

Illumination colorée et information commune pour le modèle Horn-Retinex en restauration d’images à faible luminosité

Arthur LECERT¹ Aline ROUMY¹ Renaud FRAISSE² Christine GUILLEMOT¹

¹Inria Rennes – Bretagne-Atlantique 263 Avenue Général Leclerc, 35042 Rennes Cedex, France

²Airbus Defence and Space, Toulouse, France

Résumé – Dans ce travail, nous étudions la restauration d’images à faible luminosité avec des scènes extérieures sans vérité terrain. Jusqu’à présent, les approches de la littérature évitaient d’utiliser le modèle de décomposition Horn-Retinex de manière non supervisée ou ajoutaient des a priori contraignants sur les composantes cherchées afin de faciliter la résolution de ce problème inverse. Nous proposons ici de relaxer la contrainte d’une illumination en niveaux de gris en définissant un modèle basée sur la physique de la lumière. De plus, nous formulons une architecture adaptée à ce nouveau problème et montrons que nous obtenons une décomposition de qualité qui produit de meilleurs résultats visuellement que les méthodes de l’état de l’art c’est-à-dire sans artefact, sans amplification du bruit de l’image et sans hallucination avec une restauration simple pour chacune des composantes.

Abstract – In this work, we study the restoration of low-light images with outdoor scenes without ground truth. Until now, approaches in the literature have avoided using the Horn-Retinex decomposition model in an unsupervised way or have added constraining priors on the searched components to make it easier to find a solution to this inverse problem. We propose here to relax the constraint of a grayscale illumination by defining a model based on the physics of light. Moreover, we formulate an architecture adapted to this new problem and show that we obtain a good quality decomposition that produces visually better results than the state-of-the-art methods, i.e. without artifact, without noise amplification and without hallucination with a simple restoration for each of the components.

1 Introduction

De nombreuses tâches de haut niveau en vision par ordinateur, telles que la classification, bénéficieraient d’algorithmes plus efficaces pour restaurer les images nocturnes. Dans ce papier, nous travaillons plus particulièrement sur le traitement d’images prises en extérieur à l’aide de réseaux de neurones de façon non supervisée du fait de la difficulté d’obtenir des paires images nuit/jour de strictement la même scène. De plus, nous cherchons à restaurer ces images sans halluciner afin de garder l’intégrité du contenu de la scène.

D’après la définition physique de la réflectance [10], cette propriété d’un matériau représente la fraction de la radiance réfléchiée par une surface sur la radiance reçue par cette surface. Cette définition a été utilisée par Horn [9] afin d’élaborer une interprétation de la théorie Retinex suite aux travaux de Land [11] ainsi que ceux de McCann *et al.*[14]. Cette interprétation (Horn-Retinex) affirme qu’une image est un produit de deux composantes : l’illumination et la réflectance. Il s’agit donc d’une interprétation physique de la théorie Retinex et non perceptuelle ce qui permet de l’appliquer de manière plus intuitive dans le contexte de la vision par ordinateur.

La valeur de la réflectance d’un matériau varie en fonction de la longueur d’onde de la lumière incidente de manière non linéaire et est couramment illustrée par un spectre. Pour tenir compte de cette caractéristique, nous proposons que, dans le modèle de décomposition, l’illumination soit colorée. De plus, cette capacité de réflexion est intrinsèque à chaque matériau et est indépendant de son environnement. Seule l’information renvoyée l’est. Ici, nous considérons la couleur physique liée au modèle et non la couleur perceptuelle qui est plus difficile à modéliser. Nous supposons par la suite que l’illumination

dans une scène de jour est de couleur blanche et couvre donc l’ensemble du spectre visible. Dans ce cas, les couleurs observées de jour sont les « vraies » couleurs de la scène alors que les éclairages de nuit (souvent colorés, voire monochromes) ne permettent d’obtenir qu’une information partielle de ces couleurs avec une quantité d’information variant spatialement selon l’intensité de la lumière. Ainsi, nous proposons d’interpréter la réflectance comme l’information commune aux domaines des images de jour et celles de nuit bien que dégradée si extraite de ces dernières.

Les méthodes de l’état de l’art peuvent être catégorisées en deux groupes : les méthodes supervisées pour lesquelles les paires avec vérité terrain sont disponibles et les méthodes non supervisées.

Dans ce premier ensemble de méthodes, il est naturel d’extraire l’information commune aux deux domaines et de nommer cela la réflectance. La décomposition Horn-Retinex étant un problème inverse difficile, les auteurs de KinD++ [8] ajoutent plusieurs a priori pour remédier à cela, comme contraindre l’illumination à être une image en niveaux de gris et une image lisse tout en conservant la structure de la scène.

