時空間特徴とセンサ情報を用いたパーソナル映像の自動編集

Automatic Video Editing Using Spatiotemporal Feature and Sensor Information

古川拓弥　藤吉弘亘
Takuya Furukawa, Hironobu Fujiyoshi
中部大学 工学部 情報工学科
Department of Computer Science, Chubu University
E-mail: takuya@vision.cs.chubu.ac.jp

Abstract

本稿では，カメラから得られる映像情報とセンサ情報を用いて，映像中の動きとカメラワークを捉えたパーソナル映像の自動カット編集手法を提案する．本研究では，一般ユーザによりカット編集されたパーソナル映像とその編集履歴を用いて，カットされたシーン（使用シーン）とその切り出し点（カット編集点）における傾向を調査した．調査の結果，使用シーンではカメラワークの発生率が高くなり，カット編集点では，映像中における動きの変化が少なく，前後フレームで類似した動きの変化があることを確認した．この傾向に基づき，提案手法では，映像中における動きの変化を捉えることができるST-patchから算出されるCRIM（動きの定常度）やMC（動きの類似度），そして，カメラのセンサから得られる加速度や角速度を利用し，カメラワークや被写体の動きが多く発生しているシーンをカット編集する．ユーザの編集傾向に基づいた映像編集をすることで，ユーザの嗜好に合った映像編集が可能となる．実験の結果，ST-patch特徴のみを用いた映像編集手法と比べ，視聴者の満足度が高い映像の自動編集が可能であることを確認した．

1　はじめに

パーソナルビデオカメラの低価格化と機能向上により，一般のユーザが映像を撮影する機会が増加している．また，デジタルカメラやカメラ付き携帯電話においても，ビデオカメラ（映像撮影）機能は標準装備となっている．これに伴い，YouTubeを代表とした動画共有サイトも普及してきた．これらの動画共有サイトでは，撮影した映像をインターネット上に配信し，世界中の人々と共に楽しむことができる．また，近年のカメラ付き携帯電話では，撮影した映像をその場で動画共有サイトへアップロード可能な機能も搭載されている．このような動画共有サイト上へパーソナル映像を投稿する際には，配信時における映像のデータ容量制限や第3者に対する誘目性の高い映像作成を考慮する必要があり，映像中における重要なシーンの選定や冗長なシーンの削除が重要となる．しかし，映像中のどこが必要でどこが不要であるかは，映像をすべて視聴しなければ判断することができず，多大な時間と労力が必要となる．そのため， 映像編集を容易に行える編集支援システムの需要が高まっている．

これらの問題を解決するために，映像を時間的に圧縮した映像要約手法 [1][2]，注目領域のトリミング手法 [3][4]，特定シーンの切り出し手法 [5][6] などの映像の自動編集に関する研究がされてきた．しかし，多くの先行研究では，固定カメラで撮影した映像やスポーツ，講義など特定の映像を対象とした編集であるため，撮影環境が様々に変化する個人が撮影したパーソナル映像には適応できない．

そこで本研究では，一般ユーザが撮影したパーソナル映像を対象に，映像中から必要なシーンを自動的に切り出すことができるカット編集手法を提案する．提案手法では，一般ユーザが実際にWeb上でカット編集したパーソナル映像とその編集履歴から，ユーザがカットしたシーン（使用シーン）とその切り出し点（カット編集点）における傾向を事前に調査する．このユーザの編集傾向に基づき，映像中の大域的な動き情報とセンサから得られる加速度と角速度を用いてカット編集する．これにより，映像中の動き情報からは，被写体の動きが多く発生しているシーンを捉え，センサ情報からは，カメラワークシーンを正確に捉えることが可能となる．

2　関連研究

本研究では，映像編集にかかるユーザの負担を軽減するために，映像編集の自動化を目指す．そこで本章では，映像の自動編集における従来法について述べる．

篠木らは，講義映像を対象として，高解像度画像を

トリミングすることで，講師を追跡するようなカメラワーク映像を自動生成する手法 [3] を提案している．この手法は，画像中における注目領域の自動編集ができるものの，どのシーンが重要であるかを把握するには映像を全て視聴する必要がある．Rodriguez は，映像を時間軸方向に圧縮し，異なる時間帯に発生した事象を同時に表示させた要約映像を生成する手法 [2] を提案している．映像中から特定の動きをテンプレートマッチングにより求め，特定の動作のみを圧縮した映像を生成しているが，固定カメラを対象とした編集手法である．

一般的なパーソナル映像を対象とした自動編集手法として，Simakov らは，映像の画像サイズの縮小と時間軸の圧縮を同時に行う手法 [7] を提案している．入力映像と編集後の映像の視覚的な完全性と一貫性を捉えることができる双方向類似度を定義し，双方向類似度の最適化問題を解くことで，要約映像を生成している．この手法は，視覚的に違和感のないような要約映像を生成できるが，ユーザの意図を反映させた映像編集ができないという問題がある．Girgenshon らは，カメラワークの動きや映像中における輝度などを用いて，キーフレームを自動的に抽出し映像を再構成する手法 [8] を提案している．このシステムでは，ユーザインターフェースが工夫されており，キーフレームの選択により，ユーザの意図を容易に反映することができる．

