

# Imagerie à forte gamme dynamique HDR Imaging

Yann GOUSSEAU

Télécom Paris - IP Paris



MVA  
2022-2023

# Imagerie à grande gamme dynamique (HDR)

Capture d'une scène avec de très fortes variations de luminosité...



Les limitations de l'appareils → perte de détails dans les zones claires et/ou sombres.

# Imagerie à grande gamme dynamique (HDR)

... avec un appareil photo standard.



Les limitations de l'appareils → perte de détails dans les zones claires et/ou sombres.

# Imagerie à grande gamme dynamique (HDR)

... avec un appareil photo standard.



Les limitations de l'appareils → perte de détails dans les zones claires et/ou sombres.

# Imagerie à grande gamme dynamique (HDR)

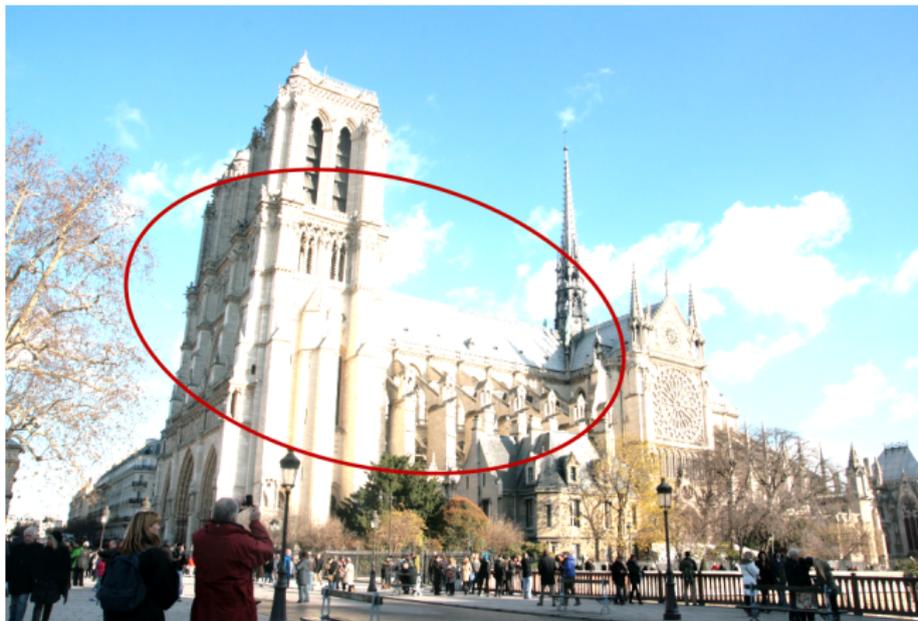
... avec un appareil photo standard.



Les limitations de l'appareils → perte de détails dans les zones claires et/ou sombres.

# Imagerie à grande gamme dynamique (HDR)

... avec un appareil photo standard.



Les limitations de l'appareils → perte de détails dans les zones claires et/ou sombres.

# Exemples d'images HDR



by J.A. Sanjurjo under CC BY-NC-ND 2.0



by Daniel under CC BY 2.0



by Marc under CC BY-NC-SA 2.0



by Daniel under CC BY 2.0



aspect artificiel ...

mais pratique ancienne !



J. Constable, *Salisbury Cathedral from the Bishop's Grounds*, 1823.

Illustration suggested by "The art and science of HDR imaging", McCann - Rizzi, 2012

- Dynamique des scènes naturelles souvent >> capacités des capteurs photographiques standards

⇒

- Zones claires saturées ("brulées")  
et / ou
- Zones sombres saturées ou très bruitées
- Dynamique également supérieure à celles des dispositifs de visualisation (écrans/projecteurs/imprimantes)
- ⇒ Comment capturer toute la plage dynamique d'une scène et la visualiser ?

- Que mesure-t-on ?
- Rappels sur l'acquisition des images
- Création d'images HDR
  - Cas statique : un problème statistique
  - Gestion des mouvements
- Visualisation (tone mapping)
- Fusion directe d'expositions différentes

# Que mesure-t-on ?

**Radiométrie** : mesure d'une intensité de radiation (energie)

**Photométrie** : mesure d'une quantité de lumière captée par notre système visuel

Pondération par des courbes de sensibilité

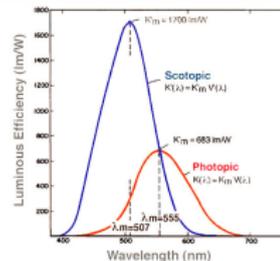
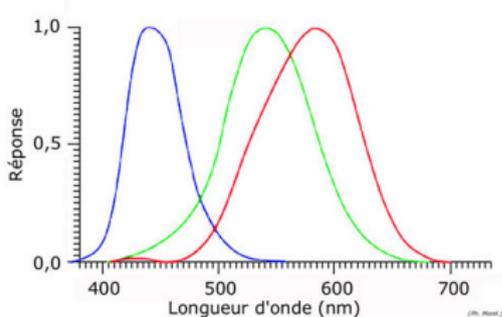


Figure 8. The scotopic and the photopic curves of spectral luminous efficacy (non-normalized values).

curves from [webvision.med.utah.edu](http://webvision.med.utah.edu)

- **Luminance** : quantité de lumière émise par une source lumineuse
  - **Flux lumineux** : émission totale en lumens ( $lm$ )
  - **Intensité lumineuse** : émission directionnelle (/ angle solide) en candela ( $cd$ )

$$cd = lm/sterd$$

- **Luminosité d'une surface** : en  $cd.m^{-2}$
- **Illuminance** : lumière reçue  
flux lumineux / surface en  $lm.m^{-2}$  (lux)



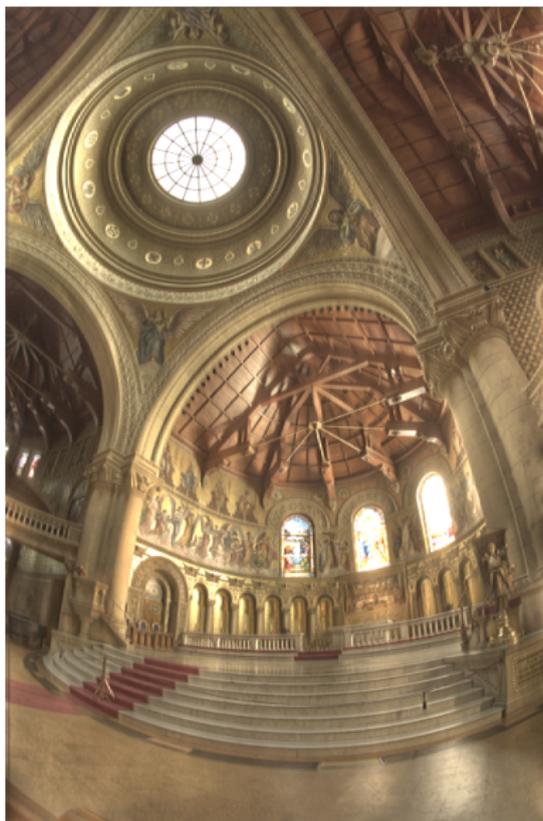
Gauche : lux-mètre ; droite luminance-mètre

- **Radiance/irradiance** : idem en radiométrie

# Quelques ordres de grandeur

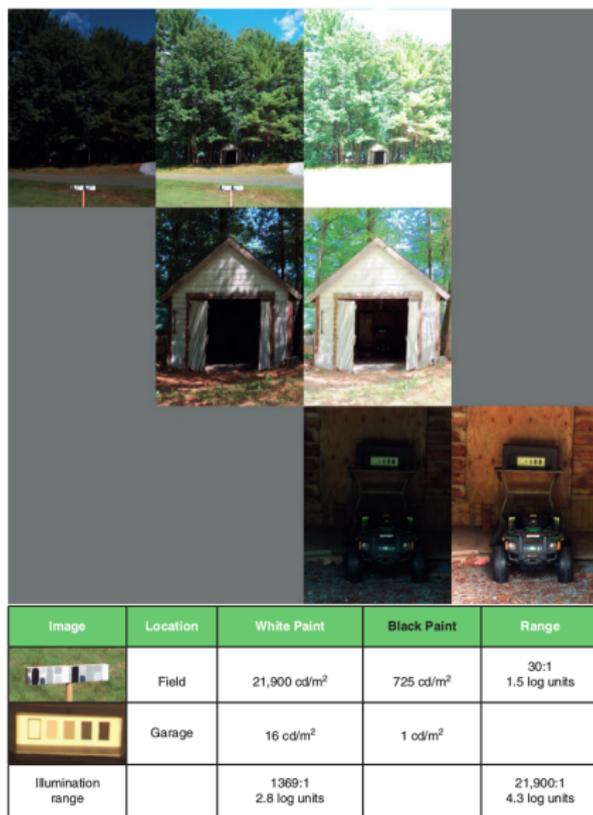
- Luminance :
  - Etoile  $10^{-3}cd.m^{-2}$
  - Neige au soleil  $10^4cd.m^{-2}$
  - Soleil  $10^9cd.m^{-2}$
- Dynamique généralement comptée en "stops" (dynamique =  $2^{stops}$ )
- Oeil humain (avec adaptation) :  $\sim 25$  stops
- Oeil humain (sans adaptation) :  $\sim 7-15$  stops
- Illuminance d'une scène à grande dynamique typiquement entre 15 et 20 stops ( $10^6$ )
- Appareil photo :
  - smartphone  $\rightarrow 10$  stops, souvent moins
  - plein ou moyen formats numériques  $\rightarrow 15$  stops , souvent 12 stops
  - Argentique noir et blanc 13-14 stops
- Ecrans
  - LCD standard .002  $\rightarrow 100cd.m^{-2}$   
 $\sim 12$  stops (5000 :1)  
Nettement plus faible si on n'est pas dans le noir
  - Ecran HDR .002  $\rightarrow 5000cd.m^{-2}$
  - Ecrans OLED : noirs plus faibles / 20 stops (1000000 :1)  
à regarder dans le noir !!

