

Imagerie Haute Dynamique par Fusion d'Exposition

Lina Mezghani & Raphaël Montaud

January 22, 2019

INTRODUCTION

En photographie, la plage dynamique d'un appareil correspond à sa capacité à retranscrire les hautes et basses lumières. Celle-ci étant souvent limitée, et inférieure à la plage de lumière présente dans la plupart des scènes réelles, il est très fréquent de rencontrer des problèmes d'exposition trop faible ou trop élevée. Pour répondre à ce problème, il est possible de capturer une suite de photographies de la même scène avec des temps d'exposition croissants afin d'acquérir l'ensemble de la plage dynamique d'une scène réelle, puis de rassembler cette suite d'images en une image dite "HDR" (High Dynamique Range). La plage dynamique de la plupart des écrans étant elle aussi limitée, il n'est pas possible de visualiser ces images HDR telles quelles. Il faut alors envoyer cette grande plage dynamique vers une plage dynamique plus petite, via un processus appelé "tone mapping".

La méthode proposée par Mertens *et al.*[4] présente une alternative au processus décrit ci-dessus qui permet de ne pas passer par une image HDR intermédiaire. Cette méthode, que nous allons étudier dans ce rapport, présente de nombreux avantages par rapport aux méthodes antérieures, notamment un temps de calcul réduit et une qualité de l'image souvent améliorée. La suite du rapport sera organisée comme suit: dans un premier temps, nous décrirons la méthode proposée par Mertens *et al.*[4]. Dans une deuxième partie, nous présenterons des commentaires sur les détails de son implémentation en pratique, et dans une dernière partie, nous exposerons et comparerons les résultats de cette méthode sur différents exemples.

1 DESCRIPTION DE LA MÉTHODE IMPLÉMENTÉE

Dans cette section, nous présenterons en détail la méthode proposée Mertens *et al.*[4]. Comme expliqué dans l'introduction, cette méthode a l'avantage de ne pas nécessiter l'implémentation d'une image HDR intérmédiaire, mais permet d'obtenir directement une image de bonne qualité, sans défaut d'exposition, et avec une plage dynamique réduite. Etant donnée une suite d'images d'une même scène avec des temps d'expositions croissants, que l'on appellera "séquence multi-exposition" dans la suite de ce rapport, l'idée est de se baser sur des mesures de qualité en chaque pixel pour garder les zones de meilleure qualité de la séquence. Les images de la séquence sont ensuite fusionnées, et pour éviter certains artefacts inhérent à la fusion d'images, celle-ci est effectuée dans un cadre multi-échelles, en utilisant des pyramides gaussiennes et laplaciennes. Il est important de remarquer que les images doivent être parfaitement alignées. Dans un premier temps, nous décrirons les différentes mesures de qualité utilisées, puis nous expliquerons le processus de fusion multi-échelle.

1.1 MESURES DE QUALITÉ

Trois mesures de qualité sont proposées dans l'article. La saturation, le contraste et l'exposition. Nous renvoyons le lecteur vers [4] pour les détails. Chacune de ces mesures permet d'estimer si un pixel est dans une région où le temps d'exposition est convenable par rapport à la luminosité de la région de l'image. Dans la figure 1, nous illustrons les trois mesures de qualité sur un ensemble de trois images ayant des temps d'exposition différents.

- La mesure de contraste, calcule le Laplacien de l'image en chaque point (dérivée seconde). Il semble en effet logique que les zones ayant un "bon" temps d'exposition sont plus contrastées. On peut ainsi voir que les rochers sont plus détaillés dans l'image la plus claire et ce sont aussi les plus contrastés.
- La mesure d'exposition permet de distinguer les pixels "bien exposés" de ceux qui sont trop exposés (valeurs proches de 255) ou sous-exposés (valeurs proches de 0). Cette mesure semble particulièrement adaptée car elle distingue bien les trois régions de l'image. Le ciel, la partie la plus lumineuse de l'image est considéré comme "bien exposé" dans l'image la plus sombre, l'eau est considérée "bien exposée" dans le seconde image et enfin on obtient la meilleure exposition pour les rochers dans l'image au temps d'exposition le plus grand. Cela correspond tout à fait à notre perception de ces images.
- Enfin, la saturation mesure si les couleurs sont vives, par exemple, un bleu ou un rouge prononcé. Dans l'exemple, le soleil est bien saturé uniquement dans la première image car c'est là qu'il a de la couleur (dans les autres images il est en fait tout blanc).

