

Étude et implémentation d'algorithmes de quantification basés sur des modèles psycho-visuels en vue de valider la qualité en temps réel de codecs vidéo.

Ing. T. JOYE
HELMo-Gramme – Liège

1. Contexte et objectifs du travail

Dans le domaine du streaming vidéo, l'objectif primordial est de fournir aux utilisateurs la meilleure expérience visuelle possible.

Cependant, atteindre cet équilibre est un véritable défi pour EVS Broadcast Equipment. D'un côté, EVS veut que chaque image soit parfaite pour le spectateur, mais d'un autre côté, l'entreprise doit minimiser l'utilisation de la bande passante et des ressources. C'est là que mon travail de fin d'études entre en jeu. Ce travail est centré sur le développement d'algorithmes de quantification basés sur des modèles psycho-visuels pour résoudre cette problématique.

Fondamentalement, le but de ce travail est de répondre à la question suivante : comment détecter les défauts d'images en temps réel ? Le terme "temps réel" signifie que cette analyse doit être effectuée pendant que la vidéo est en cours de visionnage, exactement comme lorsqu'une personne regarde un flux vidéo à la télévision. Pour rappel, une vidéo est composée d'une succession d'images. Il fallait donc inspecter chaque image pour détecter les éventuelles erreurs et le défi était de réaliser cette analyse sans ajouter de latence ou de saccade à la vidéo.

Il s'agit d'une tâche complexe qui nécessite de comprendre non seulement comment une vidéo est encodée et transmise, mais aussi comment notre cerveau et nos yeux perçoivent les images.

Les modèles psycho-visuels reproduisent la manière dont le système visuel humain traite les informations visuelles. Ils sont déployés dans les algorithmes pour détecter toute distorsion ou anomalie qui pourrait affecter la perception visuelle de l'utilisateur.

L'application de ces algorithmes sur des flux vidéo en temps réel permet à EVS d'optimiser la qualité visuelle et l'efficacité du streaming en quantifiant en temps réel et en minimisant les erreurs potentielles.

2. Défauts classiques d'une image

Commençons par comprendre ce que sont les défauts dans une image[1]. Voici deux exemples simples de défauts d'images. La version longue de cet article contiendra une liste plus exhaustive des défauts à analyser et permettra d'expliquer leur origine.

- **Le blocage** : L'effet de blocage est une distorsion visible dans les images ou vidéos compressées. Il apparaît comme des blocs carrés uniformes sur l'image (voir la *Figure 1 : Image montrant un défaut de blocage*), perturbant sa fluidité naturelle et impactant la qualité de l'expérience visuelle.



Figure 1 : Image montrant un défaut de blocage[2]

- **Le flou** : Le flou dans les images se manifeste comme une diminution de la netteté ou de la clarté, donnant à l'image un aspect brouillé.



Figure 2: Image montrant le flou d'une image[3]

3. Composition d'une image

Une image est un ensemble de pixels que mes algorithmes comparent, comme nous le montre la *Figure 3* [4]. Le mode de comparaison dépend de l'algorithme choisi, non détaillé ici.

Les pixels sont stockés sur un support physique de manière spécifique, nécessitant plusieurs informations comme la couleur du pixel ou sa luminosité. Chaque information doit être traitée différemment par le programme afin de donner un résultat fiable.

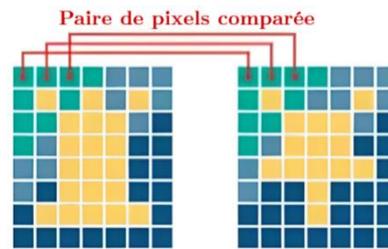


Figure 3 : Comparaison de pixels

Par exemple, une image en Full HD a une taille d'environ 6 millions de bytes, ce qui équivaut à autant de données à examiner pour chaque image. Étant donné qu'une vidéo peut inclure jusqu'à 60 images par seconde, cela représente une quantité considérable de données à traiter instantanément pour prévenir tout délai dans l'affichage de la vidéo.

4. Différents algorithmes retenus

Dans ce programme, le choix a été fait d'implémenter deux algorithmes distincts pour leurs spécificités individuelles. De plus, dans la section 4.3, j'aborde un algorithme supplémentaire non implémenté, mais qui présente un intérêt notable en termes de fonctionnalités.

4.1. Peak Signal-to-Noise Ratio (PSNR)

Le PSNR[5], une mesure de qualité populaire, se base sur le rapport signal/bruit après compression ou autre traitement. Bien qu'il offre l'avantage d'une évaluation rapide due à sa facilité de calcul, il ne reflète pas toujours fidèlement la perception visuelle humaine, conduisant parfois à des évaluations qui manquent de cohérence.

4.2. Structural Similarity Index Measure (SSIM)

Le SSIM[6] évalue la qualité d'une image en fonction des perturbations structurelles, de luminance et de contraste entre deux images. Le SSIM reflète bien la perception humaine et est assez rapide pour un usage en temps réel. Cependant, sa sensibilité aux paramètres utilisateur peut affecter les résultats.

