

Critère optimal pour la création de séquences d'images panoramiques

Henri NICOLAS

IRISA/INRIA, Projet TEMICS
Campus universitaire de Beaulieu
35042 RENNES CEDEX, FRANCE
hnicolas@irisa.fr

Résumé – Cet article propose une nouvelle méthode permettant la création d'une séquence d'images panoramiques à partir d'une séquence vidéo originale, où chaque image panoramique doit représenter au mieux le fond de la scène à l'instant correspondant. Dans la mesure où les images originales se recouvrent et où les valeurs des pixels pour un même point ne sont généralement pas identiques, il est nécessaire de définir un critère permettant de gérer efficacement ce recouvrement. En pratique, chaque nouvelle image est calculée en utilisant une somme pondérée entre l'image panoramique à l'instant précédent, et l'image originale courante recalée dans le repère du panoramique. Dans ce contexte, cet article présente un critère permettant d'optimiser le choix du coefficient de pondération en fonction de critères liés à la résolution des images panoramiques et au délai temporel associé.

Abstract – This paper proposes a new method for the creation of a sequence of mosaic images from an original video sequence. Each mosaic image is built in order to represent as close as possible the scene background for the corresponding time instant. The overlapped areas existing between the original images require the design of a criterion to determine the pixel values of the mosaic image. This is due to the fact that, for the same point, each available original image have not the same pixel values. In practice, each new mosaic image is calculated as a weighed sum between the previous mosaic and the current original image. In this context, this paper presents an analytic criterion which permits to optimize the choice of the blending coefficient based on the resolution and on the temporal delay of the computed mosaic images.

1 Introduction

Cet article présente une nouvelle méthode pour la création de séquences d'images panoramiques. Classiquement, une image panoramique représente l'arrière plan de la scène visualisé tout au long de la séquence d'images. Cette représentation suppose généralement que le fond de la scène n'est pas modifié au cours du temps, et donc, qu'une seule image panoramique peut le représenter tout au long de la séquence. En pratique, cette hypothèse n'est pas vérifiée lorsqu'il se produit des variations d'illumination, si l'arrière plan n'est pas totalement fixe, ou encore si la position de certaines images dans le référentiel du panoramique n'a pas été correctement estimée par l'algorithme d'estimation du mouvement. A l'inverse, une représentation dynamique des images panoramiques permet de s'adapter au mieux aux variations de la scène [1]. Pour chaque image originale, une image panoramique est alors créée en utilisant l'image panoramique précédente et l'image originale courante, repositionnée dans le référentiel du panorama. Les images panoramiques ainsi créées contiennent l'ensemble du fond de la scène visible depuis la première image et seront appelés 'Images Panoramiques Complètes (IPC)'. Dans ce contexte, cet article propose une nouvelle méthode de création d'images panoramiques dynamiques. De plus, à partir de chaque image IPC créée une image panoramique partielle (IPP), dont le point de vue peut varier temporellement, est extraite. En pratique, ces images cor-

respondent, par exemple, à des dimensions d'images plus larges, à une variation artificielle de focale ou encore à une modification du point de vue. Ce type de traitements est important dans de nombreuses applications liées à la post-production vidéo [2][3]. La nouvelle approche proposée contient alors les différentes phases algorithmiques suivantes:

- **Segmentation basée objet:** Cette phase de segmentation permet de sélectionner, dans les images originales, l'objet vidéo correspondant à l'arrière plan fixe de la scène. En pratique, la méthode proposée dans [4] a été utilisée.
- **Estimation des paramètres de positionnement.** Les paramètres de positionnement (un vecteur de translation et un paramètre de divergence k) d'une image originale permettent de localiser cette image dans le référentiel de l'image panoramique. Ils sont calculés ici en utilisant l'algorithme proposé dans [2], dans laquelle la position de chaque image dans le référentiel du panoramique est directement estimé (pas d'estimation explicite du mouvement caméra).
- **Description d'une séquence d'images IPP.** Le point de vue des images IPP peut varier au cours du temps, et est décrit par un vecteur de translation, un facteur d'échelle et les dimensions des images [2].