On peut ensuite combiner ces trois mesures sous forme de produit. Cela permet "d'exiger" que les trois qualités soient présentes pour que le pixel soit sélectionné. Nous appelons ce produit la fonction de poids, et elle est définie comme suit :

$$W_{i,j,k} = (C_{i,j,k})^{w_C} (S_{i,j,k})^{w_S} (E_{i,j,k})^{w_E} \quad (1)$$

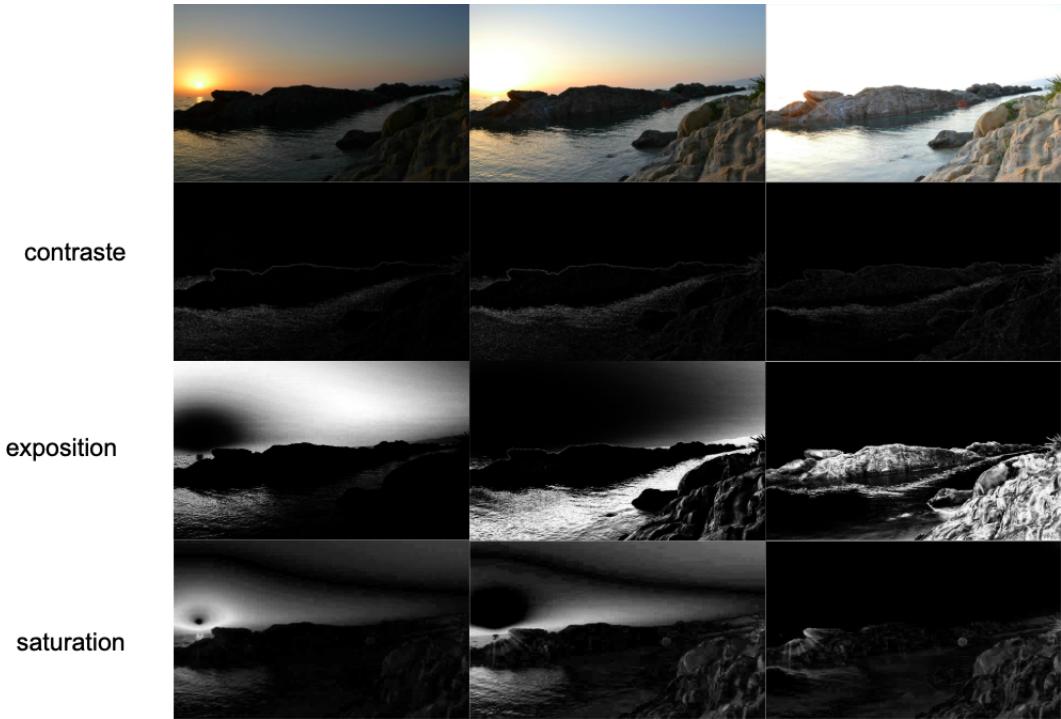


Figure 1: Matrices de contraste, d'exposition et de saturation pour trois photographies de la même scène avec des temps d'exposition croissants. Crédits : Nikolaos Mitianoudis [2]

où C , S et E représentent respectivement la mesure de contraste, de saturation et d'exposition, w_C , w_S , et w_E les poids associés à ces mesures et l'indice i, j, k fait référence au pixel en position (i, j) dans la k -ième image de la séquence d'exposition. Nous remarquons qu'il est possible de faire varier l'importance de chaque mesure de qualité dans la fonction de poids en faisant varier les paramètres w_C , w_S , et w_E .

Nous avons donc une fonction de poids pour chaque temps d'exposition. Pour fusionner les images il faut également normaliser la matrice de poids pour obtenir un résultat cohérent. Ainsi, la fonction de poids normalisée est définie par:

$$\hat{W}_{i,j,k} = \left(\sum_{k'=1}^N W_{i,j,k'} \right)^{-1} W_{i,j,k} \quad (2)$$

La figure 2 montre les fonctions de poids normalisées pour les différentes photos de la même scène. On peut bien voir les trois zones qui sont distinguées et qui vont être utilisées pour reconstruire la photo.



