

동영상에서 모양 시퀀스를 이용한 동작 검색 방법

최민석*

요약

동영상에서 객체의 동작 정보는 장면의 내용을 분류하고 구분하는 중요한 정보로 이용될 수 있다. 본 논문에서는 동영상에서 객체의 동작을 효과적으로 찾기 위한 모양기반 동작 검색 방법을 제안한다. 객체의 동작 정보는 동영상 프레임에서 객체 영역을 추출하여 연속된 2차원 모양 정보로 표현되고, 각각의 2차원 모양 정보는 모양 기술자를 이용하여 1차원 모양 특징값으로 변환된다. 순서에 따라 나열된 모양 기술자 시퀀스를 이용하여 개별 동작의 분할 과정 없이 문서에서 단어를 검색하듯이 동영상에서 객체의 동작을 검색할 수 있다. MPEG-7 모양 변화 기술자와의 성능 비교 실험을 통하여 제안된 방법이 객체의 동작 정보를 보다 효과적으로 표현할 수 있으며, 동작 검색 및 분석 응용에 적용할 수 있음을 보였다.

Movement Search in Video Stream Using Shape Sequence

Min-seok Choi*

ABSTRACT

Information on movement of objects in videos can be used as an important part in categorizing and separating the contents of a scene. This paper is proposing a shape-based movement-matching algorithm to effectively find the movement of an object in video streams. Information on object movement is extracted from the object boundaries from the input video frames becoming expressed in continuous 2D shape information while individual 2D shape information is converted into a 1D shape feature using the shape descriptor. Object movement in video can be found as simply as searching for a word in a text without a separate movement segmentation process using the sequence of the shape descriptor listed according to order. The performance comparison results with the MPEG-7 shape variation descriptor showed that the proposed method can effectively express the movement information of the object and can be applied to movement search and analysis applications.

Key words: movement search(동작 검색), Content-based video retrieval(내용기반 동영상 검색), shape sequence(모양 시퀀스)

1. 서론

디지털 방송의 시작, 디지털 카메라와 캠코더 등의 각종 촬영 장비와 카메라 내장 휴대전화 등의 보급은 UCC(User Created Contents) 또는 UGC(User Generated Contents)라 불리는 일반인들의 동영상 콘텐츠를 포함하여 디지털 형태의 동영상 정보의 급속한 증가를 가져왔다. 이러한 디지털 정보의 홍수

속에서 원하는 정보를 쉽게 찾고 접근하고 관리하는 것이 더욱 중요해졌다. 그러나 기존의 사람에 의한 문자 색인 방식의 관리 방법으로는 증가하는 디지털 멀티미디어 정보를 효율적으로 관리하는데 한계를 가진다. 이러한 상황은 디지털 멀티미디어 정보의 더욱 빠르고 편리한 검색 및 관리를 위하여 동영상 등의 디지털 데이터 자체의 내용 및 특성에 기반을 둔 검색 방법의 필요성을 낳게 되었고 다양한 내용기반

* 교신저자(Corresponding Author): 최민석, 주소: 서울시 노원구 공릉2동 26-21(134-742), 전화: (02)3399-2087, FAX: (02)3399-2027, E-mail: mschoi@hanyang.ac.kr

접수일: 2008년 8월 4일, 완료일: 2009년 2월 3일

* 정회원, 삼육대학교 경영정보학과 겸임교수

검색 방법들이 연구되고 있다.

동영상 분석 및 검색 응용에서 움직임 정보는 다른 정보들보다 중요한 의미를 가진다[1,2]. 따라서 움직임 정보를 이용하여 동영상의 내용을 분석하고 검색하는 방법들이 많이 제안되었다. 움직임 분석은 카메라의 움직임과 객체의 움직임 궤적 분석과 관련된 연구로부터 시작하여 최근에는 객체의 동작을 분석하여 장면의 내용을 분류하려는 방향으로 진행되고 있다. 특히 인간의 동작을 인식하기 위한 많은 연구가 진행되었고, 현재도 진행 중이다. 이러한 연구는 그 접근 방법에 따라 세 가지로 분류할 수 있다[3,4].

첫 번째는 모델기반 접근 방법(Model-Based Approach)으로 각 시간에 따른 사람의 자세를 3차원 모델을 이용하여 기술하고 모델 정합을 이용하여 동작을 분석한다. 두 번째는 모양기반 접근 방법(Appearance-Based Approach)으로 동작을 객체의 연속된 2차원 모양으로 기술하고 이를 이용하여 동작을 분석하는 방법이다. 세 번째는 움직임기반 접근 방법(Motion-Based Approach)으로 신체의 각 부분을 간단한 단일 덩어리 형태의 영역으로 구성하고 구조적 특징 대신 미리 정의된 영역의 움직임을 추적하여 동작을 분석한다.

