

Résumé

Le travail de thèse développé dans ce mémoire porte sur le traitement des séquences vidéos. Ceci inclut la compression pour une grande partie de la thèse, mais également le filtrage spatio-temporel et la segmentation vidéo. Un des outils d'analyse récurrent dans chacune de ces applications est la mesure du mouvement, c'est-à-dire la description de la cohérence temporelle d'une séquence vidéo.

Un système de compression vidéo comprend généralement les trois composantes suivantes :

- estimation du mouvement,
- transformations temporelle et spatiale,
- quantification et codage des coefficients.

Nous nous intéressons successivement à chacune de ces composantes.

Dans un premier temps, nous essayons d'adapter une méthode d'estimation par projection du flot optique sur une base d'ondelettes à valeur complexe à la compression vidéo. Le champ obtenu étant dense et sensible au bruit, nous introduisons dans la résolution un élément de régularisation afin de diminuer son coût de codage. En dépit d'une nette amélioration apportée par nos modifications, la technique d'estimation par projection sur une base d'ondelettes n'est pas compétitive face au block-matching qui constitue l'algorithme de référence pour la compression vidéo. Cela illustre bien le fait que le choix et la conception d'une méthode sont étroitement liés à l'usage qui en est fait : des méthodes d'estimation de mouvement optimisées pour des applications aussi diverses que la compression vidéo, le filtrage et la segmentation, ou encore l'analyse de scènes 3D ont peu de chances d'être les mêmes. Dans ces expériences, nous avons utilisé une méthode visant à satisfaire un critère qui est équivalent à un critère d'appariement optimal pour la compression sous des conditions théoriques qui ne sont en général pas vérifiées. Une telle approche est donc visiblement sous-optimale.

Forts de cette observation, nous avons ensuite développé une méthode de mesure de mouvement qui repose sur une paramétrisation du champ bilinéaire par morceaux, et qui minimise cette fois directement l'erreur quadratique moyenne qui est notre critère d'évaluation. Nous montrons qu'il est possible d'obtenir de bons résultats quand les paramètres du champ sont épars.

Un codage vidéo avec une prédiction temporelle suppose de coder des données aussi hétérogènes que des champs de mouvement ou des images d'erreur. Nous avons travaillé sur l'allocation de débit entre images d'erreur et de manière moins approfondie entre image d'erreur et champ de mouvement. Nous avons adapté un modèle de planification

de débit introduit par Mallat et Falzon qui a été initialement conçu pour des images statiques et qui est actuellement utilisé pour la compression au flot d'images satellitales. Cette approche se révèle meilleure que des approches plus classiquement utilisées en compression vidéo.

Pour pouvoir effectuer un codage par transformée du champ et des images d'erreur, nous avons cherché à concevoir de nouvelles transformations en sous-bandes non-linéaires. Pour cela, nous avons utilisé le schéma de *lifting*, qui garantit l'inversibilité formelle des transformations qu'il peut réaliser, que celles-ci soient linéaires ou non. Nous avons construit deux nouvelles décompositions non-linéaires.

La première vise à réduire un artéfact communément appelé effet de Gibbs. Cette première décomposition consiste à utiliser un prédicteur de Deslauriers-Dubuc modifié de manière à réduire ces artéfacts. La modification introduite permet effectivement de réduire les oscillations autour de discontinuités en échange d'un surcoût modeste en terme d'efficacité de représentation dans les sections régulières du signal représenté. La formulation évite le mécanisme de transition d'un filtre à l'autre relativement habituel dans ce genre d'approche en recourant à des opérateurs continus de type min et max, qui permettent de garantir la continuité de la transformation et donc sa stabilité après quantification.

L'autre se propose d'améliorer la décomposition en ondelettes du champ de mouvement en utilisant l'information qu'apporte chacune de ses composantes sur l'autre. En effet, l'intuition nous incite à penser que les discontinuités sont présentes au même endroit dans chacune des composantes du mouvement. Nous nous servons de cette co-occurrence des discontinuités pour choisir le filtre de prédiction.

Dans les deux cas, les méthodes mises au point donnent des résultats positifs sur des signaux de synthèse mais perdent en efficacité sur des signaux réels. Une des grandes difficultés est de mettre au point un étage de mise-à-jour dans le schéma de *lifting*. Par ailleurs, le schéma linéaire le plus efficace est un schéma à 4 étages pour lequel il est difficile de concevoir un concurrent non-linéaire dans la mesure où ses propriétés sont difficilement lisibles sur les étages individuels du schéma de *lifting*.

Nous avons enfin transposé des idées rencontrées en compression vidéo pour définir des opérations de filtrage morphologique vidéo intégrant la mesure du mouvement, utilisant des éléments structurants qui suivent le mouvement. L'application de ces idées donne des résultats probants en filtrage et en segmentation, en particulier grâce à une forte cohérence spatio-temporelle introduite dans les voisinages : cette approche donne des résultats de segmentation plus stables puisqu'elle impose une cohérence temporelle beaucoup plus forte aux frontières des régions que les méthodes itératives en temps. Nous discutons ensuite des possibilités d'utilisation de champs de mouvement à précision sous-pixellique.

Mots-clés Compression vidéo, estimation du mouvement, quantification, allocation de débit, transformation en ondelettes, schéma de *lifting*, morphologie mathématique, segmentation spatio-temporelle.