compte d'un contexte temps réel et ne supporte pas un système de capteur dynamique pour la détermination du point de regard de l'observateur. Le point de regard, ou fovéal est statique et pré déterminé.

Afin de créer des images fovéales, un programme Matlab à été créé implémentant les différentes caractéristiques énoncées dans les sections précédentes.

## V. Analyse des résultats et Discussion

Tel que présenté à la Figure 10, une image originale de résolution uniforme à été prise comme entrée à l'algorithme implémenté. Le Tableau I contient les divers paramètres d'entré qui ont été utilisés.

Figure 10. Image originale uniforme.

TABLEAU I
PARAMÈTRES D'ENTRÉ

| Symbole | Paramètre | Valeur |
|---|---|---|
| V | Distance | 1 m |
| $\alpha$ | Constante de fréquence spatiale | 0.106 |
| $e_2$ | Constante de la moitié de la résolution de l'excentricité | 2.3 |
| $CT_0$ | Contraste minimal | 1/64 |
| | Taille de l'image | 308x410 pixels |
| D | Résolution du moniteur | 0.23 x $10^{-3}$ m |
| | Point fovéal | 200, 140 |

Les divers paramètres d'entré à l'algorithme implémenté.

À partir d'une distance ***V*** de 1.0 m, les degrés fovéaux ont été calculés pour tous les points de l'image originale. La Figure 11 représente une matrice tri dimensionnelle des ***pyrlevel*** calculés pour les points de l'image originale.

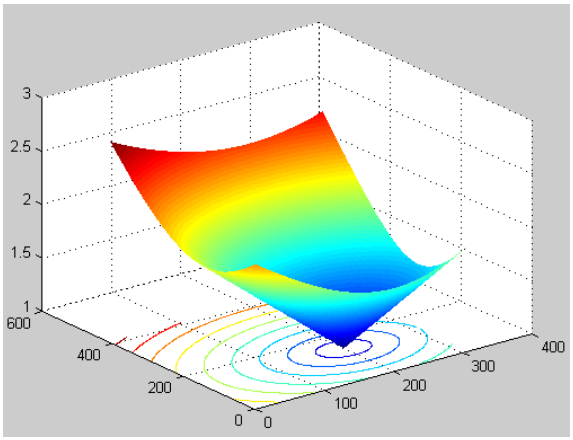


Figure 11. Niveaux pyramidaux de l'image originale.

Au centre fovéal, zone bleue de la Figure 11, la distribution des données à dû être tronquée afin d'exprimer la résolution finie du moniteur comparativement à celle de l'œil humain. Comme on peut l'observer, seulement trois niveaux sont suffisants pour traiter l'image en entrée. Comme Li *et al* [18] le propose, il est possible d'ajuster le nombre maximal de niveaux en contrôlant la création de celles-ci.

La résolution à espace variable à été représenté à la Figure 12. Comme nous pouvons le voir, la distribution s'étant du blanc au noir, le blanc indiquant la résolution maximale et le noir la résolution minimale. On peut identifier clairement que le point fovéal, situé à (200, 140) possède une résolution maximale. À partir du point fovéal la résolution chute rapidement en fonction de l'excentricité.

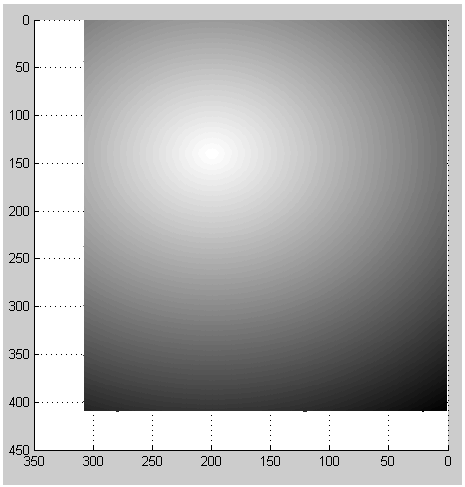


Figure 12. Distribution de la résolution à travers l'image.

La structure pyramidale à été utilisée afin de générer l'image fovéale. Grâce à l'interpolation linéaire, la Figure 13 à été produite.

Figure 13. Image finale.

Cette image montre qu'il n'y a pas de discontinuités marquantes entre les différents niveaux pyramidaux. On peut aussi remarquer qu'à partir du point fovéal l'image semble être de plus en plus pixélisée ou floue, décrivant le comportement du système visuel humain. Li *et al* [18] montre plusieurs autres cas intéressants de la réalisation de cet algorithme. On peut y voir la génération d'image sans interpolation montrant du coup les divers artefacts engendrés par chaque changement de niveaux pyramidal.

Grâce au filtrage appliqué sur l'image originale, des gains de 22% (47.1 kb à 36.9 kb) ont pu être obtenu lors de l'application de la compression JPEG, ce qui montre l'efficacité de l'algorithme implémenté et la congruence des résultats avec Li *et al* exposant le fait qu'ils avaient obtenus des compression de l'ordre de 20% à 30% [18].

