H-004

# 画像片リストに基づいたショットの統合によるシーン系列生成

## Video Scene Sequence Generation by Shot Integration based on Image Peace List

望月 貴裕† 佐野 雅規†
Takahiro MOCHIZUKI Masanori SANO

## 1．まえがき

近年，アーカイブスなどに大量の番組映像が蓄積される時代となり，それらを活用するために構造化して効率的にハンドリングする技術が必要とされている．映像構造化の基本単位としては，映像が切り替った編集点において分割した「ショット」が一般的である．しかし，映像を検索して閲覧・利用する際には，連続する複数のショットを「同じ場面（場所，状況）」などの意味的要素で統合した「シーン」単位で扱うニーズがあり，そのための技術が考案されている．

[1]では音響信号を利用してショットを統合しているが，音響信号が付与されていない，あるいは音響信号の品質が劣悪な映像は処理が困難である．また [2][3]では，ショット代表画像の色ヒストグラムの差異を，ショット統合の基準となる「場面転換」検出の尺度としている．しかしこの手法では，同じ場面のショット代表画像にもかかわらず，全体の色特徴が大きく変化した場合に，誤って場面転換点と判断してしまうケースが生じる（図 1）．このような誤検出を抑制するために，各ショット代表画像の「色リスト」とその前複数枚の「統合色リスト」との包含関係に基づき場面転換を検出する手法が提案されている[4]．しかしこの手法は，空間的に細かい色の変化を持つ映像に対して精度が落ちる可能性がある．

そこで本稿では，[4]の手法を拡張し，「色」ではなく「画像片（ブロック画像）」のリストを用いて場面転換を検出する手法を提案する．「画像片リスト」は[5]で提案された「多重スケール画像片ヒストグラム」に基づいて生成される．本提案手法により，従来手法よりも高精度な場面転換検出によるシーン系列の生成が可能となる．

## 2．多重スケール画像片ヒストグラム

本章ではまず，場面転換検出に用いる「画像片リスト」を生成するための基本特徴となる，ショット代表画像の「多重スケール画像片ヒストグラム（H-MIPW）」[5]について述べる．H-MIPW はショットやシーンの内容に基づく類似検索のための特徴であり，場面転換検出の大きな手がかりとなり得るものである．

### 2.1．画像片ワードの生成

本節では，処理対象映像に出現する画像片（ブロック画像）の種類を表す画像片ワード（IPWord）を生成する手法について述べる．IPWord は，処理対象映像から無作為に選んだフレーム画像集合を用いて生成する．

1. 各フレーム画像を，スケール 1（$n_{W1}\times n_{H1}$個），…，スケール $N_d$（$n_{WN_d}\times n_{HN_d}$個）の複数スケールにブロック分割する．
2. 各ブロック画像の特徴ベクトル（色特徴およびテクスチャ特徴により構成）を計算する．
3. 各スケール $i$ ($i$=1, …,$N_d$) において，ブロック画像集合を特徴ベクトルの類似性に基づいてクラスタリングする．このようにして生成された，各スケール $i$ における $K_i$ 個のクラスタを $C[i,1], \cdots, C[i, K_i]$とする．
4. 各クラスタ $C[i,k]$の中心ベクトル $\mathbf{w}[i,k]$ を要素（ワード）とする $\mathbf{W}=\{\mathbf{w}[1.1],\cdots,\mathbf{w}[i,k],\cdots,\mathbf{w}[N_d,K_{N_d}]\}$を IPWord とする．

