

Iフレームを用いたショット境界検出の高速化に関する性能評価

Performance Evaluation of Fast Shot Boundary Detection Using I-frames

内田 祐介 橋本 真幸 米山 暁夫
Yusuke Uchida Masayuki Hashimoto Akio Yoneyama

1. はじめに

映像ハンドリング技術の基本であるショット境界の検出に関しては、これまで様々な手法が提案されている。例えば、サポートベクターマシン (SVM) を用いた手法が、精度において優れた性能を示すことが報告されている [1]。処理速度も精度とともに重要な要素であるが、上記手法は圧縮符号化コンテンツのデコード時間がオーバーヘッドとなっており、高速化検討の余地がある。本論文では、圧縮された映像データの基本構造である GOP (Group of Pictures) に着目し、まず I フレームのデコードを行い、各 GOP 内にショット境界が存在するか否かを判定し、GOP 内にショット境界があると判定された GOP のみをデコードすることでショット境界検出を高速に実現する手法を提案し評価を行う。また、GOP 内のショット境界判定時に必要となる例外処理についても検討を行う。

2. 提案手法

図 1 に提案手法のフレームワークを示す。提案手法では、各 GOP 内にショット境界があるか否かを判定する GOP レベル境界判定を行い、GOP 内にショット境界があると判定された GOP のみフレームレベルでのショット境界を検出するフレームレベル境界判定を行う。ここでは、瞬時ショット境界のみを対象とし、フェード等のトランジションを持つショット境界は扱わない。以降では提案手法を詳述する。

2.1. GOP レベル境界判定

提案手法における GOP レベルおよびフレームレベルでの境界判定手法は、TRECVID'05 のショット境界タスクで高い精度を実現している SVM を用いた手法 [1] をベースとする。図 2 に GOP レベル境界判定に利用する特徴量の抽出法を示す。まず、判定対象 GOP の前後それぞれ N 個の GOP から I フレーム I_1, \dots, I_{2N} をデコードする。各 I フレームを $X \times Y$ 個の領域 $R_{x,y}$ ($1 \leq x \leq X, 1 \leq y \leq Y$) に等分割し、フレーム I_i の領域 $R_{x,y}$ から YCbCr 成分をそれぞれ 4 分割した 64 次元色ヒストグラム $v_{i,x,y}$ を作成する。その後、隣接する I フレーム I_i, I_{i+1} の領域 $R_{x,y}$ のヒストグラム $v_{i,x,y}, v_{i+1,x,y}$ 間の距離をカイ二乗値にて算出する。ヒストグラム v, v' のカイ二乗値は

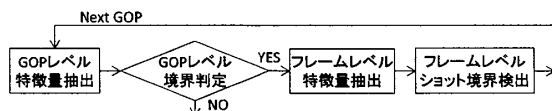


図 1 提案手法フレームワーク

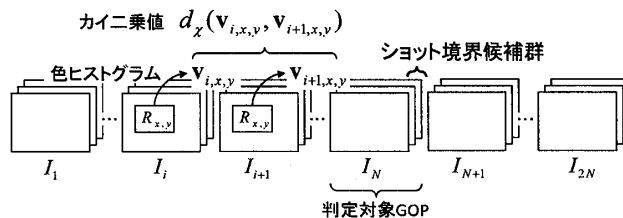


図 2 GOP レベル境界判定のための特徴量

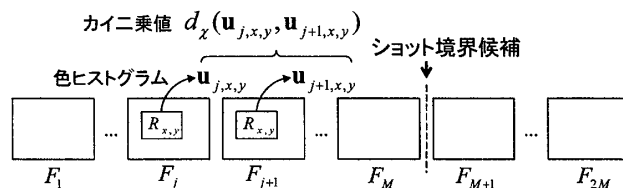


図 3 フレームレベル境界判定のための特徴量

次式で与えられる。

$$d_x(v, v') = \sum_{1 \leq b \leq 64} \frac{(v[b] - v'[b])^2}{v[b] + v'[b]} \quad (1)$$