## Scène HDR typique



Dynamique 17 stops

# Scène HDR typique



From "The art and science of HDR imaging", McCann - Rizzi, 2012

Dynamique 14,5 stops (21900 :1)

# Comment faire ?

Stratégie classique : le multi-images

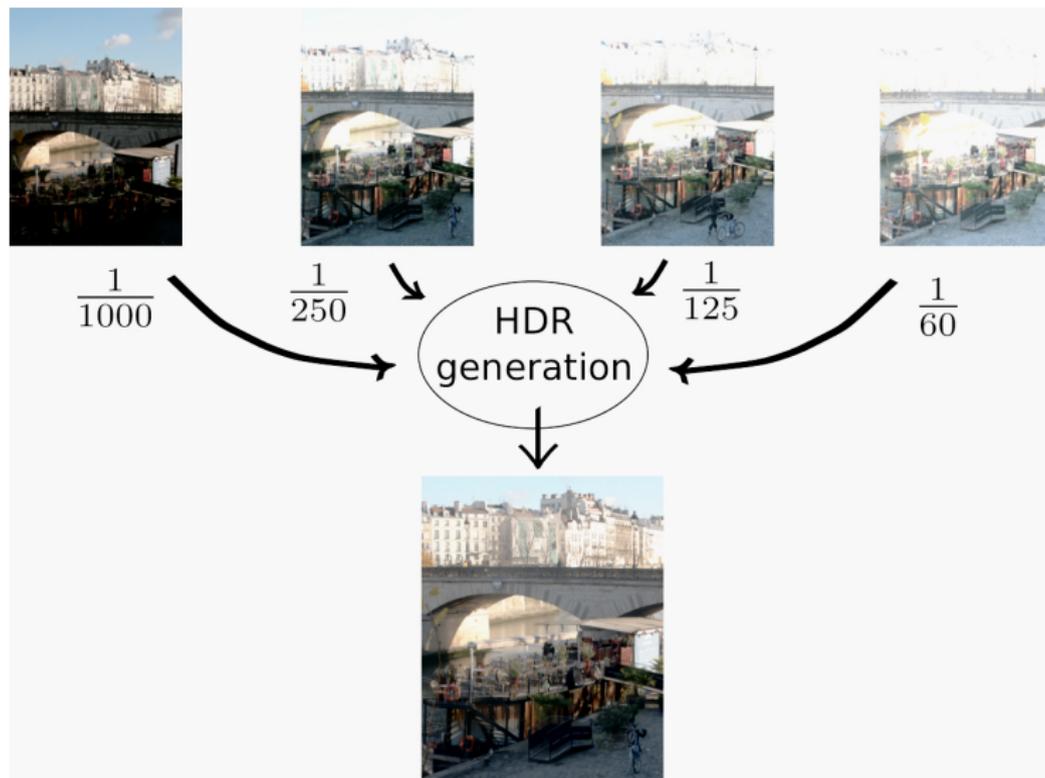


Fig. C. Aguerrebere



Gustave LeGray, 1850. Technique des *ciels rapportés*; le ciel provient d'une autre scène

Autre stratégie en photo argentique : "dodging and burning"



(application "Vivid")

Aujourd'hui pratique de poche  
Un problème réglé ?



Samsung S9



Apple Iphone 12

## Présence de fantômes (ghosting)

Illustration DxO Mark

A partir de  $I_1, \dots, I_n$  acquises avec des temps différents, deux approches :

- 1
  - **Création d'image HDR** (généralement sur 32 bits)  
Compte rendu fidèle de l'irradiance reçue.
  - Compression de la dynamique : *tone mapping* (typiquement sur 8 bits)  
Pour visualisation (écran, imprimante)
- 2 **Fusion d'exposition**  
Génère directement une image 8 bits (par canal) sans chercher à créer une image HDR.

Première étape :

$$u = Q[(h * s). \Pi_{\Gamma} + n]$$

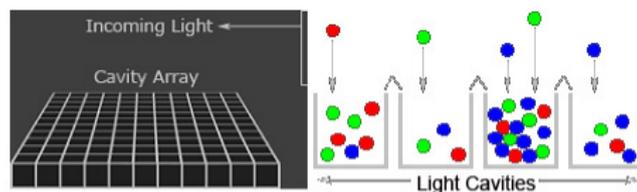
où

- $s$  scène (émission de photons),
- $h$  : réponse impulsionnelle du système optique et de l'intégration des capteurs ( $g_{ouv} * g_{flou} * g_{fil} * g_{capt}$ ),
- $\Pi_{\Gamma} = \sum_{\gamma \in \Gamma} \delta_{\gamma}$ , échantillonnage
- $Q$  opérateur de quantification
- $n$  bruit, éventuellement dépendant du signal

→ image RAW (tramée)

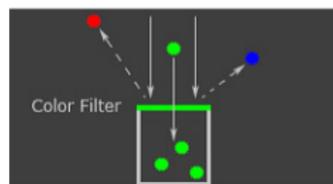
Réponse linéaire à l'illuminance

Solution classique pour l'acquisition de la couleur : trame de Bayer



*courtesy of cambridgeincolour.com*

Color Filter Array



Nécessité d'un démosaïcage (détramage) pour générer  $I \in \mathbb{R}^{3.N.N}$ .

Deuxième étape :

- Balance des blancs et changement d'espace RGB

$$I \rightarrow A.I$$

avec  $A$  matrice 3x3 (généralement diagonale pour la balance des blancs)

- Courbe de réponse  $A.I \rightarrow g(A.I)$  (non-linéaire, tone mapping)
- Compression JPEG

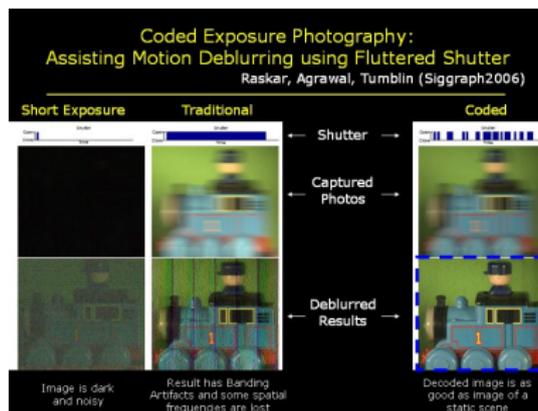
- Temps d'exposition
- Ouverture
- Sensibilité (nombre ISO)

- Peut compenser un manque de lumière ou une ouverture faible
- Doit être court pour les objets en mouvement (sinon, flou de mouvement)
- Pour les images RAW, réponse proportionnelle à  $\tau$

Exemple de flou de mouvement :



Alternative à l'obturation unique : l'ouverture papillonante (flutter shutter)



ouverture  $D$  diamètre du diaphragme ;  $f$  distance focale ; nombre d'ouverture  $N = f/D$  (généralement progression géométrique)



wikipedia.org

Quand  $N$  augmente :

- La profondeur de champs augmente (cf premier cours)
- Les défauts de diffraction augmentent
- Les défauts de vignettage diminuent

changement de profondeur de champs difficile à prendre en compte : on privilégie le multi-images à  $\tau$  variable  
(autre possibilité : changement de sensibilité, pas traitée dans ce cours)

En négligeant la quantification, on observe

$$z = C.\tau.R$$

(images RAW)

ou (en négligeant le matricage couleur et la compression),

$$z = g(C.\tau.R),$$

(images finales)

où

- $g$  fonction de réponse
- $\tau$  temps d'exposition
- $R$  irradiance
- $C$  constante (dépend de la sensibilité)

Pour retrouver l'information d'irradiance : nécessité d'estimation de  $g^{-1}$   
(méthodes itératives, moindres carrés, etc.)

... ou de travailler avec des images RAW

Images  $I_1, \dots, I_N$  acquises avec  $\tau_1, \dots, \tau_n$  (connus).

On suppose que les images sont parfaitement recalées ;

- Appareil sur un pied ou mouvement de caméra compensé
- Scène statique
- Pas de changement d'éclairage (e.g. nuages)

En chaque pixel

**Données :**  $z_1, \dots, z_N$  valeurs correspondant à  $\tau_1, \dots, \tau_N$ .

**Résultat :** l'irradiance  $R$  (à une constante près), proportionnelle au nombre moyen de photons par unité de temps atteignant le pixel

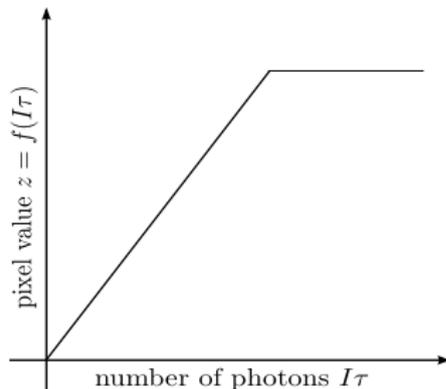
- $z_j$  = version bruitée de  $g(R\tau_j)$  où  $g$  est la fonction de réponse de la caméra ;
- Estimation de l'irradiance par **moyennage** des estimateurs pour différents  $\tau$  :

$$\hat{R} = \sum_j w(z_j) \frac{g^{-1}(z_j)}{\tau_j},$$

avec  $\sum_j w(z_j) = 1$ .