$N_d$=2 の場合の IPWord 生成の流れを図 2 に示す．

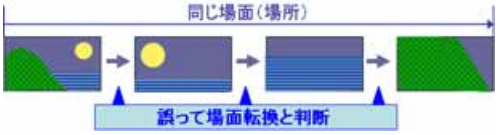


図 1　従来の場面転換検出手法の問題点

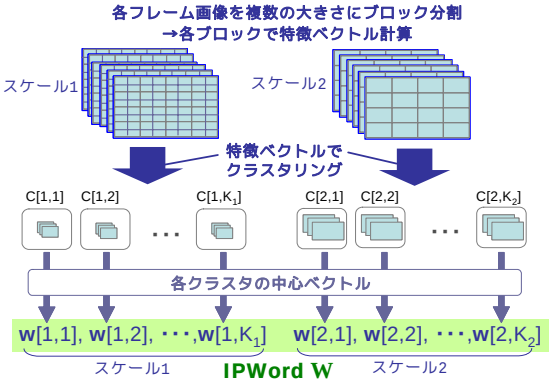


図 2　画像片ワードの作成

### 2.2．多重スケール画像片ヒストグラムの算出

本節では，2.1 節で生成した IPWord に基づき，処理対象映像の各ショット代表画像の H-MIPW を算出する手法について述べる．

1. IPWord $\mathbf{W}$ のワード数と同数のビンから成るヒストグラム $\mathbf{H}=\{\mathbf{h}[1.1],\cdots,\mathbf{h}[i,k],\cdots,\mathbf{h}[N_d,K_{N_d}]\}$を準備し，各要素を 0 とする．
2. ショット代表画像をスケール 1,…, スケール $N_d$ の複数スケールにブロック分割し，各ブロック画像について特徴ベクトルを計算する．
3. 各スケール $i$ ($i$=1, …,$N_d$)について以下の処理を行う．
   3.1 各ブロック画像の特徴ベクトルと最も類似度の高い IPWord $\mathbf{W}$ の ベクトルを $\mathbf{w}[i, k']$とし，対応する $\mathbf{H}$ の要素 $\mathbf{h}[i, k']$に 1 を加算する．
   3.2 各 $\mathbf{H}$ の要素 $\mathbf{h}[i, k]$ ($k$=1,⋯,$K_i$)をブロック画像数で除算し正規化する．
4. 計算された $\mathbf{H}=\{\mathbf{h}[1.1],\cdots,\mathbf{h}[i,k],\cdots,\mathbf{h}[N_d,K_{N_d}]\}$をこのショット代表画像の H-MIPW とする．

$N_d$=2 の場合の処理の流れを図 3 に示す．

## 3．提案手法によるシーン系列生成

本提案手法の流れを以下に示す．番組のショット分割およびショット代表画像抽出は[6]などの手法により処理済と

† NHK 放送技術研究所

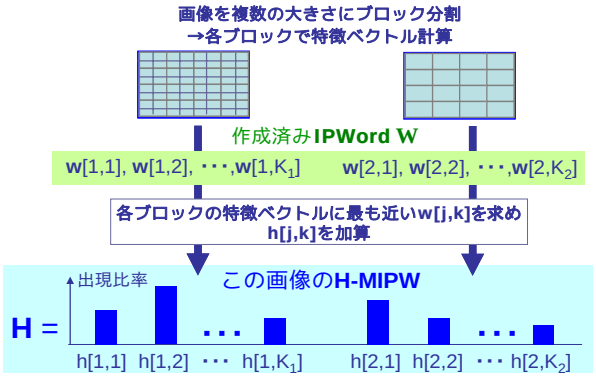


図 3　多重スケール画像片ヒストグラムの算出

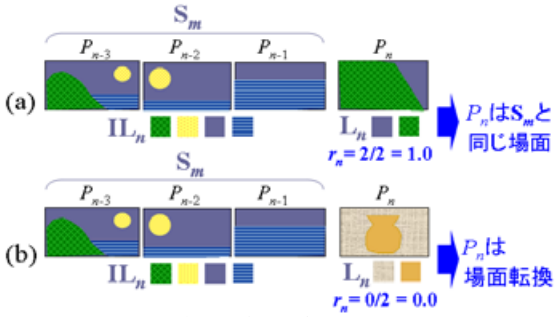


図 4　提案手法の処理の流れ

する．入力ショット代表画像系列を $P_n$（$n$=1,…,$N$），生成されるシーン系列を $\mathbf{S}_m$（$m$=1,…,$M$）とする．

1. $n$=1，$m$=1 とする．
2. $n > N$ ならば終了．そうでない場合は，H-MIPW $\mathbf{H}_n$ を，画像 $P_n$ について算出する．
3. $\mathbf{H}_n$ の要素の中で，$\mathbf{h}[i, k] \geqq Th_H$ を満たす $\mathbf{h}[i,k]$に対応した $\mathbf{w}[i, k]$の集合を $P_n$ の「画像片リスト」$\mathbf{L}_n$ とする．
4. $n$=1 の場合は $\mathbf{S}_m$ に $P_n$ を加え，$n$=$n$+1 として 2 へ．それ以外は 5 へ．
5. $\mathbf{L}_{n-1}$ から $\mathbf{L}_{n-C}$ までを統合し，現在のシーンの「統合画像片リスト」$\mathbf{IL}_n$ とする．ここで $C$=min($n$-1, $C_{CN}$)であり，$C_{CN}$ は「どこまでショットを遡って $\mathbf{IL}_n$ を生成するか」を表す固定値である．
6. $\mathbf{L}_n$ の $\mathbf{IL}_n$ への包含率 $r_n$ を以下のように計算する．

$$r_n = \frac{\mathbf{L}_n\text{の中で}\mathbf{IL}_n\text{に含まれている画像片数}}{\mathbf{L}_n\text{の画像片数}} \quad (1)$$

