

2D 動画クエリによる TVM 検索のための特徴量の検討

Feature Extraction for TVM Retrieval with 2D Video Query

葛西 大佑†
Daisuke KASAI山崎 俊彦†
Toshihiko YAMASAKI相澤 清晴†
Kiyoharu AIZAWA

1. はじめに

近年、複数台のカメラで撮影した多視点画像から高精細な動的 3 次元モデル (以降 TVM と呼ぶ) を生成する研究が盛んに行われている[1]. TVM とは、撮影対象の 3 次元形状やテクスチャの時間変化を記録したものであり、その特徴として視聴者が任意の視点から撮影対象を見ることができるとい点があげられる. 将来このような TVM が一般的になったときに、大量の TVM の中から自分が見たい動画を効率的に検索するための技術が必要になると考えられ、研究が進みつつある[2]-[5]. 我々も[5]において、従来の手法で問題であったクエリの生成コストの問題を解決し、2 次元動画画像をクエリとした TVM 検索の基礎技術を開発した.

本論文では、[5]で我々が TVM と 2 次元動画画像の類似度を定義する際に用いた特徴量である P 型フーリエ記述子、Z 型フーリエ記述子、ゼルニケモーメントの比較と検討を行う.

2. 我々の TVM 検索アルゴリズム

我々が[5]で用いた 2 次元動画画像をクエリとした TVM の検索手法について簡単に説明する.

1. 正規化: データベース中の TVM について、位置は重心を中心とする座標系に、大きさは重心から最も遠い頂点までの距離が 1 になるように正規化する.
2. TVM のレンダリング: 3 次元静止モデルの検索手法[6]を元に、それぞれの TVM のフレームについて、周囲の 20 視点からレンダリングし複数の 2 次元動画画像を生成する.
3. TVM の特徴量抽出: それらの 2 次元動画画像のそれぞれのフレームについて、人物領域の輪郭画像を取得し、画像特徴量を計算する. これにより、それぞれの視点について画像特徴量の時系列が計算される.
4. クエリの特徴量抽出: クエリとなる 2 次元動画画像についても同様に、人物領域特定ののち画像特徴量の時系列を計算する.
5. 類似度計算: クエリとデータベース中の TVM について、特徴量の時系列を DP マッチングにより比較し、両者の類似度を計算する. TVM は 20 視点分の時系列があるが、クエリに最も類似した視点との類似度をクエリと TVM の類似度とする.
6. 結果表示: TVM を類似度順にソートして返す.

2-5 の処理の流れを図 1 に示す. 本論文では 3,4 における画像特徴量として P 型フーリエ記述子、Z 型フーリエ記述子、ゼルニケモーメントについて検討する.

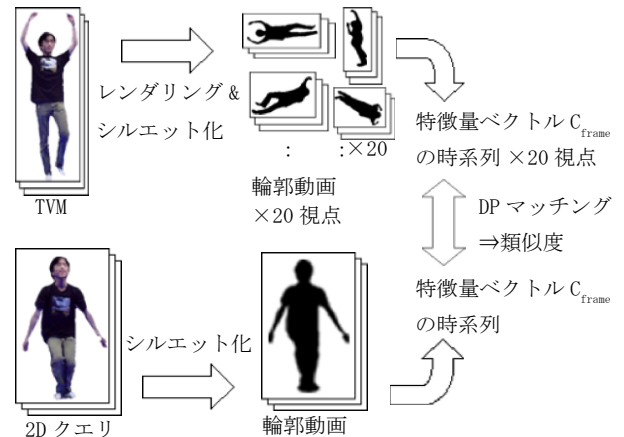


図 1 2次元動画画像をクエリとした TVM 検索手法の処理の流れ

3. 画像特徴量

3.1 P 型フーリエ記述子

P 型フーリエ記述子[8]は閉曲線にも適用可能なフーリエ記述子である. まず曲線を長さ δ の線分で折れ線近似する. この折れ線の偏角 $\theta[i]$ の指数関数を $w[i]$ とすると、複素数平面上における頂点座標 $z[i]$ との関係は次式になる.

$$w[i] = \frac{\exp(j\theta[i])}{\delta} (z[i+1] - z[i])$$

この $w[i]$ を離散フーリエ変換したものが P 型フーリエ記述子である.