そこで，本研究においても，ユーザの意図を反映させた映像編集を行うために，一般ユーザが撮影した映像を編集する際の編集傾向を調査し，編集傾向に基づいた映像の自動編集手法を提案する．一般ユーザの編集傾向を利用することで，ユーザの嗜好に合った映像編集が可能となる．

3　ユーザの編集傾向調査

本研究では，ユーザの嗜好に合ったカット編集を実現するために，一般ユーザの編集傾向を調査する．そこで本章では，傾向調査に使用する動画編集共有サイト ClipCast に投稿されるパーソナル映像について述べ，ユーザによりカット編集されたシーンにおけるカメラワークの傾向について述べる．

3.1　パーソナル映像のオンライン編集

動画編集共有サイト ClipCast[9] は，一般のユーザが映像を Web 上に投稿し，図 1 に示すような編集が Web 上で可能となっているサイトである．ClipCast に投稿されている映像は，図 2 のように人物や動物，風景，スポーツなど様々なシーンを撮影したものである．今回は，2006 年 8 月から 2008 年 8 月までに投稿された 6,518 本の映像の中で，サイト上でカット編集された 1,075 本の映像を対象に調査する．ClipCast では，ユーザが映像編集した際の履歴を保存している．本研究では，編集された映像と編集履歴を用いて一般ユーザの編集傾向を調査する．図 3 のようにカット編集されたシーンの切り出しフレームと切り終わりフレームをそれぞれ「IN 点」,「OUT 点」，切り出されたシーンを「使用シーン」，それ以外のシーンを「未使用シーン」と定義する．また，未使用シーンは使用シーンの前後で「前未使用シーン」,「後未使用シーン」と定義する．

(a) 装飾挿入画面　(b) カット編集画面

図 1　ClipCast における映像編集

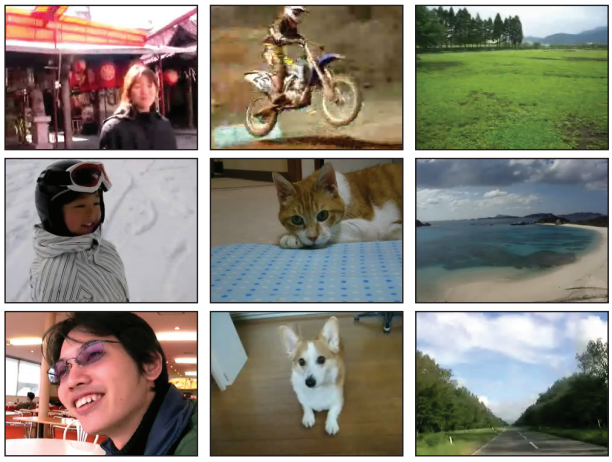


図 2　投稿映像

3.2　使用シーンにおけるカメラワークの発生傾向

文献 [10] や [8] では，カメラ操作部分には撮影者の意図が反映されているため，映像中の重要なシーンであると述べている．そこで，使用シーン，未使用シーンとカメラワーク間の関係を調査するために，それぞれのシーンにおけるカメラワークの有無を調査する．調査方法は，以下に示すカメラワークについて，フレーム単位でどのカメラワークが使用されているか目視にて調査する．

1. 「フィックス」カメラを動かさずに撮影

図 3　シーンの定義

表 1　使用シーンにおけるカメラワークの割合

	前未使用	使用	後未使用
総フレーム数	479,117	1,058,145	421,057
カメラワーク有り	36,412	285,699	34,107
割合 [%]	7.6	27.0	8.1
Follow [%]	30.2	24.1	27.4
Left pan [%]	12.8	17.8	14.5
Right pan [%]	14.9	19.0	16.2
Dolly [%]	13.8	12.7	14.0
Zoom in [%]	8.0	7.2	7.8
Zoom out [%]	9.1	8.0	8.9
Tilt up [%]	5.0	5.9	5.4
Tilt down [%]	6.2	5.2	5.8

2. 「パン」 カメラを左右に振る
3. 「チルト」 カメラを上下に振る
4. 「ズーム」レンズを操作し被写体に寄るまたは引く
5. 「フォロー」動く被写体に合わせカメラを動かす
6. 「ドリー」カメラマン自身が動いて撮影を行う

表 1 に使用シーン，未使用シーンにおける，カメラワークの発生割合の調査結果を示す．表 1 より，前未使用シーン，後未使用シーンについて比較すると，カメラワークの割合について違いが見られなかった．カメラワークの種類については多少の変動があるものの，順位は変わらず顕著な傾向はないといえる．次に，使用シーンと未使用シーンにおけるカメラワークの割合に着目すると，前未使用シーンで 7.6%，後未使用シーンで 8.1% という結果に対し，使用シーンでは 27.0% と 3 倍以上高いことがわかる．これより，ユーザはカメラワークが発生しているシーンを使用シーンとする傾向があるといえる．そこで本研究では，映像中の動きに着目し映像編集をする．

