

二層ビデオオブジェクト自動生成法と MPEG-4 符号化への適用

秦泉寺久美, 渡辺裕, 小林直樹

NTTサイバースペース研究所

〒239-0847 横須賀市光の丘 1-1

+81-468-59-3034(TEL)

+81-468-59-2829(FAX)

{kumi,hiroshi,kobayashi}@nttvdt.hil.ntt.co.jp

あらまし

新しい画像符号化標準の MPEG-4 が標準化されつつある。これは、新しいファンクショナルリティであるオブジェクト単位の符号化を提供するものである。著者らはインターネットアプリケーションなどに適用できる超低ビットレート符号化方法の研究を行っており、“VHS on 2B” (128kbps で VHS ビデオ並の品質) の符号化方式の開発を行っている。本稿では、MPEG-4 Version 1 Main profile にてサポートされている「スプライト符号化」に着目し、スプライト符号化の課題の一つであるビデオオブジェクトの自動抽出アルゴリズムを提案した。また、提案アルゴリズムを MPEG-4 符号化して従来符号化方法と比較検討を行った。同程度の画質を、従来符号化方法の 1/2 以下の符号量で実現した。

キーワード

MPEG-4, スプライト生成, オブジェクト抽出, グローバル動き推定

Automatic Two-layer VOP Generation and Its Application to MPEG-4 Video Coding

Kumi Jinzenji, Hiroshi Watanabe, Naoki Kobayashi

NTT Cyber Space Labs.

1-1 Hikari-no-oka Yokosuka-Shi, Kanagawa

239-0847JAPAN

+81-468-59-3034(TEL)

+81-468-59-2829(FAX)

{kumi,hiroshi,kobayashi}@nttvdt.hil.ntt.co.jp

Abstract

MPEG-4, Sprite generation, Object Extraction, Global motion estimation

The new video coding standard MPEG-4 provides content-based functionality and low bit-rate video compression. We focused on the “sprite coding” supported MPEG-4 Version 1 Main profile in order to achieve “VHS quality video on 2B (128kbps)” for narrow-band transmission such as the Internet. Automatic VOP (Video Object Plane) generation technologies are being studied as one of the most important issues of MPEG-4 object coding. This paper proposes a two-layer VOP generation scheme with some core algorithms such as GME (Global Motion Estimation), foreground moving object extraction, and background sprite generation. This paper also describes a shape information reduction method for foreground objects. Using this method, shape information is compressed by 90%. Experiments are conducted on VOP generation and video coding with MPEG-4. Coding efficiency is several times higher than that of typical existing video coding schemes at the same objective image quality.

key words

1. はじめに

新しい画像符号化標準 MPEG-4[1]では、オブジェクトベースの符号化を新しいファンクショナルリティとして取り入れている。著者らはインターネットアプリケーションなどに適用できる超低ビットレート符号化方法の研究を行っており、“VHS on 2B” (128kbps で VHS ビデオ並の品質) を達成する符号化方式の開発を行っている。本稿では、ソリューションの一つとして、MPEG-4 Version 1 Main profile でサポートされている「スプライト符号化」に着目した。スプライト符号化およびオブジェクト符号化の課題の一つとして、オブジェクトの自動生成がある。スプライト自体の自動生成法はさまざまな提案[2-4]がなされているが、動物体の無いビデオシーケンスからの生成法であるなど、背景スプライトと前景動物体オブジェクトのトータルな自動生成はまだなされていない。また、動物体の抽出法には手動もしくは半自動の抽出法[5][6]が提案されている。また、動物体の自動抽出法[7-9]も提案されているが、オブジェクトが唯一存在する場合であり、多数のオブジェクトが存在する場合の異なるフレーム間での同一オブジェクトの対応問題が解決されていない。