3.2 Z 型フーリエ記述子

Z 型フーリエ記述子[7]は閉曲線の記述法である. まず閉曲線上の任意の点を始点 S とし、閉曲線上の点における接線の偏角 θ を S からの (閉曲線上での) 距離 d の関数 $\theta(d)$ として表現する. $\theta(d)$ を、閉曲線の全長 D を用いて次式のように正規化し周期関数とする.

$$\hat{\theta}(d) = \theta(d) - \theta(0) - 2\pi \frac{d}{D}$$

この $\hat{\theta}(d)$ の離散フーリエ変換が閉曲線を表す Z 型フーリエ記述子となる.

3.3 ゼルニケモーメント

ゼルニケモーメントは回転不変性、直交性を併せ持つ、画像特徴量である. 画像 $f(x, y)$, オーダ n , 反復数 m について次式で定義される.

$$A(n, m) = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) [V(n, m, x, y)]^*$$

where $x^2 + y^2 \leq 1$

ここで $V(n, m, x, y)$ は直交基底画像である. この $A(n, m)$ の低次の項が画像の概形を表す特徴量となる.

† 東京大学大学院情報理工学系研究科, Department of Information and Communication Engineering, The University of Tokyo

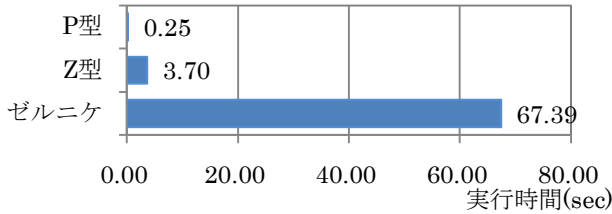
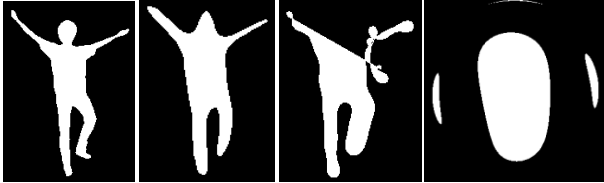


図2 特徴量の計算時間



(a)元の画像(b)P型 Fourier(c)Z型 Fourier (d)ゼルニケ

図3 再構成した輪郭画像

4. 実験

4.1 計算速度

それぞれの特徴量の計算時間を比較する。300 × 300 ピクセルの2値pngファイル40フレームからなる輪郭動画画像について、それぞれの特徴量ベクトルシーケンスを計算し実行時間を計測した。一つの特徴量につき5回計測し1計測当たりの平均をとった結果が図2である。ここで特徴量の数は、ゼルニケモーメントはオーダー $n=6$ まで、Z型フーリエ記述子は輪郭のピクセル数相当(約700個、ただし後半半分は前半の複素共役)、P型フーリエ記述子は折れ線の数を100としたため100係数計算した。ただし、フーリエ記述子の計算で用いるFFTのライブラリにはFFTW¹を用い、ゼルニケモーメントを計算するライブラリにはXITE²を用いた。

ゼルニケモーメントは輪郭画像に特化した特徴量ではないため非常に長い時間がかかっている。一方、輪郭線に特化し、しかもFFTによる高速演算を行うフーリエ記述子は短時間で計測できているのがわかる。また、折れ線近似でデータ量を削減するため、Z型よりP型のほうが計算量が少ないことがわかる。

4.2 元画像の再構成

図3(a)の輪郭画像について、P型フーリエ記述子の最初と最後の10項計20項(図3(b))、Z型フーリエ記述子の低次の20項(図3(c))、ゼルニケモーメントのオーダー10まで(図3(d))のそれぞれを用いて元画像を再構成したものを図3に示す。ただし外側の黒のみの領域はトリミングで除いている。

