

Réduction du débit et de la complexité de décodage par l'analyse de scène pour la télé surveillance basée sur Motion JPEG 2000

J. MEESSEN, C. PARISOT, X. DESURMONT et J.-F. DELAIGLE

Multitel A.S.B.L., Avenue Copernic 1, B-7000 Mons, Belgique.

jerome.meessen@multitel.be

Résumé – Dans ce papier, nous proposons un système de codage/transmission vidéo orienté objet utilisant le standard Motion JPEG 2000 [1] pour le stockage et la diffusion efficaces de vidéo surveillance sur les canaux bas débit. La méthode présente les mêmes avantages que les techniques actuelles de codage de régions d'intérêt tout en améliorant considérablement le rapport débit/qualité moyen de la vidéo transmise, lorsque les caméras sont statiques. Nous proposons ici de transmettre le flux vidéo en transmettant d'une part un flux Motion JPEG 2000 contenant une estimation du fond rafraîchie régulièrement et, d'autre part, un second flux Motion JPEG 2000 ne contenant que les objets mobiles de la scène. Cette méthode fournit une meilleure qualité de la vidéo décodée, qui se rapproche de celle obtenue avec un codeur MPEG-4, mais aussi une complexité réduite pour le client par rapport à un encodage Motion JPEG2000 standard, avec un surcoût de stockage négligeable. Les flux de surveillance stockés sur le serveur sont complètement conformes au standard Motion JPEG 2000.

Abstract – In this paper, we propose a new object-based video coding/transmission system using the emerging Motion JPEG 2000 standard for the efficient storage and delivery of video surveillance over low bandwidth channels. The method provides the advantages of state-of-the-art techniques for Regions-Of-Interest (ROI) coding while significantly improving the average bitrate/quality ratio of delivered video when cameras are static. We transmit only ROIs of each frame as well as an automatic estimation of the background at a lower frame rate in two separate Motion JPEG 2000 streams. This method provides both better video quality and reduced client CPU usage with negligible storage overhead when compared to standard Motion JPEG 2000 encoding. Video surveillance streams stored on the server are fully compliant with existing Motion JPEG 2000 decoders.

1. INTRODUCTION

Dans ce papier, nous proposons un système de codage/transmission vidéo orienté objet utilisant le standard Motion JPEG 2000 [1] pour le stockage et la diffusion efficaces de vidéo surveillance sur les canaux bas débit. Les techniques de compression JPEG 2000 basées sur les Régions d'Intérêt (ROI) [2][3] permettent d'assurer une qualité supérieure pour les régions d'intérêt par rapport au reste lorsque le débit disponible est trop faible pour transmettre chaque image avec une qualité suffisante. Dans les systèmes à caméra fixe, des parties du fond sont cependant transmises de façon redondante à chaque image de la séquence Motion JPEG 2000. Nous proposons ici de transmettre le flux vidéo en transmettant d'une part un flux Motion JPEG 2000 contenant une estimation du fond rafraîchie régulièrement et, d'autre part, un second flux Motion JPEG 2000 ne contenant que les objets mobiles de la scène.

La figure suivante présente l'architecture générale du système proposé :

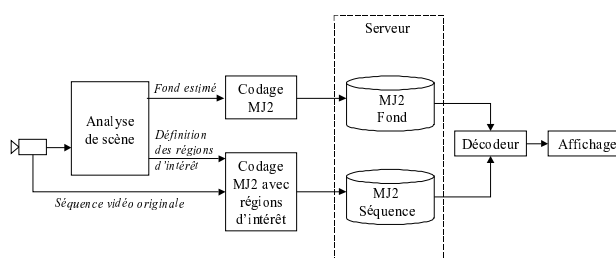


Figure 1: Architecture générale du système

Le flux vidéo d'entrée est analysé de façon automatique par un module réalisant une estimation du fond de la séquence mis à jour à chaque image. La segmentation des objets mobiles de la scène (ROI) est aussi réalisée par ce module. La séquence originale est codée en Motion JPEG 2000, les données relatives aux objets mobiles étant placées dans une première couche de qualité du flux compressé, le reste des images dans une deuxième couche. Le décodage de cette séquence permet la reconstruction de chacune des images originales. L'estimation du fond est codée, elle aussi,

dans un flux Motion JPEG 2000 à intervalles réguliers, par exemple une fois toutes les quatre secondes. Le serveur envoie au décodeur l'estimation du fond et une partie des données de la première couche de la séquence compressée. Le client reconstruit alors la séquence à partir des deux flux sans nécessiter une description explicite des zones d'intérêt.

Le paragraphe 2 traite de la segmentation automatique de la scène. Le paragraphe 3 décrit l'approche proposée pour la diffusion optimisée des flux vidéo segmentés. Avant de conclure, le paragraphe 4 présente les résultats obtenus.