본 논문에서는 세 가지 접근 방법 중 모양기반 접근 방법을 이용한 객체의 동작 표현 및 검색 방법을 제안한다. 모양 정보는 객체의 인식, 정합, 등록 등의 응용에서 매우 중요한 역할을 한다. 따라서 객체의 모양을 효과적으로 표현하고 그 특징을 특징값으로 정의하는 많은 연구가 진행됐다. 이를 기반으로 동영상 분석에서 객체의 모양 변화를 효과적으로 분석하여 동영상의 핵심 정보를 얻어내는 것이 가능하다 [5]. 모양기반 접근 방법에서 동작은 객체의 연속된 2차원 모양으로 표현되고 많은 방법이 배경과 분리된 객체의 정규화 된 이미지를 이용한다. 또한, 동작에 의한 모양 변화를 하나의 계조(gray scale) 영상 형태로 표현하는 방법을 많이 이용하였다[4,6-10]. 그림 1은 Bobick과 Davis[4]가 모양 변화를 표현하기 위해 사용한 계조 영상인 시간적 템플릿(temporal template)의 예를 보여준다.

1.1 MPEG-7 모양 변화 기술자

이 절에서는 먼저 기존의 모양 기반 접근 방법 중 MPEG-7의 모양 변화 기술자에 대하여 설명하고 이

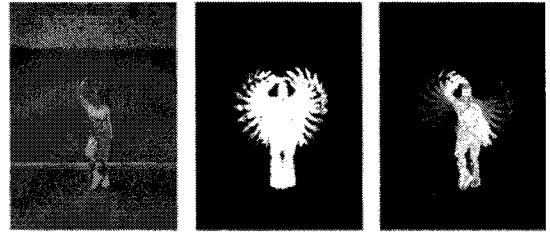


그림 1. Bobick과 Davis가 사용한 temporal template

를 이용한 동작 검색 방식의 문제점을 설명하고자 한다. MPEG-7 표준화 과정에서 많은 기술이 제안되었고 객관적 성능 검증 과정을 거쳐 선택된 기술과의 비교를 통하여 새롭게 제안된 방법의 성능을 객관적으로 검증하기 위함이다.

내용기반 미디어 검색을 위한 국제 표준인 MPEG-7에는 움직임 정보 표현을 위한 여러 방법이 제안되었다. 초기 MPEG-7 표준에는 카메라의 3차원적인 움직임을 표현하기 위한 카메라 움직임(camera motion) 기술자, 시간에 따른 객체의 움직임 궤적을 표현하기 위한 움직임 궤적(motion trajectory) 기술자, 시간에 따른 영역의 2차원 기하학적 변화를 표현하기 위한 매개 변수 움직임(parametric motion) 기술자, 그리고 동영상 내의 움직임 정도를 표현하기 위한 움직임 활동(motion activity) 기술자가 포함되었다[11,12]. 이들 중 움직임 궤적과 매개 변수 움직임 기술자를 이용하여 객체의 광역적 움직임 정보를 표현할 수 있지만, 세부적 움직임의 내용적 분석 및 표현은 불가능하다. 예를 들면, 사람이 걸어가는 것과 동물이 걸어가는 것은 내용상으로 다른 것이지만 위의 움직임 기술자만을 이용하면 구분이 불가능해진다. 움직임 궤적 기술자는 움직이는 객체의 궤적만을 표현하기 때문에 객체의 종류와 움직이는 동작을 구분할 수 없으며, 매개 변수 움직임 기술자는 사람의 움직임 같은 비정형 모양 변화(non-rigid deformation)는 표현할 수 없다. 그러나 동영상의 내용을 특성화하는데 객체의 동작 정보가 중요한 역할을 하는 경우가 많아서 객체의 동작 정보를 기술하기 위한 모양 변화 기술자가 제안되어 채택되었다.

모양 변화 기술자의 목적은 동영상에서 비슷한 동작을 검색하는 것이다. 따라서 다양하고 복잡한 형태의 동작들을 의미상 비슷한 그룹으로 분류하기 위하여 3차원 동작 정보를 간단한 2차원 영상형태의 모양 변화 측정도(shape variation map)를 이용하여 표현