- De très nombreux choix proposés pour les pondérations  $w$  :
  - Selon  $g$  Mann & Picard 1995, Debevec & Malik 1997
  - dépendant du SNR Mitsunaga & Nayar 1999
  - critères de variance Robertson et al. 1999, Kirk & Andersen 2006, Granados et al. 2010, Hasinoff et al. 2010
- Pour couleur : traitement par canal  
Alternative : création dans le domaine luminance-chrominance (Pirinen et al. 2007)

- Utilisation des images RAW (fonction de réponse  $g$  linéaire et saturation)



- Modèle statistique pour les  $z_i$  dépend du paramètre  $R$  et des particularités de l'appareil photographique (bruit)
- Trouver  $R$  revient à un problème d'**estimation** à partir des  $N$  échantillons  $z_1, \dots, z_N$ .

# Retour sur la modélisation du bruit

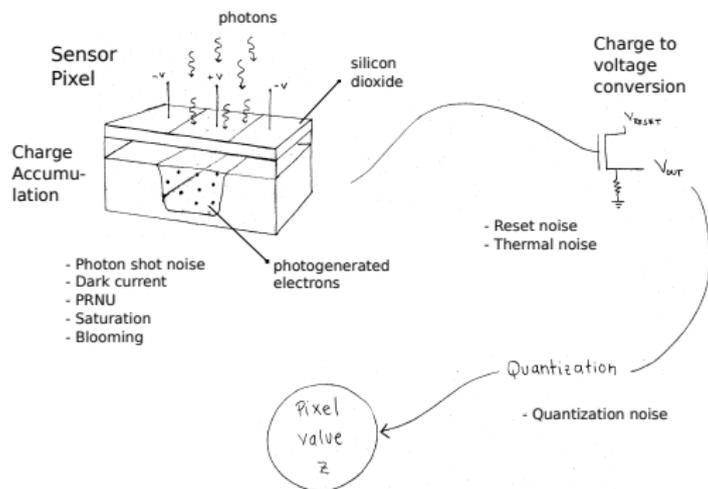
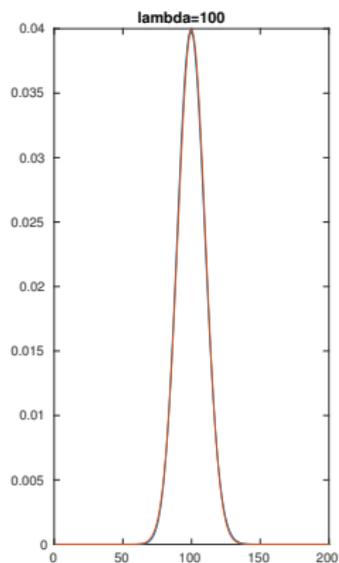
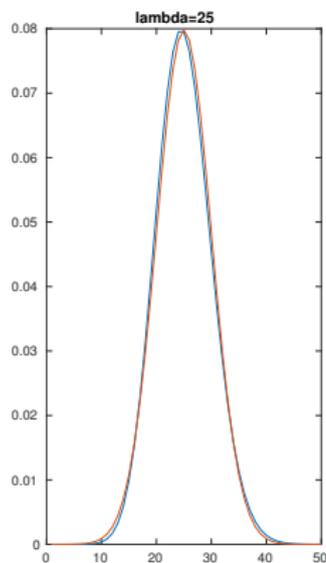
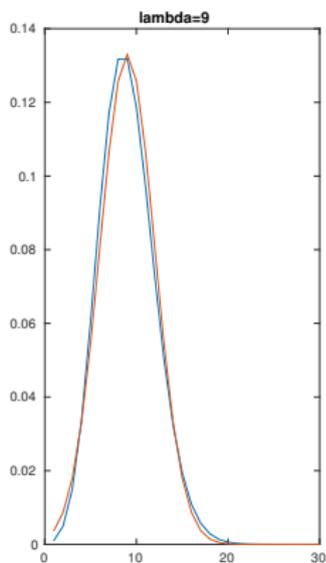


Illustration C. Aguerrebere

- Bruit photonique ("shot noise")  $\rightarrow$  Poisson  
Approchable par une gaussienne  $N(\lambda, \lambda)$  pour  $\lambda$  grand ( $\lambda > 20$ )
- Bruit thermique résiduel (dark current) négligeable (pour temps d'acquisition  $< 1s$ )
- Bruit de lecture Gaussien

# Approximation gaussienne



Bleu : Poisson ; rouge : normale

- $Z_1, \dots, Z_N$  v.a. indépendantes (en un pixel donné), mais pas identiquement distribuées,
- $Z_k \sim g\mathcal{P}(R\tau_k) + \mathcal{N}(\mu_R, \sigma_R^2)$
- Approximation gaussienne :  
 $Z_k \sim \mathcal{N}(gR\tau_k + \mu_R, g^2R\tau_k + \sigma_R^2)$ .

Avec

- $g$  gain de l'appareil
- $\mu_R, \sigma_R$  caractéristiques du bruit de lecture

On s'intéresse à des estimateurs de  $R$  de la forme

$$\hat{R} = \sum_j w_j \frac{z_j - \mu_R}{g\tau_j},$$

- $\rightarrow$  quel est le meilleur estimateur ?
- quelles sont ses performances ?

## Modèle de bruit

- $Z_1, \dots, Z_N$  v.a. indépendantes,
- $Z_k \sim \mathcal{N}(g\mathbf{R}\tau_k + \mu_R, g^2\mathbf{R}\tau_k + \sigma_R^2)$ .

## Borne de Cramer-Rao pour les estimateurs non-biaisés

variance(tout estimateur non-biaisé de  $\mathbf{R}$ )  $\geq$  CRLB,

où

$$\frac{1}{\text{CRLB}} = \frac{1}{2} \sum_{j=1}^N \frac{g^2 \tau_j^2}{(g^2 \mathbf{R} \tau_j + \sigma_R^2)^2} (2(g^2 \mathbf{R} \tau_j + \sigma_R^2) + g^2)$$

On peut montrer que

*Dans le cadre du modèle précédent, il n'existe pas d'estimateur efficace (estiateur non-biaisé atteignant la borne CRLB) pour  $R$ .*

Experimentalement, on observe

- L'estimateur du maximum de vraisemblance (MLE) a de bonnes performances
- la variance du MLE est très proche de CRLB ;
- le biais du MLE est très faible ;

En pratique ,

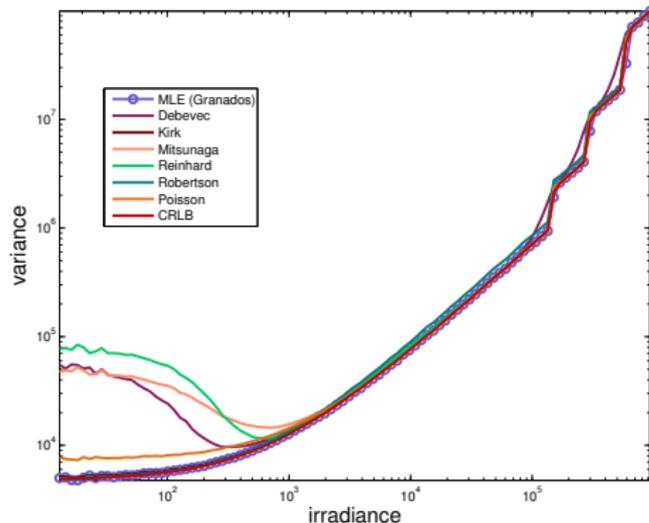
- Pas de formule close pour le MLE
- Bonne approximation avec

$$\hat{C} = \frac{\sum_{j=1}^N \frac{x_j}{\sigma_j^2}}{\sum_{j=1}^N \frac{1}{\sigma_j^2}}, \quad (1)$$

avec

$$x_j = \frac{z_j - \mu_R}{g\tau_j}, \quad \sigma_j^2 = \frac{g(z_j - \mu_R) + \sigma_R^2}{(g\tau_j)^2}. \quad (2)$$