Figure 2: Fonctions de poids normalisées pour la séquence multi-exposition de la figure 1

1.2 MÉTHODES DE FUSION D'IMAGES

FUSION NAÏVE



Figure 3: Fusion par moyenne naïve pour la séquence multi-exposition de la figure 1

La figure 3, présente le résultat si on fait une moyenne naïve pixel à pixel des trois images de la séquence multi-exposition (sans calculer de fonction de poids) et la figure 4 présente le résultat de la moyenne pixel à pixel pondérée par les fonctions de poids normalisées en chaque pixel. Pour clarifier, la figure 4 a été obtenue en calculant, en chaque pixel (i, j) , la moyenne:

$$R_{i,j} = \sum_{k=1}^N \hat{W}_{i,j,k} I_{i,j,k} \quad (3)$$

où $I_{i,j,k}$ représente le pixel (i, j) de la k -ième image de la séquence multi-exposition. Le premier constat est que la moyenne naïve est satisfaisante et permet de visualiser les détails dans chaque zone. Ensuite, le résultat de la fusion avec la fonction de poids est catastrophique. Il semble que la fonction de poids oscille trop vite et donne cet aspect "pixelisé" à l'image.

Mertens *et al.*[4] proposent alors de lisser les fonctions de poids avec un filtre gaussien afin de rendre le résultat plus uniforme. Le résultat est présenté dans la figure 5. À première vue



Figure 4: Fusion par moyenne pondérée par les fonctions de poids pour la séquence multi-exposition de la figure 1



Figure 5: Fusion par moyenne pondérée par les fonctions de poids pour la séquence multi-exposition de la figure 1 sur laquelle a été appliqué un filtre gaussien au préalable

le résultat est meilleur, les couleurs sont vives et l'on a gardé du détail dans chaque zone. Cependant, on peut constater des effets étranges de "halo" au niveau des frontières entre les zones. Les frontières abruptes ont été lissées par le filtre gaussien et donc des zones débordent les unes sur les autres alors que l'on souhaiterait avoir une coupure nette. C'est pourquoi Mertens *et al.*[4] proposent de recourir à une méthode multi-échelle.

FUSION MUTLI-ÉCHELLE

La "base" de la pyramide gaussienne [1] est l'image de départ, et les étages suivants sont cette image où la résolution a été divisée par deux successivement. La pyramide laplacienne est la différence entre les étages de la pyramide gaussienne. Formellement, en notant $L(A)^l$ (resp. $G(A)^l$) le l -ième niveau de la pyramide laplacienne (resp. gaussienne) de l'image A , on a :

$$G\{A\}_{i,j}^l = \begin{cases} A_{i,j} & \text{si } l = 0 \\ \text{Downsample}(G\{A\}_{i,j}^{l-1}) & \text{si } l > 0 \end{cases} \quad (4)$$

$$L\{A\}_{i,j}^l = \begin{cases} G\{A\}_{i,j}^l - \text{Upsample}(G\{A\}_{i,j}^{l+1}) & \text{si } l < l_{\max} \\ G\{A\}_{i,j}^l & \text{si } l = l_{\max} \end{cases} \quad (5)$$

où l_{\max} représente le nombre d'étage des pyramides laplaciennes et gaussiennes. Ici, la fonction Downsample (resp. Upsample) divise (resp. multiplie) la taille de la matrice donnée en entrée par deux.

Ainsi définie, la somme des étages de la pyramide laplacienne reconstruit l'image de départ (annulation de termes en cascade). La pyramide gaussienne est donc l'image de plus en plus grossière, et la pyramide laplacienne représente les "frontières" des plus détaillées aux plus grossières (derniers étages de cette pyramide).

Mertens *et al.* [4] proposent donc de calculer les pyramides laplaciennes des images ainsi que les pyramides gaussiennes des fonctions de poids. On multiplie terme à terme les pyramides gaussiennes et laplaciennes avant de reconstruire l'image de départ. Ainsi, avec les notations précédentes, la pyramide laplacienne de l'image fusionnée R est obtenue en calculant :

$$L(R)_{i,j}^l = \sum_{k=1}^N G(\hat{W})_{i,j,k}^l L(I)_{i,j,k}^l \quad (6)$$

où $I_{i,j,k}$ représente le pixel (i, j) de la k -ième image de la séquence d'exposition. Pour obtenir l'image fusionnée, il suffit alors de sommer les niveaux de la pyramide laplacienne de R définie ci-dessus. On obtient ainsi une fusion multi-échelle.