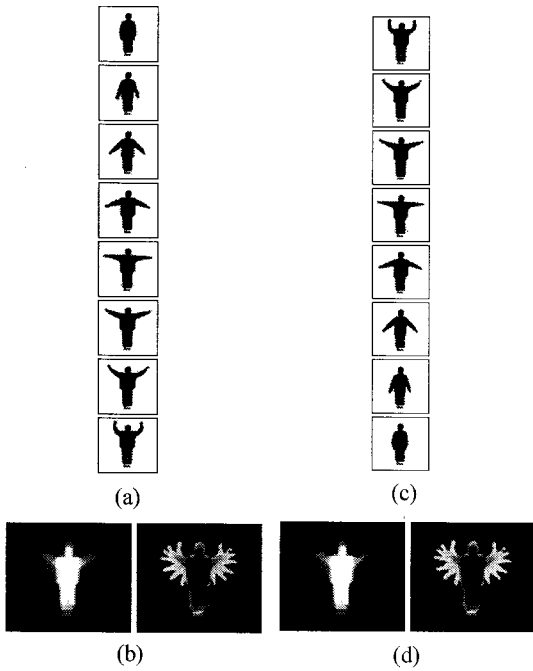


그림 2. 두 가지 동작의 모양 변화 축적도 비교 (a) 손 올리기, (b) 손 올리기 동작에서 추출된 모양 변화 축적도, (c) 손 내리기, (d) 손 내리기 동작에서 추출된 모양 변화 축적도

하였다[10]. 이 과정에서 동작의 순서 정보를 유지하지 않기 때문에 그림 2와 같이 손을 올리는 동작과 손을 내리는 동작 등과 같이 동작의 정확한 순서를 구분하지는 못하게 된다.

모양 변화 축적도가 생성되고 나서, 생성된 모양 변화 축적도로부터 모양 기술자인 ART(Angular Radial Transform) 기술자를 추출하여 질의 동작과의 유사도를 계산한다. ART는 극좌표상에서 직교하는 정현파(sinusoidal) 함수로 구성된 기저함수를 이용하여 단위 원안에서 정의되는 직교 변환이다.

$$F_{nm} = \langle V_{nm}(\rho, \theta), f(\rho, \theta) \rangle \quad (1)$$

$$= \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) f(\rho, \theta) \rho d\rho d\theta$$

여기서, F_{nm} 은 차수가 n 과 m 인 ART 계수를 나타내며, $f(\rho, \theta)$ 는 극좌표상의 영상 함수, $V_{nm}(\rho, \theta)$ 은 ART 기저함수로 식 (2)와 같이 정의된다.

$$V_{nm}(\rho, \theta) = \frac{1}{2\pi} \exp(jm\theta) R_n(\rho) \quad (2)$$

$$R_n(\rho) = \begin{cases} 1, & n = 0 \\ 2\cos(\pi n\rho), & n \neq 0 \end{cases}$$

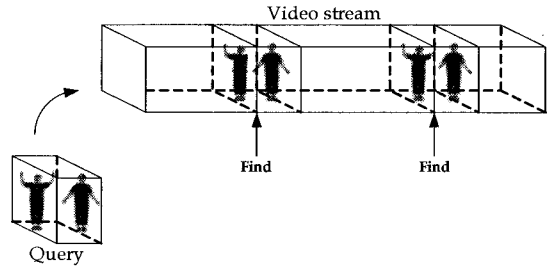


그림 3. 동영상에서 유사 동작 검색

ART 기술자는 추출된 ART 계수를 1차원 형태로 배열하여 모양 특징값으로 이용한다.

모양 변화 기술자를 이용하여 객체의 동작 정보를 기술하려면 입력된 동영상을 같은 동작 특성이 있는 단위 구간의 동영상 클립으로 나누는 동작 분할 과정이 전처리로 요구되어 진다. 이렇게 나누어진 단위 동작에 대해서만 모양 변화 기술자 등의 동작 기술자들을 추출하고 검색할 수 있기 때문이다. 그러나 실제 동영상 응용에서 사용자는 잘린 동영상 클립 단위가 아닌 전체 동영상에서 해당 장면(shot)이나 신(scene)을 찾기를 원한다. 따라서 그림 3과 같이 연속적인 동영상 스트림에서 동작 분할 과정 없이 빠르고 정확하게 동작을 검색하는 방법이 요구되고 있다.

본 논문에서는 동영상에서 동작 분할의 전처리 과정 없이 질의 동작과 유사한 동작의 장면을 찾는 방법을 제안한다. 우선 2장에서는 기존의 MPEG-7 모양 변화 기술자를 응용한 동작 검색을 설명하고, 3장에서는 모양 시퀀스를 이용한 동작 검색을 제안한다. 4장에서는 실험을 통하여 제안된 방법의 성능을 검증하고 5장에서는 결론을 맺는다.

2. 모양 변화 기술자를 이용한 검색

1장에서 언급하였듯이 모양 변화 축적도를 이용하는 모양 변화 기술자는 동영상에서 해당 동작이 존재하는 특정 구간에서만 추출할 수 있다. 따라서 동작 분할과정을 거친 후 분할된 각 동작 구간에 대하여 모양 변화 기술자를 추출하여 검색에 이용할 수 있다. 그러나 모양 변화 기술자를 동영상의 처음부터 끝까지 고정된 윈도우 크기로 연속적으로 추출하여 비교하면 동작 분할 없이 전체 동영상에서 특정 동작이 존재하는 부분을 찾는 것이 가능하다. 모양 변화 축적도의 특성상 재귀적(recursive) 생성이 가

능하여 이러한 작업의 계산을 효과적으로 줄일 수 있다.