本稿では、ビデオオブジェクトを自動生成するアルゴリズムを提案し、提案アルゴリズムによって生成したビデオオブジェクトを MPEG-4 符号化して従来符号化方式との比較検討を行った。はじめに、オブジェクトの対応問題を回避するために、前景オブジェクトと背景スプライトの二層からなる VOP(Video Object Plane)を定義する。この方法は背景以外の動物体とみなされた領域すべてをひとつの前景とするものである。例えば、「サッカー」を想定すると、選手、審判員、ボールなどはすべて一つの前景オブジェクトとして扱われる。従来方法は異なるオブジェクトとして扱うため、フレーム毎のオブジェクト対応が未解決という問題を生じている。二層ビデオオブジェクト生成アルゴリズムは、スプライト生成に特化したグローバルモーション (GM) 算出法、高品質背景スプライト生成法、前景オブジェクト抽出法、前景形状情報近似法からなる。背景スプライトを生成するためには GM はカメラ操作を反映したものであることが望まれる。これは、従来法[2]の GM が主に GM による予測画像と原画像の最小二乗誤差(Mean Square Error: MSE)を与えるものとして定義されるものとは大きく異なる。従来法では、カメラ操作を反映した大局的な動きが GM として必ずしも算出されるわけではい。このことがスプライトの劣化につながる。高品質スプライト生成法では、従来の時間メディアン法、オ

ーバライツ法[10]の長所を生かした。時間メディアン法は時間方向にメディアン (中央値) をとってスプライトの画素値とするものである。そのため、前景オブジェクトを消し去ったスプライトを生成することができるが、スプライトの画質は若干ぼける傾向があった。上書き法は、画素をそのまま基準フレームに貼りつけていくので画質の良いスプライトが生成できる。反面、一番手前に来るフレームの前景と、各フレームの端の前景オブジェクトが残るといった問題があった。本手法では、前景のなかつ高品質なスプライト生成法を提案する。前景オブジェクト抽出法は、原画像と背景スプライトの差分を基本とし、GM のズレに対してロバストな方法を提案する。また、自動生成における前景オブジェクト形状は輪郭が複雑であること、孤立点が多数含まれること、形状の予測範囲が±1.6画素であることなどからほとんど予測があたりはず、その結果イントラ符号化されて符号量が増大する。この問題を回避するために形状をマクロブロックで近似する方法を提案する。この手法は形状符号量を 1/10 程度に削減できる。さらに、本稿では、提案アルゴリズムを用いて生成した二層オブジェクトを MPEG-4 Version 1 Main profile で符号化し、VOP 構造を持たない従来符号化方法と比較検討を行った。

以降、2章では二層オブジェクトについて提案する。3章では2層 VOP 自動生成アルゴリズムを提案する。4章では MPEG-4 符号化実験について述べ、5章で結論を述べる。

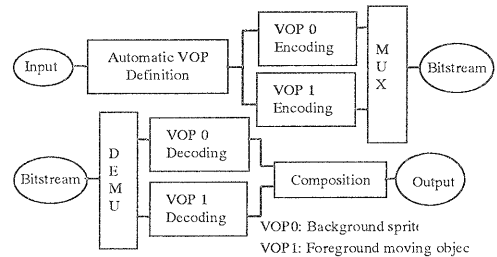


Fig. 1 Two-layer VOP Model.

2. 二層ビデオオブジェクト

本稿における、二層ビデオオブジェクト生成アルゴリズムの大きな特徴は、①二層ビデオオブジェクトを生成すること、②ビデオオブジェクトの一つをスプライトで表現することである。前者は、多数の前景領域を一つの前景オブジェクトとして扱うことにより、多数の前景オブジェクトのフレーム間対応問題を回避する目的がある。また後者においては、スプライト符号化に特徴的なドラスティックな符号量削減[3][4]を実現する目的がある。

図1に二層ビデオオブジェクトを有する符号化復号モデルの概念図を示す。まず、符号化側で、画像を前景オブジェクトと背景スプライトの二層のビデオオブジェクトに分割する。背景スプライトはカメラモーションを反映した背景部分である。前景オブジェクトは背景以外のあらゆる動領域で、これらはすべて一つの前景として扱われる。前景オブジェクトと背景スプライトは独立なビデオオブジェクトとして別々に符号化、多重化されて一つのビットストリームで送信される。受信側では、ビットストリームを逆多重し、ビデオオブジェクト毎に復号され、合成されて表示される。

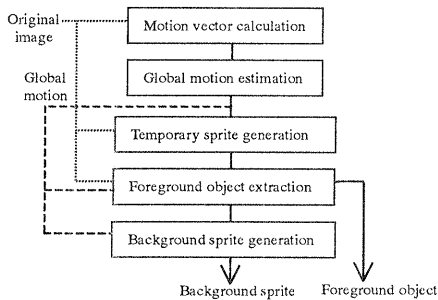


Fig. 2 Flowchart of Automatic Two-layer VOP Generation Scheme.

