

ロバスト潜在意味解析に基づくビデオの類似検索 Similarity Search of Videos Based on Robust Latent Semantic Analysis

井上 光平[†]
Kohei Inoue

浦浜 喜一[†]
Kiichi Urahama

あらまし

ベクトル量子化によってビデオを記号列で表して、潜在意味解析によって類似検索する手法において、ベクトル量子化をロバスト化することによって検索性能が向上することを示す。

1. まえがき

筆者らは前に [1], 文書検索で開発された潜在意味解析 (Latent Semantic Analysis: LSA) をビデオの類似検索に応用した。文書検索では文書は単語列であり, LSA では各文書での単語の頻度をデータとする行列を特異値分解して文書間の類似度を計算する。文献 [1] では, フレーム集合の階層クラスタリングに基づいてビデオを記号列に変換して, 各ビデオでの記号の頻度を求め, LSA によってビデオ間の類似度を求めた。本論文では, この手法での階層クラスタリングを k 平均法に変え, ベクトル量子化をロバスト化することにより検索性能が向上することを示す。

2. ビデオの記号列への変換

本論文では, ビデオの記号列への変換に文献 [2] に倣ってベクトル量子化 (vector quantization: VQ) を用い, ロバスト化して例外値 (outlier) の影響を低減する。

まず, 全データベースビデオから 10 フレームおきにサンプリングしたフレーム集合をトレーニングデータとする。このトレーニングデータを各フレームのカラーヒストグラムに基づいて k 平均法でクラスタリングして, 得られた代表フレームをコードワードとする。コードワードの集合がコードブックであり, VQ に倣って各ビデオのフレーム列を記号列に変換する。すなわち各フレームを自分に最も近い代表フレームの記号に置換える。これをそのまま行なうと

変換法 1:

ビデオの各フレームについて, 全代表フレームとの距離を計算し, 最も近い代表フレームに置換える。

となる。しかし次のようにして時間的な連続性を利用すれば計算量を削減できる。

変換法 2:

Step 0: 各クラスタの半径 (そのクラスタの代表フレームから最も近い代表フレームまでの距離の $1/2$: 図 1 参照) を計算する。

Step 1: ビデオの第 1 フレームと各代表フレームとの距離を計算し, 最も近い代表フレームに置換える。

Step 2: 第 2 フレーム以降については, 1 つ前のフレームが置換えられた代表フレームとの距離を計算し, それがその代表フレームの半径以下であればそれに置換え, そうでなければ他の全代表フレームとの距離を計算し, 最も近い代表フレームに置換える。

このようにして各データベースビデオを記号 (クラスタの識別子, 例えば番号) 列に置換え, 各ビデオでの各記号の出現頻度を要素とするデータ行列を構成する。

しかしこの方法では, どの代表フレームからも遠いようなフレームもどれかの記号に置換えられ, そのような例外値によって検索性能が落ちる可能性がある。そこで, 最も近い代表フレームとの距離がその代表フレームの半径の $\alpha (\geq 1)$ 倍より大きいフレームは無視することにする。

そのためにまず, 代表フレームを求める k 平均法をロバスト化する。すなわち k 平均法では, 各代表点を自分に属すデータの重心に移動するのを繰り返すが, そのとき自分の半径の $\alpha (\geq 1)$ 倍よりも遠いデータは自分に属さないとして無視する。

またビデオを記号列に置換える手順 (上記の変換法 2) も次のように変える。

変換法 3:

Step 0: 各クラスタの半径を計算する。

Step 1: 第 1 フレームと各代表フレームとの距離を計算し, 最も近い代表フレームに仮に置換える。

Step 2: 第 2 フレーム以降については, 1 つ前のフレームが置換えられた代表フレームとの距離を計算し, それがその代表フレームの半径以下であればそれに仮に置換え, そうでなければ他の全代表フレームとの距離を計算し, 最も近い代表フレームに仮に置換える。

Step 3: 上で仮に置換えられた代表フレームとの距離がその半径の α 倍以内 (inlier) なら, その置換えを確定し, α 倍より大きければ (outlier) そのフレームは無視する (図 2 参照)。

3. 潜在意味解析による類似検索

以上によって各ビデオは記号列に置換えられる。記号の数を m , データベースビデオの数を n とする。記号 i のビデオ j での出現頻度を d_{ij} とすると, ビデオ j はベクトル $d_j = [d_{1j}, \dots, d_{mj}]^T$ (これは記号のヒストグラムである) で表され, データベース全体は $m \times n$ 行列 $D = [d_{ij}]$

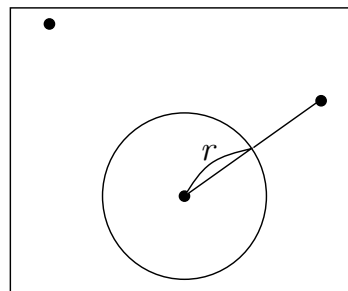


図 1: クラスタの半径 r

[†]九州芸術工科大学画像設計学科

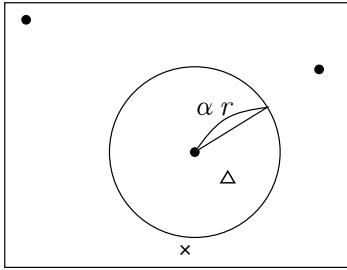


図 2: inlier() と outlier(x)