3.3　自動編集に向けての問題点

使用シーンにおけるカメラワークの調査から得られた傾向に基づき，カメラワークシーンを捉えてカット編集する．ここで，カメラワークシーンを捉えるには，映像中の動き特徴量を用いる方法と，カメラのセンサ情報を用いる方法が考えられる．前者では，カメラワークの動きに加えて被写体の動きも同時に捉えることが可能であるが，それらの動きを峻別できない．後者では，カメラワークの動きのみを正確に捉えることが可能であるが，被写体の動きを捉えることはできない．そこで，本研究では，これら 2 つの方法を統合した映像編集手法を提案する．

4　提案手法

本研究では，被写体の動きを捉えつつ正確なカメラワークの推定をするために，時空間特徴とセンサ情報を用いたカット編集手法を提案する．提案手法におけるカット編集の流れを図 4 に示す．提案手法では，時空間特徴である ST-patch 特徴から閾値処理により，映像中における動き変化が多いシーンを推定し，加速度センサとジャイロセンサのセンサ情報からベクトル量子化によりカメラワークシーンを推定する．そして，最後に，これら 2 つのシーンを統合することで，被写体の動き変化が多いシーンやカメラワークシーンを多く含むシーンをカット編集することができる．以下では，それぞれの方法と 2 つのシーンの統合方法について述べる．

4.1　時空間特徴を用いたカット編集点の検出

映像中における動きの変化を捉えた映像編集をするために，ユーザがカット編集した IN 点と OUT 点における動きの傾向を調査し，その傾向に基づきカット編集点を検出する．以下では，映像中の動きを捉えることができる ST-patch と IN 点，OUT 点における動き変化の傾向調査について述べ，ユーザの編集傾向に基づいた ST-patch 特徴によるカット編集点の検出方法について述べる．

4.1.1　Space-Time Patch

画像上の動きの解析を行う手法として，物体の動きの移動量をベクトルデータとして表現したオプティカルフロー [11] が挙げられる．しかし，オプティカルフローは，非剛体の物体に対する動きの解析やテクスチャの異なる物体同士における動きの評価が困難であるなどの問題がある．この問題に対して Shechtman らは，テンプート映像と入力映像の時空間画像から抽出された 2 つの Space-Time Patch(ST-patch) の相関関係を複数用いることにより，テンプレート映像の物体と同じ動きをしている物体を入力映像から識別する手法 [12] を提案している．この手法で用いられている ST-patch の相関関係は，テクスチャが異なる対象物体の動きや，非剛体の物体のように複雑に動く物体に対しても動きを評価することができる．

4.1.2　ST-patch 特徴

ST-patch 特徴は，画像を時間方向に重ねた時空間画像の局所領域である ST-patch から算出する．図 5 に ST-patch の概要図を示す．x, y は画像の座標軸，t は時間軸，3 本の線は個々の画素の動き，$[u\ v\ w]^{\mathrm{T}}$ は ST-patch 内の動きベクトルを表している．この ST-patch 内において，個々の画素 i における x, y, t 軸の勾配値は $\nabla P_i = [P_{xi}\ P_{yi}\ P_{ti}]^{\mathrm{T}}$ となる．この ∇P_i から ST-patch 特徴を式 (1) により算出することができる．

$$\mathbf{M} = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y & \sum P_x P_t \\ \sum P_y P_x & \sum P_y^2 & \sum P_y P_t \\ \sum P_t P_x & \sum P_t P_y & \sum P_t^2 \end{bmatrix} \quad (1)$$

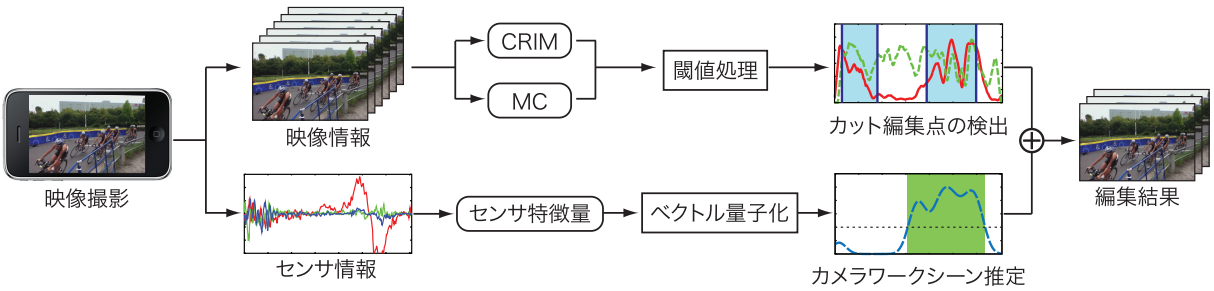


図 **4**　時空間特徴とセンサ特徴を用いたカット編集

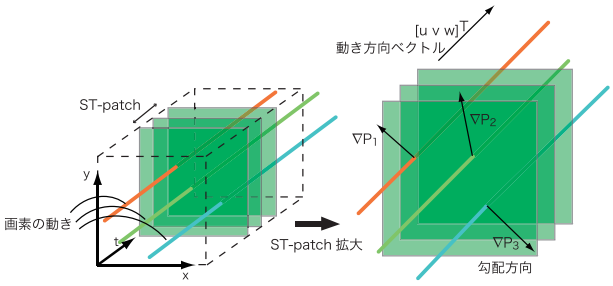


図 **5**　**ST-patch** の概要図

式 (1) より求められる行列 $\mathbf{M}$ が 1 つの ST-patch から得られる ST-patch 特徴となり，左上の 2×2 の行列にはアピアランスの情報，3 行目と 3 列目にはモーションの情報を持つ．この ST-patch 特徴を用いて，ST-patch 内における動きの定常度を表す CRIM（Continues Rank-Increase Measure） Δr と異なる 2 つの ST-patch 間における動きの類似度を表す MC（Motion Correlation） m_{12} を算出することができる．