3. 二層 VOP 自動生成アルゴリズム

図2に二層 VOP(Video Object Plane)自動生成アルゴリズムの全体像を示す。本アルゴリズムは主にGM算出、背景スプライト生成、前景オブジェクト抽出、前景形状マクロブロック近似の4つのパートからなる。

まず、隣接フレーム間にて動きベクトルを算出する。次に、その動きベクトルを用いて基準フレームからの GM を算出する。この GM によって原画像を変形して基準座標にマップし、動物体を消去した仮スプライトを生成する。さらに、仮スプライトと原画像の差分画像から前景候補画像と背景画像を生成する。前景候補画像はマクロブロック近似されて最終的な前景画像となる。一方で、背景画像は背景スプライト生成に用いられる。以下にその詳細を述べる。

3.1 GM 算出

スプライト生成のための、カメラ操作を反映した GM を算出するために、本稿では以下の2点を特徴とする GM 算出法を用いた。

1. カメラモーションを反映した局所的動きベクトルにおいて、そのx方向ならびにy方向の変化量を軸とする特徴空間で独特の分布を示すことを利用

2. カメラモーションの候補となる複数のクラスタをそれぞれ検出し、最適なものを GM として算出(クラスタ検定)

なお、文献[12]に GM 算出の具体的な手法と定量的な実験を行っているので参照されたい。

3.2 背景スプライト生成

前景のないかつ高品質なスプライトを生成するために、本稿では従来の時間メディアン法、上書き法[10]の長所を生かした方法を提案する。生成されるスプライトは2種類ある。仮スプライトと本スプライトである。

スプライトとは画像を模様がつながるように並べてパノラマ画像にしたものである。前節で算出された基準フレームからの GM を用いて各々の画像を変形し基準座標系の格子点上にマッピングする。基準座標系の任意の座標に注目すると、複数の画素が重なっている。これらのメディアンをとってその座標の値とする。一般に動物体領域がその座標値画素の半分より少なければ、メディアンをとることで動物体を反映する画素は選択されない。よって、動領域のないスプライトを生成することができる。しかし、時間メディアン法によるスプライトは若干ぼける傾向があった。そこで、このメディアン値によるスプライトを仮スプライトとして、前景抽出に利用する¹。上書き法は、画素をそのまま基準フレームに貼りつけていくので画質の良いスプライトが生成できるが、一番手前に来るフレームの前景と、各フレームの端の前景オブジェクトがのこるといった問題があった。よって、動物体を抽出した後、画像の残りの部分を背景画像とし、それを基準座標に順次上書きし、最終的な高画質スプライトを生成する。これを本スプライトという。

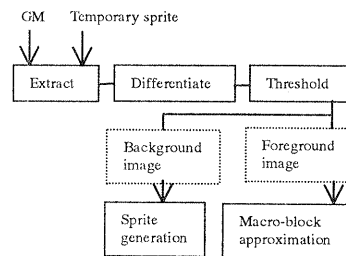


Fig. 3 Foreground object extraction.

3.3 前景オブジェクト抽出

図3に前景オブジェクト抽出アルゴリズムを示す。まず、仮スプライトからGMによって切り出された画像と原画像の差分画像を作る。この差

¹ GM が高精度で求まっているときは画質の損なわないスプライトを生成できる。

分画像を任意の閾値で二値化して前景候補画像と背景画像に分離する。前景候補画像は次節の前景形状マクロブロック近似手法で最終前景画像とされる。

3.4 前景形状のマクロブロック近似

自動生成における前景オブジェクト形状は輪郭が複雑であること、孤立点が多数含まれること、形状の予測範囲が±16画素であることなどからほとんど予測があたりずイントラ符号化され、その結果、符号量が増大する。本稿ではこの問題を回避するために、オブジェクト形状をマクロブロックで近似する方法を提案する。図4に前景形状のマクロブロック近似の様子を示す。

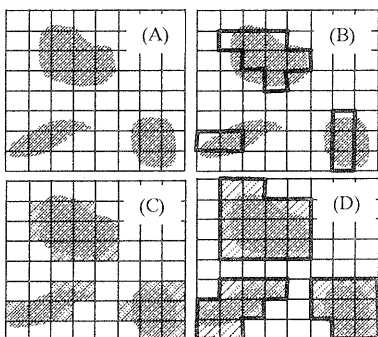


Fig.4 Macro-block approximation. (A)original contours; (B)most lossy mode; (C)first macro-block approximation; (D)Final contours.

