

움직임에 의한 모양 변화를 위한 모양 시퀀스 기술자

공 영 민^o, 최 민 석, 김 회 율
한양대학교 전자통신전파공학과

Shape Sequence Descriptor for Describing Shape Variation by Object Movement

Young-Min Kong^o, Min-Seok Choi and Whoi-Yul Kim
Division of Electrical and Computer Engineering, Hanyang University
ymkong@vision.hanyang.ac.kr

요 약

본 논문에서는 객체의 움직임에 의한 모양 변화를 표현하기 위하여 MPEG-7 에 제안된 모양 시퀀스 기술자 (Shape Sequence Descriptor)에 대하여 설명하고, 모양 시퀀스 기술자 추출에 필요한 Shape Variation Map 생성을 위한 두 가지 방법을 비교하였다. 기존의 방법은 물체의 평균적인 모양에 가중치를 두어 생성되며, 새로운 방법은 물체의 움직임에 의해 변화되는 부분에 더 가중치를 두는 방법으로 생성된다. 또한 최종적으로 사용되는 특징 값으로 Zernike moment 를 이용하는 방법과 ART(Angular Radial Transform)을 이용하는 방법을 비교하여 모양 시퀀스 검색을 위한 가장 효율적인 방법을 제안하였다.

1. 서 론

비디오 이미지의 검색에 있어서 움직임 정보는 중요한 특징량이다. MPEG-7 에서는 움직임의 기술자로 Camera Motion, Motion Trajectory, Parametric Motion, Motion Activity 를 사용하고 있다[1].

이 중에서 Motion Trajectory 는 모양 정보가 없이 객체의 움직임 궤적만 기술하고 있기 때문에, 다른 모양을 갖는 객체의 동일한 움직임에 대해서 같은 특징량을 나타낸다. 예를 들면 동물이 걸어가는 것과 사람이 걸어가는 것은 인간의 인지적 관점에서는 서로 다른 움직임이지만 Motion Trajectory 의 특징량은 같은 값으로 표현된다[2]. 이와는 달리 모양 시퀀스 기술자는 물체의 움직임에 의한 모양 변화를 표현할 수 있는 Shape Variation Map 을 이용해 서로 다른 물체의 서로 다른 움직임에 대한 특징량을 표현할 수 있다.

본 논문에서는 모양 시퀀스 기술자 추출에 필요한 Shape Variation Map 생성을 위해 물체의 평균적인 모양에 가중치를 두어 생성되는 기존의 방법과 물체의 움직임에

의해 변화되는 부분에 더 가중치를 두는 새로운 방법을 비교하여 모양 시퀀스 검색을 위한 가장 효율적인 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2 절에서는 모양 시퀀스 기술자에 대한 설명과 그 추출 방법에 대해 설명하고, 3 절에서는 본 논문에서 실험한 방법과 그 결과를, 4 절에서는 실험 결과에 대한 결론을 내린다.

2. 모양 시퀀스 기술자

2.1 모양 시퀀스 기술자

비디오 세그먼트에서 물체의 모양 변화는 세그먼트의 내용을 결정하는 중요한 요소로 작용한다. 예를 들어 골프의 스윙 장면을 검색하고자 할 때나 농구에서 슛 장면만을 검색하고자 할 때 물체의 모양 변화에 대한 정보가 필요하게 된다. 이러한 연속된 프레임에서의 물체의 모양 변화 정보를 나타내기 위해 모양 시퀀스 기술자가 제안되었다.

그림 1 은 실험에 사용된 비디오 클립을 예로 들어 움직임에 의한 모양 변화를 보여준다.

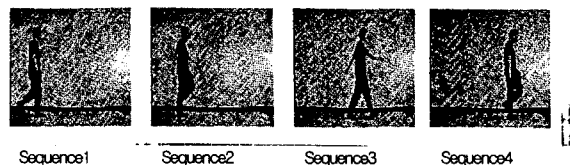


그림 1. Image Sequences

2.2 Shape Variation Map

모양 시퀀스 기술자는 Shape Variation Map 이라는 중첩된 이미지를 이용하여 물체의 모양 변화 정보를 표현하게 된다. Shape Variation Map 의 생성 방법은 다음과 같다.