7. $r_n \geqq Th_R$ の場合は，「$P_n$ は現在のシーンと同じ場面の画像（場面転換なし）」と判断し（図 4(a)），$\mathbf{S}_m$ に $P_n$ を加え，$n$=$n$+1 として 2 へ．それ以外は 8 へ．
8. 「$P_n$ は現在のシーンと違う場面のショット代表画像（場面転換点）」とし（図 4(b)），$m$=$m$+1 として $\mathbf{S}_m$ に $P_n$ を加え，$n$=$n$+1 として 2 へ．

## 4．シーン系列生成実験

本手法の有効性を検証するために，「場面」の定義が比較的容易である自然番組映像を用いて，従来手法との場面転換点の検出精度を比較した．

- **使用映像**: 15 分～30 分の NHK 自然番組映像 20 本．
- **提案手法の各パラメータ**: スケール数 $N_d$=2，$C_{CN}$=3
  - スケール 1：ブロック数 16×16，$K_1$ =750，$Th_H$=1/$K_1$
  - スケール 2：ブロック数 8×8，$K_2$ =750，$Th_H$=1/$K_2$


図 5　正解場面転換の例

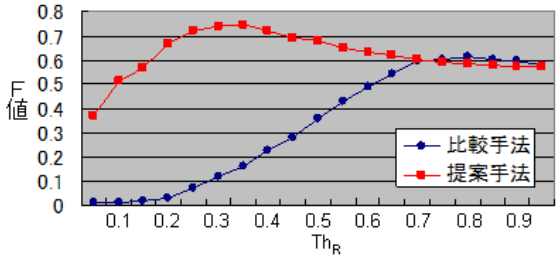


図 6　精度比較

- **比較手法**: [4]で提案された「色リスト」による手法
  - RGB ヒストグラムのビン数：1000（RGB 各 10 分割）
  - $Th_H$ =1/1000，$C_{CN}$=3
- **正解設定**: 街の中→海岸，小屋の中→森など「大きく場所が切替ったショット代表画像」を場面転換点とする（例を図 5 に示す）．使用映像における正解場面転換数は 304 であった．
- **評価**: 再現率 $R$ と適合率 $P$ の調和平均である F 値で評価する．$R$, $P$, F 値は次のように算出する．

  - $R$ = 全ての正解場面転換のうち検出されたものの比率
  - $P$ = 全ての検出場面転換のうち正解であるものの比率
  - F 値 = $2RP/(R+P)$ (2)

提案手法および比較手法において，3 章の処理 7 で用いる $Th_R$ を 0.05～0.95 まで 0.05 刻みに変化させて場面転換検出処理を行い，閾値ごとに全映像トータルでの F 値の算出を行なった．図 6 にその結果を示す．横軸が $Th_R$，縦軸が F 値である．比較手法の F 値の最高値 0.61 に対し提案手法は 0.75 となり，提案手法の有効性が示された．

## 5．あとがき

本稿では，番組の各ショット代表画像の画像片リストとその前複数枚のショット代表画像の統合画像片リストとの包含関係に基づき場面転換を検出し，同じ場面のショットを統合してシーン系列を生成する手法について述べた．今後は，精度改善に取り組むとともに，放送局の映像アーカイブスに対する実用を視野に入れ，本手法を用いたシーン単位での検索システムの構築を進めていく．

## 参考文献

[1] 山本, 長谷山：“映像の構造に基づいたシーン分割に関する一検討,” 信学技報, IE2007-229, pp. 61-66 (2008)
[2] 谷澤, 上原: “動画像の特徴量を用いた意味的構造の自動検出,” 情処研報, データベース・システム研究会報告, 2000(10), pp. 75-82 (2000)
[3] 青木, 下辻, 堀：“映像ブラウジングのための類似ショット統合,” 情処研報, ヒューマンインタフェース研究会報告, 96(62), pp. 43-50 (1996)
[4] 福田, 望月, 佐野, 藤井: “統合色リストに基づく番組映像のシーン系列生成,” 映像情報メディア学会年次大会予稿集, 23-8 (2012)
[5] 望月, 佐野, 藤井: “多重スケール画像片ワードヒストグラムを用いた映像検索,” 信学技報,vol.112, No.385, PRMU2012-89(MVE2012-54),2013, pp.75-80 (2013)
[6] 河合, 住吉, 八木: “逐次的な特徴算出によるディゾルブ,フェードを含むショット境界の高速検出手法,” 信学論(D), Vol.J91-D, No.10, pp. 2529-2239 (2008)