• **Rafraîchissement d'une image IPC.** Une méthode doit être définie pour le calcul de l'image panoramique dans les régions de recouvrement entre la nouvelle image, repositionnée dans le référentiel du panoramique, et l'IPC déjà existante. En fonction de la méthode retenue, de nombreuses images différentes peuvent être obtenues. Par exemple, on peut systématiquement conserver l'information déjà existante (méthode souvent utilisée pour la création d'une image panoramique statique), ou, à l'inverse, systématiquement remplacer l'information ancienne par la nouvelle, ou encore calculer une somme pondérée entre les deux images. Dans cet article, un nouveau critère permettant de déterminer la valeur optimal du coefficient de pondération en fonction de contraintes liées à la qualité des images, au délai temporel et à la résolution des images originales est proposé.

• **Calcul d'une image IPP.** Chaque image IPC représente une portion de l'image panoramique complète. Elle est donc simplement extraite de l'image IPP courante en utilisant les paramètres définis ci-dessus.

2 Critère d'estimation

Si M_t et I_t représentent l'image panoramique totale et l'image originale, compensée dans le référentiel du panoramique, à l'instant t , M_t est calculé en utilisant l'équation suivante:

$$M_t(p) = \alpha_t I_t(p) + (1 - \alpha_t) M_{t-1}(p), \quad (1)$$

où α_t , le coefficient de pondération, est déterminé en fonction des contraintes suivantes:

Prise en compte de l'image courante. Il est souhaitable d'avoir une image panoramique la plus proche possible de l'image originale courante. Pour cela, l'Erreur Quadratique Moyenne correspondante est introduite dans le critère d'estimation.

Minimisation des variations temporelles. Des modifications trop rapides de la séquence créées peuvent entraîner des défauts visuels gênants, notamment parce que seulement une partie de l'image panoramique est rafraîchie à chaque instant. Ce problème est pris en compte en limitant les modifications temporelles des images panoramiques par l'introduction de l'EQM entre deux images panoramiques successives.

On obtient alors le critère suivant:

$$C(\alpha_t) = \sum_{p \in \Omega_t} [(M_t(p) - I_t(p))^2 + (M_t(p) - M_{t-1}(p))^2].$$

où Ω_t est le support d'estimation correspondant à l'intersection entre M_{t-1} et I_t .

Minimisation du délai temporel. Une image panoramique n'est pas homogène temporellement car elle est calculée à partir de plusieurs images originales. Un délai temporel $T_t(p)$ peut alors être associé à chaque pixel à

l'instant t . Dans le but de minimiser ce délai, l'équation précédente est modifiée de la manière suivante:

$$C(\alpha_t) = \sum_{p \in \Omega_t} [(M_t(p) - I_t(p))^2 + \frac{(M_t(p) - M_{t-1}(p))^2}{S_t T_{t-1}(p)}]$$

S_t est le coefficient de pondération de la contrainte temporelle. $T_t(p)$ est rafraîchi pour chaque image comme suit:

$$T_t(p) = (1 - \alpha_t)(T_{t-1}(p) + 1).$$

Maximisation de la résolution. La dernière contrainte est liée à la maximisation de la résolution des images panoramiques. En effet, si la distance camera-scène et/ou la focale varie, la résolution des images originales est modifiée. Dans ce but, le rapport $K_r(p) = \frac{K_t(p)}{K_{t-1}(p)}$, représentant le gain en résolution entre deux images panoramiques successives, est utilisé pour tenir compte de la résolution. $K_t(p)$ représente la résolution au pixel p de M_t , et peut être calculé comme suit: $K_t(p) = \alpha_t K_i + (1 - \alpha_t) K_{t-1}(p)$, où K_i , la résolution de l'image I_t est directement obtenue en utilisant le paramètre k tel que $K_i = k + 1$. On obtient alors:

$$K_r(p) = \alpha_t \frac{\Delta K(p)}{K_{t-1}(p)} + 1 = \alpha_t K_d(p) + 1. \quad (2)$$