## VI. Conclusion

La motivation première du filtrage d'image en appliquant le processus fovéal est la discrétisation d'information redondante dans les régions périphériques du point de regard d'un observateur. Une représentation bien plus efficience des images peut donc être obtenue en supprimant ou réduisant ces informations. La vision humaine ne peut distinguer les images originales des images filtrées par le processus fovéal pour un point de regard situé aux mêmes coordonnées. Li *et al* [18] ont pu obtenir des taux de compressions très impressionnants de l'ordre de 5.9 en appliquant l'algorithme présenté à la section III. Plusieurs autres recherches ont montrées un potentiel de réduction de bande passante de l'ordre de 94.7% [3] en appliquant les méthodes de filtrage présenté à la section II.

De telles données poussent les recherches de façon continu et leur applications diverses se fait déjà voir dans plusieurs standards de compression d'image (JPEG) et vidéo (H.26x et MPEG).

La simulation du comportement du système visuel humain est un domaine assez complexe puisque qu'elle est sujette à des connaissances multi disciplinaires. Les différentes implémentations sont souvent complexes et doivent être ciblées pour un contexte d'application particulier (contexte temps réel ou générique, *Voir* section I).

Dans cet article nous avons couvert les différentes techniques d'acquisition de zones d'intérêts, nous avons exploré les différentes approches pour l'encodage spatial variable et nous avons montré l'efficacité d'une méthode multi résolution dans le contexte de traitement d'image ce qui a permis la génération d'images ayant une compression de l'ordre de 22%.

## VII. Références


[1] L. Itti. "Automatic Attention-Based Prioritization of unconstrained Video for Compression". University of Southern California. 2004.

[2] F. Mohsen, F. Kurugollu, F. Murtagh "Adaptive Wavelet Eye-Gaze Based Video Compression". School of Computer Science, Queen's University, Belfast, UK. 2002.

[3] O. Komogortsev, J. Khan. "Predictive Perceptual Compression for Real Time Video Communication". Kent State University. 2004.

[4] W. S. Geisler and J. S. Perry, "A real-time foveated multi-resolution system for low-bandwidth video communication", in Proc. SPIE, pp. 294-305, 1998.

[5] P. J. Burt, "Smart sensing within a pyramid vision machine," Proc. IEEE, vol. 76, pp. 1006-1015, Aug. 1988.

[6] A. T. Duchowski and B. H. McCormick, "Pre-attentive considerations for gaze-contingent image processing," in Proc. SPIE, vol. 2411, pp. 128-139, 1995.

[7] A. T. Duchowski and B. H. McCormick, "Modeling visual attention for gaze-contingent video processing," in Ninth Image and Multidim.l Signal Proc. (IMDSP) Workshop, pp. 130-131, 1996.

[8] W. Osberger and A. J. Maeder, "Automatic identication of perceptually important regions in an image using a model of the human visual system" in Int. Conf. Patt. Recogn., pp. 701-704, Aug 1998.

[9] Wang Zhou, Bovik Alan, "Digital Video Image Quality and Perceptual Coding", Marcel Decker Series, Signal Processing and Communications, 2005.

[10] Wallace Richardm Ong Ping-Wen, Bederson Benjamin, Schwartz Eric, "Space Variant Image Processing", International Journal of Computer Vision, 13:1, 71-90, 1994.

[11] T. Kuyel, W. Geisler and J. Ghosh, "Retinally reconstructed images: digital images having a resolution match with the human eyes," IEEE Trans. System, Man and Cybernetics, Part A: Systems and Humans, vol. 29, no. 2, pp. 235-243, Mar. 1999.

[12] P. T. Kortum and W. S. Geisler, "Implementation of a foveated image-coding system for bandwidth reduction of video images" in Proc. SPIE, vol. 2657, pp. 350-360, 1996.

[13] S. Liu, "DCT domain video foveation and transcoding for heterogeneous video communication", Ph.D. dissertation, Dept. of ECE, University of Texas at Austin, May 2002.

[14] H. R. Sheikh, S. Liu, Z. Wang and A. C. Bovik, "Foveated multipoint videoconferencing at low bit rates," IEEE Inter. Conf. Acoust., Speech, and Signal Processing,

[15] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Communications, vol. 31, pp. 532-540, Apr. 1983.vol. 2, pp. 2069-2072, May 2002.

[16] Z. Wang, L. Lu and A. C. Bovik, "Foveation scalable video coding with automatic fixation selection," IEEE Trans. Image Processing, vol. 11, no. 2, pp. 243-254, Feb. 2003.

[17] H. Sun, W. Kwok and J. W. Zdepski, "Architectures for MPEG compressed bitstream scaling," IEEE Trans. Circuits and Systems for Video Technology, vol. 6, no. 2, pp. 191-199, Apr. 1996.

[18] Li Zuojin, Shi Weiren, Zhong Zhi, "Simulated Distribution of the Retinal Photoreceptors for Space Variant Resolution Imaging", Information Technology Journal 8, 717-725, 2009.