

Algorithmes optimaux pour la génération d'images HDR

Une étude des bornes de performance

Cecilia AGUERREBERE^{1,2}, Julie DELON¹, Yann GOUSSEAU¹, Pablo MUSÉ²

¹LTCI CNRS & Télécom ParisTech,
75634 Paris Cedex 13, France

²Instituto de Ingeniería Eléctrica
Universidad de la República, Montevideo, Uruguay

aguerreb@telecom-paristech.fr, delon@telecom-paristech.fr
gousseau@telecom-paristech.fr, pmuse@fing.edu.uy

Résumé – Nous posons le problème de l'estimation de l'irradiance comme un problème d'estimation statistique et en calculons la borne de performance (borne de Cramér-Rao). Nous montrons qu'elle ne peut être atteinte, mais qu'une méthode récente [5] permet de l'approcher. Nous montrons ensuite qu'en tenant compte des échantillons saturés, les résultats d'estimation de l'irradiance sont améliorés. Les résultats obtenus justifient l'introduction d'un nouvel estimateur qui, contrairement aux méthodes de la littérature, prend en compte les échantillons saturés.

Abstract – We state the problem of irradiance estimation as a statistical parameter estimation problem and compute the performance bounds of the estimation (Cramér-Rao lower bound). We show that no efficient estimator exists under the stated hypothesis. However, we find that a recently proposed method [5], performs very close to this bound. Moreover, we show that releasing one of the working hypothesis (including the saturated samples) the irradiance estimation can be improved. We introduce a new estimation method that, unlike existing methods, takes into account the information provided by the saturated samples.

1 Introduction

L'œil humain a la capacité de percevoir des scènes ayant une très forte dynamique, comportant des détails aussi bien dans les régions sombres que dans les régions claires. Au contraire, les capteurs des appareils photographiques classiques sont incapables de capturer ces très fortes dynamiques en raison du phénomène de saturation. En présence de très fortes lumières, les photosites saturent et l'information est perdue. À l'opposé, réduire le temps d'exposition pour éviter la saturation réduit considérablement le nombre de photons capturés dans les régions sombres, quantité qui sera alors masquée par le bruit d'acquisition. En conséquence, une exposition unique d'une scène à forte dynamique avec un appareil classique contiendra soit des pixels saturés soit des pixels très bruités.

L'imagerie haute dynamique (HDR pour l'anglais High Dynamic Range) est la discipline dont le but est de capturer fidèlement puis de reproduire des scènes ayant une plage dynamique aussi étendue que possible. Pour dépasser les limitations physiques des capteurs actuels, l'approche la plus classique consiste à combiner plusieurs vues ayant chacune une dynamique limitée, acquises avec des temps de pose différents τ_1, \dots, τ_T . En effet, pour une irradiance donnée C et un temps de pose τ_i , la valeur au pixel correspondant est une fonction de l'énergie lumineuse reçue $\tau_i C$. Dans ces conditions, avoir recours à plusieurs temps d'exposition permet d'échantillonner la

fonction de réponse de l'appareil photographique en différents points et ainsi d'éviter la saturation au moins pour un sous-ensemble des temps d'exposition, ce qui permet de conserver les détails à la fois dans les régions claires et sombres. Depuis les travaux précurseurs de Mann et Picard [8], la plupart des méthodes estiment l'irradiance au pixel p comme une combinaison linéaire de $f^{-1}(z_i^p)/\tau_i$, $i = 1, \dots, T$, où z_i^p représente la valeur de l'image au pixel p pour un temps de pose τ_i et f est la fonction de réponse de la caméra.

Dans cette communication, nous nous concentrons sur le problème de l'estimation de la carte d'irradiance (c'est à dire l'irradiance en chaque pixel) à partir d'un petit nombre de photographies dont les temps de poses sont fixés *a priori*. Plus précisément, notre but est d'étudier les limites théoriques du problème d'estimation correspondant et la capacité des méthodes de la littérature à atteindre ou approcher ces bornes. Nous nous placerons donc dans le cas où les photographies sont parfaitement recalées et où les changements radiométriques potentiels ont été compensés. Une présentation plus détaillée de ces travaux est disponible [1].