아래 식 (3)은 모양 변화 축적도 중 낮은 변화 축적도(low variation map)를 생성하는 방법이다.

$$LVM(x,y) = \frac{1}{N} \sum_{k=1}^N V_i(x,y) \quad (3)$$

$$V_i(x,y) = \begin{cases} 1, & \text{if } (x,y) \in \text{object region} \\ 0, & \text{others} \end{cases}$$

여기서 V_i 는 i 제 프레임에서 추출된 객체 영역을 무제 중심이 영상의 중심에 오도록 정렬한 것이며 N 은 생성에 사용된 전체 프레임 수이다. 식 (4)는 LVM을 연속된 동영상에서 재귀적으로 계산하는 방법을 나타낸다.

$$LVM_i(x,y) = LVM_{i-1}(x,y) + \frac{V_{i+(N-1)}(x,y) - V_{i-1}(x,y)}{N} \quad (4)$$

여기서 LVM_i 는 i 번째 프레임을 시작 프레임으로 하는 낮은 변화도 이다. N 은 낮은 변화도가 생성되는 프레임 윈도우의 크기이다. 높은 변화 축적도(high variation map)의 경우 식 (5)에 의해 낮은 변화 축적도로부터 생성되어 진다[7].

$$HVM(x,y) = \begin{cases} 1 - LVM(x,y), & \text{if } LVM(x,y) \neq 0 \\ 0, & \text{others} \end{cases} \quad (5)$$

그림 4는 모양 변화 축적도가 재귀적으로 생성되는 과정을 보여주고 있다.

2.1 검색 시뮬레이션

실험을 위한 동영상은 5명의 사람이 22가지 동작을 하는 것을 촬영하고 수작업으로 동작 분할을 하여 생성한 110개의 동영상 클립을 통합하여 생성하였다. 사람의 동작 특성상 움직임이 빠르지 않으며 실험의 편의를 위하여 동영상에서 초당 3프레임을 추

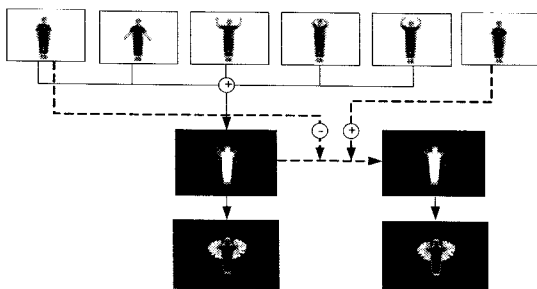


그림 4. 모양 변화 축적도의 재귀적 생성 과정

출하고 각 프레임에서 객체를 분리한 2788프레임에 대하여 모양 변화 기술자를 이용한 검색 시뮬레이션을 수행하였다. 실제 적용에서 특징 추출을 위한 초당 프레임 수 선택은 응용 대상에 따라 다르게 적용할 수 있다. 각 동작을 구성하는 총 프레임 수는 동작의 종류나 사람의 특성에 따라 20에서 80프레임 사이로 변화한다.

그림 5는 전체 동영상에서 각 프레임을 시작 프레임으로 하여 추출된 모양 변화 기술자들과 질의 동영상에서 추출된 모양 변화 기술자 사이의 거리를 그래프로 표현한 검색 시뮬레이션 결과를 보여준다. 그림에서 5개의 수직선은 질의 동작과 유사한 동작이 동영상에서 실제 시작되는 위치를 나타낸다. 검색에서는 계산된 거리 그래프에서 지역 최솟값(local minima)의 위치를 유사 동작의 위치로 판단하였다. 그림 5(a)는 지역 최솟값 중 거리가 짧은 순서로 선택된 5개의 위치가 5개의 수직선과 잘 일치하는 경우로 검색이 잘 된 결과를 보여준다. 반면 그림 5(b)는 거리가 짧은 순서로 선택된 5개의 지역 최솟값이 5개의