MPEG-4において、オブジェクト形状符号化はロスレス、ロシーの二通りのモードが用意されている。その粗さはパラメータ α_{Th} で表される。 α_{Th} は1から256の値を取り、256の時はロスレスモード、それ以外はロシーモードである。特に値1の時は一番粗い形状表現であり、マクロブロックで近似される。具体的には、マクロブロック内に前景情報が半数以上あればマクロブロック全体を前景、半数より少なければマクロブロック全体を背景とするものである。前景オブジェクトがかなり侵食されて視覚的に妨害になるという問題があった。そこで、本稿では、2段階のマクロブロック近似法を提案する。図4-aに第1のマクロブロック近似、図4-bに第2のマクロブロック近似の様子を示す。まず、マクロブロック内に前景が閾値 Th_1 以上ある場合、マクロブロック全体を前景と判断する。それ以外の場合は背景とする。次に、第1次マクロブロック近似では前景となったマクロブロックに隣接する背景マクロブロックにおいて、そのマクロブロック内に前景画素が閾値 Th_2 ($Th_2 < Th_1$) 以上あるものを前景と再判断する。

最適な閾値 Th_1, Th_2 は画像に依存する。例えば、複雑な背景をもつ画像では、GMが誤抽出された場合に影響を受けやすく、差分による前景候補領域は必然的に多くなる。しかし、一樣なフィールドを背景に持つような画像では、GM誤抽出によって抽出される前景候補領域は少ないと考えられる。そこで、閾値 Th_1, Th_2 を自動で決定するために、図5に示すようなアルゴリズムを提案する。ある程度の前景比率以上はスプライト符号化しても符号量削減にはならないと考えられるので、前景比率を制御に使用する。 Th_1, Th_2 は1画素に相当するパーセンテージに初期設定され、前景比率が知見上から得られる閾値 R_{max} を上回るまで値が増加して更新される。

セグメンテーションマスクを有する標準画像"Stefan"において、形状情報をマクロブロック近似をしたところ、形状符号量が1/10程度に抑えられた。マクロブロック近似の他のメリットとして、ソフトウェアデコーダ実装時の処理コスト軽減があげられる。形状情報の復号、パディングといった処理をしないのでリアルタイム復号が求められるアプリケーションに好都合と考えられる。

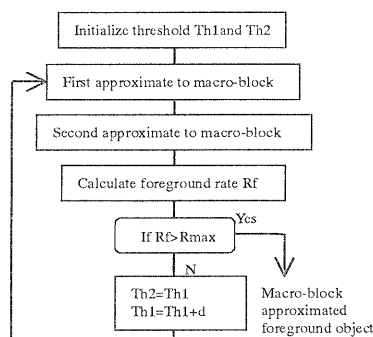


Fig.5 Flowchart of macro-block approximation.

4. MPEG-4 符号化実験

4.1 実験

以下に示すカメラ操作を含む3つのSIF画像において、提案アルゴリズムにて生成されたビデオオブジェクトをMPEG-4 Version 1 Main profileで符号化(便宜上、Sprite Modeという)し、ビデオオブジェクトを生成しない従来符号化法(同、Rectangle mode)と比較した。

1. Horsesrace: 複数の競争馬をカメラで追った画像
2. Soccer: 試合における選手とボールの動きをカメラが追った画像
3. Athlete: 複数の走者をカメラで追った画

像

フレームレートは 5,10,15,30 と変化させた。フレーム数はそれぞれ 150 枚、符号量は比較のために、1 秒あたりのビット数を用いている。マクロブロック近似で用いられる閾値 R_{max} を経験的に 10% と設定する。

4.2 結果ならびに考察

図 6 に本アルゴリズムによって生成された背景スプライトならびに前景オブジェクト、合成画像の例 (Horsrace) を示す。前景の消去されたスプライトが生成されていることがわかる。また、前景オブジェクトは前景の画面に対する比率が 10% になるまでマクロブロック単位に抽出されるアルゴリズムなので本来は背景に属する領域も前景オブジェクトとして抽出されている。しかし、背景スプライトから該当背景を切りだし、前景と合成すると違和感の無い画像が生成されていることがわかる。