で表される．LSA に倣って D を $U\Sigma V^T$ と特異値分解し， k 行， k 列だけ残してランク k の近似行列 $U_k\Sigma_k V_k^T$ とする．ベクトル d_j を $\hat{d}_j = d_j^T U_k \Sigma_k^p$ で $k (< m)$ 次元ベクトルに射影する (p は可変な指数として導入した)．クエリビデオについても上記の変換法によって記号列に置換えて記号の出現頻度を計算し，ベクトル (記号ヒストグラム) $q = [q_1, \dots, q_m]^T$ で表し， $\hat{q} = q^T U_k \Sigma_k^p$ によって k 次元ベクトルに射影する．ビデオ間の類似度を射影後のコサイン $\cos\theta_j = \hat{q}\hat{d}_j^T / \|\hat{q}\|\|\hat{d}_j\|$ で求め，大きい順にビデオ j を出力する．なおコサインの計算では，オフラインで $\hat{d}_j / \|\hat{d}_j\|$ を計算しておき， q が入力されたら $q^T U_k \Sigma_k^p / \|q^T U_k \Sigma_k^p\|$ を計算して $\hat{d}_j / \|\hat{d}_j\|$ との内積を計算すればよい (これは k 次元での内積なので m 次元よりも計算量が少ない)．

4. 実験

フレームサイズ 320×240 ，フレーム数 300 のビデオ 40 本 (海辺，夕景，サッカー，雪景色の 4 グループ各 10 本ずつ) を用い，各ビデオをクエリとしたときに，同じグループのビデオが上位 10 本のうちに何本含まれるか調べた．クラスタ (記号) の数 m は 20 とした．

まず最初にベクトル (記号ヒストグラム) q, d_j 間のインタセクションで類似検索したところ，上位 10 本中平均 7 個が一致した．検索時間は 500 回当たり 0.92 秒であった．

次に LSA で検索した．射影行列 $U_k \Sigma_k^p$ の p を -1 から 3 まで変えて実験した結果， $p = 2$ のときが最もよかったので $p = 2$ とした．ロバストでない場合 (変換法 2) の

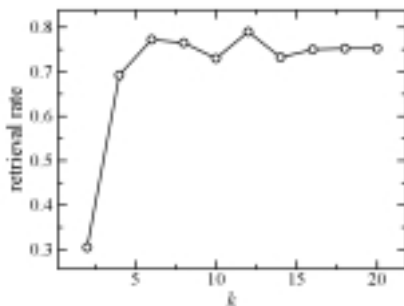


図 3: 上位 10 位中のグループの一致率

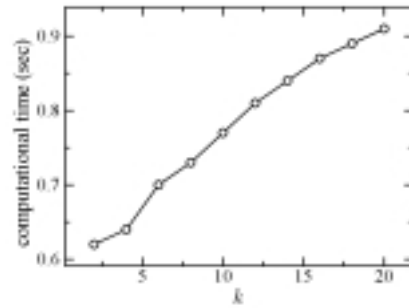


図 4: 検索時間

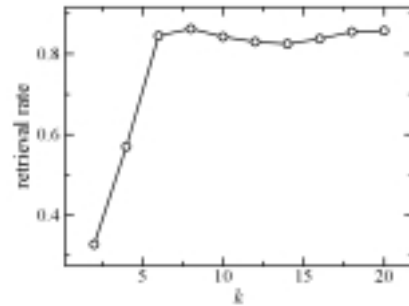


図 5: ロバスト VQ でのグループの一致率

一致率を図 3 に，500 回当たりの検索時間を図 4 に示す．横軸はランク k である． k を 6 や 8 辺りにすれば一致率もインタセクションの 0.7 よりも大きく，検索時間も 0.92 秒よりも短い．次にロバストにした場合 (変換法 3) の一致率を図 5 に示す．棄却半径の α は 2 とした．図 3 よりも更に一致率が向上している．検索時間は図 4 とほとんど同じである．

5. むすび

ビデオの類似検索の 1 手法として，ベクトル量子化によりビデオを記号列で表して潜在意味解析で検索する手法を提案し，ベクトル量子化をロバスト化することにより検索性能が向上することを示した．本手法では記号は単純なフレームクラスタであるが，記号表現を高度化するのが今後の課題である．

謝辞 本研究の一部は電気通信普及財団からの研究助成による．

参考文献

- [1] 堀田政二，宮原末治，井上光平，浦浜喜一，“潜在意味解析に基づくビデオの類似検索，” 画像の認識・理解シンポジウム (MIRU 2002)，vol.2，pp.203-208，2002.
- [2] 井上光平，浦浜喜一，“時間コレログラムによるビデオの類似検索，” 映情誌，vol.56，no.10，pp.1651-1656，2002.