H-019

テクスチャ特徴を利用した動物番組からの被写体抽出に関する検討

A study of object detection method for an animal TV program based on texture features

河合 吉彦 †　住吉 英樹 †　八木 伸行 †
Yoshihiko Kawai　Hideki Sumiyoshi　Nobuyuki Yagi

1 まえがき

大量の映像を効率的に検索するためには，意味内容に基づいた有効なインデクスが必要不可欠である．このような映像の索引付けにおいて，どの区間にどのような被写体が映っているかということは，最も重要な情報のひとつである．そこで本研究では，動物番組を対象に，映像中に出現する動物被写体を画像処理によって検出することを試みる．提案手法では，各キーフレームにおけるブロック領域に対して様々なテクスチャ特徴を算出し，ランダムフォレスト法によって被写体とそれ以外の背景領域とに分類を試みる．実験では，実際の放送番組に対して提案手法を適用し，被写体領域および被写体出現フレームの検出精度を評価する．

2 動物番組映像からの被写体抽出

提案手法の概要を図 1 に示す．まず，各ショットの中間位置にあるフレームをキーフレームとして抽出する．提案手法ではこのキーフレームをショットの代表画像として使用することで，計算負荷の軽減を図る．次に，抽出されたキーフレームを複数のブロック領域に分割し，各領域に対して様々な画像特徴量を算出する．続いて，ランダムフォレスト識別器 [1] では，算出された特徴量を用いて，各領域が被写体領域であるかどうかを判定する．最後に，ブロック領域の分類結果を統合しフレームに被写体が映っているかを判定する．以降では，提案手法で利用するテクスチャ特徴と，ランダムフォレスト，被写体フレームの判定について説明する．

2.1 テクスチャ特徴

テクスチャ特徴としては，映像特徴解析の従来研究 [2] において，有効性が示されている特徴量を考慮して選択した．具体的には，カラーモーメント特徴，エッジ方向ヒストグラム，ガボール（Gabor）特徴，ローカルバイナリパターン（LBP: Local Binary Pattern）[3] の 4 種類を用いた．加えて，提案手法では元のフレームにおけるブロックの位置も特徴量として利用した．それぞれのテクスチャ特徴について簡単に説明する．

2.1.1 カラーモーメント特徴（18 次元）

入力画像を HSV 色空間，Lab 色空間のそれぞれに変換し，各コンポーネントに対して，画素値の平均 μ，標準偏差 σ，歪度の立方根 s を算出する．

2.1.2 エッジ方向ヒストグラム（37 次元）

-90 度から +90 度の範囲を 5 度ごとに区切った 36 方向と，非エッジ点について頻度ヒストグラムを求め，特徴量とする．検出には Sobel フィルタを用いる．

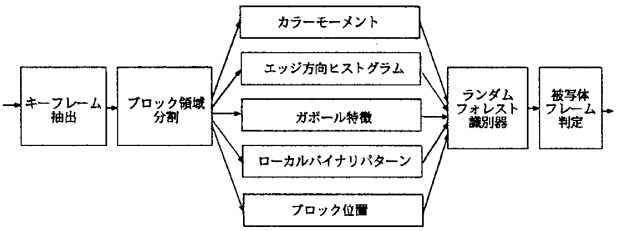

図 1　手法の概要

2.1.3 ガボール特長（48 次元）

ガボールフィルタは濃淡特徴の方向と周期を特徴量として抽出できるフィルタである．解像度 m，方向 n のガボールフィルタを式（1）に示す．

$$g_{mn}(x,y) = \frac{k_m^2}{\sigma^2}\exp\left\{-\frac{k_m^2(x^2+y^2)}{2\sigma^2}\right\} \times\left[\exp\{jk_m(x\cos\theta_n+y\sin\theta_n)\}-\exp\left(-\frac{\sigma^2}{2}\right)\right] \quad (1)$$

ここで，$k_m = a^m (0 \le m \le S-1)$，$\theta_n = n\pi/K\ (0 \le n \le K-1)$ である．提案手法では，$\sigma = 2.5, a = \sqrt{2}, S = 4, K = 6$ とした．上記のフィルタを入力画像に畳み込み，その結果における平均と標準偏差を特徴量として利用する．

2.1.4 ローカルバイナリパターン（54 次元）

LBP[3] は注目画素に対する周辺画素の濃度の大小パターンを表した特徴量である．半径 R の位置にある P 個の画素の LBP は式（2）で算出できる．

$$LBP_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(g_p), & if\ U(L_{P,R}) \le 2 \\ P+1, & otherwise \end{cases} \quad (2)$$

$$s(g_p) = \begin{cases} 1, & g_p - g_c \ge 0 \\ 0, & otherwise \end{cases} \quad (3)$$

ここで，g_c は注目画素の画素値を表し，g_p は座標 $(R\sin(2\pi p/P), R\cos(2\pi p/P))$ の画素値を表す．また，U は 0 と 1 が変化する箇所の数を表し，下式で算出される．

$$U(L_{P,R}) = |s(g_{P-1})| + \sum_{p=1}^{P-1}|s(g_p) - s(g_{p-1})| \quad (4)$$

解像度変化に耐性を持たせるため，提案手法では $(P,R) = (8,1),(16,2),(24,3)$ の三種類の組み合わせを使用する．各 LBP の頻度ヒストグラムを求め特徴量とする．