このヒストグラム間距離の組

$$V_{\text{inter}} = \{d_x(v_{i,x,y}, v_{i+1,x,y}) \mid 1 \leq i < 2N, 1 \leq x \leq X, 1 \leq y \leq Y\} \quad (2)$$

を特徴量として SVM を利用しショット境界か否かを判定する。

2.2. フレームレベル境界判定

図 3 にフレームレベル境界判定に利用する特徴量の抽出法を示す。これはほぼ GOP レベル境界判定の場合と同じである。判定対象となる境界の前後それぞれ M フレーム F_1, \dots, F_{2M} に関して、フレーム F_j の領域 $R_{x,y}$ からヒストグラム $u_{j,x,y}$ を作成し、ヒストグラム間距離の組 V_{intra} を特徴量とする。

$$V_{\text{intra}} = \{d_x(u_{j,x,y}, u_{j+1,x,y}) \mid 1 \leq j < 2M, 1 \leq x \leq X, 1 \leq y \leq Y\} \quad (3)$$

3. 評価実験

3.1. 実験概要

パラメータ 提案手法の有効性検証のため、評価実験を行う。学習およびテストには、映画コンテンツに対して、人手によりラベル付けを行った映像コーパスを利用する。表 1 に詳細を示す。提案手法のパラメータは、予備実験より $N=2, M=2, X=2, Y=2$ とした。SVM ライブラリとして SVM-Light を利用し、RBF カーネルを利用する以外はデフォルトのパラメータで運用した。実験に利用した PC の性能は、OS: Windows XP, CPU: Core 2 Quad 3GHz, メモリ: 8GB である。実験は単一スレッドで行ったため、処理時間は 1 コアのみを利用した測定値となる。

学習 本実験では、GOP レベルおよびフレームレベルの境

表1 実験に用いた映像データ情報

(i) 学習/テスト用の映像データ				
	映像長 [秒]	フレーム数	正解ショット境界数	
学習	41,534	1,244,778	9,342	
テスト	161,816	4,849,636	33,105	
(ii) ストリーム情報				
Codec	Profile	Framerate	Bitrate	GOP size
MPEG-2	MP@ML	30[fps]	4.0[Mbps]	15(IBBP)

界判定用の2種類のSVMを利用する。GOPレベル境界判定用SVMの学習は、学習用映像データから15フレーム間隔で $2N=4$ 枚のフレームを選択し、 I_2, I_3 フレームの間に1つ以上のショット境界が存在する場合には正例、そうでない場合には負例として学習データに追加することを、ウィンドウを1フレームずつずらしながら行った。フレームレベル境界判定用SVMの学習には、学習用映像データから連続する $2M=4$ 枚のフレームを選択し、 F_2 と F_3 フレームの間がショット境界である場合には正例、そうでない場合には負例として学習データに追加することを、ウィンドウを1フレームずつずらしながら行った。

3.2. 実験結果

本実験では、(a)Frame、(b)GOP、(c)GOP+Frame(提案手法)の3つの手法を比較する。Frameは、フレームレベル境界判定を全フレームに対して行う手法(従来手法[1]の簡易版)、GOPは、GOPレベル境界判定のみを行う手法、GOP+Frameは、GOPレベル境界判定後にフレームレベル境界判定を行う手法(提案手法)である。精度はPrecision(適合率)およびRecall(再現率)で評価する。

$$\text{Precision} = \frac{\text{検出された正解ショット境界数}}{\text{検出されたショット境界数}}$$

$$\text{Recall} = \frac{\text{検出された正解ショット境界数}}{\text{正解ショット境界数}} \quad (4)$$

精度 表2(i)に各手法の精度を示す。(a),(c)を比較すると、GOPレベル境界判定を基にフレームレベルの境界判定を省略したことでF値で4%程度の精度低下が発生していることが分かる。詳細には、適合率が若干向上し、再現率が大きく低下している。適合率の向上は、GOPではGOPレベルとフレームレベル両方の判定でショット境界と判定されなければショット境界と判定されないため、(過)検出数が減少したことに起因する。再現率の低下は、GOPレベル判定での再現率が低いためであり、改善の必要がある。