Les méthodes d'estimation de l'irradiance peuvent être classées selon plusieurs critères. L'un d'eux est la linéarité [11, 7, 5] ou non-linéarité [8, 3, 9, 12, 10] de la fonction de réponse de la caméra. Les premières méthodes sont développées pour les images RAW, c'est à dire sans post-traitement (démosaïquage, balance des blancs, etc.). Les méthodes de la deuxième catégorie

nécessitent l'estimation de la fonction de réponse de l'appareil. Dans cette communication, nous travaillerons à partir des images RAW : avec l'augmentation des capacités de stockage et de calcul, il est aujourd'hui parfaitement raisonnable de supposer que ces images brutes sont disponibles ou que les traitements seront appliqués dans les appareils photographiques. Dans ce qui suit, si l'on fait abstraction du bruit, les valeurs des pixels sont donc proportionnelles au nombre de photons.

2 Modèle d'acquisition

Une analyse détaillée du processus d'acquisition des images numériques par un capteur CMOS ou CCD permet de déduire le modèle aléatoire suivi par les valeurs brutes \mathbf{Z}_i des pixels avant saturation [2]. Ce modèle inclut les différentes sources de bruit (bruit photonique *shot noise*, courant noir, bruit de lecture) et les non-uniformités spatiales (non-uniformité de la réponse des photosites (PRNU), non-uniformité du courant noir (DCNU)) :

$$\mathbf{Z}_i \sim \mathcal{N}(ga\tau_i C + \mu_R, g^2 a \tau_i C + \sigma_R^2), \quad (1)$$

où C est l'irradiance, g le gain de l'appareil, a le facteur PRNU, τ_i le temps d'exposition et \mathcal{N} la distribution gaussienne. Des modèles très similaires sont considérés dans [5, 4].

3 Bornes de performances

Rappelons les bases du problème. En chaque pixel, nous observons $\mathbf{z} = z_1, \dots, z_T$ correspondant aux temps de pose τ_1, \dots, τ_T . Chaque z_i est la réalisation d'une variable aléatoire \mathbf{Z}_i suivant le modèle (1). Nous souhaitons alors estimer l'irradiance C en ce pixel, c'est à dire le nombre moyen de photons atteignant ce pixel par unité de temps. Cette estimation se fait pixel à pixel, supposant donc que les images sont parfaitement alignées. Nous supposons dans un premier temps que les pixels saturés ont été éliminés, une hypothèse systématique dans la littérature.

Nous calculons dans ce cadre la borne de Cramér-Rao qui donne la variance minimale atteignable par un estimateur pour un biais donné. En pratique, nous avons observé que pour tous les estimateurs de C proposés dans la littérature, le rapport $\frac{(\mathbb{E}[|\hat{C}-C|])^2}{C^2}$ reste extrêmement proche de 0 quelque soit la valeur de C [1]. On peut donc supposer que pour tous ces estimateurs, le terme dominant dans le calcul de l'erreur quadratique moyenne (que l'on notera EQM) est le terme de variance. Nous nous concentrons donc dans ce qui suit sur les estimateurs non biaisés de C (ou quasiment non biaisés). Pour le modèle (1), la borne de Cramér-Rao dans le cas non-biaisé est

$$\text{CRLB} = \left[\sum_{i=1}^T \frac{(ga\tau_i)^2}{g^2 a \tau_i C + \sigma_R^2} + \frac{(g^2 a \tau_i)^2}{2(g^2 a \tau_i C + \sigma_R^2)^2} \right]^{-1}, \quad (2)$$

où T est le nombre de temps de pose considéré. L'EQM de tout estimateur non-biaisé est bornée par cette expression. La question qui se pose est alors celle de l'existence d'un estimateur

efficace, c'est-à-dire d'un estimateur non-biaisé atteignant la borne pour toute valeur de C . D'après le théorème de Cramér-Rao, un tel estimateur existe si et seulement s'il existe des fonctions h et I telles que la dérivée de la log-vraisemblance puisse être factorisée sous la forme

$$\frac{\partial \ln p(\mathbf{z}, c)}{\partial c} = I(c)(h(\mathbf{z}) - c), \quad (3)$$

Dans ce cas, l'estimateur efficace est $\hat{C} = h(\mathbf{z})$, et sa variance est $1/I(C)$ [6]. L'Equation (3) doit être vérifiée pour toutes valeurs de c et de \mathbf{z} . Or, si l'on fixe \mathbf{z} et que l'on étudie la limite des deux termes de l'équation (3) lorsque c tend vers l'infini (voir [1] pour plus de détails), le terme de gauche tend vers $-\frac{1}{2} \sum_{i=1}^T a \tau_i$ alors que celui de droite tend vers $-\sum_{i=1}^T a \tau_i$. Les deux limites étant différentes, il n'existe pas d'estimateur efficace pour ce problème.