2. SÉPARATION AUTOMATIQUE DU FOND ET DES OBJETS MOBILES

Pour tenir compte de fonds variables (bruit d'acquisition, mouvements périodiques...) nous utilisons un algorithme de segmentation statistique temps réel modélisant le fond par une mixture de Gaussiennes en chaque pixel [4][5][6].

A chaque image, la valeur du pixel est comparée à l'état courant de la mixture de Gaussiennes correspondante. Si cette valeur appartient à l'une des Gaussiennes les plus souvent observées, le pixel est classé comme faisant partie du fond. Sinon il est segmenté comme faisant partie d'un objet mobile. Les paramètres de la Gaussienne (moyenne, fréquence...) contenant la valeur du pixel sont mis à jour. Si le pixel n'appartient à aucune Gaussienne, c'est la Gaussienne la moins souvent utilisée qui est mise à jour.

Des opérateurs morphologiques de dilatation et d'érosion ainsi que des critères prenant en compte les caractéristiques des zones segmentées permettent d'extraire les objets d'intérêt de la scène. A tout instant une image du fond estimé peut être obtenue en prenant pour chaque pixel la moyenne de la Gaussienne la plus probable.

3. CODAGE ET DÉCODAGE DIFFÉRENCIANT FOND ET OBJETS MOBILES

Les résultats de l'analyse sont utilisés pour encoder d'une part un flux Motion JPEG 2000 du fond estimé à intervalles réguliers (par exemple une image toutes les quatre secondes) et d'autre part un flux Motion JPEG 2000 contenant la séquence originale. Dans ce deuxième flux, les codeblocks relatifs aux régions d'intérêt sont isolés dans une première couche de qualité tandis que les autres codeblocks sont placés dans une deuxième couche de qualité. Les deux flux sont par exemple codés sans perte pour une utilisation éventuelle en justice. A qualité identique pour les deux flux, le surcoût d'encodage et de stockage est alors seulement de 1% pour une séquence à 25 images par seconde.

Lorsqu'un client avec peu de ressources, ex. un PDA, accède à la vidéo, nous lui envoyons le flux du fond et les données de la première couche du flux de la séquence en tronquant ces flux, par exemple, au même débit par image. Davantage de débit peut être alloué à la séquence comprenant deux couches. Les régions d'intérêt sont donc

transmises avec une qualité supérieure (dans le cas de petites régions d'intérêt) ou égale (si les régions d'intérêt couvrent la totalité de l'image) au fond de la séquence.

Pour chaque image du fond reçue, le client décode l'image sans effectuer la transformée en ondelettes inverse. A chaque image du flux de la séquence reçue, le client décode le flux binaire. Les codeblocks contenant de l'information sont des codeblocks relatifs aux régions d'intérêt. Les codeblocks vides de toute information sont supposés ne pas avoir été transmis car faisant partie du fond de la séquence (deuxième couche de qualité). Pour les coefficients d'ondelettes de chacun de ces codeblocks vides, nous utilisons alors ceux des codeblocks de même position précédemment décodés depuis le flux Motion JPEG 2000 du fond (cf. Figure 2).

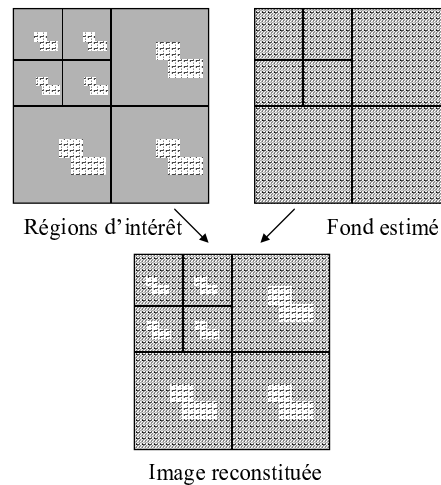


Figure 2: Reconstruction d'une image dans le domaine JPEG 2000 au décodeur

Le surcoût maximum de débit et de complexité de décodage est de 1% tandis que le gain par rapport à une séquence Motion JPEG 2000 standard à qualité identique peut atteindre, lorsqu'il n'y a pas de régions d'intérêt, un facteur 100 à la fois en débit et en complexité de décodage. Le gain est directement lié à la taille et au nombre de régions d'intérêt détectées dans la séquence.

4. RÉSULTATS

Nous avons évalué notre système sur la séquence *speedway* en CIF à 25 images par seconde du consortium Motion JPEG 2000. Le module d'analyse ne requiert qu'environ 15% du temps CPU sur un Dell Precision 3 GHz, ce qui est faible par rapport à la complexité d'encodage de la vidéo (huit fois plus). Notons que les résultats issus de cette analyse peuvent également servir à la détection automatique d'événements anormaux en vidéo surveillance.