CRIM（Continues Rank-Increase Measure）

行列 $\mathbf{M}$ の左上の行列を行列 $\mathbf{M}^\diamond$ とし，行列 $\mathbf{M}$ と行列 $\mathbf{M}^\diamond$ より決められる固有値から Continues Rank-Increase Measure（CRIM） Δr を算出する．$\lambda_1 \geq \lambda_2 \geq \lambda_3$ を行列 $\mathbf{M}$ の固有値，$\lambda_1^\diamond \geq \lambda_2^\diamond$ を行列 $\mathbf{M}^\diamond$ の固有値とする．行列 $\mathbf{M}$，行列 $\mathbf{M}^\diamond$ は共に対称行列であり，行列 $\mathbf{M}$，行列 $\mathbf{M}^\diamond$ の固有値はすべて実数となる．また，それぞれの固有値の関係は $\lambda_1 \geq \lambda_1^\diamond \geq \lambda_2 \geq \lambda_2^\diamond \geq \lambda_3$ となる．このような関係から，CRIMΔr を式 (2) のように定義される．

$$\Delta r = \frac{\lambda_2 \cdot \lambda_3}{\lambda_1^\diamond \cdot \lambda_2^\diamond} \qquad (0 \leq \Delta r \leq 1) \qquad (2)$$

ここで求められた Δr の値は，ST-patch 内の動きベクトルが一定の時に $\Delta r \approx 0$ となり，動きベクトルの変化が激しいときに $\Delta r \approx 1$ となる．これにより，ST-patch 内の動きベクトルの定常度を求めることができる．

MC（Motion Correlation）

式 (2) から求められた CRIM Δr より，2 つの ST-patch P1 と P2 の動きの類似度を求めることができる．P1 の CRIM を Δr_1，P2 の CRIM を Δr_2，P1 と P2 を合わせた ST-patch を P_{12} とし，P_{12} の CRIM を Δr_{12} とすると，式 (3) により Motion Correlation（MC） m_{12} を算出する．

$$m_{12} = \frac{\min(\Delta r_1, \Delta r_2)}{\Delta r_{12}} \qquad (0 \leq m_{12} \leq 1) \qquad (3)$$

ここで求められた m_{12} は，異なる 2 つの ST-patch 間における動きの類似度が高い場合 $m_{12} \approx 1$ となり，動きの類似度が低い場合 $m_{12} \approx 0$ となる．

以上の ST-patch より算出される CRIM（動きの定常度）と MC（動きの類似度）を用いてカット編集点における動き変化の傾向を調査する．

4.1.3　カット編集点における CRIM と MC の傾向

ST-patch から算出される CRIM と MC を用いて，カット編集点における動き変化の傾向を調査する．ここで，一般ユーザは使用シーンにおいてカメラワークの発生割合が高いという傾向から，映像中の大域的な動きの変化に着目することが重要であると考えられる．そこで，提案手法では画像全体を ST-patch として，図 6 のように CRIM と MC を算出する．CRIM は図 6(a) のようにカット編集点において生成される ST-patch から算出し，MC は図 6(b) のようにカット編集点の前後で生成される ST-patch から算出する．また，提案手法における ST-patch のフレーム数は 11 とする．

図 7 に，IN 点，OUT 点の ST-patch から算出された CRIM と MC の分布を示す．図 7 より，IN 点と OUT 点の分布に違いは見られない．これより，IN 点，OUT 点において CRIM と MC の傾向には差がないと判断できる．次に，図 7(a) に着目すると，CRIM は低い値により多く分布をしていることがわかる．これよりユーザは，動きの変化が少ないフレームをカット編集点とする傾向があるといえる．また，図 7(b) に着目すると，MC は高い値により多く分布していることがわかる．これよりユーザは，前後フレームにおける動き変化が類似しているフレームをカット編集点とする傾向があるといえる．

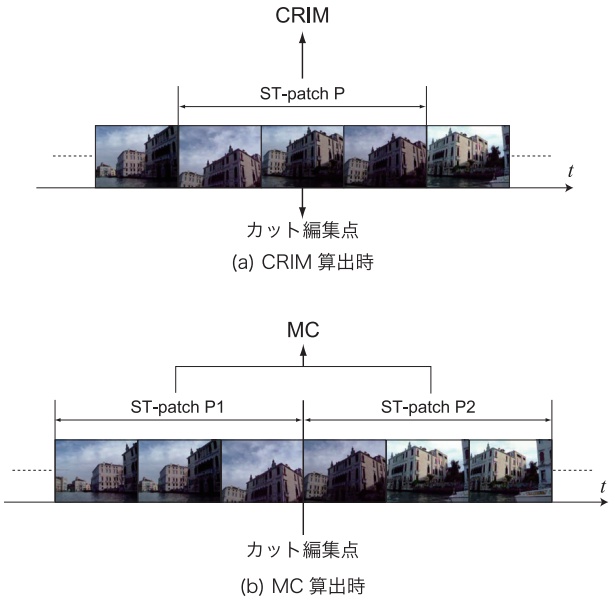


図 **6**　カット編集点における **CRIM** と **MC** の算出方法

以上より，一般ユーザは，映像中の被写体やカメラワークによる動き変化が少なく，前後フレームにおいて類似した動きがあるフレームをカット編集点にする傾向があるといえる．