3.1 Performances des estimateurs classiques

S'il n'existe pas d'estimateur efficace de l'irradiance sous le modèle (1), il est néanmoins intéressant d'étudier les performances des méthodes proposées dans la littérature par rapport à la borne CRLB. Dans ce qui suit, nous évaluons sept méthodes d'estimation de C : le maximum de vraisemblance (EMV) proposé par Granados et al. [5], les méthodes de Kirk et Andersen [7], Robertson et al. [11], Debevec et Malik [3], Mitsunaga et Nayar [9], Reinhard et al. [10], ainsi qu'un estimateur très simple obtenu en prenant des poids proportionnels aux temps τ_i . Ce dernier estimateur, utilisé dans l'industrie de l'imagerie HDR, est l'estimateur non biaisé de variance minimale dans le cas où les échantillons suivent simplement une loi de Poisson. Une modification triviale de l'approche par Kirk et Andersen (l'inclusion des facteurs PRNU), nous permet d'avoir une version non biaisée de l'estimateur. L'analyse expérimentale est menée sur des données synthétiques, car la connaissance précise de la vérité terrain est indispensable au calcul de la borne CRLB.

Protocole Des échantillons synthétiques sont générés suivant le modèle (1), à partir d'une image HDR prise comme vérité terrain (c.f. Figure 1). Trois combinaisons de temps d'exposition sont testées : une composée de temps d'exposition longs¹ une de temps d'exposition courts² et une de temps d'exposition intermédiaires³. L'appareil photographique simulé est un Canon 7D réglé à ISO 200. Il est d'abord calibré [5] pour obtenir ses paramètres ($g = 0.87$; $\sigma_R^2 = 31.6$; $\mu_R = 2046$; $z_{sat} = 14042$). Cet appareil simulé est appelé Camera A. D'autres configurations de caméra et de temps d'exposition ont été testés et donnent des résultats similaires. Les facteurs PRNU en chaque pixel sont simulés selon une loi Gaussienne de moyenne 1 et d'écart-type 0.01 [5]. La plage dynamique de la scène est de 12.7 stops. Les estimations sont répétées 1000 fois pour chaque

1. $\tau_{4L} = (1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8})s$.
2. $\tau_{4S} = \frac{1}{50} \cdot (1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8})s$.
3. $\tau_{4M} = (\frac{1}{12.4}, \frac{1}{25}, \frac{1}{50}, \frac{1}{100})s$.

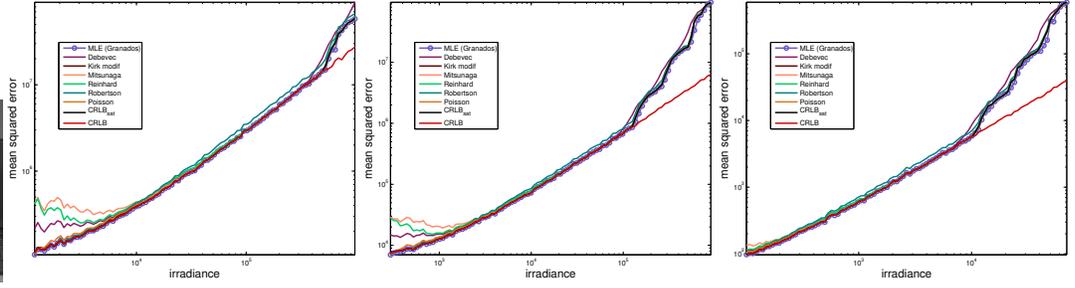


FIGURE 1 – Comparaison des courbes d’erreur quadratique moyenne et de la borne CRLB pour la Camera A avec différentes combinaisons de 4 temps d’exposition, de gauche à droite : τ_{4S} , τ_{4M} , τ_{4L} . L’erreur de l’EMV est toujours très proche de la borne.

valeur d’irradiance. Les EQM des estimateurs sont calculées à partir de ces 1000 valeurs.