Figure 6: Fusion multi-échelle par pyramide laplacienne en ne tenant compte que d'une seule mesure de qualité

Dans la figure 6, on propose les images fusionnées en ne prenant en compte qu'une seule mesure de qualité. Cela signifie que le triplet de paramètres (w_C, w_E, w_S) est à valeurs dans l'ensemble $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ dans l'équation (1). On peut déjà constater que les trois images sont de bonne qualité, il n'y a pas d'artefacts dans l'image (comme la pixelisation, ou bien les halos). Les frontières sont nettes et chaque zone de l'image est détaillée. On peut remarquer que la photo obtenue avec les poids de contraste est proche de la fusion par moyenne naïve (cf. figure 3). On peut ensuite voir que les deux autres mesures permettent de conserver les couleurs du soleil présentes dans l'image la plus sombre.



Figure 7: Fusion multi-échelle par pyramide laplacienne en tenant compte des trois mesures de qualité.

Dans la figure 7, on présente le résultat en utilisant les trois mesures de qualité (c'est-à-dire en prenant $(w_C, w_E, w_S) = (1, 1, 1)$). Le résultat est tout à fait satisfaisant et on ne saurait prendre une telle photo en une seule prise avec les appareils photos actuels.

2 COMMENTAIRES SUR LA MÉTHODE

Il est ici important de faire deux remarques sur la méthode de Mertens *et al.* [4]:

- l'article ne commente pas le fait que les valeurs résultantes ne sont pas nécessairement comprises entre 0 et 255. Cela est problématique car l'un des avantages vanté par les auteurs est de ne pas avoir besoin de recourir à du "tone mapping". Nous avons constaté que les pixels peuvent prendre des valeurs entre -50 et 300 et que le nombre de pixels en dehors des bornes est généralement entre 1% et 3%. Il se trouve que "clipper" les images fonctionne très bien et ne présente pas de problèmes visuels.
- dans l'article, la pyramide gaussienne est construite en divisant la résolution par deux à chaque étage (4 pixels sont remplacés par 1 pixel qui vaut leur moyenne). Au moment d'écraser la pyramide pour la reconstruction de l'image, il est nécessaire de remettre les images à la bonne taille pour les sommer. Nous avons préféré une autre méthode. Notre pyramide gaussienne est composée d'images de même taille, nous appliquons simplement des filtres gaussiens successifs. Cela s'est avéré être beaucoup plus facile à implémenter et donne des résultats satisfaisants. Par ailleurs, il ne faut pas gérer la divisibilité par 2 des tailles d'images et un filtre gaussien est plus précis, dans le sens où l'influence des pixels voisins dépend de leur distance. De plus dans le dernier étage de la pyramide gaussienne, on garde une résolution égale à celle de départ, donc deux pixels voisins peuvent avoir des valeurs différentes. Il serait intéressant dans une potentielle continuation de ce travail de comparer la qualité des résultats entre notre implémentation et celle de Mertens *et al.* [4].
- La table 1 présente une comparaison du temps d'exécution entre notre méthode et celle de Mertens *et al.* [4]. On constate que bien que notre implémentation semble un peu plus longue que celle d'origine, les temps d'exécution sont tout de même du même ordre de grandeur. Cet écart peut certainement s'expliquer par la différence soulignée au point précédent.

$w \times h \times N$	Mertens <i>et al.</i> [4]	Notre implémentation
$864 \times 576 \times 3$	1.6	2.3
$1227 \times 818 \times 3$	3.2	4.5
$1728 \times 1152 \times 3$	6.2	10.4
$864 \times 576 \times 7$	3.0	4.6
$1227 \times 818 \times 7$	6.1	8.9
$1728 \times 1152 \times 7$	12.0	18.1

Table 1: Comparaison du temps d'exécution entre notre implémentation et celle de Mertens *et al.* [4]. Les temps sont affiché en secondes. w, h et N désignent respectivement la largeur, la hauteur et le nombre d'images utilisées pour la fusion.