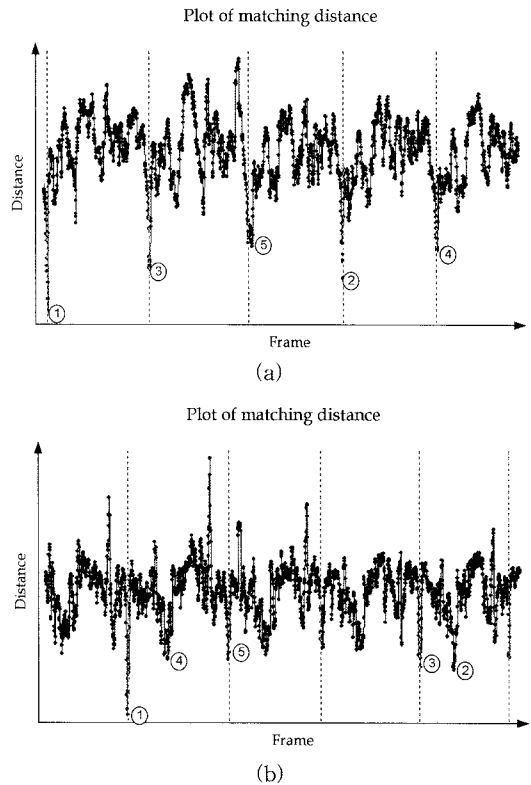


그림 5. 동작 검색 시뮬레이션 결과

수직선과 잘 일치하지 않는 경우로 잘못된 검색 결과를 보여준다.

3. 모양 시퀀스를 이용한 검색

만약 객체 분할에 의해 각각의 동영상 프레임에서 객체와 배경이 분리되어 진다면 각 프레임에서 모양 기술자를 이용하여 객체 영역의 모양 정보를 추출하는 것이 가능하며 그림 6과 같이 동영상으로부터 2차원의 모양 기술자 시퀀스가 얻어질 수 있다.

생성된 모양 시퀀스를 이용하면 마치 문서 내에서 단어를 찾는 것과 유사하게 질의 동작의 모양 시퀀스를 전체 동영상에서 생성된 모양 시퀀스 내에서 찾는 것이 가능해진다. 먼저 검색 대상 동영상에서 일정 간격으로 선택(sub sampling)된 각 프레임에서 객체 영역을 추출하고, 모양 기술자를 이용하여 모양 시퀀스를 추출한다. 본 논문에서는 모양 시퀀스 생성을 위한 모양 기술자로 MPEG-7 모양 기술자인 ART 기술자를 사용하였다. 질의 동영상에 들어오면 질의 동영상에서도 같은 과정으로 모양 시퀀스를 추출하고 질의 영상의 모양 시퀀스를 검색 대상 영상의 모양 시퀀스와 비교하여 유사한 시퀀스가 존재하는 위치를 찾는다. 그림 7은 이러한 시퀀스 검색의 개요를 보여준다.

3.1 시퀀스 정합

모양 시퀀스 검색에서 시퀀스의 정합(matching)에는 두 가지 방식이 있다. 첫 번째는 모양 변화 기술자와 같이 동작의 순서를 고려하지 않는 비순서 정합 방식이며, 두 번째는 동작의 순서를 고려하여 정확한

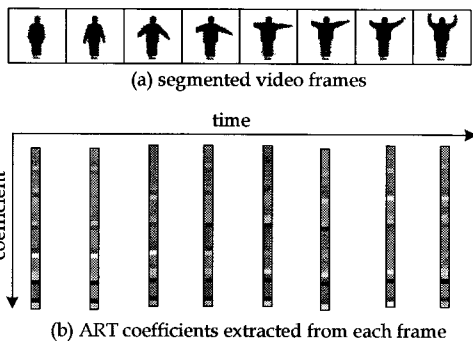


그림 6. 모양 기술자 시퀀스

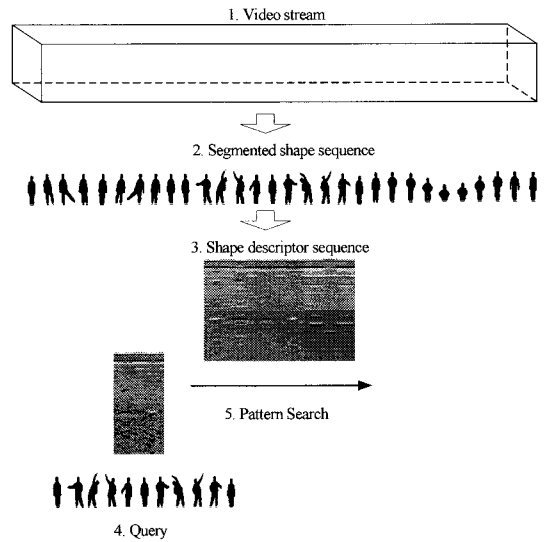


그림 7. 모양 시퀀스 검색의 개요

동작을 찾는 순서 정합 방식이다. 두 가지 방식은 비슷한 동작을 찾고자 하는 경우와 정확한 동작을 찾고자 하는 경우 등의 대상 응용에 따라 선택적으로 적용할 수 있다.

다양한 상황에 의해 발생하는 동작의 속도 변화는 동작 검색에서 어려운 문제 중의 하나이다. 그림 8과 같이 같은 동작이 여러 요인에 의해 서로 다른 프레임 수를 가질 경우 두 동작의 유사도 비교가 어렵게 된다. 기존의 모양 변화 기술자를 이용하면 생성되는 모양 변화 축적도의 모양이 서로 달라져 유사도가 떨어지는 문제를 보인다. 모양 시퀀스 검색에서는 이러한 속도 변화의 영향을 줄이고자 큰 검색 윈도우를 사용하고 최소 오류 정합(minimum error matching)을 도입하였다.