## Borne de Cramér Rao pour l'estimation d'irradiance



III. Aguerrebere et al., 2014  
(Au plus 4 échantillons par pixel)

## Exemple - génération HDR - log tone mapping



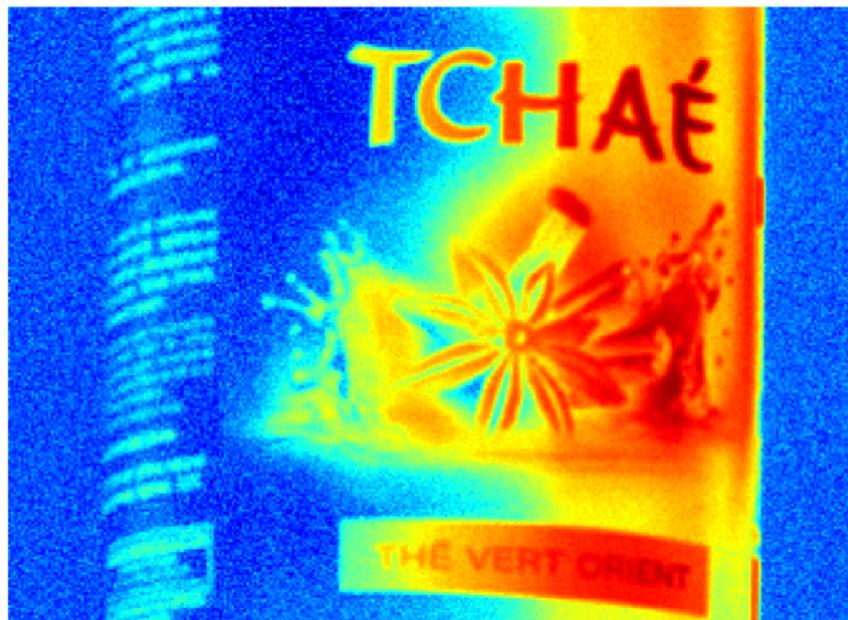
Vérité terrain

## Exemple - génération HDR - log tone mapping



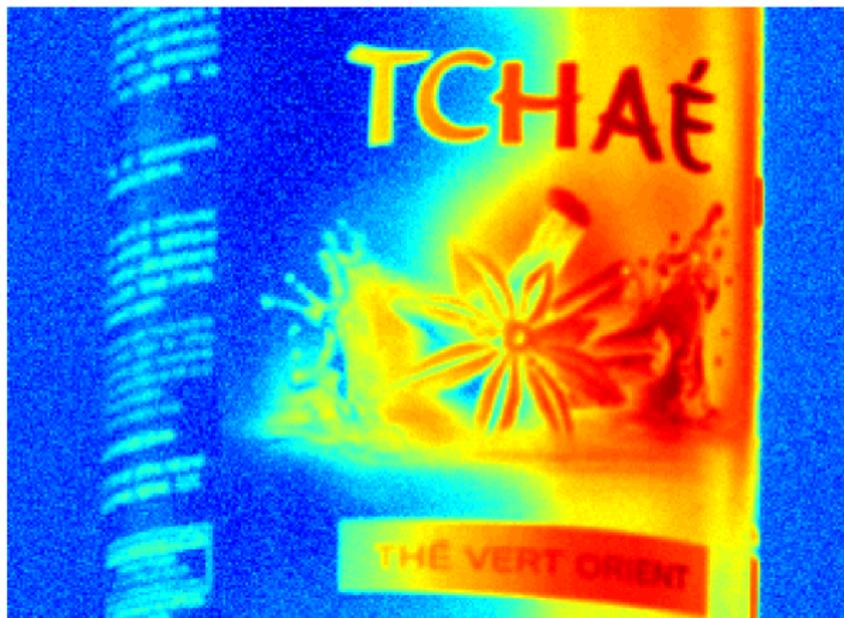
MLE

## Exemple - génération HDR - log tone mapping



Mann - Picard

## Exemple - génération HDR - log tone mapping



Mitsunaga - Nayar



## Difficultés de l'HDR dynamique

noise



# Difficultés de l'HDR dynamique

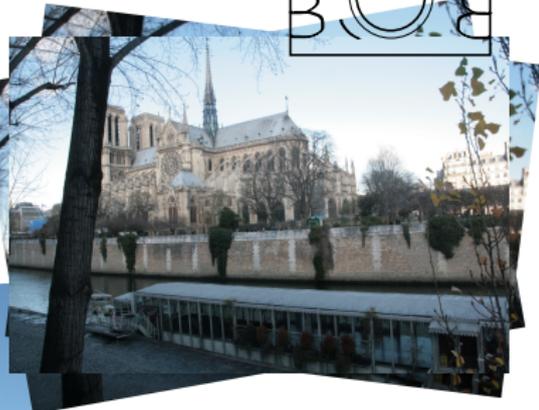
noise



camera  
motion

# Difficultés de l'HDR dynamique

noise



moving  
objects



camera  
motion

## Recalage

- Extraction de points clés (e.g. SIFT)
- Mise en correspondance de points clés
- Algo. estimation robuste (typiquement RANSAC) pour estimer une transformation compatible avec les mises en correspondances.
  - Transformations affines
  - Transformations homographiques

Puis HDR sur les images alignées (e.g. Tomaszewska-Mantiuk 2007)

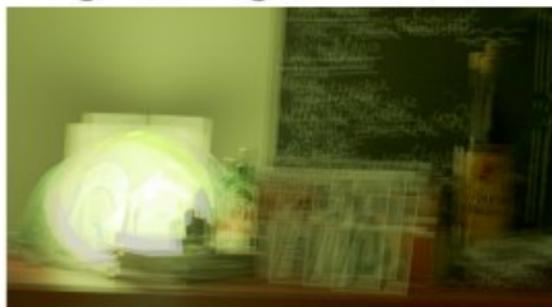
L'approche précédente est insuffisante si

- Le recallage est insuffisant (e.g. scène non plane et mouvement du centre optique)
- Mouvement d'objets
- Changement de luminosité

camera + object motion



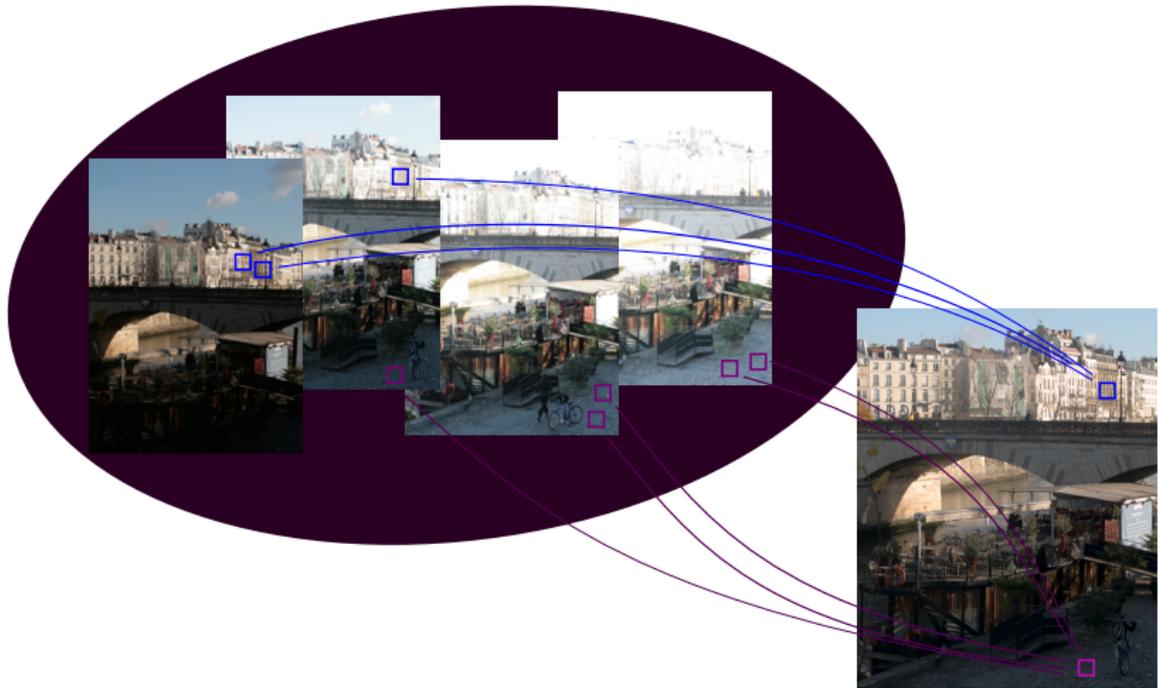
ghosting artifacts



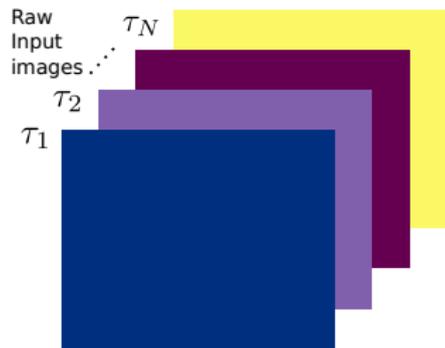
- **Compensation du mouvement** (e.g. Zimmer et al. 2011)  
estimation du flot optique /problème mal posé
- **Détection du mouvement** (e.g. Jacobs et al. 2008)  
recallage, puis détection des pixels en mouvement, ignorés pour la reconstruction finale
- Méthodes par patches (e.g. Aguerrebere et al. 2013)  
On fusionne les pixels ayant un voisinage semblable