3 RÉSULTATS ET DISCUSSION

3.1 IMPACT DE LA TAILLE DE LA PYRAMIDE LAPLACIENNE



Figure 8: Fusion multi-échelle par pyramide laplacienne pour différentes valeurs du nombre d'étage de la pyramide n .



Figure 9: Fusion multi-échelle par pyramide laplacienne avec une pyramide à deux étages ($n=2$) et en utilisant uniquement la mesure d'exposition ($(w_C, w_E, w_S) = (0, 1, 0)$)

L'un des paramètres à optimiser dans cette méthode est le nombre d'étages que l'on met dans la pyramide. Nous présentons en figure 8 les résultats pour différents valeurs du nombre d'étages, que l'on note n . On peut voir qu'un halo apparaît dans le ciel pour $n = 2$ et $n = 3$. Cela correspond à une zone moins saturée (une bande blanche entre le passage du rouge au bleu). Il faut noter que ce halo n'existe pas dans les images originales et son apparition est

due à la mesure de saturation. Augmenter le nombre d'étages dans la pyramide tend à faire disparaître cet artefact. On présente en figure 9 le résultat avec $n = 2$ en utilisant uniquement la mesure d'exposition, on constate que le halo disparaît.

Pour la suite des résultats, on prendra $n = 4$, car c'est la première valeur de n qui corrige le problème du halo.

3.2 AUTRES RÉSULTATS

RÉSULTATS AVEC NOS PROPRES IMAGES



Figure 10: Séquence de sept photographies de la même scène avec des temps d'exposition croissants.

On s'intéresse maintenant à un autre jeu de photographies (que nous avons pris nous-mêmes). Comme on peut le voir sur la figure 10. Il s'agit de sept photographies d'une même scène avec des temps d'expositions différents. Nous avons choisi de prendre nos propres photos pour pouvoir tester plusieurs aspects: des objets colorés sont posés sur le bureau et on souhaite garder leurs couleurs vives. Nous avons disposé un pull rayé dans la scène ainsi qu'un cahier avec du texte manuscrit. On souhaite que le résultat de la fusion permette de distinguer les rayures du pulls ainsi que le texte sur le cahier.

Nous comparons quatre résultats obtenus avec des méthodes différentes:

- Notre implémentation
- La fusion par moyenne naïve décrite à la section 1.2
- Le résultat obtenu avec le logiciel Photomatix [3]
- Le résultat obtenu avec le logiciel inclus dans l'appareil photo de l'ordinateur utilisé pour capturer les images de la figure 10

Les résultats pour ces quatre méthodes sont présentés dans la figure 11.

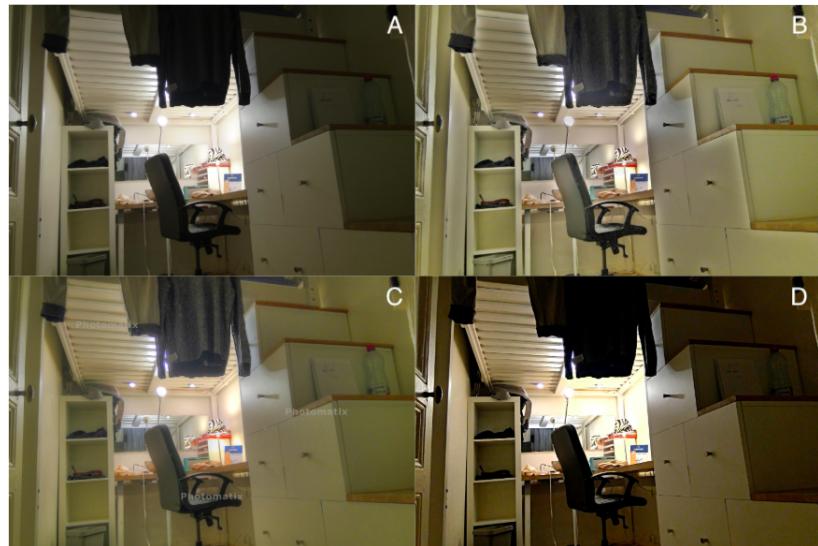


Figure 11: Comparaison de différentes méthodes de fusion d'expositions pour les images de la figure 10. L'image A a été obtenue par moyenne naïve, l'image B par notre implémentation, l'image C avec le logiciel Photomatix [3] et l'image D avec le logiciel intégré de Microsoft.