処理時間 表2(ii)に各手法の処理時間を示す。ここでは、デコード、ヒストグラム作成、特徴量作成、SVM判定、それぞれに必要な処理時間、およびそれらの合計時間を記載している。(c*)はショット境界を含むと判定されたGOPのみに対してフレームレベル境界判定を行った際の処理時間、(c)は(b)と(c*)の合計時間である。(a),(c)を比較すると、提案手法では全体の処理時間が1/3程度に抑えられており、効果的に処理時間を削減することができている。(a),(b)を比較すると、(b)では判定すべき事例数が約1/15(Iフレームの割合)であるにもかかわらず、SVM判定に必要な時間が増加している。これはIフレームのみからショット境界判定を行う問題が、全てのフレームを利用する場合よりも難しく、サポートベクターが増加したためと考えられる。

端処理の考察 フレームレベル判定では、図4に示す例のよ

表2 実験結果

(i) 検出精度					
	Precision	Recall	F-measure		
(a) Frame	0.9071	0.8935	0.9002		
(b) GOP	0.9081	0.8087	0.8555		
(c) GOP+Frame	0.9654	0.7729	0.8585		
(ii) 処理時間 [秒]					
	Dec.	Hist.	Feat.	SVM	Total
(a)	14,937	1,748	61	3,930	20,676
(b)	130	121	4	4,939	5,194
(c*)	1,288	139	5	328	1,760
(c)	1,418	260	9	5,267	6,954

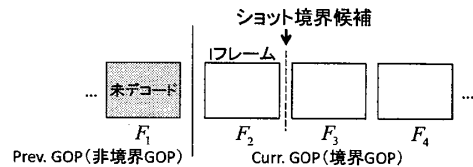


図4 端処理が必要な例

表3 端処理の精度

	Precision	Recall	F-measure
(a) Frame	0.9071	0.8935	0.9002
(d) Frame	0.7659	0.9219	0.8367
(e) Frame	0.7790	0.8977	0.8342
(f) Frame	0.8513	0.7428	0.7933

うに、ヒストグラム間距離を算出する際にデコードされていないフレームを参照することがある。すなわち、図4の場合には F_1 がデコードされていないため、 F_1 と F_2 間の距離が算出できない。 F_1 をデコードすることは F_1 を含むGOPをデコードすることとほぼ等しいため非効率的である。このため、(d) F_1, F_2 間の距離を0とする、(e) F_1, F_2 間の距離を F_3, F_4 間の距離とする、(f) F_2, F_3, F_4 間の距離のみを利用するSVMを個別に用意する、3手法を検討した。表3にそれぞれの精度を示す。比較のため、ここでは端処理を全フレームに対して行った際の精度を記載している。各手法とも実際の F_1 フレームを利用する(a)よりも精度が低下しているが、(d)が最もF値に関して精度低下が抑えられているため、提案手法(c)では(d)を利用している。 F_4 がデコードされていない場合も同様の検討を行ったが、精度は F_1 がデコードされていない表3とほぼ同じであった。

4. まとめ

本論文では、SVMを利用したショット境界検出手法について、GOPレベルの境界判定とフレームレベルの境界判定の2段階の判定を行うことで、デコード時間等の処理を削減し、高速にショット境界検出を行う手法を提案した。提案手法は、全てのフレームを判定する手法と比較して、処理時間を66%削減する一方、精度低下を4%程度に抑えることができた。今後はトランジションを持つショット境界への応用を検討する。

参考文献

- [1] K. Matsumoto, M. Naito, K. Hoashi, and F. Sugaya, "SVM-Based Shot Boundary Detection With a Novel Feature," in *Proc. of ICME'06*, pp. 1837-1840, 2006.