Un débit de 634 kbps a été alloué au flux de la séquence et 26 kps pour le flux du fond (0,25 image par seconde). La Figure 3 montre les résultats de la reconstruction de l'image 197. Le pic du rapport signal sur bruit (PRSB) sur la

séquence de 200 images est de 34,7 dB contre 30,6 dB pour Motion JPEG 2000 standard, soit un gain supérieur à 4 dB. La qualité subjective de la séquence est aussi améliorée. Le fond restitué est de meilleure qualité et est beaucoup plus robuste aux artefacts générés par le bruit d'acquisition contenu dans la séquence. En effet, les fluctuations des erreurs de compression au cours du temps n'existent plus pour le fond alors qu'elles peuvent être importantes d'une image à l'autre dans le cas de Motion JPEG 2000 standard. Même dégradé, le fond restitué fluctue donc peu au cours du temps, ce qui rend la visualisation plus naturelle et moins fatigante pour l'opérateur.

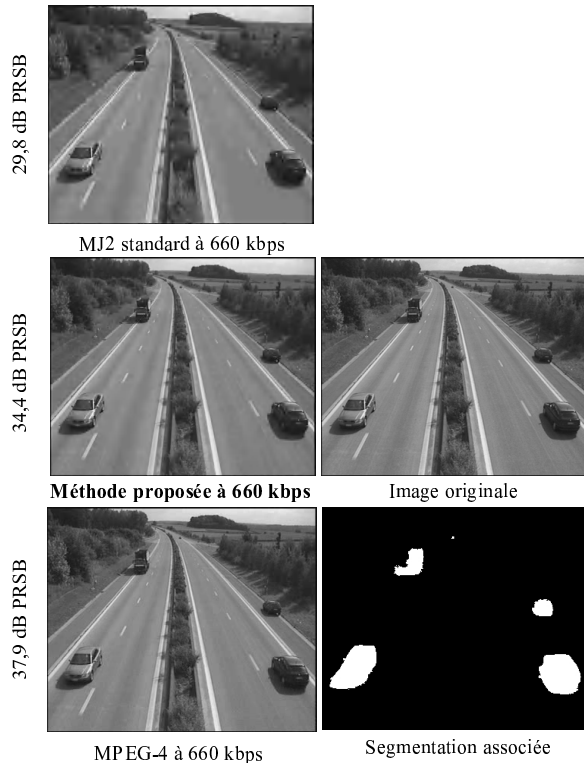


Figure 3: Comparaison de l'approche proposée à Motion JPEG 2000 (MJ2) à 660 kbps.

5. CONCLUSION

L'approche Motion JPEG 2000 basée sur des régions d'intérêt produit un fond dégradé par rapport au Motion JPEG 2000 standard. Ici, nous utilisons deux flux conformes au standard Motion JPEG 2000. L'un des deux contient la séquence d'origine encodée avec les régions d'intérêt dans sa première couche de qualité tandis que le second flux contient une estimation statistique du fond des images. Le surcoût en terme de stockage et d'encodage est négligeable tandis que le gain en terme de débit et de complexité de décodage peut être très important lorsqu'il y a peu d'activité dans la scène. Le gain en PRSB atteint 4 dB pour la séquence *speedway* à 660 kbps. La qualité visuelle de restitution de la séquence est

proche celle obtenue avec un encodeur MPEG-4, tout en conservant les propriétés de scalabilité de JPEG 2000.

6. REMERCIEMENTS

Ce travail a été financé par la commission européenne dans le cadre du projet WCAM (FP6 IST-2003-507204)[7].

7. RÉFÉRENCES

- [1] ISO/IEC 15444-3 "JPEG 2000 part 3: Motion JPEG 2000"
- [2] V. Sanchez, A. Basu, M. Mandal, "Prioritized Region Of Interest Coding in JPEG2000", *IEEE trans. on CSVT*, vol. 14(9), pp. 1149-1155, Septembre 2004.
- [3] J. Meessen, C. Parisot, C. Le Barz, D. Nicholson and J.-F. Delaigle, "WCAM: Smart Encoding for Wireless Surveillance", *Proc. of SPIE Image and Video Communications and Processing (IVCP 05)*, San Jose, USA, Janvier 2005.
- [4] C. Stauffer, W.E.L. Grimson, "Adaptive Background mixture models for real-time tracking", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 246-252, Juin 1999.
- [5] K. Kim, T. Horprasert, D. Harwood, L. Davis, "Codebook-based Background Subtraction and Performance Evaluation Methodology", 2003.
- [6] X. Desurmont, C. Chaudy, A. Bastide, C. Parisot, J.F. Delaigle, B. Macq, "Image analysis architectures and techniques for intelligent systems", *IEE proc. on Vision, Image and Signal Processing, Special issue on Intelligent Distributed Surveillance Systems*, vol 152(2), pp. 224-231, Avril 2005.
- [7] FP6 IST-2003-507204 WCAM, "Wireless Cameras and Audio-Visual Seamless Networking.", <http://www.ist-wcam.org>.