비디오 클립에서 특정 객체의 움직임이 존재하는 연속된 N 개의 프레임 F_i 를 추출하여 프레임 셋 S 를 만든다.

$$S = \{F_1, F_2, F_3, \dots, F_N\} \quad (1)$$

여기서 T 개의 프레임만큼 건너뛰면서 이미지를 선택하여 다시 M 개의 이미지로 재구성한다.

$$S' = \{F_1, F_2, F_3, \dots, F_M\} \quad M \approx N/T \quad (2)$$

이렇게 얻어진 각 프레임의 이미지에서 배경을 제거하는 방법으로 대상 객체를 추출하여 이진화 된 이미지인 V_i 를 만든다.

객체의 모양 변화 정보를 생성하기 위해 객체의 무게중심을 이미지의 중심으로 정렬한 후 V_i 를 누적시켜 하나의 Shape Variation Map I 를 만든다[3].

$$I = V_1 \oplus V_2 \oplus V_3 \oplus \dots \oplus V_M \quad (3)$$

그림 2는 위의 과정에 의해 Shape Variation Map이 생성되는 과정을 본 논문의 실험에 사용된 Data Set의 일부를 이용해 보여준다.

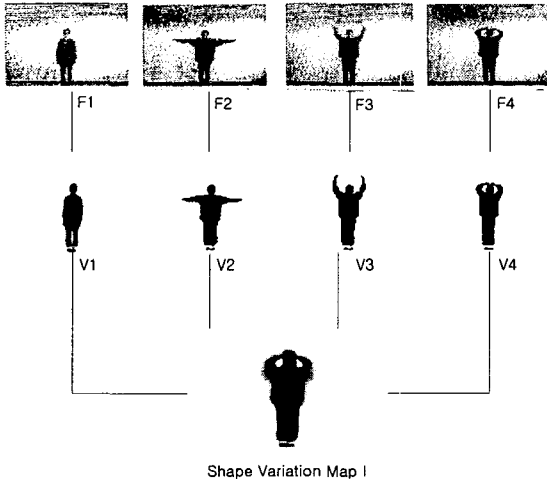


그림 2. Shape Variation Map

2.3 Negative Shape Variation Map

기존의 방법으로 만들어진 Shape Variation Map은 물체의 중심 부분이 중첩도가 가장 높게 나타난다. 따라서 모멘트를 취함에 있어서 물체의 움직인 부분보다 움직임이 없는 부분에 가중치가 더해진다고 볼 수 있다. 본 논문에서는 앞서 제시된 이러한 Shape Variation Map에서 물체의 움직인 부분에 가중치를 주기 위해 Negative Shape Variation Map을 제시한다. Negative Shape Variation Map에서는, 물체의 중첩되는 부분을 빼는 방법을 통해 가장 큰 중첩도를 가지는 움직임이 없는 부분에 가장 적은 값을 할당하게 된다. 이를 도식적으로 그림 3에 나타내었다.

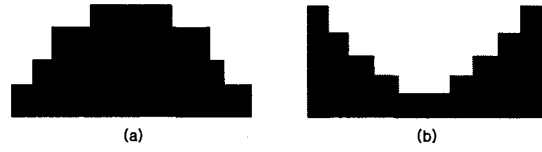


그림 3. (a) Original SVM (b) Negative SVM

2.4 Feature Extraction

모양 시퀀스의 특징량은 생성된 Shape Variation Map에 모양 기술자를 이용하여 누적된 모양 특징량을 추출하게 되는데 본 논문에서는 Zernike moment[4]와 ART[5] 모양 기술자를 이용하였다.

두 기술자는 모두 이미지의 회전과 크기에 불변한 성질을 나타내며 얻어진 각 계수들의 차에 대한 합으로 유사도를 나타낸다.

그림 4는 모양 시퀀스 기술자를 이용한 특징량 추출의 전과정을 나타내는 블록 다이어그램이다.

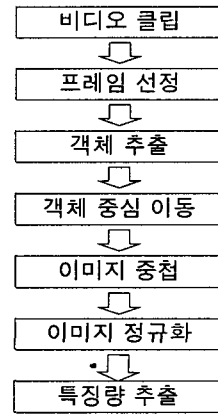


그림 4. Shape Sequence Descriptor 추출 과정

3. 실험 및 결과

3.1 실험 방법

본 논문의 실험을 위해 10명의 사람을 대상으로 20가지 동작을 재현하여 데이터 셋을 만들었다. 이를 실험을 위해 구현된 검색 시스템에 적용하여 Original Shape Variation Map을 사용한 경우와 Negative Shape Variation Map을 사용한 경우에 대한 검색 효율을 비교하였다. 검색을 위해 사용된 모양 기술자 Zernike moment와 ART 계수는 각각 35까지 사용하였다.