Le logiciel de Microsoft ne permet pas de voir les détails dans les zones sombres, cependant l'image a un aspect naturel et les contrastes de lumière sont intacts. La fusion par moyenne naïve ne permet pas de distinguer les détails et ne conserve pas les couleurs vives. Enfin notre méthode conserve les couleurs vives et les détails. Cependant, on perd les effets de lumière et l'image ne semble pas naturelle.

IMPACT DU NOMBRE D'IMAGES UTILISÉES



Figure 12: De gauche à droite : fusion multi-échelle par pyramide laplacienne avec 2, 3 et 7 images dans la séquence multi-exposition. Les images ont été choisies de façon à garder les temps d'expositions extrêmes.

Dans cette section, nous analysons l'impact du nombre d'images utilisées pour la fusion. La figure 12 présente les résultats pour notre méthode avec deux, trois puis l'ensemble des

sept images de la séquence multi-exposition. Afin d'avoir la plage d'exposition la plus grande possible, nous choisissons les photographies aux temps d'expositions extrêmes (le plus long et le plus court) pour les exemples avec deux et trois photographies. A l'inverse, dans la figure 13, on fusionne deux photographies ayant des temps d'exposition plus proches.



Figure 13: fusion multi-échelle par pyramide laplacienne avec deux photographies capturées avec des temps d'exposition proches.

On constate alors deux choses :

- Avec peu d'images et des temps d'exposition trop éloignés, la fusion ne se passe pas bien et des artefacts apparaissent (les ombres sur les tiroirs par exemple).
- Avec des temps d'exposition plus proches, on perd du détail mais le résultat est plus naturel et ne contient pas d'artefacts (notamment le résultat apparaît plus naturel qu'en utilisant l'ensemble des sept photos de la figure 10).

Il faut donc faire attention à bien choisir le temps d'exposition et le nombre de photographies à utiliser dans la fusion pour que cela convienne dans le cas le plus général possible.

RÉSULTATS POUR LA FUSION D'IMAGES AVEC ET SANS FLASH

Dans leur article, Mertens *et al.* [4] expliquent que leur méthode est suffisamment générale pour qu'elle fonctionne pour fusionner des images avec et sans flash. Nous avons donc essayé de reproduire les résultats qu'ils ont obtenus avec des images provenant du travail de Petschnigg *et al.* [5].

Dans la figure 14, nous comparons une fusion par moyenne naïve et la méthode de fusion multi-échelle. On constate que le résultat n'est pas satisfaisant. L'effet "yeux rouges" est



Figure 14: De gauche à droite : image d'origine sans flash, image d'origine avec flash, fusion multi-échelle par pyramide laplacienne, fusion par moyenne naïve

conservé, et l'aspect de l'image est étrange, comme si l'image était une image de synthèse. On remarque cet effet à la frontière entre le haut blanc et la veste noire. De plus, le visage apparaît pixelisé, notamment aux endroits les plus texturés (par exemple, les tâches de rousseur). Nous déconseillons donc l'utilisation de cette méthode pour la fusion avec et sans flash et renvoyons vers le travail de Petschnigg *et al.* [5] pour effectuer cette tâche.

REFERENCES

- [1] P. Burt and E. Adelson. “The Laplacian Pyramid as a Compact Image Code”. In: *IEEE Transactions on Communications* 31.4 (Apr. 1983), pp. 532–540. ISSN: 0090-6778. DOI: 10.1109/TCOM.1983.1095851.
- [2] *HDR Exposure fusion dataset - Nikolaos Mitianoudis*. <http://utopia.duth.gr/nmitiano/download.html>. Dernière connexion: 17/01/2019.
- [3] *Logiciel Photomatix*. <https://www.hdrsoft.com/>. Dernière connexion: 16/01/2019.
- [4] Tom Mertens, Jan Kautz, and Frank Van Reeth. “Exposure Fusion”. In: *Pacific Graphics* (Oct. 2007), pp. 382–390. DOI: 10.1109/PG.2007.17.
- [5] Georg Petschnigg et al. “Digital Photography with Flash and No-flash Image Pairs”. In: *ACM Trans. Graph.* 23.3 (Aug. 2004), pp. 664–672. ISSN: 0730-0301. DOI: 10.1145/1015706.1015777. URL: <http://doi.acm.org/10.1145/1015706.1015777>.