# Méthode par patches

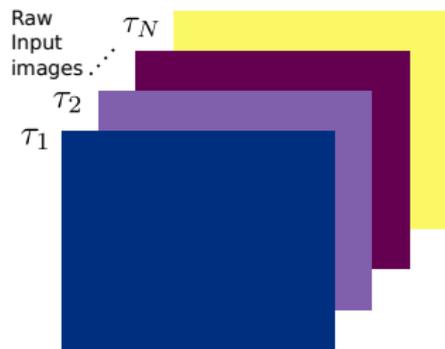
Aguerreberre et al. 2013



# création HDR et débruitage simultanés



# création HDR et débruitage simultanés

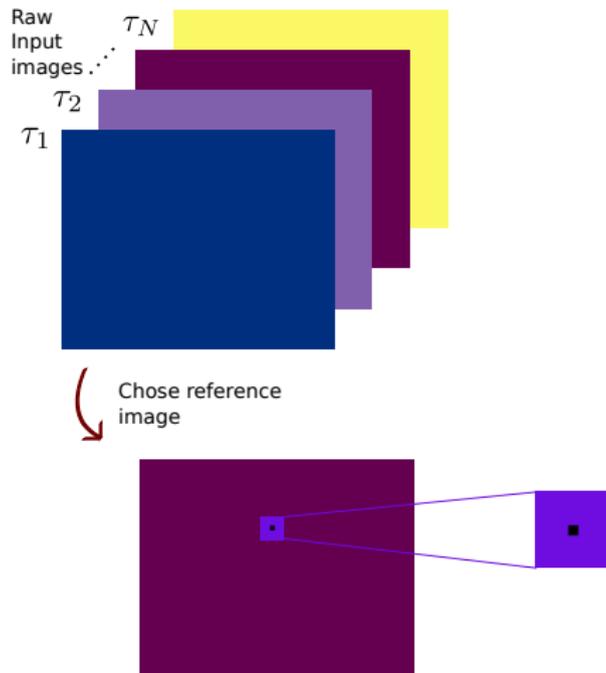


Chose reference image

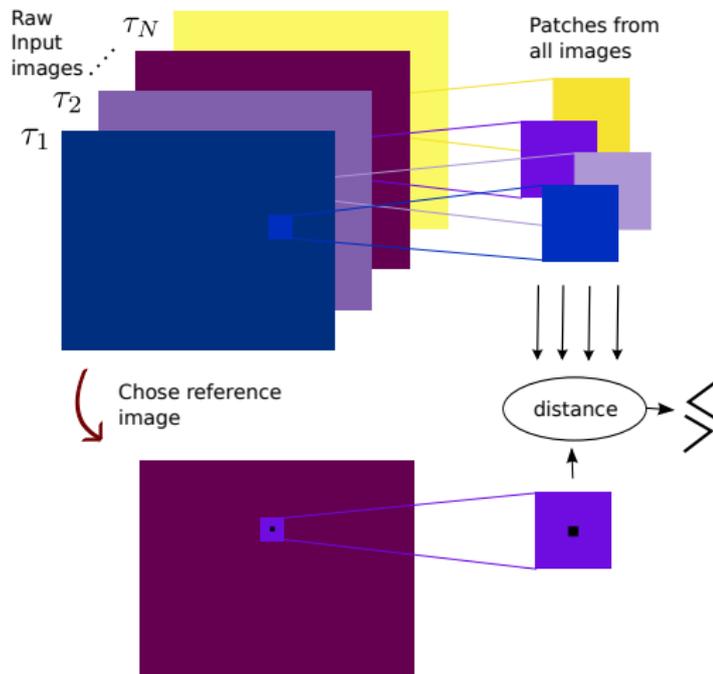
A red arrow points from the text 'Chose reference image' to the reference image below.



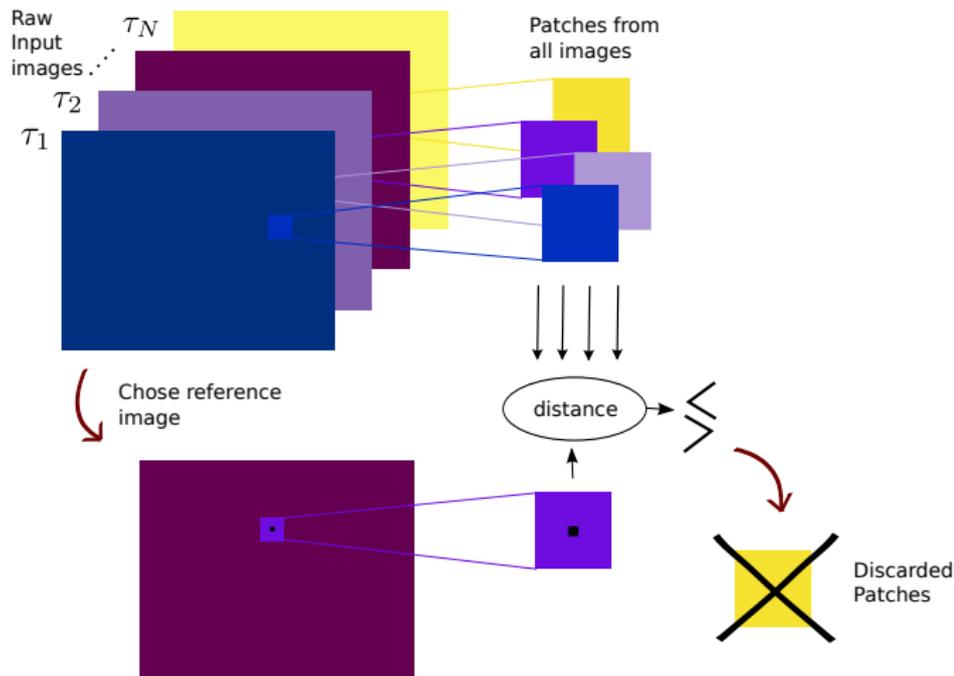
# création HDR et débruitage simultanés



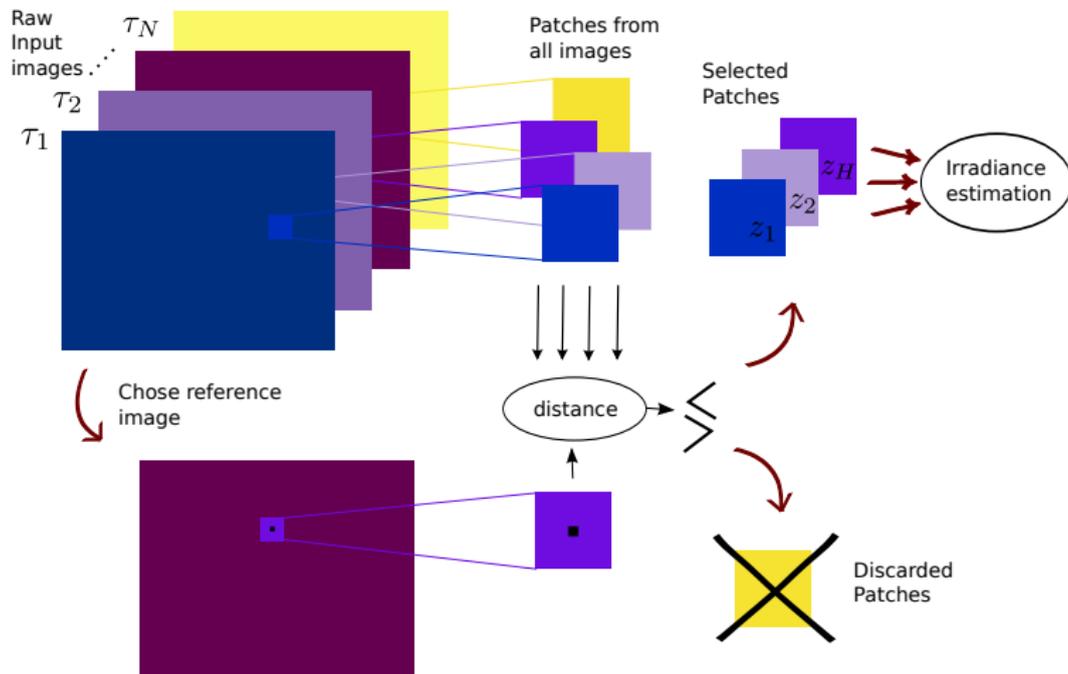
# création HDR et débruitage simultanés



# création HDR et débruitage simultanés



# création HDR et débruitage simultanés

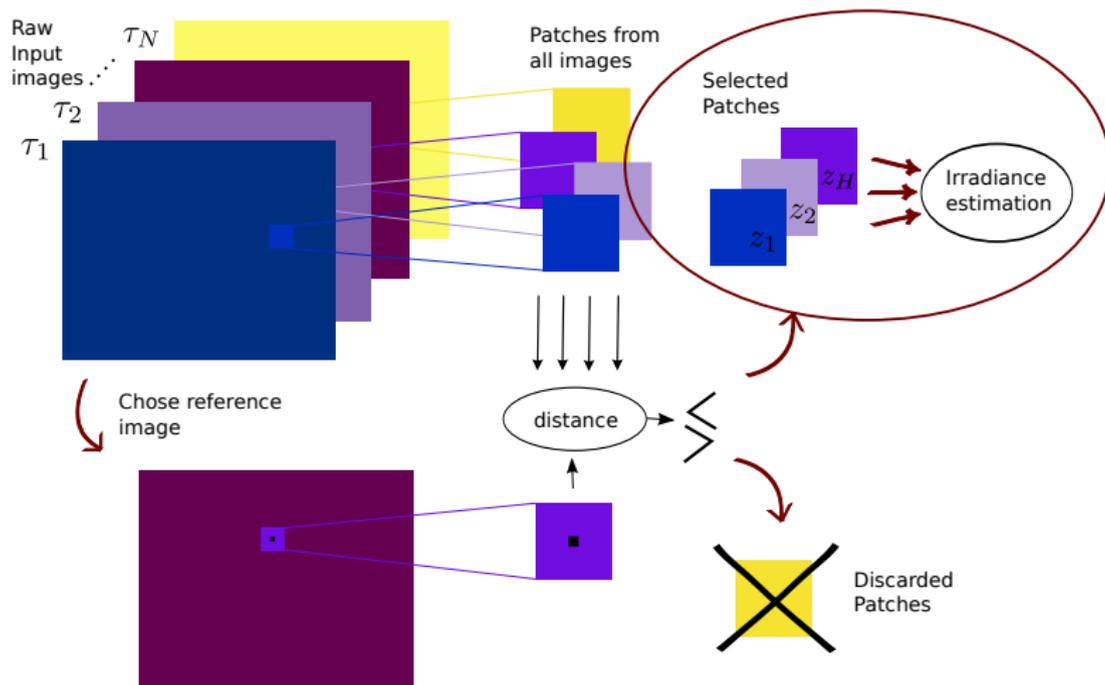


### Avantages :

- Pas besoin de recalage global (sauf pour accélérer les calculs)
- Gère automatiquement les mouvements locaux
- Peu sensible au bruit dans la comparaison des pixels (patches)
- Prend en compte les redondances inter et intra-images
- Résultat final débruité

# Estimation de l'irradiance

Comme dans le cas statique en utilisant les pixels similaires



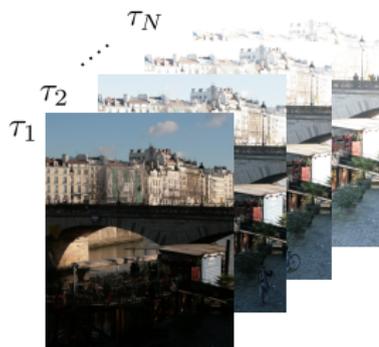
Une fois sélectionnés les pixels similaires,  
que l'on suppose suivre la loi  $Z \sim \mathcal{N}(\mu(R), \sigma^2(R))$   
On estime  $R$  comme dans le cas statique :

$$R_{\text{MLE}} = \frac{\sum_{h=1}^H w_{\text{MLE}}^h x_h}{\sum_{h=1}^H w_{\text{MLE}}^h}$$

$$w_{\text{MLE}}^h = \frac{1}{\text{var}(x_h)} \simeq \frac{g^2 \tau_h^2}{g^2 \tau_h x_h + \sigma_R^2}$$

avec  $\{x_h\}_{h=1, \dots, H}$ , les pixels semblables selon  $d(p, q)$  (seuil).