이상의 조건들을 기준으로 다음 4가지 경우에 대해 실험을 행하였다.

Case 1 : Shape Variation Map에 Zernike moment를 적용하여 특징량을 추출

Case 2 : Shape Variation Map에 ART를 적용하여 특징량을 추출

Case 3 : Negative Shape Variation Map에 Zernike moment를 적용하여 특징량을 추출

Case 4 : Negative Shape Variation Map에 ART를 적용하여 특징량을 추출

3.2 데이터 셋

본 논문에서는 앞에서 언급된 4 가지 특징량 추출 방법을 20 개의 그룹으로 분류된 사람의 움직임을 나타내는 200 개의 비디오 클립에 적용시켜 그 성능을 비교하였다. 10 개의 비디오 클립이 하나의 그룹으로 분리 되었으며 하나의 그룹은 하나의 동작을 나타낸다. 한 그룹 내에는 해당 그룹이 나타내는 동작을 각기 다른 10 명이 취한 비디오 클립들이 포함되었다. 표 1 은 그룹 별 비디오 클립의 개수와 그룹이 나타내는 움직임에 대한 기록이다.

표 1. 실험 그룹

그룹	비디오클립 수 (Shape Variation Map)	움직임
1	10	목운동
2	10	허리 숙이기
3	10	팔 들기 1
4	10	옆차기
5	10	옆구리운동
6	10	팔 들기 2
7	10	걷기
8	10	뛰기
9	10	앉았다 일어나기
10	10	권투
11	10	팔 들기 3
12	10	돌기
13	10	경례
14	10	막고 찌르기
15	10	손 흔들기
16	10	허리 돌리기
17	10	어깨 운동
18	10	팔 허리 돌리기
19	10	원 그리기
20	10	온몸 운동

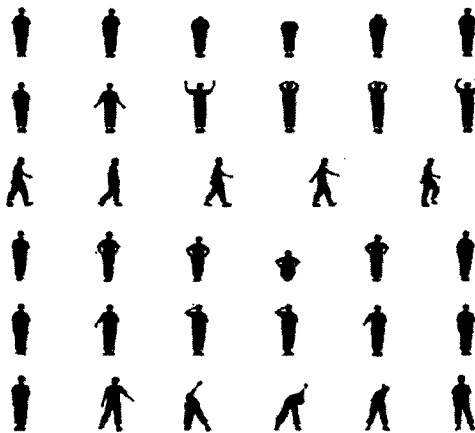


그림 5. Sample motions in the data set

3.3 실험 결과

앞의 방법으로 실험을 한 결과 각 성능평가 척도의 수치는 다음 표 2 와 같이 나타났다. 여기서 평가 척도로는 NMRR, NDS[6]가 사용되었으며 0 에서 1 사이의 값을 가지며 그 값이 작을수록 검색 효율이 좋다. 표 2 에서는 각 그룹에 대해 구해진 NMRR, NDS 들의 평균을 나타낸다.

표 2. Result

추출 방법	ANMRR	ANDS
Case 1	0.405194	0.548286
Case 2	0.340871	0.491857
Case 3	0.260194 **	0.411905 **
Case 4	0.262645	0.419286

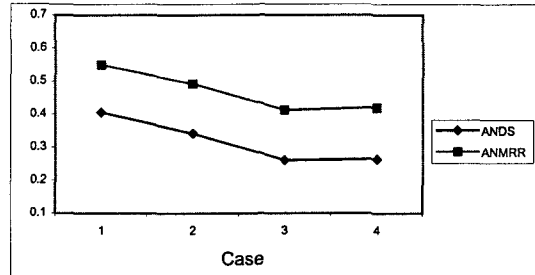


그림 6. Result

위의 실험에 의하면 특징량 추출을 위해 Zernike moment 와 ART 두 가지 경우에 대해 모두 물체의 중심에 가중치를 둔 경우(Case 1, 2)보다 물체의 움직인 부분에 가중치를 둔 경우(Case 3, 4) 주목할만한 검색 성능의 향상을 가져왔음을 볼 수 있다.