Lorsque les paires d’images avec vérité terrain ne sont pas disponibles, il est compliqué d’extraire l’information commune. C’est typiquement le cas lors de la restauration d’images en extérieur. Il est en effet très ardu de constituer une base de données de taille conséquente. La majorité des méthodes non supervisées évitent ou s’affranchissent du modèle Horn-Retinex pour pallier ce problème. Jiang *et al.* essaient de restaurer directement les images de nuit à l’aide de réseaux génératifs adversariels dans leur approche EnlightenGAN [4]. Avec la méthode Zero-DCE [5], les auteurs considèrent qu’un mappage non local est suffisant pour restaurer les images et

estiment directement les courbes à appliquer depuis les bases de données d'images. Les méthodes non supervisées qui utilisent la théorie Horn-Retinex nécessitent des a priori pour contraindre l'optimisation ce qui limite l'exploration de nouvelles solutions. Les auteurs de LIME [2] proposent d'explorer les solutions autour d'une initialisation de la composante d'illumination en tant que maximum de l'image sur ces canaux de couleur. Cependant, la méthode ne permet pas de s'éloigner de cette solution et n'explore que son voisinage. En contraignant la recherche de composantes à l'espace de paramètres de réseaux de neurones convolutifs, RetinexDIP [13] permet d'obtenir deux composantes tout en conservant le modèle de décomposition. Cet espace pourrait, cependant, ne pas contenir la solution optimale à notre problème.

Dans l'ensemble des méthodes de la littérature, l'a priori d'une illumination en niveaux de gris est utilisé afin de réduire le nombre de degrés de liberté et faciliter la recherche de solutions de ce problème inverse. Nous proposons dans ce papier de relâcher cette contrainte en définissant un modèle de décomposition basée sur la physique de la lumière qui reprend le modèle Horn-Retinex avec une illumination en couleur. L'extraction des composantes devient alors un problème plus difficile et pour le résoudre, nous proposons d'une part l'idée d'extraire une information commune, et d'autre part une architecture de réseaux de neurones adversariels, permettant, in fine, une restauration des images à faible illumination de manière non supervisée.

2 Modèle Horn-Retinex généralisé et a priori

2.1 Modèle avec illumination colorée

L'étude du système visuel humain qui a conduit à la théorie Retinex remonte aux travaux fondamentaux de Land *et al.* [11], rapidement suivis par ceux de Barrow *et al.* [1] ainsi que les nombreuses contributions de McCann *et al.* [14]. Tout au long de son histoire, cette théorie a connu diverses interprétations basées sur des chemins, des techniques centre/voisinage ou la physique de la lumière. Nous renvoyons le lecteur au papier [17] pour une étude plus approfondie. Cependant, le modèle de décomposition d'image Horn-Retinex que l'on trouve couramment dans la littérature a été défini pour la première fois par Horn dans [9] pour des images en niveaux de gris à un unique canal. Dans le contexte de la restauration d'images à faible luminosité [2, 7, 13], ce modèle est étendu aux images RVB $I \in \mathbb{R}^{3n}$ de la manière suivante :

$$I = L \cdot R + \epsilon \quad (1)$$

où \cdot est le produit élément par élément, $L \in \mathbb{R}^n$ la composante dépendant de la lumière connue sous le nom de carte d'illumination, $R \in \mathbb{R}^{3n}$ la composante complémentaire appelée réflectance, et ϵ un bruit additif gaussien. Ainsi, l'illumination est considérée comme une image en niveaux de gris qui pondère la réflectance avec un facteur commun pour les différents canaux de couleur. La réflectance est supposée lambertienne (c'est-à-dire que la surface en tout point de la scène réfléchit de manière diffuse les rayons lumineux). L'angle d'incidence du rayon lumineux peut être ignoré et la fonction de réflectivité bidirectionnelle (BRDF dans la littérature) [15] n'est pas utilisée. En outre, toute composante spéculaire ou

ambiante est également négligée, ce qui invalide l'utilisation de modèles tels que le modèle de Phong de la réflexion de l'illumination sur une surface [16].

Au lieu de cela, nous définissons la réflectance comme l'information commune entre deux distributions d'images (qui sont supposées avoir des scènes et des objets similaires, mais pas de paires d'images). En ce sens, la réflectance est réellement indépendante du domaine, tandis que l'illumination est la composante dépendante de la lumière et contient, par exemple, le bruit spécifique des images à faible luminosité.