# Sélection de l'image de référence



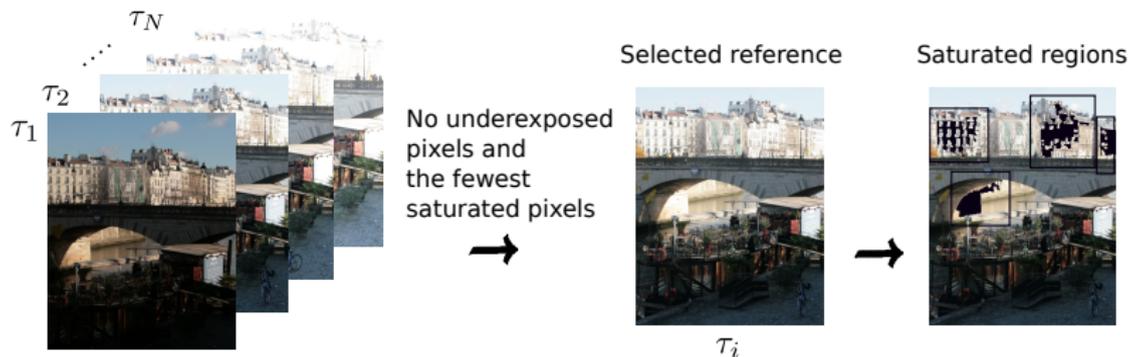
- L'information est perdue dans les zones saturées
- Nécessité de les remplir : e.g. par inpainting

# Sélection de l'image de référence



- L'information est perdue dans les zones saturées
- Nécessité de les remplir : e.g. par inpainting

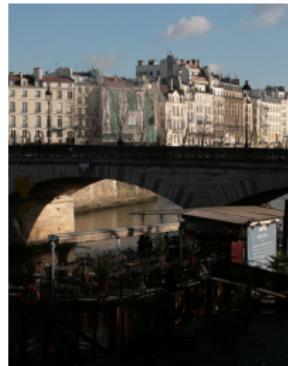
# Sélection de l'image de référence



- L'information est perdue dans les zones saturées
- Nécessité de les remplir : e.g. par inpainting

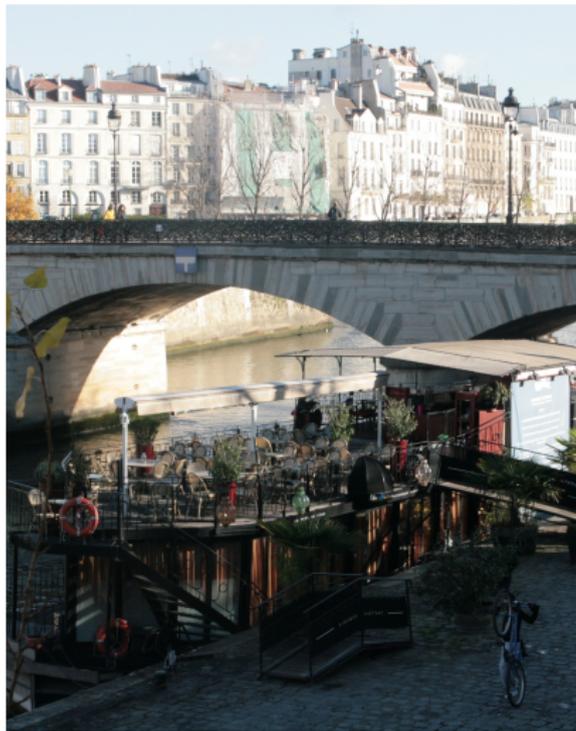
# Résultats : Exemple 1

## Entrées

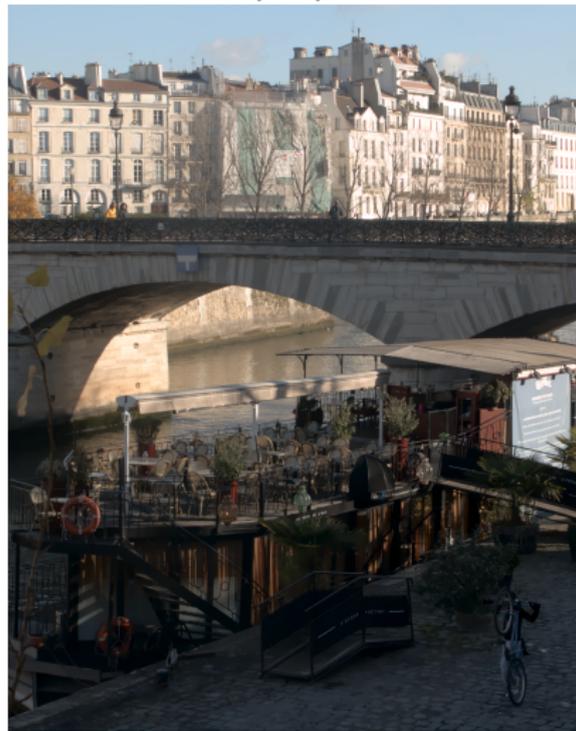


# Résultats : Exemple 1

Référence



HDR par patches



Pas de fantomes



Référence



HDR par  
patches



# Résultats : Exemple 2

Entrées



# Résultats : Exemple 2

Référence



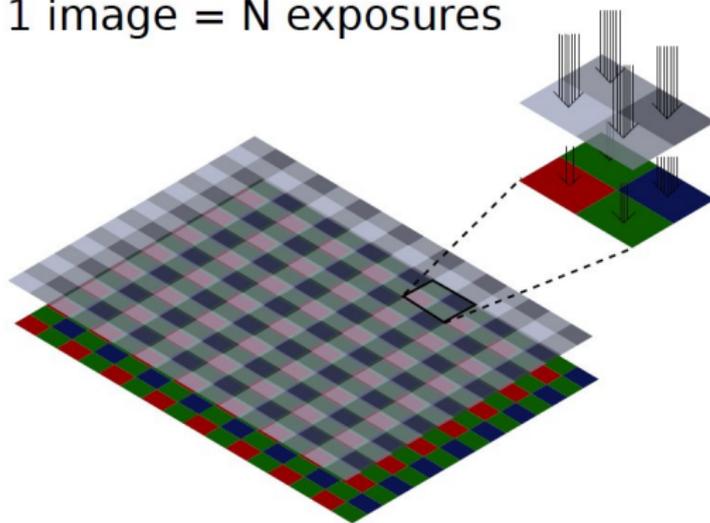
HDR par patches



# HDR en une prise de vue (single shot)

Une alternative au multi-images :  
expositions spatialement variables [Nayar and Mitsunaga 2000, Aguerrebere  
et al. 2014, 2017]

1 image = N exposures

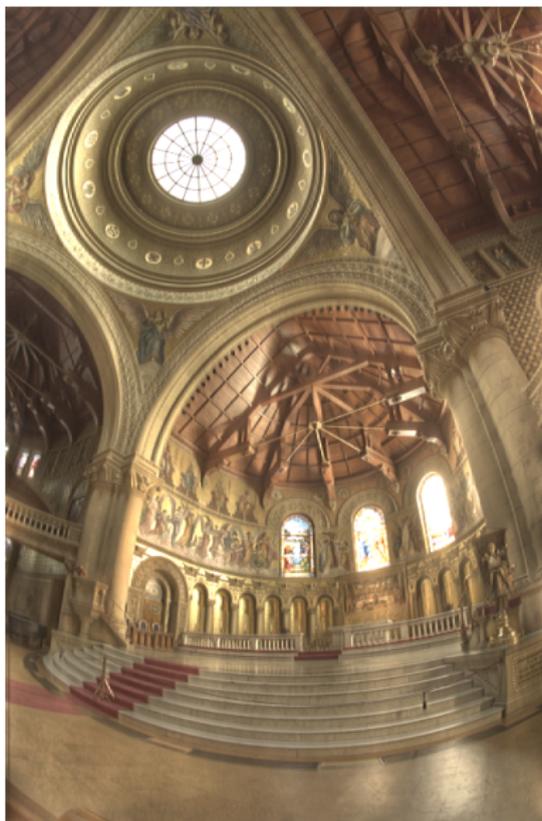


Comment visualiser une image HDR  
(potentiellement 15-20 stops)  
avec un dispositif traditionnel ?  
(écran LCD inférieur à 12 stops)  
⇒ réduction du contraste

Première option, on modifie  $I$  en  $h \circ I$ , avec  $h$  un changement de contraste (fonction croissante).