4.1.4　CRIM と MC を用いたカット編集点の検出

カット編集点における CRIM の値は低く，MC の値は高い値により多く分布しているという傾向に基づき，カット編集点を検出する．検出は以下の 3 ステップで行う．

Step 1 入力映像から各フレームにおける CRIM $R(t)$ と MC $C(t)$ を算出し，移動平均フィルタを用いて平滑化

Step 2 平滑化後の CRIM $R_a(t)$ と閾値が交差するフレーム t をカット候補点 e として式 (4) により算出

$$e = t \qquad (t|R_a(t) = th) \tag{4}$$

Step 3 カット候補点 e の周囲 $\pm d$ フレーム中において，平滑化後の MC $C_a(t)$ が最も高いフレーム t をカット編集点 E として式 (5) により算出

$$E = \underset{t}{\operatorname{argmax}}\, C_a(t) \qquad ((e-d) \leq t \leq (e+d)) \tag{5}$$

検出したカット編集点の間を使用シーンとしてカット編集する．この手法でカット編集されたシーンは，CRIM が高い値に分布しているシーンとなり，映像中の動き変化が激しいカメラワークシーンや被写体の動き変化が多く含まれた映像となる．

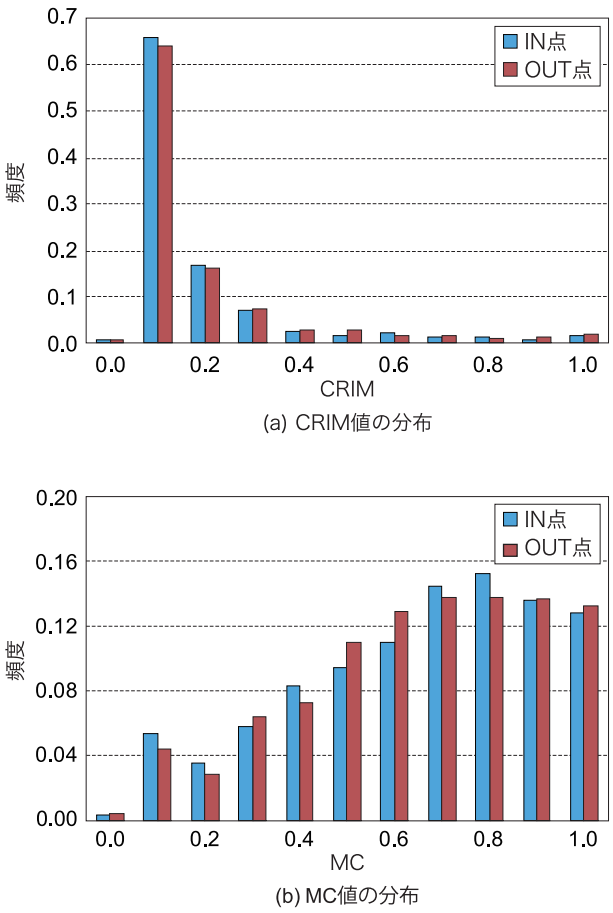


図 **7**　**IN** 点，**OUT** 点における **CRIM** と **MC** の分布

4.2　センサ情報を用いたカメラワークの推定

映像中における動き特徴を用いてカメラワークを推定すると，被写体の動きにより正しい推定が困難となる．そこで，携帯電話に搭載されている加速度センサとジャイロセンサからセンサ情報を取得し，映像撮影時のカメラワークを推定する．本研究におけるセンサ情報の取得には iPhone4[13] を用いる．iPhone4 には，3 軸加速度センサと 3 軸ジャイロセンサが搭載されており，x, y, z 軸の加速度と x, y, z 軸の角速度を取得することができる．図 8 に，iPhone 上におけるセンサ座標系を示す．これらの値を用いて，映像撮影時のカメラワークを推定する．以下では，センサ情報を用いたカメラワーク推定方法と，推定結果を用いた映像編集手法について述べる．

4.2.1　ベクトル量子化によるカメラワーク推定

映像撮影時に取得できる加速度と角速度を特徴量として，ベクトル量子化によりカメラワークを推定する．推定するカメラワークは，パンライト，パンレフト，チルトアップ，チルトダウン，フィックスの 5 種類とし，