4.3. Video Multimethod Assessment Fusion (VMAF)

Le VMAF[3], développé par Netflix, combine plusieurs algorithmes, dont le PSNR et le SSIM, pour évaluer la qualité vidéo. Il imite étroitement la perception humaine grâce au Deep Learning, mais nécessite énormément de temps pour l'entraînement des modèles et le calcul de la métrique. De plus, il inclut l'analyse d'images pré- et post-traitement, complexifiant l'analyse globale.

5. Explication des résultats

Voici un exemple illustrant les résultats obtenus. Une image originale a été utilisée comme référence et juxtaposée à ses variantes contenant des défauts, comme illustré dans la *Figure 4* : Comparaison d'une image avec deux types de défauts. Ces deux images ont été soumises à une analyse via le programme développé qui utilise divers algorithmes, et voici en résumé les conclusions tirées de cette analyse.

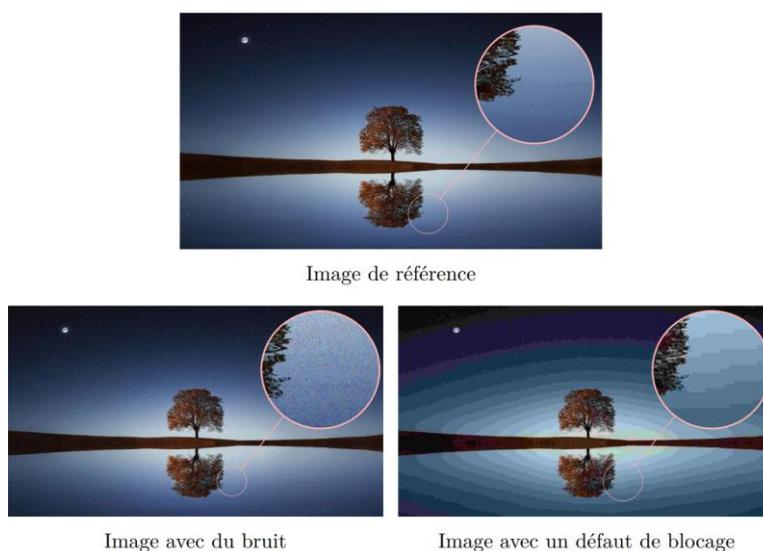


Figure 4 : Comparaison d'une image avec deux types de défaut

Le *Tableau 1 : Résultats de l'analyse des images*, nous montre les résultats. Plus le PSNR est élevé, plus l'image comparée est fidèle à l'image de référence. Pour le SSIM, la valeur doit être le plus proche de 1.

Algorithme	Valeur (Image bruit)	Valeur (Image blocage)
PSNR	27.681002 dB	30.470774 dB
SSIM	0,719698	0,372772
Evaluation subjective	1	2

Tableau 1 : Résultats de l'analyse des images

Le SSIM semble mieux refléter la perception humaine des images. En effet, nous préférons généralement celle avec du bruit (à gauche). Pourtant, le PSNR suggère que l'image la plus fidèle à l'originale est celle de droite, contrairement à nos impressions.

Dans le calcul du SSIM, une analyse des défauts globaux de l'image a été privilégiée (ce qui est paramétrable dans ce programme), comme le ferait l'œil humain, plutôt que la recherche de petites erreurs. Le PSNR ne permet pas ce choix. Ainsi, le PSNR interprète les pixels de couleurs¹ très différentes dans l'image bruyante comme de grandes imperfections, alors que l'œil humain est moins affecté par ces variations.

Voici un autre exemple de résultats provenant d'une étude[7] effectué par Mr Delleji et Mr Bouhleh.