그림 9(a)와 같은 순서 정합에서 가능한 경우의 수는 식 (6)과 같다.

$$C(m,n) = \frac{m!}{n!(m-n)!} \tag{6}$$

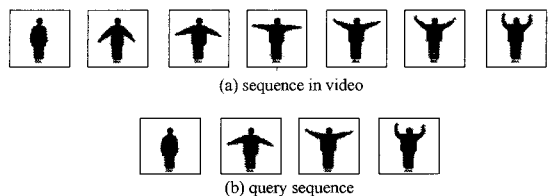


그림 8. 프레임 수 차이에 따른 비교 문제

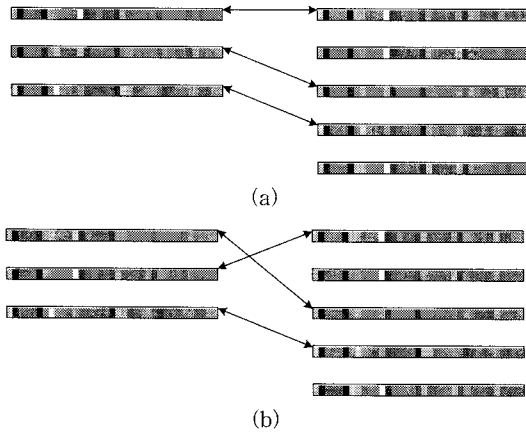


그림 9. 시퀀스 정합 개요 (a) 순서 정합 (b) 비순서 정합

여기서 n 은 질의 시퀀스의 길이이고 m 은 검색 윈도우의 길이이다. 그림 9(b)와 같은 비순서 정합의 가능한 경우의 수는 식 (7)과 같다.

$$P(m,n) = \frac{m!}{(m-n)!} \tag{7}$$

3.2 빠른 정합 방법

앞 절의 식 (6)과 (7)에서 알 수 있듯이 질의 시퀀스와 검색 윈도우의 길이가 커질수록 정합 과정의 계산 복잡도는 $O(m!)$ 형태로 증가한다. 따라서 두 가지 방식에 대하여 빠른 정합 방법이 도입되었다.

비순서 정합. 만약 시퀀스의 순서를 고려하지 않는다면 시퀀스 정합 문제는 일반적인 최소 비용 문제로 생각될 수 있다. 그러면 최소 오류를 찾기 위한 계산 복잡도를 줄이고자 이분 그래프 정합(bipartite graph matching)을 적용할 수 있다[13,14].

질의 시퀀스 q 와 대상 시퀀스 p 사이의 비용 집합을 C_{ij} 라 하면 정합 비용은 식 (8)과 같이 줄일 수 있다.

$$H(\pi) = \sum_i C_{q_i, p_{\pi(i)}} \tag{8}$$

여기서 π 는 순열(permutation)이다. 이것은 가중치 이분 정합 문제이며, 헝가리안(Hungarian) 방법으로 $O(m^2)$ 시간으로 풀릴 수 있다[14,15]. 본 논문의 실험에서는 Jonker와 Volgenant가 제안한 정합 알고리즘이 사용되었다[16].

순서 정합. 순서 정합은 그리디(greedy) 알고리즘을 적용하였다. 그리디 알고리즘은 매 순간 가장 좋은 것처럼 보이는 것을 선택하는 것이다[17]. 즉, 지역 최적 값의 선택이 전체 최적 해로 이끌어 줄 것이라고 보는 것이다. 따라서 그리디 알고리즘은 항상 최적 해를 얻지는 못한다. 그러나 그리디 알고리즘은 단순하고 효과적이며 다양한 분야의 문제에 적용되고 있다[17].

3.3 시퀀스 검색 방법

본문의 실험에서는 동영상에서 유사한 동작을 찾고자 그림 10와 같이 검색 윈도우의 길이는 질의 시퀀스 길이의 2배로 설정하였고 정합은 검색 윈도우 길이의 절반씩 겹쳐서 이동하며 수행하였다. 이러한