2.1.5 ブロック位置（2 次元）

フレーム内でのブロック位置を x，y 座標で表す．

† NHK 放送技術研究所

図 2　検出された被写体フレームの例

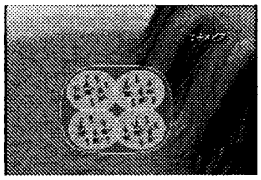

図 3　誤検出の例

2.2　ランダムフォレスト法

ランダムフォレスト法 [1] は，アンサンブル学習の一種であり，多数の決定木（CART）を組み合わせることによって高い分類精度を実現する手法である．バギングやブースティングなどと比較して性能が優れているとされている．アルゴリズムを以下に示す．

(1) 学習データから，作成する木の数だけのブートストラップサンプルを作成する．
(2) 作成されたブートストラップサンプルを用いて決定木を学習する．各ノードの分岐においては，M 個の全特徴量の中から m 個（$m < M$）をランダムに選択し，m の中から最良のものを分岐に使用する．木は最大まで成長させ、枝刈りはしない．
(3) 識別器の出力結果は，全ての決定木の出力の多数決によって決定する．

ランダムフォレストでは，ノード分岐に用いる特徴量をランダム抽出するため，高次元の特徴ベクトルに対しても，短時間で学習することができる．

2.3　被写体フレームの判定

ランダムフォレスト法によって被写体領域と判定されたブロック領域について，近接する被写体領域を再帰的に統合していく．最終的に，フレーム内に閾値以上の面積を持つ領域が存在すれば，被写体フレームと判定する．

3　評価実験

提案手法を用いて，実際に放送された動物番組に対して被写体の抽出を試みた．実験には，ふしぎ大自然「大絶壁をヒヒが登る」を利用した．番組長は 43 分である．今回の実験では，図 2 に示すような“ヒヒ”を検出対象の被写体に設定した．番組映像に含まれるキーフレームの総数は 239 フレームであり，そのうち番組前半の 120 フレームを学習データとして使用し，残りの 119 フレームを検出精度の評価に使用した．学習データについては，キーフレームの各ブロック領域に対して人手で正解を付与した．なお，キーフレームの解像度は 720×480 画素であり，各フレームを 64×64 画素のブロック領域に分割した（合計 77 ブロック）．また，ランダムフォレスト法における木の総数は 500 本に設定した．評価には次式で表される再現率および適合率を用いた．

$$\text{再現率} = N_b/N_g, \qquad \text{適合率} = N_b/N_o \qquad (5)$$

表 1　ブロック領域単位での検出精度

	再現率	適合率
被写体以外	90% (6592/7308)	95% (6592/6872)
被写体	84% (1575/1855)	68% (1575/2291)

表 2　フレーム単位での検出精度

	再現率	適合率
被写体以外	76% (38/50)	95% (38/40)
被写体	97% (67/69)	85% (67/79)

N_g は正解数，N_o は提案手法による検出数，N_b は正解のうち，提案手法でも検出できた数を表す．

表 1 に，ブロック領域単位での検出結果を示す．被写体以外については，再現率，適合率とも 90% 以上の非常に高い結果となった．それに対して，被写体のブロック領域については，再現率、適合率とも精度が低下した．これは，被写体領域に対する学習データが，被写体以外に対する学習データと比較して，約 1/4 と少なかったことが原因のひとつとして考えられる．さらなる精度向上のためには，分類精度の高い有効な特徴量を検討する必要があると考える．

次に，フレーム単位での評価結果を表 2 に示す．実験では，被写体が一定の大きさ以上に映されているフレームを正解とした．実験の結果，被写体検出の再現率が 97%，適合率が 85% と，ブロック単位の結果と比較して精度が向上した．被写体の出現フレームでは，ブロック単位で未検出があっても，その他の領域が正しく検出されていれば未検出とならなかったため再現率が向上したと考えられる．また，誤検出される被写体領域は，フレーム内に散らばって存在し，大きな領域を形成することが少なかったため，フレーム単位での誤検出が軽減された．しかし，動物被写体と類似したテクスチャを持つ領域がフレームの大きな部分を占める場合には誤検出となった．図 3 に誤検出の例を示す．この例では，フレームの右側の領域が被写体と誤検出された．被写体の形状を考慮するなどの検討が必要である．

4　あとがき

本稿では，テクスチャ特徴に基づいた画像解析による動物被写体の検出手法を提案した．提案手法では，映像から抽出したキーフレームのブロック領域に対して，様々なテクスチャ特徴を算出し，ランダムフォレスト法によって被写体とそれ以外の領域とに分類を試みた．実際に放送された動物番組に対する実験では，再現率 97%，適合率 85% という精度で被写体出現フレームを検出することができ，提案手法の有効性が確認できた．今後は，他の動物に対する実験によって本手法の汎化性を検証するとともに，動物以外の被写体の検出手法などについても検討をすすめたい．

参考文献

[1] L. Breiman, “Random Forests,” Machine Learning, vol.45, pp.5–32, 2001.
[2] TREC video retrieval evaluation online proceedings, http://www-nlpir.nist.gov/projects/tvpubs/tv.pubs.org.html
[3] T. Ojala M. Pietikaninen and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern analysis and machine intelligence, vol.24, no.7, pp.971–987, 2002.