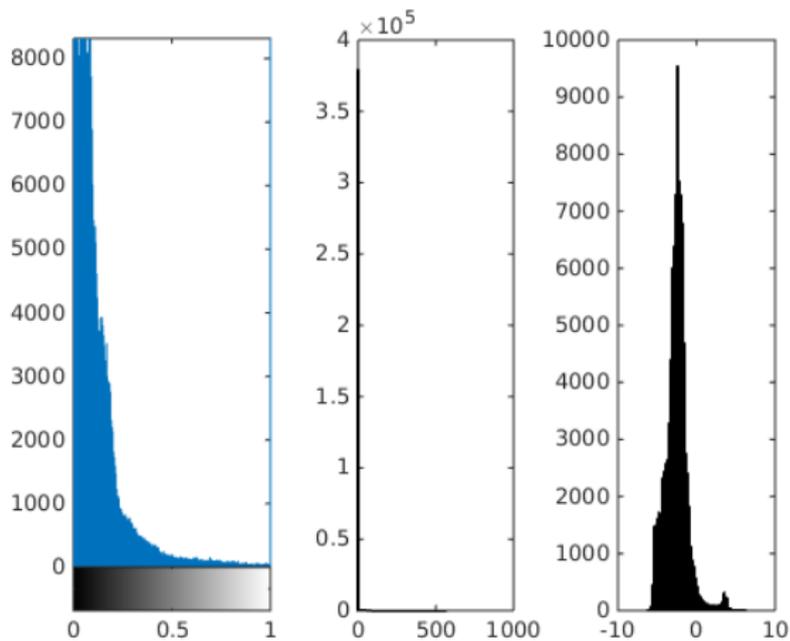
Exemples : transformations gamma (avec  $\gamma < 1$ ), logarithmiques, etc. voir e.g. (Tumblin-Rushmeier 93)

## Scène HDR typique

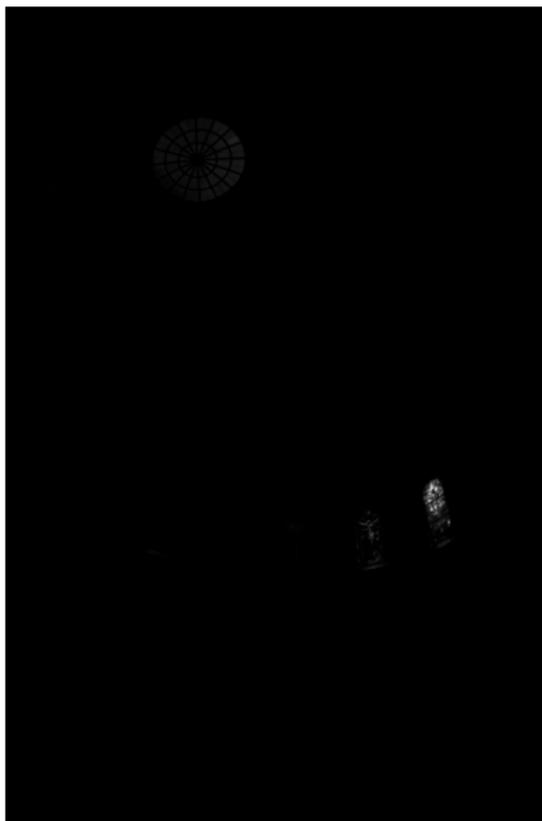


Dynamique 17 stops

## Tone mapping global



Histogramme des valeurs 0-1 ; histogramme 0-570 ; histogramme du log



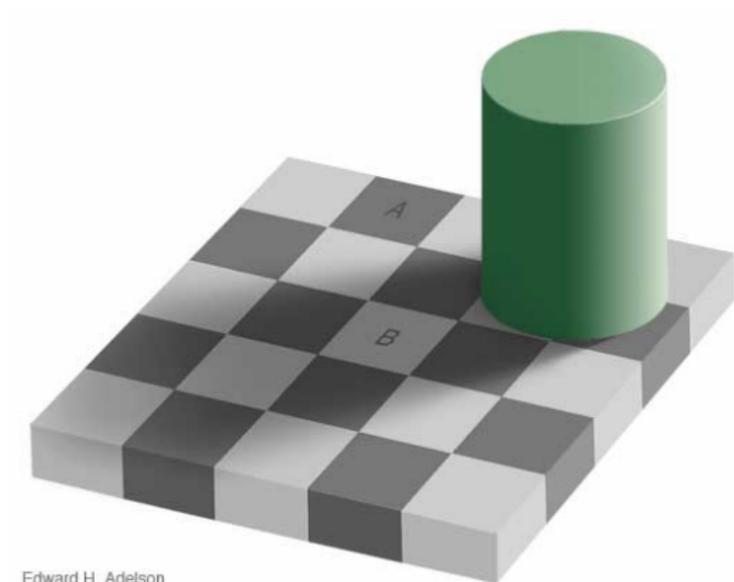
Tone mapping linéaire 0-max





logarithmique

Pour obtenir une plus grande réduction de la dynamique : préservation du contraste local



Edward H. Adelson

- Principe : atténuer le contraste d'une version "cartoon" de l'image tout en préservant le contraste des détails (Durand-Dorsey 2002)
- Soit  $B = F(I)$  un filtrage de l'image qui préserve les bords ;  
On considère  $\tilde{I} = k.F(I) + (I - F(I))$ ,  
 $k$  jouant le rôle d'un facteur d'atténuation.
- Calculs effectués dans le domaine de la log-luminance
- Proposition initiale :  $F$  est un *filtre bilatéral* :

$$F(I)(x) = C(x)^{-1} \sum_{y \in \Omega} g_{\sigma_1}(x - y) \cdot g_{\sigma_2}(I(x) - I(y)) I(y).$$

avec

$$C(x) = \sum_{y \in \Omega} g_{\sigma_1}(x - y) \cdot g_{\sigma_2}(I(x) - I(y))$$

- En pratique, de nombreuses alternatives efficaces (décomposition variationnelles cartoon-textures, TV-L1)

Algorithme :

Pour une image couleur  $I = (R, G, B)$

- Calcul de l'intensité  $L = (a.R + b.G + c.B)$
- $L_{base} = F(\log L)$ , avec  $F$  filtre bilatéral
- $L_{detail} = \log(L) - L_{base}$
- $L_{out} = k.L_{base} + L_{detail} - C$ ,  $C$  constante de normalisation
- $R_{out} = (R/L) \cdot \exp(L_{out})$ ,  
idem pour  $G$  et  $B$ .
- Possiblement saturation des plus grandes valeurs (ou d'une proportion)

# Décomposition base-détail



Base



Detail



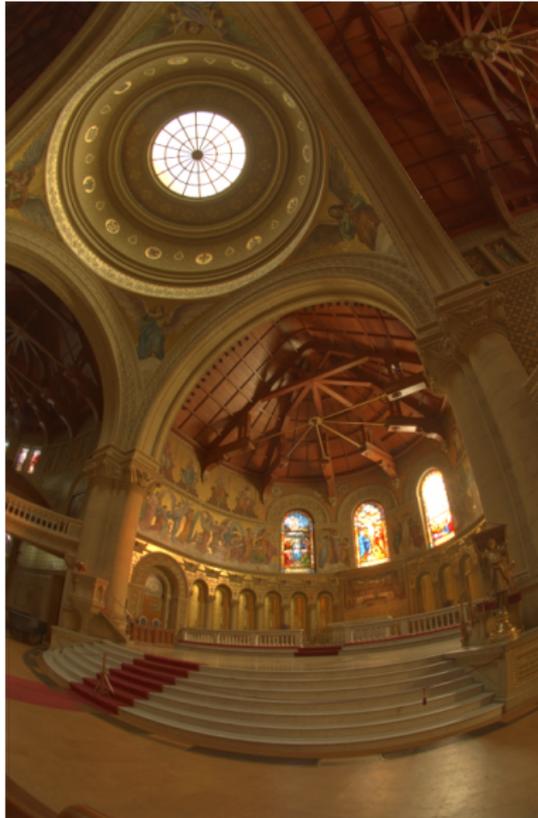
Color

From Durand-Dorsey 2002

# Décomposition base-détail



From Durand-Dorsey 2002





# Décomposition base-détail



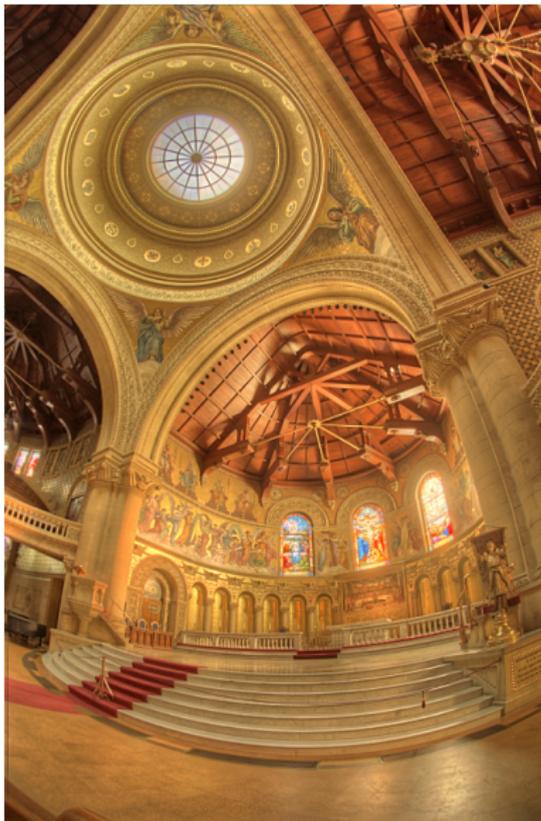
logarithmique

Principe : atténuation de l'image des gradients

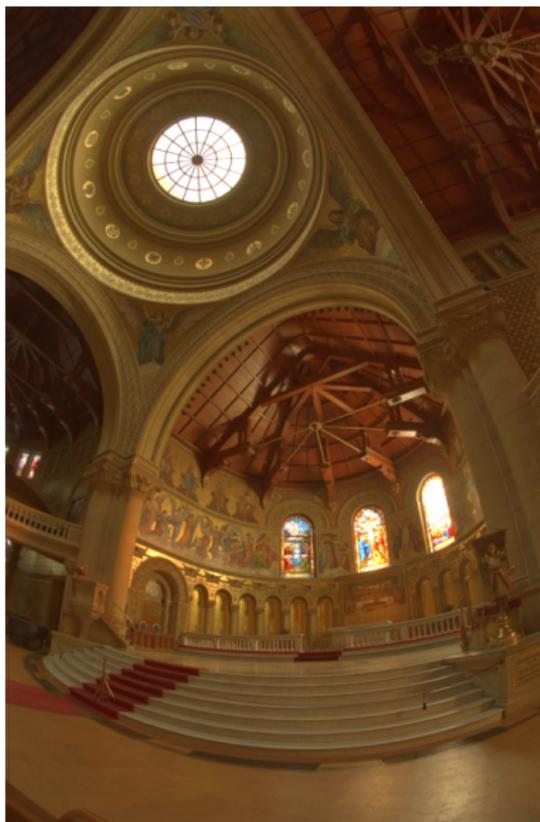
- Calcul de la carte de gradient ( $\nabla I$ )
- Atténuation de la dynamique des gradients
- Reconstruction de l'image par équation de Poisson