Avec $\Delta K(p) = K_i - K_{t-1}(p)$. Si $\Delta K(p) > 0$ (resp. $\Delta K(p) < 0$), cela signifie que la résolution au pixel p de la nouvelle image est plus élevée (resp. plus faible) que pour M_{t-1} , de ce fait la minimisation du terme $(M_t(p) - I_t(p))^2$ (resp. $(M_t(p) - M_{t-1}(p))^2$) doit être favorisée. Cette dernière contrainte donne alors le critère final:

$$C(\alpha_t) = \sum_{p \in \Omega_t} (M_t(p) - I_t(p))^2 [1 + S_r S(p) K_d(p) \alpha_t] + \sum_{p \in \Omega_t} \frac{(M_t(p) - M_{t-1}(p))^2 [1 - \alpha_t K_d(p) S_r (1 - S(p))]}{S_t T_{t-1}(p)} \quad (3)$$

où $S(p) = 1$ si $\Delta K(p) \geq 0$ et $S(p) = 0$ sinon. S_r est le coefficient de pondération de la contrainte de résolution.

2.1 Minimisation analytique du critère

Le critère C peut être minimisé en résolvant l'équation: $\frac{\partial C(\alpha_t)}{\partial \alpha_t} = 0$. Les équations (1) et (3) donnent alors:

$$\alpha_t^2 (3\beta_2 + \beta_4) + \alpha_t (\beta_1 + \beta_3 - 4\beta_2) - \beta_1 + \beta_2 = 0, \quad (4)$$

avec

$$\beta_1 = 2 \sum_{p \in \Omega_t} (M_{t-1}(p) - I_t(p))^2$$

$$\beta_2 = S_r \sum_{p \in \Omega_t} S(p) K_d(p) (M_{t-1}(p) - I_t(p))^2$$

$$\beta_3 = 2 \sum_{p \in \Omega_t} \frac{(M_{t-1}(p) - I_t(p))^2}{S_t T_{t-1}(p) + 1}$$

$$\beta_4 = -3S_r \sum_{p \in \Omega_t} (1 - S(p)) (M_{t-1}(p) - I_t(p))^2 \frac{K_d(p)}{S_t T_{t-1}(p)}$$

Si $\Delta = (\beta_1 + \beta_3 - 4\beta_2)^2 - 4(3\beta_2 + \beta_4)(\beta_2 - \beta_1) \geq 0$, cette équation du second degré admet les solutions:

$$\alpha_{opt} = \frac{4\beta_2 - \beta_1 - \beta_3 \pm \sqrt{\Delta}}{2(3\beta_2 + \beta_4)} \quad (5)$$

Par construction, $0 \leq \alpha_t \leq 1$. Si les deux solutions sont dans cet intervalle, la solution retenue est celle minimisant le critère. Sinon, les valeurs obtenues avec $\alpha_t = 0$ et 1, et avec la solution dans l'intervalle $[0, 1]$ (si elle existe) sont comparées. α_t est appliqué uniquement sur Ω_t . Pour les autres pixels définis dans I_t , nous avons $M_t(p) = I_t(p)$. La contribution de I_t (en %) dans la création de M_t peut alors s'exprimer de la manière suivante:

$$\bar{\alpha}_t = 100 * \frac{Card[\Omega_t]\alpha_t + Card[I_t] - Card[\Omega_t]}{Card[M_t]} [\%].$$

3 Expérimentations et conclusions

Les expérimentations ont été réalisées sur la séquence "Stefan", laquelle contient un large mouvement de caméra. Les mêmes points de vue de la scène sont visibles à différent instants et avec des résolutions différentes ce qui permet de tester la robustesse de l'algorithme. La figure 3 montre plusieurs exemples d'images IPC. Il est à noter que les mouvements d'aller-retours de la caméra ne dégradent pas la qualité des images panoramiques obtenues. La figure 1 donne des exemples d'images panoramiques partielles pour lesquelles le point de vue et la taille des images ont été modifiés. La qualité des séquences d'images obtenues est bonne et sans artefact temporel gênant. Les courbes de la figure 2 présentent les variations temporelles du paramètre α ainsi que la résolution et le délai temporel associé aux images panoramiques créées. Des augmentations significatives de résolution ou une baisse du délai temporel peuvent ainsi être obtenues en modifiant les valeurs des coefficients S_r et S_d , qui sont alors choisis en fonction de l'application visée.

La méthode présentée dans cet article permet de calculer efficacement les images panoramiques en tenant compte de diverses contraintes, liées à la résolution, à la qualité et au délai temporel, et pouvant être pondérées selon l'application visée. De plus, la minimisation analytique du critère permet d'avoir un temps de calcul très faible. L'utilisation d'un tel critère est dès lors tout à fait prometteur dans le cadre des applications liées à la manipulation d'objets vidéo ou à la composition de scène, pour lesquelles des images panoramiques doivent être utilisées.