Résultats La Figure 1 montre les courbes d’EQM pour toutes les méthodes testées. Aucune des méthodes comparées ici ne tient compte des pixels saturés. On calcule donc deux bornes grâce à l’équation (2) : $CRLB_{SAT}$, borne pour laquelle les temps d’exposition correspondant aux pixels saturés ne sont pas utilisés, et CRLB, borne idéale prenant en compte toutes les expositions, en supposant qu’aucun pixel ne sature.

Dans les cas de faible irradiance, les résultats obtenus par Debevec et Malik [3], Mitsunaga et Nayar [9] et Reinhard et al [10] sont très loin de la borne $CRLB_{SAT}$. Ces schémas sont très sensibles au bruit et leur performance est sévèrement dégradée pour les faibles valeurs d’irradiance où le SNR est relativement bas.

Les résultats montrent que l’estimateur du maximum de vraisemblance (EMV) proposé par Granados et al. [5] est en pratique très proche de la borne de Cramér-Rao. Quelques pistes théoriques sont présentées dans la section suivante pour expliquer cette proximité. Rappelons en effet que si l’EMV est asymptotiquement efficace, ses performances sont inconnues dans le cas non-asymptotique (ce qui est le cas ici, avec seulement 4 échantillons par pixel). La conclusion de cette étude est qu’il y a peu de place pour améliorer l’EMV sous les mêmes hypothèses (images parfaitement alignées, modèle d’acquisition connu, rejet des échantillons saturés).

3.2 Performance de l’EMV

L’absence d’une formule close pour l’EMV sous le modèle (1) a motivé Granados et al. [5] à proposer un algorithme itératif pour le calcul de l’EMV. A l’itération $(j+1)$ l’irradiance $\hat{C}^{(j+1)}$ est estimée de la manière suivante⁴

$$\hat{C}^{(j+1)} = \frac{\sum_{i=1}^T w_i(\hat{C}^{(j)}) x_i}{\sum_{i=1}^T w_i(\hat{C}^{(j)})}, \quad (4)$$

$$\text{avec } x_i = \frac{z_i}{ga\tau_i} \quad w_i = \frac{(ga\tau_i)^2}{g^2 a \tau_i \hat{C}^{(j)} + \sigma_R^2}. \quad (5)$$

4. Contrairement à Granados et al., nous négligeons ici les courants noirs de l’appareil car leur contribution au bruit global est minimal pour les expositions généralement utilisées dans les méthodes de création d’images HDR.

On peut vérifier qu’en pratique, si les poids sont correctement initialisés, l’estimateur reste presque inchangé après la première itération. Une initialisation correcte est obtenue en calculant les poids directement à partir des échantillons observés, ce qui donne $\hat{C}^{(0)} = \frac{z_i}{ga\tau_i}$. Or, au premier ordre (voir [1]),

$$\text{var}(\hat{C}^{(1)}) = \left[\sum_{i=1}^T \frac{(ga\tau_i)^2}{g^2 a \tau_i C + \sigma_R^2} \right]^{-1} + E(o((\mathbf{z})^2)). \quad (6)$$

Par suite, la variance de $\hat{C}^{(1)}$ est proche de la borne CRLB si

$$\frac{g^2 a^2 \tau_i^2}{g^2 a \tau_i C + \sigma_R^2} \gg \frac{g^4 a^2 \tau_i^2}{2(g^2 a \tau_i C + \sigma_R^2)^2} \quad \forall i = 1, \dots, T, \quad (7)$$

ce que l’on peut réécrire $a\tau_i C \gg \frac{1}{2} - \frac{\sigma_R^2}{g^2}$ $\forall i = 1, \dots, T$. Cette inégalité est vérifiée en pratique, même pour les petites valeurs de $a\tau_i C$, car le terme $\frac{\sigma_R^2}{g^2}$ est beaucoup plus grand que $1/2$ pour la plupart des appareils. Ceci suggère que l’algorithme itératif proposé par Granados et al. atteint quasiment la borne CRLB après une seule itération.

4 Inclusion des pixels saturés

A notre connaissance, toutes les méthodes d’estimation de l’irradiance ignorent les échantillons saturés. Néanmoins, nous avons observé que la borne de Cramér-Rao était nettement meilleure dans le cas où l’on inclut tous les échantillons (voir les courbes CRLB et $CRLB_{SAT}$ sur la Figure 1). Nous proposons une méthode permettant d’inclure l’information des échantillons saturés dans l’estimation de l’irradiance. Bien que la valeur exacte menant à la saturation soit inconnue, le fait que ces valeurs soient au-dessus d’un seuil z_{sat} est porteur d’information. Nous développons donc une variante de l’EMV incluant cette information supplémentaire.