4. 결론

본 논문에서는 모양 시퀀스 기술자를 이용하여 특징량을 추출하는 방법으로 기존에 제안되었던 Shape Variation Map 을 이용하는 방법과 새로이 제안된 Negative Shape Variation Map 을 이용하는 방법을 비교하였다. 각각의 경우에 특징량 추출을 위해 Zernike moment 와 ART 두 모양 기술자를 사용하였으며 두 모양 기술자 모두에 대해 제안된 Negative Shape Variation Map 이 더 좋은 검색 효율을 보였다. 이 같은 실험 결과는 인지적 관점에서 볼 때 물체의 모양 변화 정보를 나타내기 위한 기술자이므로 물체의 중심 보다는 물체가 취한 모양의 변화, 즉 물체의 형태와 움직임에 대한 정보가 표현된 부분에 가중치를 두었을 때 보다 뛰어난 검색 성능을 보일 것이라는 사실에 일치함을 알 수 있다.

참고문헌

- [1] M. bober, S. Jeannin, "Description of Core Experiments for MPEG-7 Moion/Sape," *ISO/IEC/JTC1/SC29/WG11/N2818*, Vancouver, July 1999.
- [2] Sylvie Jeannin, Leszek Cieplinski and Jens, "MPEG-7 Visual part of eXperimentation Model Version 8.0," *ISO/IEC JTC1/SC29/WG11/N3673*, La Baule, October 2000.
- [3] Chang-Hoon Hyun, Min-Seok Choi, and Whoi-Yul Kim "Motion-Shape Descriptor for Image Sequence using Zernike Moment," 제 13 회 영상처리 및 이해에 관한 워크샵 발표 논문집, pp441-444, Jan 2001.
- [4] Yongsung Kim, Whoiyul Kim, " A region-based descriptor using Zernike moments," *Signal Processing: Image Communication*, Vol. 16, pp. 95-102, September 2000.
- [5] Yong-Sung Kim, Min-Seok Choi and Whoi-Yul Kim "An Efficient Shape Descriptor for Contents-Based Image Retrieval-Angular Radial Transform (ART)," 제 13 회 영상처리 및 이해에 관한 워크샵 발표 논문집, pp271-278, Jan 2001.
- [6] Chang-Duck Suh and Whoi-Yul Kim, "A New Metric to Measure the Retrieval Effectiveness for Evaluating Rank-based Retrieval Systems," *한국방송공학회 논문지*, 제 5 권 제 1 호, pp. 68-81, June 2000.
- [7] Aaron F. Bobick and James W. Davis "The Recognition of Human Movement Using Temporal Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, VOL.23, NO.3, March 2001.
- [8] Gyu-Tae Baek, Sang-Youn Lee, Young-Sik Choi, Sang-Hong Lee, Hae-Kwang Kim and Seoung-Jun Oh, "Shape-Sequence Descriptors for a spatio-temporal region segment," *ISO/IEC JTC1/SC29/WG11/M6425*, La Baule, October 2000.

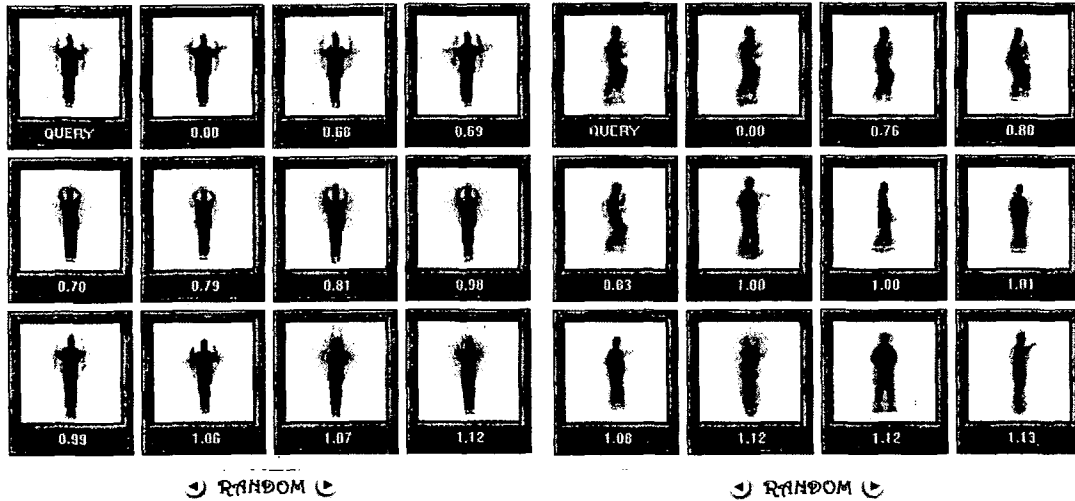


그림 7. The examples of query result