他の特徴量に比べP型フーリエ記述子は少ない項でも元の輪郭をよく表現していることがわかる。また、フーリエ記述子(特にZ型)では、始点と終点がずれるという現象が起きている。

4.3 検索の精度

4人の被験者による4つの盆踊りのTVMを、動作の種類別に短いクリップに区切ったもの計107クリップについ

¹ <http://www.fftw.org/>

² <http://www.ifi.uio.no/research/groups/dsb/software/xite/>

表1 P型フーリエ記述子の検索トップ正解率

被験者 ID	#1	#2	#3	#4	平均
正面	93%	100%	100%	89%	95%
斜め1	93%	81%	100%	78%	88%
斜め2	93%	92%	100%	85%	93%

表2 Z型フーリエ記述子の検索トップ正解率

被験者 ID	#1	#2	#3	#4	平均
正面	82%	73%	85%	67%	77%
斜め1	79%	85%	96%	70%	83%
斜め2	71%	77%	88%	67%	76%

表3 ゼルニケモーメントの検索トップ正解率

被験者 ID	#1	#2	#3	#4	平均
正面	75%	80%	88%	48%	73%
斜め1	75%	81%	92%	78%	82%
斜め2	57%	62%	77%	56%	63%

て、検索の実験を行った。クエリはこれらのTVMクリップを正面からレンダリングした2次元動画画像1つ、斜めからレンダリングしたもの2つで計321クエリについて行った。このときのnearest neighbor(検索最上位の正解率)の平均を表1~表3に示す。これによるとP型フーリエ記述子、Z型フーリエ記述子、ゼルニケモーメントの順に精度がよいことがわかった。

5. おわりに

本論文では2次元動画画像クエリによるTVMの検索における特徴量について、P型フーリエ記述子、Z型フーリエ記述子、ゼルニケモーメントの比較検討をした。その結果、計算速度、元画像の情報の集約性、実際に検索に用いた際の検索精度の面でP型フーリエ記述子が優れていることがわかった。

謝辞

本研究は一部文部科学省「知的資産の電子的な保存・活用を支援するソフトウェア基盤技術の構築」プロジェクトの支援により行われた。

参考文献

- [1] K. Tomiyama, Y. Orihara, M. Katayama, and Y. Iwadate, "Algorithm for dynamic 3D object generation from multi-viewpoint images," in Proceedings SPIE, vol. 5599, pp. 153-161, (2004).
- [2] T. Yamasaki, K. Aizawa, "Motion Segmentation and Retrieval for 3D Video Based on Modified Shape Distribution," EURASIP Journal on Applied Signal Processing, vol. 2007, Article ID 59535, 11 pages, (2007)
- [3] T. Yamasaki, K. Aizawa, "Motion Segmentation for Time-Varying Mesh Sequences Based on Spherical Registration," EURASIP Journal on Applied Signal Processing, (in press)
- [4] 唯野隆一, 山崎俊彦, 相澤清晴, "スケルトンを用いた Time-Varying Meshからの動き抽出と類似動作検索," 映像情報メディア学会誌, vol. 62, no. 5, pp. 729-742, (2008).
- [5] 葛西大佑, 山崎俊彦, 相澤清晴, "2次元動画画像をクエリとする Time-Varying Meshの検索," 信学技報, 画像工学研究会(2008).
- [6] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, "On Visual Similarity Based 3D Model Retrieval," Computer Graphics Forum, vol. 22, no. 3, pp. 223-232, (2003).
- [7] C. T. Zahn and P. Z. Roskies, "Fourier descriptors for plane closed curves," IEEE Trans. Computer, vol. C-21, no. 3, pp. 269-281, 1972.
- [8] 上坂吉則, "開曲線にも適用できる新しいフーリエ記述子," 信学論, vol. 67-A, no. 3, pp. 166-173, (1984).