Références

- [1] Z. Sun and M. Tekalp. Trifocal motion modeling for object-based video compression and manipulation. *IEEE Transactions on circuits and systems for video technology*, Vol. 8, No. 5, pp. 667-685, 1998.
- [2] H. Nicolas. Mosaic representation and video object manipulations for post-production applications. In *Proceedings of ICIP, Vol II, pp. 451-455, Chicago*, September 1998.
- [3] R. Kumar M. Irani, P. Anandan J. Bergen and S. Hsu. Efficient representations of video sequences and their applications. *Signal Processing: Image communications*, Vol. 8, pp. 327-351, 1996.
- [4] M. Lorei A. Smolis and T. Sikora. Adaptive kalman filtering for prediction and global motion parameter tracking of segments of video. *PCS, Mars* 1997.



FIG. 1: Images panoramiques partielles 150 et 237

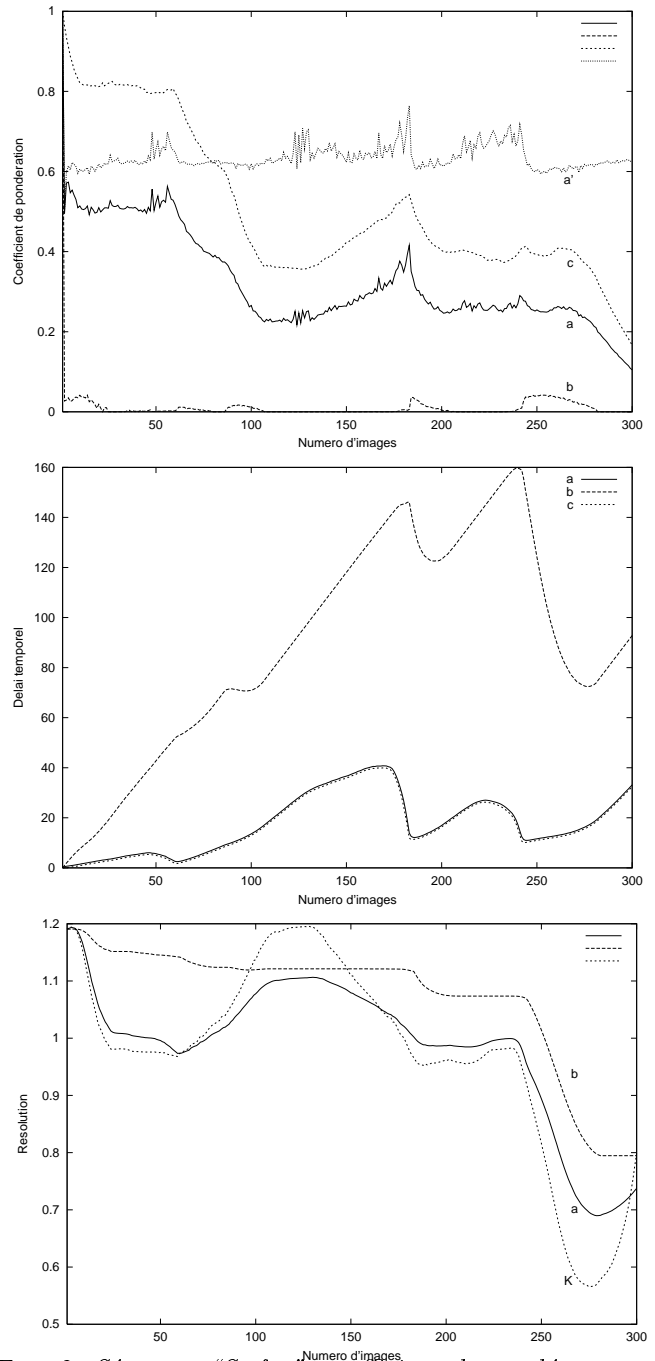


FIG. 2: Séquence "Stefan": coefficient de pondération α_t (a') et coefficient $\bar{\alpha}_t$ (a) obtenus avec: $S_r = S_t = 1$. Coefficient $\bar{\alpha}_t$ avec $S_r = +\infty$ et $S_t = 1$ (b), et $\alpha_t = 1$ (c). (K): Résolution K_i des images originales.