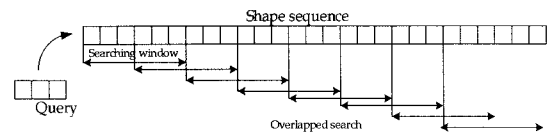


그림 10. 시퀀스 검색 방법

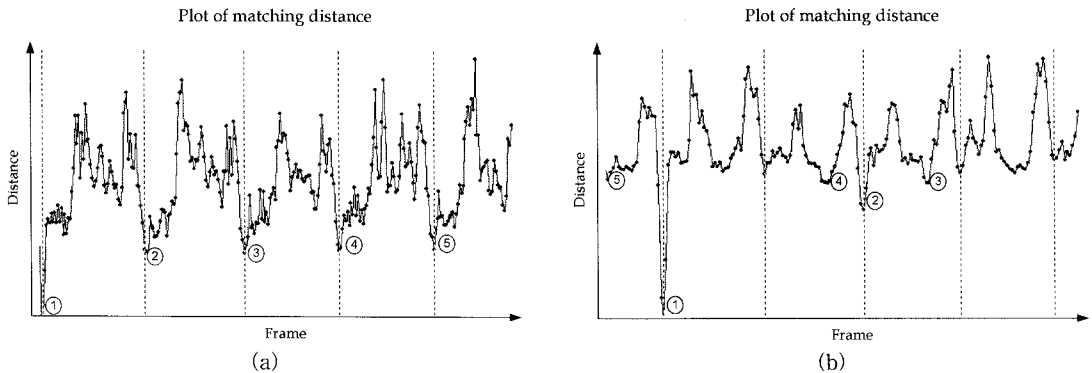


그림 11. 이분 정합을 이용한 검색 시뮬레이션 결과

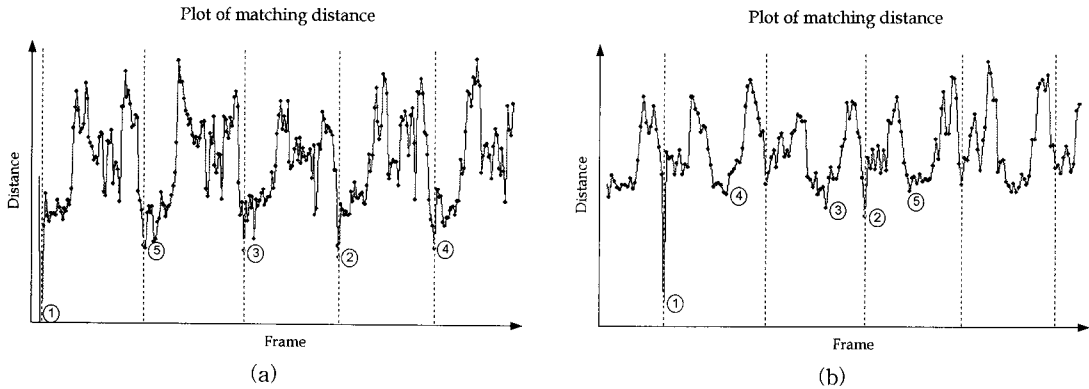


그림 12. 그리디 방법을 이용한 검색 시뮬레이션 결과

설정은 검색의 정확도 및 속도 요구에 따라 적절히 변경할 수 있다.

3.4 검색 시뮬레이션

검색 시뮬레이션은 2절의 모양 변화 기술자를 이용한 검색에서 사용된 것과 같은 동영상에서 수행되었다. 그림 11과 12은 각각 이분 정합을 사용한 것과 그리디 방법을 사용한 검색의 정합 오류를 그래프로 표현한 결과를 보여준다. 2절에서와 마찬가지로 그림 11(a)와 그림 12(a)는 검색이 잘 된 결과를 보여주며, 그림 11(b)와 그림 12(b)는 잘못된 검색 결과를 보여준다.

4. 성능 평가

4.1 실험 데이터

성능 평가 실험은 3장의 시뮬레이션에 사용된 동영상을 사용하여 진행하였다. 아래 표 1은 22가지 동작의 예를 보여주고 있다.

4.2 실험 과정

성능 평가 실험은 5명의 22가지 동작 전부를 질의하여 평균 검색 효율을 측정하였다. 제시된 각 검색 방법의 검색 효율에 대한 정량적 비교를 위하여 평가 척도로 MPEG-7 표준화 과정에서 검색 성능 평가 척도로

표 1. 22가지 동작 예제

Group	Sample image sequence	Group	Sample image sequence
m01		m12	
m02		m13	
m03		m14	
m04		m15	
m05		m16	
m06		m17	
m07		m18	
m08		m19	
m09		m20	
m10		m21	
m11		m22	

사용된 NMRR(Normalized Modified Retrieval Rank)을 사용하였다[11,18].

$$NMRR = \frac{MRR}{K+0.5-0.5R} \tag{9}$$

$$MRR = AVR - 0.5 - 0.5R$$

$$AVR = \sum_{i=1}^R Rank(i)/R$$

여기서 $Rank(i)$ 는 i 번째 관련 영상의 순위(i 번째 관련 영상이 검색되지 않았을 경우: $K+1$), R 은 전체 관련 영상의 개수, K 는 측정범위로 $min(4R, 2*max\{R_j\})$, R_j 는 j 번째 그룹의 R 을 나타낸다. NMRR은 0과 1 사이의 값을 가지며, 0에 가까울수록 높은 검색 성능을 나타낸다. 즉, 0일 경우 모든 관련 영상이 앞 순위에서 다 찾아진 경우이고, 1일 경우 측정범위 안에서 하나도 찾지 못한 경우이다. 검색 성능 평가는 아래 과정을 거쳐 진행되었다.