Avec ces définitions, nous pouvons formuler le problème de décomposition de Horn-Retinex comme suit :

$$I = \frac{L}{R} \quad R > 0 \quad (2)$$

où $I \in [0; 1]^{3n}$ l'image RVB, $L \in [0; 1]^{3n}$ l'illumination, $R \in \mathbb{R}^+$ un facteur d'échelle car la décomposition conduit à une infinité de solutions. $R \in [0; 1]^{3n}$ puisque le seul objet physique qui absorbe l'intégralité de la lumière est un trou noir. En revanche, les miroirs parfaits ne sont pas encore commercialisés à grande échelle et risquent de ne pas apparaître fréquemment dans notre vie quotidienne. $\epsilon = 1e^{-8}$ dans nos expériences. Ainsi, nous assouplissons le modèle original précédent de la contrainte des niveaux de gris de l'illumination. La composante a maintenant une chrominance locale en plus d'une valeur de luminance locale. Étant donné que l'illumination est censée contenir le bruit et les dégradations en basse lumière, il serait inefficace d'ajouter un a priori de lissage de cette composante (par exemple, celui de [7]).

2.2 A priori Horn-Retinex

2.2.1 Ambiguïté d'échelle (A priori de réflectance élevée)

Le facteur R introduit dans le modèle met en évidence le problème d'ambiguïté d'échelle dans la décomposition. Toute valeur réelle non nulle peut conduire à une solution plausible. Pour réduire encore davantage l'ensemble des solutions, nous proposons un nouvel a priori défini comme suit :

$$L_{HR} = \frac{1}{R} \quad (3)$$

$R = 1$ comme valeur initiale avant le processus d'optimisation. Comme nous travaillons avec des images à faible luminosité, nous supposons que l'illumination doit avoir la valeur la plus faible possible. D'autre part, minimiser cet a priori équivaut à rechercher la réflectance qui a la valeur la plus élevée sous contrainte du modèle. Intuitivement, cela revient à considérer un canal V élevé dans l'espace de couleurs HSV pour la réflectance, avec une valeur faible pour l'illumination. Cela signifie également que le processus d'optimisation est biaisé en défaveur des corps noirs.

2.2.2 A priori d'exposition pour images RVB

Dans cette section, nous étendons l'a priori d'exposition [13] aux images RVB. Tant que le capteur de la caméra n'est pas saturé par le rayon lumineux ($I_{c2FR;G;Bg} \notin 1$), l'illumination ne peut pas être saturée non plus ($L_{c2FR;G;Bg} \notin 1$). Cet a priori est défini pour éviter la solution triviale où $L = 1$ et ainsi $I = R$. Seules les sources lumineuses ou les régions surexposées de l'image d'entrée doivent conduire à de telles valeurs

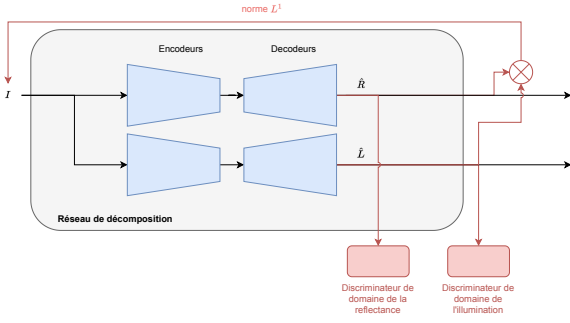


FIGURE 1 : Une illustration de notre architecture composée de deux branches : la branche supérieure extrait la réflectance de l'image, la branche inférieure permet d'obtenir l'illumination. Chaque composante Horn-Retinex a son discriminateur respectif. Celui de la réflectance guide le générateur pour dégager l'information commune de l'image tandis que celui de l'illumination fait ressortir l'information complémentaire qui permet de différencier les domaines à luminosité faible/normale.

dans les composantes. Nous définissons l'a priori comme suit

$$L_E = \prod_{c \in \{R, G, B\}} g(L_c) \quad (4)$$

où g est une fonction de seuil $g(x) = \begin{cases} x; & x > 1 \\ 0; & \text{sinon} \end{cases}$ avec

$L_c = \max_{c \in \{R, G, B\}} I_c$, une approximation de l'illumination suite au travail LIME de Guo *et al.*[2] et $\lambda = 1e^{-2}$.

2.3 La fonction de coût

Dans cette section, nous décrivons l'architecture présentée dans la figure 1 que nous avons construite suite au modèle (2). Comme nous n'avons pas accès aux images de vérités terrain et que nous voulons un temps d'exécution faible, nous utilisons une architecture basée sur les réseaux génératifs adversariaux. Pour guider l'extraction des deux composantes, nous utilisons deux discriminateurs, un pour chaque composante respective.