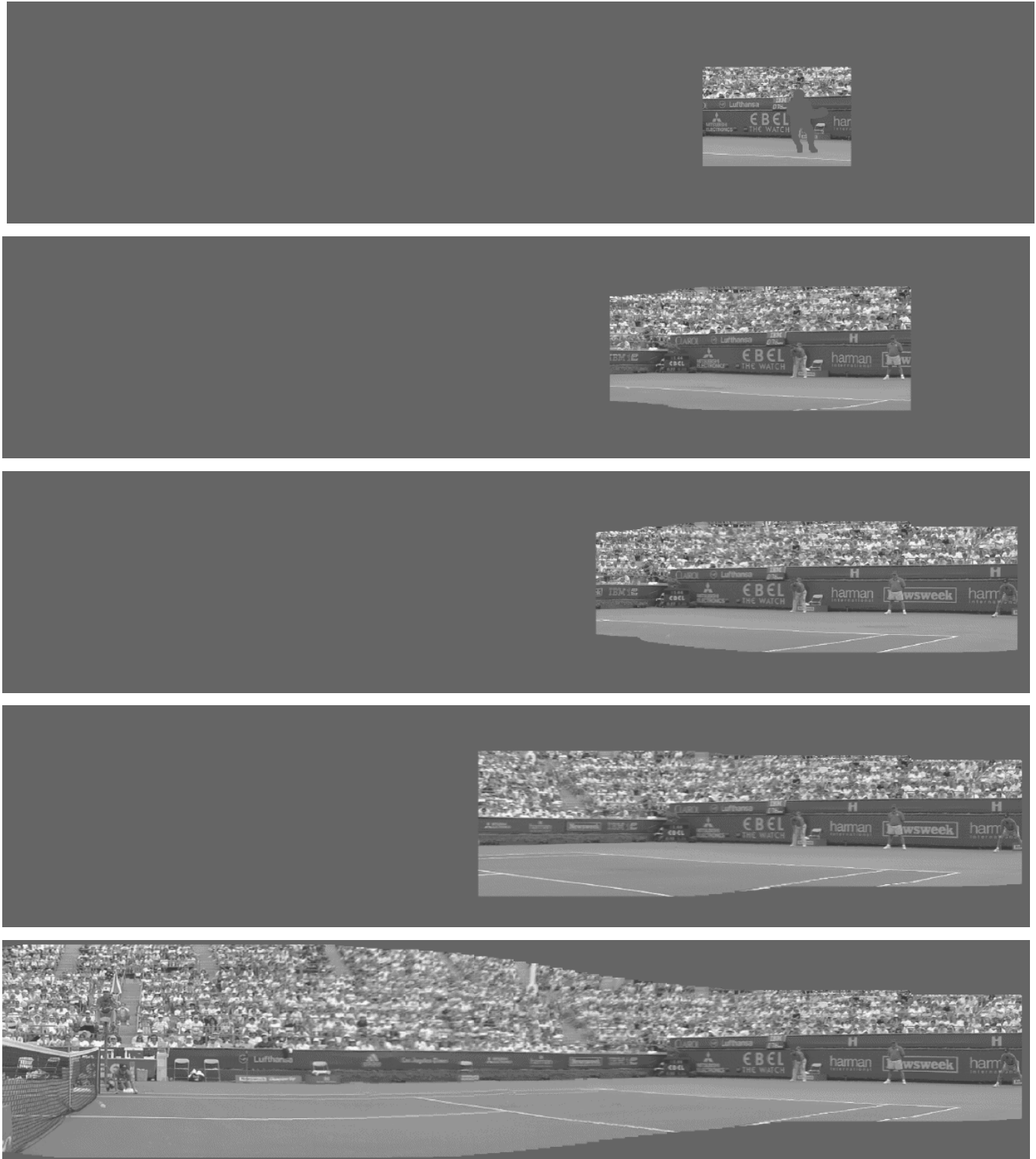


FIG. 3: Séquence "Stefan". Images panoramiques 1, 100, 200, 250 et 300 ($S_r = 1$ et $S_t = 1$).