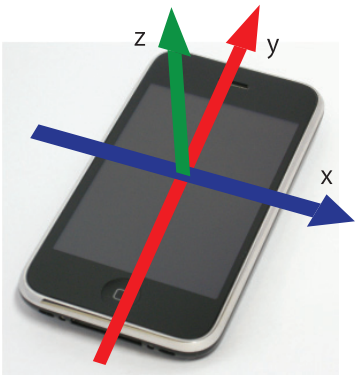


図 8　iPhone 上におけるセンサ座標系

以下の 3 ステップで行う．

Step 1 入力センサ情報から各フレームにおける加速度 $Sa(t)$ と角速度 $Sv(t)$ を算出し，移動平均フィルタを用いて平滑化

Step 2 平滑化後の加速度 $Sa_a(t)$ と角速度 $Sv_a(t)$ を合わせてセンサ特徴量 $S(t)$ とし，予め作成されたコードブックを用いてベクトル量子化により，クラス I の代表ベクトル V_I とのユークリッド距離が最小となるクラス I を式 (6) により算出

$$I = \underset{I}{\operatorname{argmin}} |S(t) - V_I| \tag{6}$$

Step 3 算出されたクラス I のコードブックからクラス I に属するサンプルのカメラワークラベル c よりカメラワーク確率 $p(c|I)$ を計算し，カメラワーク確率が最大となるカメラワークラベル C^{max} を式 (7) により算出

$$C^{max} = \underset{c}{\operatorname{argmax}}\, p(c|I) \tag{7}$$

これにより，被写体の動きに左右されずに各フレームにおけるカメラワークを推定することができる．

4.2.2　カメラワーク推定結果を用いたカット編集

センサ情報によるカメラワーク推定結果を用いて映像をカット編集する．まず，カメラワーク推定結果から，カメラワークが有るフレームでは 1，カメラワークが無いフレームでは 0 の値をとるカメラワーク有無 $F(t)$ を算出する．

次に，このデータに対してカーネル密度推定を行い，確率密度関数 $k(x)$ を算出する．バンド幅 h，N 個のデータ数の中にデータ点 x があるとき，カーネル密度推定で算出される確率密度関数 $k(x)$ は式 (8) となる．

$$k(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\sqrt{2\pi h^2}} \exp\left\{-\frac{|x - x_i|^2}{2h^2}\right\} \tag{8}$$

カーネル密度推定により，カメラワーク有無 $F(t)$ の疎な分布は密な分布として扱うことができる．生成された確率密度関数 $k(x)$ が閾値以上のシーンをカメラワーク発生確率が高いシーンとしてカット編集する．

4.3　時空間特徴とセンサ情報を用いた映像編集

映像中における被写体の動きを捉えつつカメラワークシーンを正確に推定したカット編集をするために，時空間特徴とセンサ情報を用いてカット編集する．

まず，映像情報から CRIM $R(t)$ と MC $C(t)$ を平滑化した $R_a(t)$ と $C_a(t)$ を算出し，センサ情報である加速度 $Sa(t)$ と角速度 $Sv(t)$ を平滑化した $Sa_a(t)$ と $Sv_a(t)$ から $S(t)$ を算出する．

次に，平滑化後の CRIM $R_a(t)$ に対して閾値処理によりカット候補点 e を算出し，カット候補点の間を使用シーンとする．さらに，センサ特徴量 $S(t)$ のベクトル量子化により各フレームのカメラワークを推定し，その結果に対してカーネル密度推定を行い，得られた確率密度関数 $k(x)$ に対して閾値処理により，カメラワークシーンを算出する．そして，時空間特徴を用いて推定した使用シーンとセンサ情報を用いて推定したカメラワークシーンの OR をとり，推定結果を統合する．

最後に，推定結果を統合したシーンの IN 点，OUT 点をカット候補点 e として，式 (5) により平滑化後の MC $C_a(t)$ が最も高いフレーム t をカット編集点 E として検出しカット編集する．

これにより，時空間情報により被写体の動きがあるシーンを推定し，センサ情報によりカメラワークシーンを正確に推定したカット編集が可能となる．また，ユーザの編集傾向に基づいてカット編集点の検出をしているため，ユーザの嗜好に合った映像編集が可能となる．図 9 にカット編集例を示す．

5　実験

提案手法の有効性を示すために，カメラワーク有無の推定精度の比較実験とアンケート調査によるカット編集映像の評価実験を行う．

5.1　カメラワーク有無の推定実験

センサ情報を用いたカメラワーク推定法の有効性を示すために，カメラワーク有無の推定精度を比較する．実験は，センサ情報を特徴量としてベクトル量子化によりカメラワークを推定する手法と，ST-patch 特徴により算出される CRIM が閾値以上のフレームをカメラワーク有りとする手法の推定精度を比較する．評価に使用する映像は，カメラワーク無しが 7,151 フレーム，カメラワーク有りが 7,756 フレームである．表 2 にカメラワーク有無の検出結果を示す．

表 2 より，センサ情報を用いた検出率は，ST-patch 特徴のみに比べ 34.5%向上している．これは，ゆっくり

表 **2**　カメラワーク有無の推定結果 **[%]**

	検出率	誤検出率
センサ情報	93.0	6.1
ST-patch	58.5	13.2

と変化するカメラワークシーンに対して CRIM の値が低くなり，検出ができないためである．誤検出率においては，ST-patch 特徴を用いた手法は 7.1%増加している．これは，センサ情報を用いることで，被写体の動きに左右されずに，カメラ自身の動きを捉えることが可能となったためである．以上より，カメラワークの有無の判定におけるセンサ情報の有効性を確認できる．