Afin de simplifier la notation dans les équations suivantes, le facteur d'échelle est inclus implicitement dans les deux composantes estimées ($\hat{L}; \hat{R}$). Pour s'assurer que les sorties générées ($\hat{L}; \hat{R}$) permettent de reconstruire l'image d'entrée suivant le modèle Horn-Retinex, nous utilisons une norme L^1 selon le terme suivant,

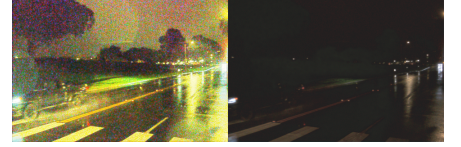
$$L_{recon} = \| I - \hat{L} \cdot \hat{R} \|_1 \quad (5)$$

Notre architecture s'appuyant sur des termes de coût adversariaux pour trouver les composants de manière non supervisée, nous définissons les deux générateurs et discriminateurs comme suit,

$$G_R: I_d \mapsto \hat{R}_d \in \mathbb{R}^{3n}; \quad G_L: I_d \mapsto \hat{L}_d \in \mathbb{R}^{3n} \quad (6)$$

$$D_R: \hat{R}_d \mapsto \hat{d}; \quad D_L: \hat{L}_d \mapsto \hat{d}; \quad \hat{d} \in \{0, 1\} \quad (7)$$

où \hat{d} est le label du domaine estimé résultant de chacune des composantes soit 1 pour les images à luminosité normale et 0 pour celles à faible luminosité. Les discriminateurs doivent être capables d'identifier le domaine du composant d'entrée



(a) (b) (c)

FIGURE 2 : De gauche à droite : l'image d'entrée, l'illumination extraite et la réflectance obtenue. L'illumination contient le bruit spécifique à la dégradation de faible luminosité.

qu'ils reçoivent (c'est-à-dire de séparer chaque composant en fonction de son domaine). Nous constatons empiriquement que l'entraînement du réseau adversarial est plus stable avec une fonction de coût des moindres carrés [6]. Les paramètres de $(G_R; G_L)$ sont fixes pour cette première passe.

$$L_{D_R} = \prod_{d \in \{0, 1\}} \mathbb{E}_{I_d}^h D_R(G_R(I_d)) \quad d^{2^i} \quad (8)$$

La fonction du discriminateur D_L est définie de façon similaire. Pour entraîner les générateurs $(G_R; G_L)$, nous fixons les paramètres de $(D_R; D_L)$. Le générateur d'illumination doit extraire la composante de l'image et le domaine doit être identifié avec précision par le discriminateur correspondant. Au contraire, nous cherchons à extraire des informations qui ne peuvent pas être classées par le discriminateur de réflectance entre les domaines de faible luminosité et de luminosité normale. Par conséquent, nous l'optimisons pour aligner la réflectance des zones de faible luminosité sur celle des zones de luminosité normale. Cela conduit aux équations suivantes,

$$L_{G_L} = \prod_{d \in \{0, 1\}} \mathbb{E}_{I_d}^h D_L(G_L(I_d)) \quad d^{2^i} \quad (9)$$

$$L_{G_R} = \prod_{d \in \{0, 1\}} \mathbb{E}_{I_d}^h D_R(G_R(I_d)) \quad 1^{2^i} \quad (10)$$

Par conséquent, nous obtenons les problèmes d'optimisation suivants pour entraîner l'ensemble des réseaux,

$$\hat{G}_R; \hat{G}_L = \underset{G_R; G_L}{\operatorname{argmin}} L_{recon} + HR L_{HR} + E_{L_E} + adv(L_{G_L} + L_{G_R}) \quad (11)$$

$$\hat{D}_R; \hat{D}_L = \underset{D_R; D_L}{\operatorname{argmin}} L_{D_R} + L_{D_L} \quad (12)$$

3 Résultats

3.1 Visualisation et restauration des composantes

Un exemple de composantes que nous obtenons est illustré dans la figure 2. Le bruit spécifique à la dégradation à faible luminosité est bien extrait et se retrouve dans la composante d'illumination. Afin de réduire le bruit dans l'illumination selon l'information que contient la réflectance, nous appliquons le réseau de neurones [12] pour débruiter. En effet, ce réseau autorise une carte de bruit qui varie spatialement. Nous pondérons donc le niveau de bruit par le gradient spatial de la réflectance. Etant donné que la réflectance correspond à l'information commune entre les deux domaines mais dégradée, nous