また、図 7 に Rectangle mode と Sprite mode においてフレームレートを変化させた場合の符号量の推移を示す。符号量は、1 秒あたりのビット数で示した。同程度の画質²を達成するのに、Sprite mode による符号化法は Rectangle mode の方法に比べて 1/2 から 1/3 の符号量を実現している。この傾向はどのフレームレートにも当てはまる。Athlete を例にとると、200kb/s 相当の符号量では 5 枚以下のフレームレートしか達成できないのに対し、Sprite mode では 20 枚以上を達成可能である。スプライトを使用すると、Internet のような狭帯域環境下でもフレームレートを稼げることがわかる。

処理時間は、モーションベクトル算出、画像処理 (GM 算出以降)、符号化全体の処理時間の合計を 100% とすると、それぞれ 78%, 18%, 4% であった。画像処理、符号化共々、モーションベクトル算出時間に比べ小さな処理量である。

5. おわりに

“VHS on 2B” (128kbps で VHS ビデオ並の品質) の符号化方式の開発を目的として、本稿では、MPEG-4 Version 1 Main profile にてサポートされている「スプライト符号化」に着目し、スプライト符号化の課題の一つであるビデオオブジェクトの自動抽出アルゴリズムを提案した。また、提案アルゴリズムを MPEG-4 符号化して従来符号化方法と比較検討を行った。同程度の画質を、

² 提案アルゴリズムによる高品質な背景スプライトが生成されているために QP が同じ場合は同程度の画質の符号化画像が生成されると考えられる。

従来符号化方法の 1/2 以下の符号量で実現した。

このようにスプライト符号化はカメラの動きがあつて、前景割合が画像全体の 10% 以下程度に収まる場合にはドラステックな符号量削減が期待できる。しかし、すべてのビデオシーケンスがこの条件を満たしているわけではなく、スプライト符号化できるショットの判定方法、シームレスな復号方法、配信を考慮した場合のビデオオブジェクト毎符号量割り当ての検討が今後の研究課題である。

謝辞

本研究を進めるにあたって計算機シミュレーションに多大な協力をいただいた (株) NTT ソフトウェアの米原紀子氏に感謝いたします。また、有益なディスカッションを賜った NTT サイバースペース研究所メディア通信プロジェクト画像符号化技術グループの皆様に感謝いたします。

参考文献

- [1] Final Draft of International Standard, ISO/IEC 14496-2.
- [2] M. Irani, S. Hsu, and P. Anandan, “Video Compression Using Mosaic Representation,” *Signal Processing: Image Communication*, Vol. 7, pp. 529-552, 1995.
- [3] K. Jinzenji, H. Watanabe, N. Kobayashi, “Global Motion Estimation for Static Sprite Production and Its Application to Video Coding,” *IEEE ISPACS'98*, pp.328-332, November 1998.
- [4] K. Jinzenji, S. Takamura, H. Watanabe, N. Kobayashi, “Automatic VOP Production Scheme for Very Low Bit Rate Coding,” *PCS'99*, pp.299-302, 1999.
- [5] M. Kass, A. Witkin, D. Terzopoulos, “SNAKES: Active Contour Models,” *Proc. 1st ICCV*, pp.259-268, 1987.
- [6] J. G. Choi, S. Lee, S. Kim, “Spatio-Temporal Video Segmentation Using a Joint Similarity Measure,” *IEEE Trans. on CSVT*, Vol. 7, No. 2, pp.279-286, April 1997.
- [7] A. Neri, S. Colonnese, G. Russo, “Automatic Moving Objects and Background Segmentation by Means of Higher Order Statistics,” *IEEE ISCAS'97*, June 1997.
- [8] R. Mech, M. Wollborn, “A Noise Robust Method for Segmentation of Moving Objects in Video Sequence,” *IEEE ICASSP'97*, April 1997.
- [9] T.Meier, K.N. Ngan, “Automatic Segmentation of Moving Objects for Video Object Plane Generation,” *IEEE Trans. on CSVT*, Vol. 8, No. 5, pp.525-538, September 1998.
- [10]阿久津明人, 外村佳伸, “投影法を用いた映像の解析法と映像ハンドリングへの応用,” *信学会論文誌 Vol. J79-D-II No.5* pp.675-686 May 1996.