Soit $\mathbf{Z}_1, \dots, \mathbf{Z}_T$ des variables indépendantes correspondant aux valeurs brutes en un pixel donné p pour des temps de pose τ_1, \dots, τ_T . En raison de la saturation, chaque \mathbf{Z}_j s’écrit $\mathbf{Z}_j = \min(\mathbf{X}_j, z_{sat})$, avec \mathbf{X}_j une variable aléatoire suivant la loi $\mathcal{N}(\mu_j(C), \sigma_j^2(C))$, où $\mu_j(C) = ga\tau_j C$ et $\sigma_j^2(C) = g^2 a \tau_j C + \sigma_R^2$. Notons $p_j(z; \phi)$ la densité de la loi $\mathcal{N}(\mu_j(\phi), \sigma_j^2(\phi))$. Les approches usuelles reposent sur la maximisation de la vraisemblance $\prod_{z_j < z_{sat}} p_j(z_j; \phi)$ en fonction de ϕ . Afin de prendre en compte la saturation, nous proposons d’utiliser la vraisemblance complète, correspondant à l’ensemble des échantillons

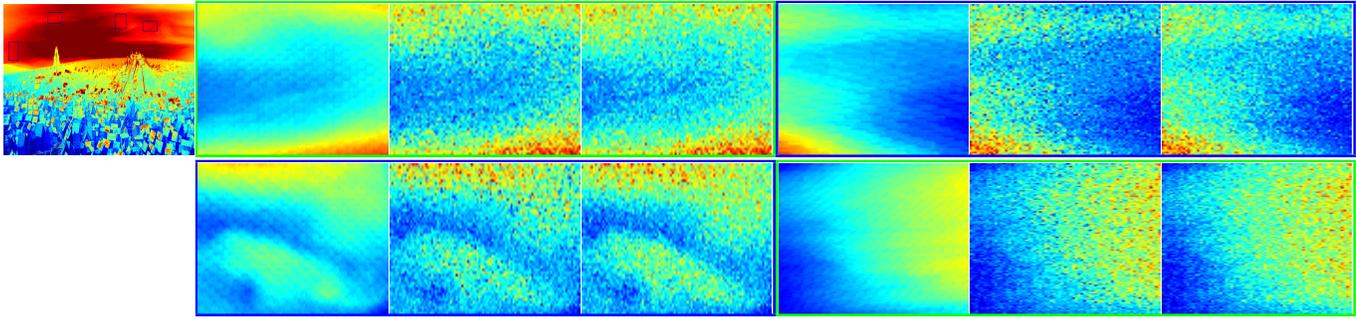


FIGURE 2 – **En haut à gauche** : vérité terrain et sous-régions traitées (en bleu). **Pour chaque ligne**, résultats pour deux sous-régions. De gauche à droite : vérité terrain, résultat de l’estimateur du maximum de vraisemblance classique, résultat obtenu en tenant compte des échantillons saturés.

$\mathbf{z} = (z_1, \dots, z_T)$, qui s’écrit

$$g(\mathbf{z}|\phi) = \prod_{j=1}^T p_j(z_j; \phi)^{k_j} P_j(\phi)^{1-k_j}, \quad (8)$$

avec $k_j = 1$ pour les échantillons non-saturés et 0 sinon. La probabilité $P_j(\phi)$ que l’échantillon j sature vaut $\int_{z_{\text{sat}}}^{\infty} p_j(z; \phi) dz$. Remarquons que si aucun des pixels ne sature $g(\mathbf{z}|\phi)$ est la vraisemblance maximisée par l’EMV classique.

Une approche EM nous permet de maximiser (8). La vraisemblance complète du problème est $h(\mathbf{x}|\phi) := \prod_{j=1}^T p_j(x_j; \phi)$. On ne peut pas maximiser h en pratique car il dépend de données censurées, mais un résultat classique de l’algorithme EM montre que la valeur ϕ maximisant $E_{\mathbf{x}}[\log h(\mathbf{x}|\phi)|\mathbf{z}, \phi]$ est exactement la même que celle qui maximise $g(\mathbf{z}|\phi)$. L’algorithme consiste donc à calculer à l’itération $p+1$ l’espérance $Q(\phi|\phi^{(p)}) := E_{\mathbf{x}}[\log h(\mathbf{x}|\phi)|\mathbf{z}, \phi^{(p)}]$, puis à trouver $\phi^{(p+1)} = \max_{\phi} Q(\phi|\phi^{(p)})$.

