

ラフ集合理論を用いたクエリー映像からイベント検索モデルの導出

Defining Event Retrieval Models from Query Videos using Rough Set Theory

杉原 ちえり[†] 白浜 公章[‡] 上原 邦昭[†]
Chier Sugihara Kimiaki Shirahama Kuniaki Uehara

1 はじめに

近年のマルチメディア技術の進歩に伴って、個人のハードディスクや Web 上に、大量の映像が“映像アーカイブ”として蓄積されるようになってきた。こうした背景のもと、映像アーカイブから、所望のイベントが表現された映像（以下，“イベント”）を効率的に検索する技術に関する研究が活発に行われている。

従来から、映像検索では、事前に定義しておいたモデルとマッチングをとることによりイベントを検索する“モデルベース”の手法が提案されている [1]。しかしながら、映像アーカイブに存在する全てのイベントを事前に想定し、モデルを定義しておくことは困難である。そのため、モデルベースの手法は、映像アーカイブの検索には現実的には適用できない。一方、ユーザから与えられたクエリー映像と類似した映像を検索する“類似度ベース”の手法が提案されている [2]。この手法では、事前にモデルを定義する必要がなく、クエリー映像を用いて任意のイベントを定義できる強みがある。しかしながら、類似度だけでは分類に必要な特徴と不必要な特徴を区別できないため、イベントを適切に定義できないという問題点がある。例えば、「車が走る」というイベントを検索する場合、赤い車のショットをクエリーとすると、赤い服を着た人などの赤色が写っているショットが検索されてしまい、白い車の走っているショットは検索されにくい。

そこで、本論文では類似度ベースの検索を拡張した、クエリーベースのイベント検索手法を提案する。具体的には、ユーザの要求する映像の例（正例）とユーザの要求に関係しない映像の例（負例）を用いて、どの特徴がイベント検索に有効かを学習する。

クエリーベースのイベント検索には、重要な点が 2 つ存在する。1 つ目は正例、負例の選択方法である。この点に関しては別に研究を行っており [3]、本論文では述べない。2 つ目は、分類器の選択である。本論文では、

イベントの多様性に注目し、ラフ集合理論 [4] を用いる。映像におけるイベントは、カメラの動きや角度、イベントの行われる場所などにより大きく異なる。たとえば、同じ「街を車が走る」というイベントでも、車が大きく写っているものや、ロングショットで街全体を撮っているものなど、撮り方ひとつで色やエッジなどの特徴が大きく変わってしまう。また「ドアを開ける」というイベントであれば、屋外から屋内、屋内から屋外、屋内同士での移動など、場所によって特徴が異なってしまう。そこで、ラフ集合理論を用いて、1 つのイベントに対して複数のルールを抽出し、イベントの多様性に対応させる。

ただし、従来のラフ集合理論では、シンボルなど離散化されたデータを対象とすることがほとんどだった。これに対して、映像からは離散値、連続値、ヒストグラムなどの様々な特徴量が抽出できる。そこで、本論文では [5] の手法を拡張して、様々な特徴量に対応できる類似度を用いたラフ集合理論を提案した。

2 イベント検索モデルの導出

本章では、各ショットから取り出す特徴量と、ラフ集合理論に基づくイベント検索モデルの導出手法について説明する。

2.1 特徴量

まず、イベント中に出現するオブジェクトの空間的な情報を得るため、キーフレームを図 1(a) に示す 30 種類の異なる領域に分割する。そして、各領域に対して次の 3 種類の特徴量を求める。

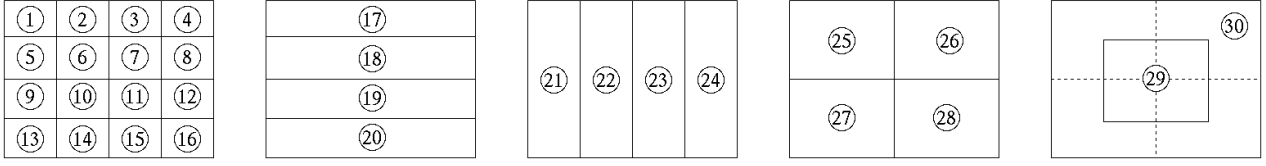
色：領域の色を 36 個のピンを持つ色ヒストグラムで表現する [6]。これにより、空、植物など特徴的な色を持つオブジェクトの認識が可能となる。

エッジ：縦、横、斜め 45°、斜め 135°、無方向の 5 種類のエッジ分布を表すヒストグラムを作成する。これにより、建物なら上部に縦のエッジが多い、芝生なら下部に縦または斜めのエッジが多いなど、オブジェクトの輪郭、テクスチャの特徴を認識することができる。