[11]秦泉寺久美, 石橋聡, 小林直樹, “カメラモーション抽出によるスプライト自動生成,” 信学会論文誌 Vol.J82-D-II No.6, pp.1018-1030, June 1999.

[12]秦泉寺久美, 渡辺裕, 小林直樹, “スプライト生成のためのグローバルモーション算出法と符号化への適用,” 信学会論文誌 (to appear)

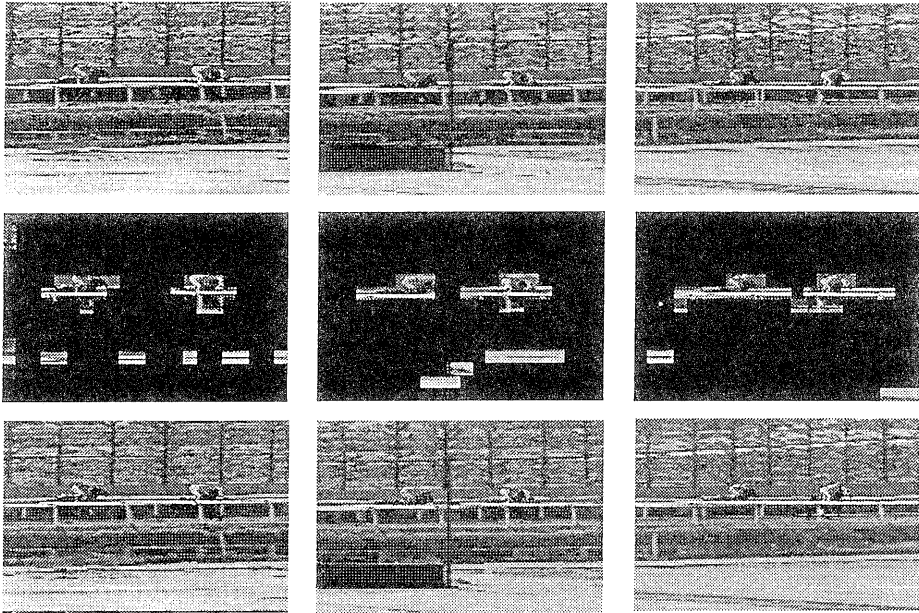
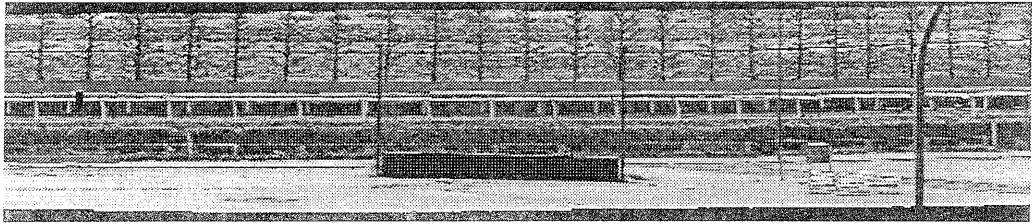


Fig.6 Generated background sprite and foreground objects. First law: background sprite, second law: original images, third law: foreground objects, lower law: composed images(Coded by QP=12).

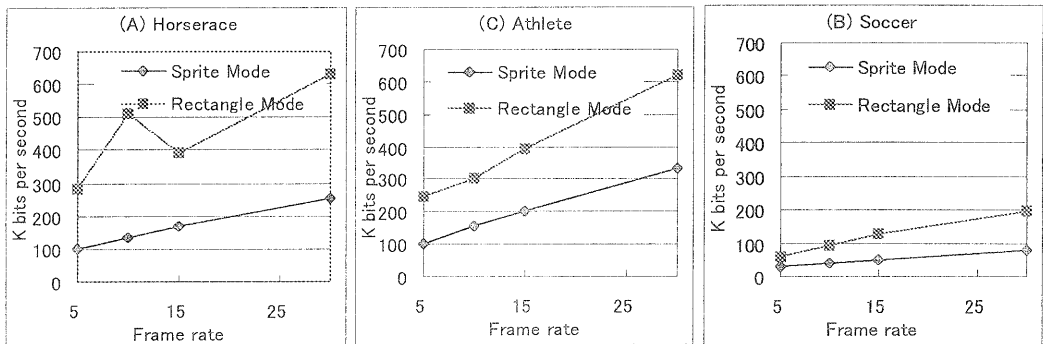


Fig. 7 Coding results.