Dans un cadre multi-échelles (pyramide gaussienne)

# Tone mapping



Gradient



Alternative à (création HDR) + (tone mapping)

On crée directement une image visualisable (e.g. 8 bits / canal) par fusion des images finales (JPEG)  $I_1, \dots, I_N$  **recalées**

En chaque pixel  $x$

$$R(x) = \sum_{i=1}^N W_i(x) I_i(x)$$

avec

$$W_i(x) = C_i(x)^{\alpha_c} E_i(x)^{\alpha_E} S_i(x)^{\alpha_S},$$

et

- contraste

$$C_i(x) = \Delta I_i(x)$$

- exposition

$$E_i(x) = \exp\left(-\frac{(I(x) - 0,5)^2}{2\sigma^2}\right)$$

- saturation

$$S_i(x) = \text{std}(R(x), G(x), B(x))$$



(a) input sequence



(b) Naive



(c) Blurred

From Mertens et al.

Zones plates bruitées + transitions abruptes

Solution : la fusion est effectuée dans un cadre multi-échelle

$$\mathcal{L}_l(R)(x) = \sum_{i=1}^N \mathcal{G}_l(W_i)(x) \mathcal{L}_l(I_i)(x),$$

Avec  $\mathcal{L}_l$ ,  $\mathcal{G}_l$  les pyramides laplacienne et gaussienne pour une série d'échelles  $l$ .



(e) Multiresolution

# Fusion multi-échelles



(a) Exposure bracketed sequence



Pour les scènes dynamiques ... mêmes problèmes que pour la création



Fig. 2: Fused images performed by, (a) non-local exposure fusion (NLEF), and (b) classical exposure fusion [7].

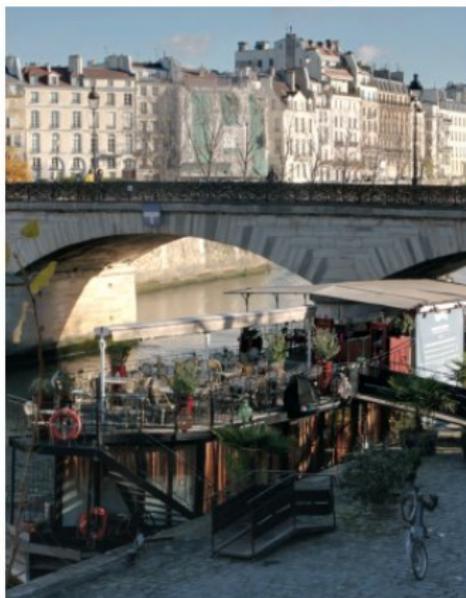
Pour traiter les scène dynamiques : fusion des pixels ayant même voisinage (patch)



Fig. 2: Fused images performed by, (a) non-local exposure fusion (NLEF), and (b) classical exposure fusion [7].

Fig. Ocampo et al. 2016

# Fusion non locale



# Fusion non locale



# Image HDR à partir d'une seule image LDR



Fig. Eilertsen et al. 2017

# Image HDR à partir d'une seule image LDR

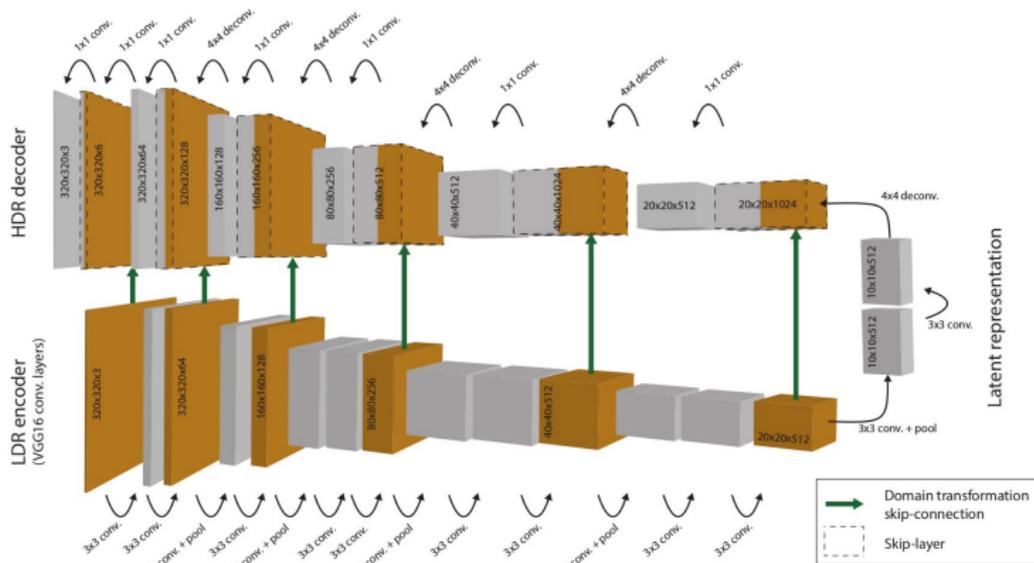


Fig. Eilertsen et al. 2017

# Image HDR à partir d'une seule image LDR

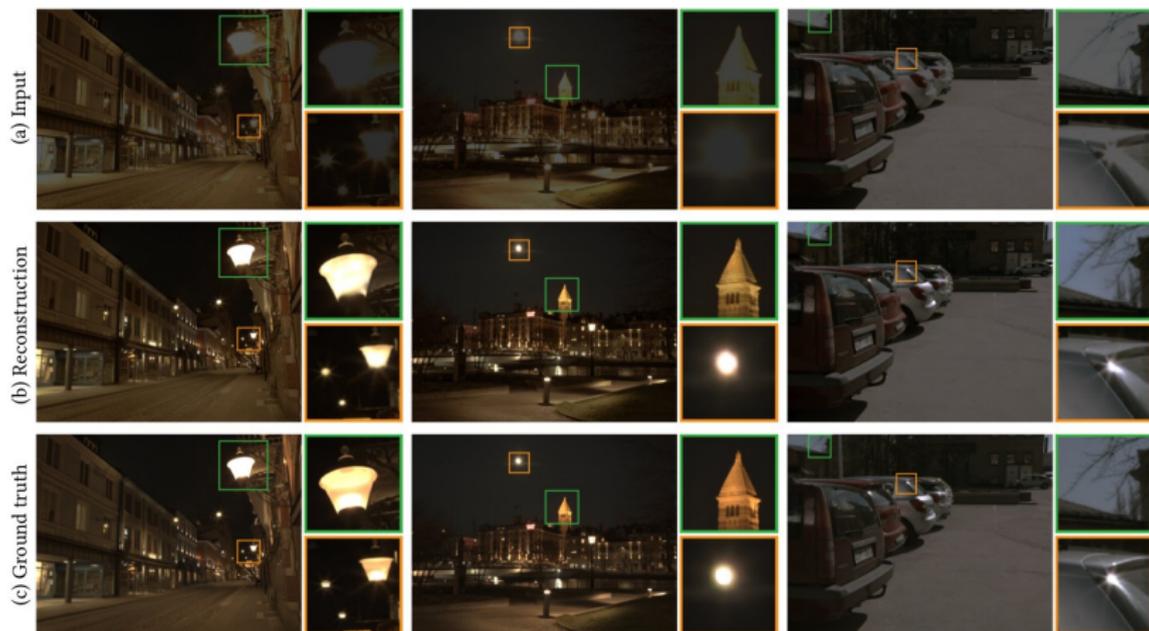
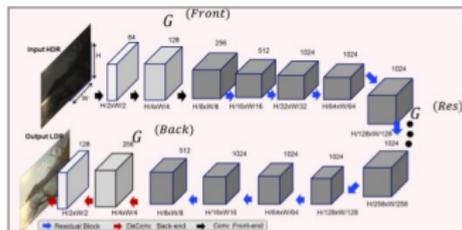
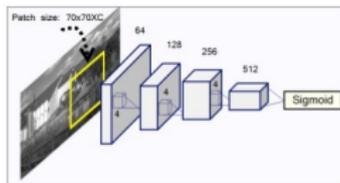


Fig. Eilertsen et al. 2017

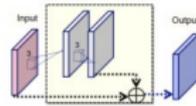
# Deep Tone Mapping



(a) Generator (Single Scale).



(b) Discriminator (Single Scale).



(c) Residual Blocks

Générateur : auto-encodeur

Discriminateur : PatchGAN / en entrée : concaténation LDR/HDR

Fig. Rana et al. 2019

# Deep Tone Mapping



Fig. Rana et al. 2019

Apprentissage profond pour la fusion (avec ou sans alignement) :  
Kalantari et al. 2017, Prabhakar et al. 2017 (DeepFuse), Yan et al. 2019,  
Yang et al. 2021 (GanFuse), etc.