[†] 神戸大学大学院工学研究科、Graduate School of Engineering, Kobe University

[‡] 神戸大学大学院経済学研究科、Graduate School of Economics, Kobe University

a) Partition of a keyframe into 30 equal-sized regions



b) 90-dimensional shot representation

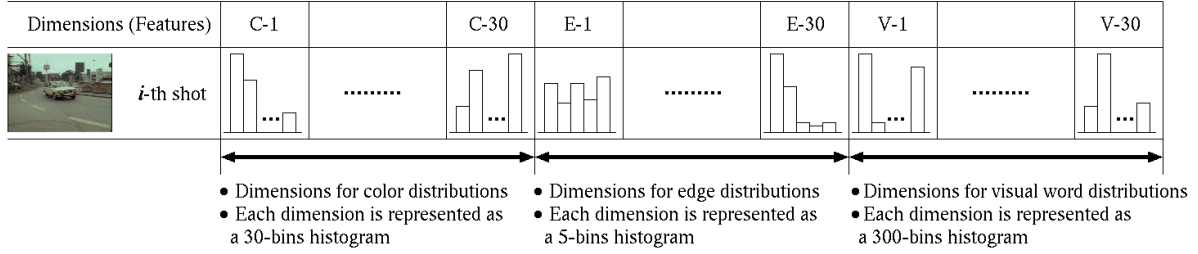


図 1: 1 キーフレームの領域分割と抽出された特徴量

visual-word: visual-word とは、オブジェクトの局所的な形状を表すパターンである。車のフロントガラスやヘッドライト、または家具のコーナーなどが表現される。このような visual-word は、特徴点周りの勾配情報をベクトル量子化して得られる。最終的に、各領域における visual-word の出現頻度ヒストグラムを作成する。この論文では、visual-word の個数を 300 個とした。

上記の特徴量を求めることにより、図 1 (b) のように各キーフレームを 90 次元の特徴量で表現できる。ここで、各次元はそれぞれ 1 つのヒストグラムで表現されている。このような形式で表現されたショットにラフ集合理論を適用し、イベント検索モデルの導出を行う。

2.2 ラフ集合理論

ラフ集合理論は、訓練例同士の識別可能性を学習し、ルールを生成する手法である。あるクラスの例について、他のクラスと分類するために有用な特徴量を発見することで、そのクラスの決定ルールを生成する。

ラフ集合理論では、以下の手順でクラス分類のルールを生成する。

1. 識別行列を作成する M 個の正例 p_i と N 個の負例 n_j を識別するために有効な特徴量の集合 $f_{i,j}$ を求める。まず、図 2 (a) のように、 p_i と n_j の k 番目の特徴量 p_i^k と n_j^k の類似度 $\text{sim}(p_i^k, n_j^k)$ を計算する。そして、類似度の閾値 β^k に基づいて各特徴量が $f_{i,j}$ に含まれるかを判定し、図 2 (b) に示す $f_{i,j}$

を要素とする行列を作成する。すなわち、 $f_{i,j}$ は以下の式で表される。

$$f_{i,j} = \{k | \text{sim}(p_i^k, n_j^k) < \beta^k\}$$

ゆえに、 $f_{i,j}$ に含まれる要素の少なくとも 1 つを考慮すれば、 p_i と n_j は識別できたことになる。

2. 識別関数を作成する 識別行列から、各正例 p_i と全ての負例を分類するためのルールを導出する。具体的には、 p_i に対して以下の識別関数 df_i を求める。

$$df_i = \bigwedge \{f_{i,j} | 1 \leq j \leq N\}$$

すなわち、 p_i と全ての負例を分類するためには $f_{i,j}$ 中の任意の特徴量の連言をとれば良い。例えば図 2 (b) では、 $f_{1,1}$ と $f_{2,1}$ から任意の特徴量をそれぞれ選び、連言をとれば、 p_1 と全ての負例を分類できる。

3. 決定ルールを生成する まず、 df_i を簡単化し、選言標準形 (連言節を選言で結合した形) に変換することで、 p_i に対する決定ルール R_i が得られる。これにより、 R_i 中の少なくとも 1 つの連言節を用いることで、 p_i を全ての負例と識別できる。次に、全ての正例、負例に対する決定ルールを求めると、

$$R = \bigvee R_i \{1 \leq i \leq M\}$$

と表現できる。 R_i と同様に、 R 中の少なくとも 1 つの論理積を用いることで、全ての正例と負例が識別できたことになる。

a) Decision table

	k=1 (C-1)	k=2 (C-2)		k=90 (V-30)	Class
p ₁					P
p ₂					P
n ₁					N
n ₂					N

b) Discernibility matrix

	p ₁	p ₂
n ₁	f _{1,1}	f _{1,2}
n ₂	f _{2,1}	f _{2,2}

$df_1 = (\bigvee f_{1,1}) \wedge (\bigvee f_{2,1})$
 $df_2 = (\bigvee f_{1,2}) \wedge (\bigvee f_{2,2})$

c) Decision rules

IF C-20 is similar to , THEN Class = P
 IF E-17 is similar to , THEN Class = P
 IF C-17 is similar to  and E-20 is similar to , THEN Class = P