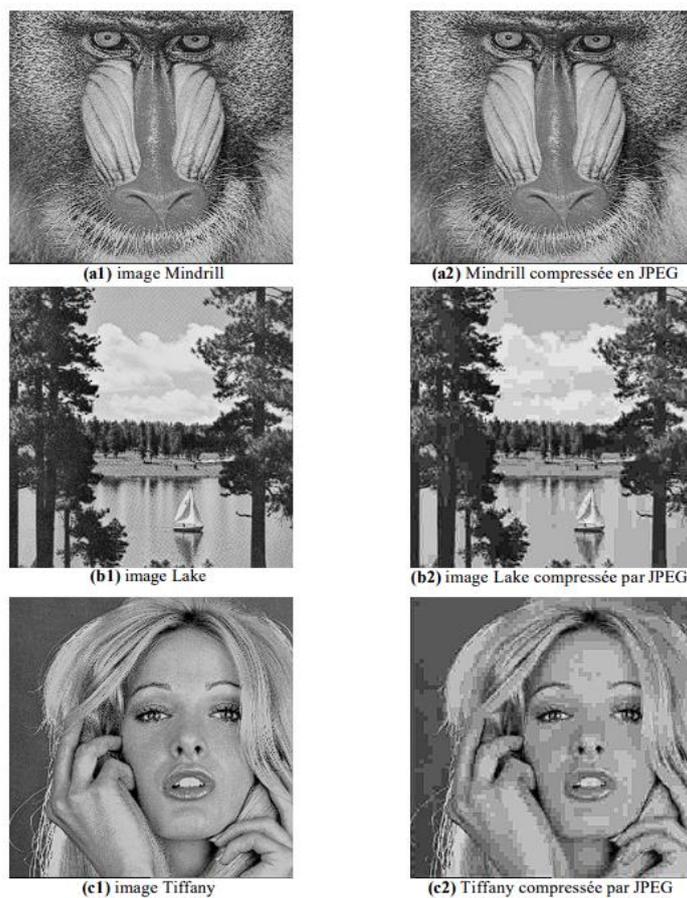


Figure 5 : Couples d'images compressées par JPEG[7]

¹ La luminance impacte davantage l'œil humain, mais la logique est identique.

Couple d'image	Type de distorsion	PSNR	ZQ ²	SSIM	Evaluation subjective
(a1)-(a2)	Compression	25.9906	2.0770	0.8646	1
(b1)-(b2)	Compression	25.8677	26.225	0.7773	2
(c1)-(c2)	Compression	25.8854	40.114	0.7112	3

Tableau 2 : Mesure de qualité d'images compressées par la norme JPEG

Le SSIM correspond bien à la perception visuelle et le ZQ a évalué efficacement les images compressées en JPEG, tandis que le PSNR demeure inchangé, ne reflétant pas correctement les différences de qualité visuelle entre les paires d'images.

6. Conclusion

Malgré la complexité d'analyser un flux vidéo en temps réel, il est possible d'obtenir des résultats exploitables et en adéquation avec les attentes des utilisateurs pour des cas d'analyses spécifiques.

Bien que le PSNR soit meilleur que le SSIM dans certains cas exceptionnels, le SSIM permet une analyse plus approfondie[8],[9]. Il peut s'adapter à diverses situations, que ce soit pour analyser l'image dans sa totalité, des parties spécifiques ou même de très petits défauts. Le SSIM peut être paramétré en fonction des besoins.

En outre, ce programme offre la possibilité d'exporter diverses données calculées ainsi que de sauver les images problématiques, pour une analyse plus poussée avec d'autres logiciels de traitement.

7. Remerciements

Je remercie EVS pour l'opportunité de stage qu'ils m'ont offerte, en particulier monsieur Frédéric Lefranc qui m'a proposé le sujet et m'a accueilli dans son équipe. Un grand merci à ma tutrice Patricia Bourmanne pour son aide, et à toute l'équipe EVS composée de Florence B., Michel C., Frédéric G., Taha K., Thibault L., Frédéric L., Serge M., Samuel M., Fabrice T., et Bernard T. pour leur assistance.

² ZQ signifie « Zipf Quality » qui est une métrique qui mesure la qualité des images JPEG.

8. Références

- [1] A. Robinson, « The difference between JPEG artifacts and image noise », Welcome to the Topaz Labs Learning Center. Consulté le: 26 mai 2023. [En ligne]. Disponible sur: <https://www.topazlabs.com/learn/the-difference-between-jpeg-artifacts-and-image-noise>
- [2] « Comment faire une belle photo noir et blanc ? », Quora. Consulté le: 8 février 2024. [En ligne]. Disponible sur: <https://fr.quora.com/Comment-faire-une-belle-photo-noir-et-blanc>
- [3] « 10 Best Types of Drought-Tolerant Lawn Grass », The Spruce. Consulté le: 8 février 2024. [En ligne]. Disponible sur: <https://www.thespruce.com/drought-tolerant-lawn-grass-2153119>
- [4] S. Viquez, « The Benefits of Visual AI over Pixel-Matching & DOM-Based Visual Testing Solutions », Automated Visual Testing | Applitools. Consulté le: 15 février 2024. [En ligne]. Disponible sur: <https://applitools.com/blog/visual-ai-vs-pixel-matching-dom-based-comparisons/>
- [5] A. Tanchenko, « Visual-PSNR measure of image quality », *J. Vis. Commun. Image Represent.*, vol. 25, n° 5, p. 874-878, juill. 2014, doi: 10.1016/j.jvcir.2014.01.008.
- [6] M.-J. Chen et A. C. Bovik, « FAST STRUCTURAL SIMILARITY INDEX ALGORITHM ».
- [7] T. Delleji et M. S. Bouhleb, « Evaluation de deux métriques pour la mesure de la qualité des images compressées par la norme JPEG ».
- [8] U. Sara, M. Akter, et M. S. Uddin, « Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study », *J. Comput. Commun.*, vol. 7, n° 3, Art. n° 3, mars 2019, doi: 10.4236/jcc.2019.73002.
- [9] A. Horé et D. Ziou, « Image Quality Metrics: PSNR vs. SSIM », in *2010 20th International Conference on Pattern Recognition*, août 2010, p. 2366-2369. doi: 10.1109/ICPR.2010.579.