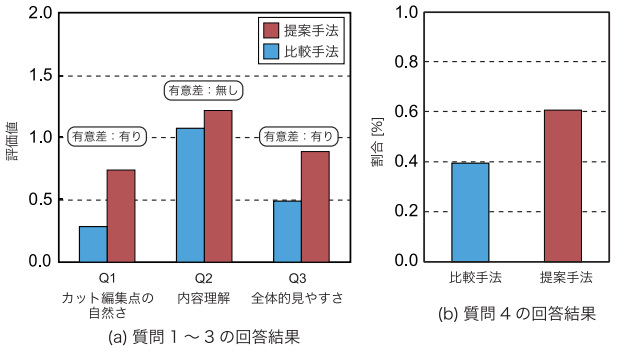
5.2　アンケート調査による評価実験

提案手法によりカット編集した映像が，どの程度見やすい映像であるのか視聴者に対するアンケート調査により評価する．比較手法として，ST-patch 特徴のみを用いた映像編集手法を用いる．評価方法は，編集前の映像を視聴した後に，比較手法により編集した映像と提案手法により編集した映像を順不同で被験者に提示し，以下に示す項目に対するアンケートをとる．

Q1 カット編集点は自然であるか
Q2 内容は理解できたか
Q3 全体的に見やすい映像であるか
Q4 2 つ映像のうちどちらの映像がよかったか

評価は 5 段階評価 [とても良い : 2，良い : 1，どちらでもない : 0，悪い : -1，とても悪い : -2] で，被験者の人数は 23 人である．図 10 にアンケート調査の結果を示す．

図 10(a) より提案手法は，カット編集点の自然さと全体的な見やすさという点において比較手法に比べて高い評価を得ることができた．これは，t-検定 (有意水準 5%) の結果，両手法間に有意差があることを確認した．提案手法では，動き変化が少くなく，前後フレームで動き変化が類似しているフレームをカット編集点として検出しているため，自然なカット編集点となり，ユーザの意図が含まれたカメラワークシーンをセンサ情報により正確に捉えているため，見やすい映像となったと考えられる．さらに，図 10(b) においても，提案手法の映像が良いという回答の割合が 0.21 多いことがわかる．以上より，提案手法は，センサ情報を用いてカメラワークを正確に捉え，ST-patch 特徴により映像中における被写体の動きを捉えることで，視聴者の満足度の高い映像編集が可能であることを確認した．



(a) 質問 1 ～ 3 の回答結果　　(b) 質問 4 の回答結果

図 **10**　アンケート評価実験結果

6　おわりに

本研究では，センサ情報と時空間特徴を用いて一般ユーザの編集傾向に基づいた映像の自動編集手法を提案した．センサ情報と時空間特徴を用いることにより，カメラワークと映像中における被写体の動きを同時に捉えることで，自然な映像編集を可能とし，アンケートによる評価実験により提案手法の有効性を確認した．今後は，手ぶれシーンを考慮した映像の自動編集を検討する．

参考文献

[1] Y. Pritch, A. Rav-Acha, and S. P. and: “Nonchronological Video Synopsis and Indexing”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 30, No. 11, pp. 1971–1984, 2008.

[2] M. Rodriguez: “CRAM: Compact Representation of Actions in Movies”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR’10)*, 2010.

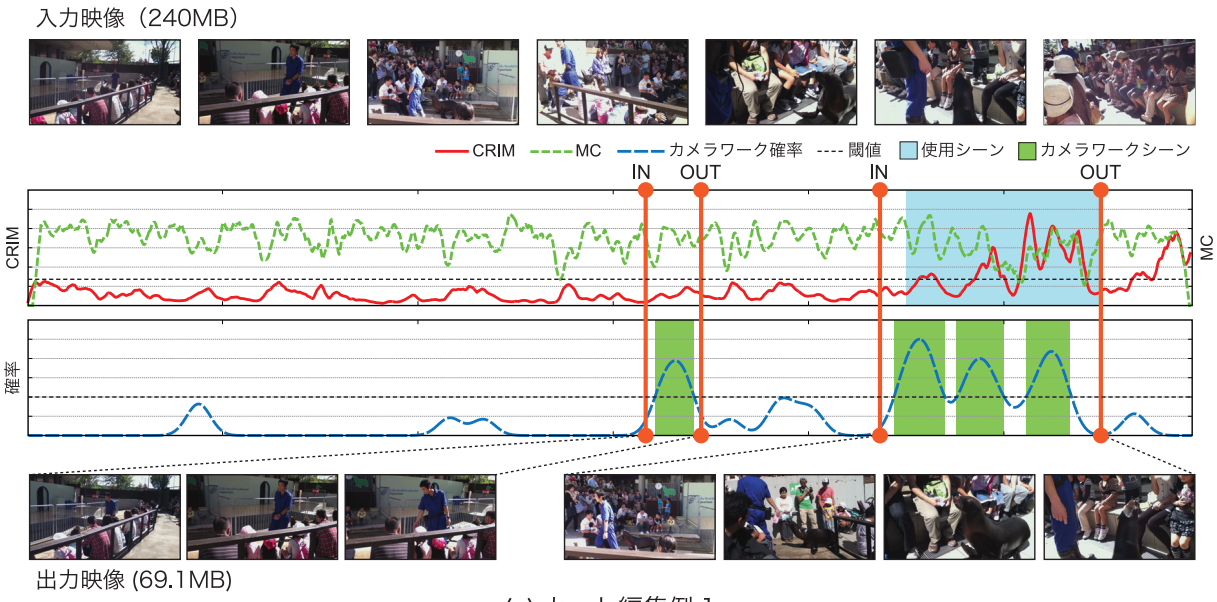
[3] 篠木雄大, 藤吉弘亘: “高解像度映像からの視聴者の注目点を考慮した講義映像の自動生成”, 映像情報メディア学会誌, Vol. 62, No. 2, pp. 240–246, 2008.