図 2: ラフ集合理論

4. ルールを用いてランク付けを行う 検索対象となる各データに対して、決定ルールに当てはまるかどうかを調べる。各決定ルールの特徴量と検索対象の特徴量を比較し、類似度が閾値以上であればそのルールに当てはまるとする。当てはまるルール数の合計によって、検索されたショットのランク付けを行う。

3 実験

実験には TRECVID2008 の訓練、テストデータを用いた。訓練データの中から正例、負例を選択し、テストデータに対して検索を行った。

検索したイベントは「車が街を走る」「街中でのインタビュー」「ドアを開ける」の3種類である。検索結果について、上位100位、300位、500位、1000位を取り出したときの precision を表1に示す。

表 1: 各イベントの検索結果

検索数	車が街を走る	街中でのインタビュー	ドアを開ける
正例数	9	11	9
負例数	14	16	16
p@100	0.340	0.100	0.090
p@300	0.216	0.087	0.070
p@500	0.168	0.078	0.068
p@1000	0.139	0.068	0.063

表1より、ラフ集合を用いることで同じイベントでも、異なる状況下で撮影されたショットを検索することができた。図3に「車が街を走る」における検索結果の一例を示す。カメラアングルや、ビルや車などオブジェ



図 3: 「車が街を走る」の検索結果

クトの大きさの違いに関わらず、検索できていることが分かる。また「街中でのインタビュー」と「ドアを開ける」イベントは、NIST 後援の世界的な映像解析プロジェクト TRECVID 2008 のサーチタスクで取り上げられている [8]。そして、inferred Average Precision (infAP) という共通の評価尺度を用いて、世界の各研究チームで開発された手法の検索結果が比較されている。ここで、「街中でのインタビュー」「ドアを開ける」に対する本手法の infAP は、それぞれ 0.022, 0.221 であり、前者は TRECVID 2008 に投稿された 34 個の結果のうち上位 10 位、後者はトップという非常に良好な結果であることが分かった。

4 まとめ

映像アーカイブからのイベント検索において、ラフ集合理論に基づく、クエリベースのイベント検索を提案した。実験では、本手法を用いることで、同一のイベントでも、撮影技法などにより特徴量の異なる多様なショットを検索できることを確認した。

しかし、現在の手法では動きの特徴を考慮していないため「ドア自体は写っているが開かれてはいない」とい

うような誤ったショットが検索されてしまう．そのため，今後の課題として，オブジェクトやカメラの動きを特徴量として取り入れることが挙げられる．また，この手法では，類似度を計算する際の距離のとり方が非常に重要となる．検索精度を高めるために，映像の特徴を分類するのに最適な距離関数の導入も必要である．

5 参考文献

- [1] C. Snoek and M. Worring. Multimedia event-based video indexing using time intervals. *IEEE Transactions on Multimedia*, 7(4):638-647, 2005.
- [2] K. Kashino, T. Kurozumi and H. Murase. A quick search method for audio and video signals based on histogram pruning. *IEEE Transactions on Multimedia*, 5(3):348-357, 2003.
- [3] 松村佳奈，白浜公章，上原邦昭．部分教師付学習を用いた映像検索．電子情報通信学会 2009年総合大会
- [4] J. Komorowski, Z. Pawlak, L. Polkowski and A. Skowron. Rough Sets: A Tutorial. In S. Pal and A. Skowron, editor, *Rough-Fuzzy Hybridization: A New Trend in Decision Making*, pages 3-98. Springer, 1999.
- [5] Y. Leung, M. Fischer, W. Wu and J. Mi. A rough set approach for the discovery of classification rules in interval-valued information systems. *International Journal of Approximate Reasoning*, 47(2):233-246, 2008.
- [6] L. zhang, F. lin and B. Zhang. A CBIR method based on color-spatial feature. In *Proc. of TEHNCON 1999*, pages 166-169, 1999.
- [7] Moving Picture Coding Group(MPEG). MPEG-7 Overview (version 10). <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.html>, 2004.
- [8] A. Smeaton, P. Over and W. Kraaij. Evaluation campaigns and TRECVID. In *Proc. of MIR 2006*, pages 321-330, 2006.