Nous avons montré dans [1] qu’avec cette nouvelle vraisemblance $g(\mathbf{z}|\phi)$, la borne CRLB décroît pour les valeurs d’irradiance proches de la limite entre deux niveaux de saturation (on appelle niveau de saturation une plage d’irradiances pour laquelle le nombre d’échantillons saturés est constant).

Expériences Les échantillons synthétiques sont générés à partir d’une image HDR, qui sert de vérité terrain (c.f. Figure 2), en suivant le modèle (1). L’appareil simulé est la Camera A avec les temps d’exposition $\tau = (1/4.2, 1/16.8, 1/67.2, 1/268.8)s$. La Figure 2 montre les résultats pour quatre sous-régions. Les pixels de ces sous-régions ont 2 ou 3 échantillons saturés. Dans tous les cas, l’information fournie par les échantillons saturés permet d’améliorer les performances. Le gain en comparaison de l’EMV classique est de l’ordre de 1dB.

5 Conclusion

Dans cet article, nous avons présenté une étude des bornes statistiques pour le problème de l’estimation de l’irradiance et nous avons analysé les performances des estimateurs de l’état de l’art. Cette étude montre qu’au premier ordre, l’EMV est un estimateur efficace, même avec un nombre réduit d’échantillons.

Nous avons par ailleurs montré comment exploiter les échantillons saturés, ce qui n’avait jamais été proposé dans la littérature, grâce à une approche de type EM avec données censurées. Les résultats obtenus confirment que ces échantillons contiennent une information non négligeable pour l’estimation de l’irradiance, notamment aux alentours des limites entre deux niveaux de saturation, sans pour autant dégrader les résultats pour les autres valeurs d’irradiance.

Références

- [1] C. AGUERREBERE, J. DELON, Y. GOUSSEAU, AND P. MUSE, *Best algorithms for HDR image generation. a study of performance bounds*, Preprint HAL http://hal.archives-ouvertes.fr/docs/00/73/38/53/PDF/best_hdr_algo_hal.pdf, (2012).
- [2] ———, *Study of the digital camera acquisition process and statistical modeling of the sensor raw data*, Preprint HAL http://hal.archives-ouvertes.fr/docs/00/73/35/38/PDF/camera_model.pdf, (2012).
- [3] P. E. DEBEVEC AND J. MALIK, *Recovering high dynamic range radiance maps from photographs*, in SIGGRAPH, 1997, pp. 369–378.
- [4] A. FOI, M. TRIMECHE, V. KATKOVNIK, AND K. EGIAZARIAN, *Practical poissonian-gaussian noise modeling and fitting for single-image raw-data*, IEEE Transactions on Image Processing, 17 (2008), pp. 1737–1754.
- [5] M. GRANADOS, B. AJDIN, M. WAND, C. THEOBALT, H. P. SEIDEL, AND H. P. A. LENSCH, *Optimal HDR reconstruction with linear digital cameras*, in CVPR, 2010, pp. 215–222.
- [6] S. M. KAY, *Fundamentals of statistical signal processing : estimation theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.
- [7] K. KIRK AND H. J. ANDERSEN, *Noise characterization of weighting schemes for combination of multiple exposures*, in BMVC, 2006, pp. 1129–1138.
- [8] S. MANN AND R. W. PICARD, *On being ‘undigital’ with digital cameras : Extending dynamic range by combining differently exposed pictures*, in Proceedings of IS&T, 1995, pp. 442–448.
- [9] T. MITSUNAGA AND S. K. NAYAR, *Radiometric self calibration*, in CVPR, 1999, pp. 1374–1380.
- [10] E. REINHARD, G. WARD, S. N. PATTANAIK, AND P. E. DEBEVEC, *High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting*, Morgan Kaufmann, 2005.
- [11] M. A. ROBERTSON, S. BORMAN, AND R. L. STEVENSON, *Estimation-theoretic approach to dynamic range enhancement using multiple exposures*, J. Electronic Imaging, 12 (2003), pp. 219–228.
- [12] Y. TSIN, V. RAMESH, AND T. KANADE, *Statistical calibration of the CCD imaging process*, in ICCV, 2001, pp. 480–487.