[4] 篠木雄大, 藤吉弘亘: “キーフレーム間の双方向物体追跡結果に基づく仮想カメラワーク映像の生成法”, 映像情報メディア学会誌, Vol. 63, No. 2, pp. 209–215, 2009.

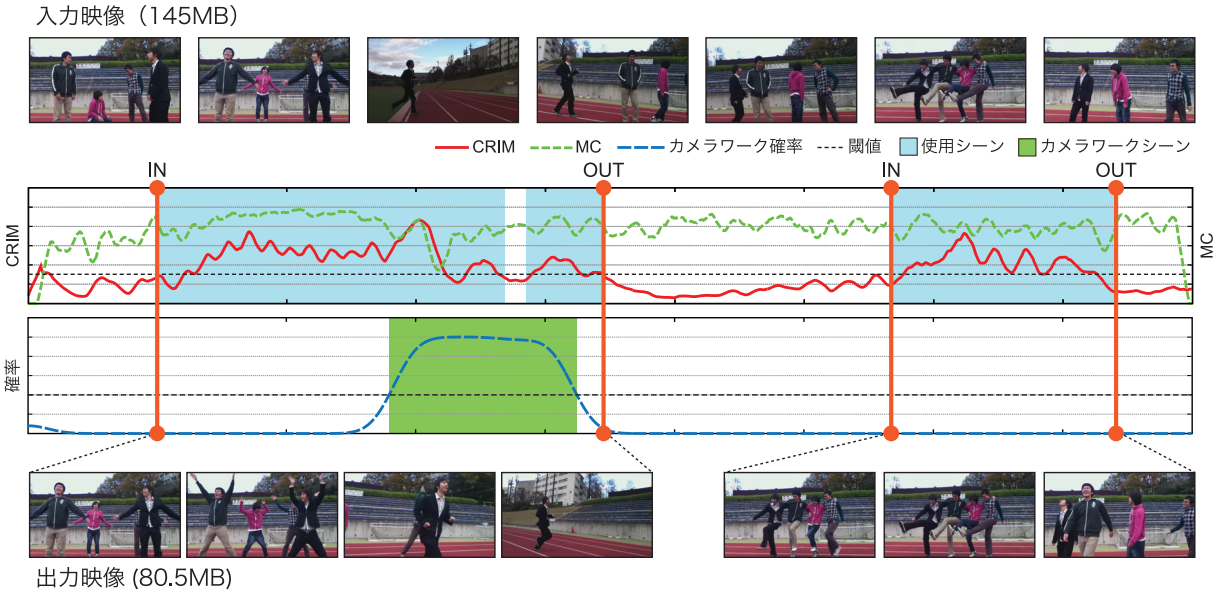
[5] M. Kumano, Y. Ariki, K. Tsukada, S. Hamaguchi, and H. Kiyose: “Automatic extraction of PC scenes based on feature mining for a real time delivery system of baseball highlight scenes”, *IEEE International Conference on MULTIMEDIA & EXPO (2004)*, pp. 277–280, 2004.

[6] 尾関基行, 中村裕一, 大田友一: “注目換気行動に基づいた机上作業映像の編集”, 電子情報通信学会論文誌 D-II, Vol. J88-D-II, No. 5, pp. 844–853, 2005.

[7] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani: “Summarizing Visual Data Us-



(a) カット編集例 1



(b) カット編集例 2

図 **9**　カット編集例

ing Bidirectional Similarity", *IEEE Conference on Computer Vision and Pattern Recognision (CVPR'08)*, 2008.

[8] A. Girgensohn, J. Boreczky, P. Chiu, J. Doherty, J. Foote, G. Golovchinsky, S. Uchihashi, and L. Wilcox: "A Semi-automatic Approch to Home Video Editing", *Proc. User Interface Software and Technology 2000, ACM Press, New York,*, pp. 81–89, 2000.

[9] "ClipCast". http://clipcast.jp.

[10] M. Smith and T. Kanade: "Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques", *International Workshop on Content-Based Access of Image and Video Databases (CAIVD'98)*, pp. 61–70, 1998.

[11] B. Horn and B. Dchunck: "Determining optical flow", *Artificial Intelligence*, Vol. 17, pp. 185–203, 1981.

[12] E. Shechtman and M. Irani: "Space-time behavior based correlation -or- how to tell if two underlying motion fields are similar without computing them?", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 29, No. 5, pp. 2045–2056, 2007.

[13] Apple. "iPhone". http://www.apple.com/jp.