- a. 질의 동작과 동영상의 검색 윈도우 사이의 유사도를 계산
- b. 검색 윈도우를 동영상의 다음 위치로 이동 후 (a) 과정 반복
- c. (a)와 (b) 과정을 검색 윈도우가 동영상의 끝에 도달할 때까지 반복
- d. (a)~(c) 단계에서 계산된 유사도 그래프에서 지역 최솟값의 위치를 찾음
- e. 찾아진 지역 최솟값들을 정렬하여 검색 결과 도출
- f. 정렬된 검색결과를 이용하여 NMRR 값 계산
- g. 110개의 질의 동작 전부를 (a)~(f)를 반복하여 평균 NMRR 값 계산

4.2 검색 성능 비교

검색 성능 평가는 모양 변화 기술자를 이용한 검색, 비순서 정합을 이용한 모양 시퀀스 검색, 순서 정합을 이용한 모양 시퀀스 검색의 세 가지 방법에 대하여 비교하였다. 보다 객관적 비교를 위하여 동작 분할된 110개의 동영상 클립에 대하여 모양 변화 기술자와 Bobick과 Davis가 제안한 시간적 템플릿[4]을 이용한 검색 효율도 같이 비교하였다. 각 방법에 따른 검색 효율 측정 결과는 아래 표 2에 정리되어 있다.

표 2에서 보이듯이 모양 시퀀스를 이용한 검색이 모양 변화 기술자를 이용한 검색보다 더 좋은 효율을 보여준다. 모양 변화 기술자에서 이용되는 모양 변화 측정도가 동작 분할 과정 없이 고정된 크기의 프레임

표 2. 각 방법들에 대한 검색 효율 비교

검색 방법	검색 효율(NMRR)
모양 변화 기술자	0.263
모양 시퀀스(비순서 정합)	0.112
모양 시퀀스(순서 정합)	0.230
모양 변화 기술자(동작 분할)	0.271
시간적 템플릿 (동작 분할)	0.353

윈도우를 사용하여 생성되면 동작의 다양한 속도 변화가 반영되지 않아 인접 프레임의 영향을 받아 해당 동작의 정확한 모양 변화 측정도의 생성이 어렵기 때문이다. 동작 분할 후 모양 변화 기술자를 이용한 경우도 개별 동작의 속도 변화에 대한 정확한 정규화가 이루어 지지 않기 때문에 동작 분할을 하지 않은 경우와 유사한 검색 효율을 보여주었다. 반면에 모양 시퀀스를 이용한 검색은 검색 윈도우를 넓히고 정합 오류를 최소화함으로써 동작의 속도 변화에 따른 영향을 최소화할 수 있다. 또한, 모양 변화 기술자는 동작의 순서를 구분하지 못하지만 모양 시퀀스를 이용하면 정합 방법에 따라 동작의 순서 구분 여부를 선택할 수 있다. Bobick과 Davis의 방법은 시간적 템플릿에서 Hu의 모먼트를 추출하여 특징을 비교하기 때문에 가장 낮은 검색 성능을 보여주었다.

모양 시퀀스 검색에서는 이분 정합을 이용한 비순서 정합 방법이 그리디 방법을 이용한 순서 정합보다 더 우수한 성능을 보여주었다. 앞서 언급하였듯이 그리디 방법이 항상 최적의 해를 찾아주지 못하기 때문이다. 그러나 동작의 순서를 고려한 검색이 필요한 응용에서는 순서 정합이 필요하다. 그리디 방법보다 최적화된 알고리즘을 도입하면 순서 정합의 검색 성능을 보다 향상시킬 수 있을 것이다.

5. 결 론

본 논문에서는 동영상에서 객체의 동작을 검색하기 위한 간단하고 효과적인 방법을 제안하였다. 제안된 방법은 객체의 동작을 시간에 따른 객체의 모양 변화를 이용하여 표현하고 검색한다. 객체의 모양 정보는 모양 기술자를 이용하여 1차원 특징 벡터로 표현되고, 최종적으로 연속된 프레임에서 추출된 모양 정보는 벡터 열의 형태인 모양 시퀀스로 표현된다. 모양 시퀀스 사이의 유사성 비교를 통하여 유사한

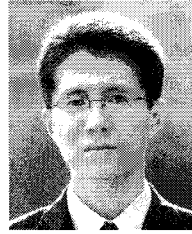
동작의 검색이 가능하다. 동작의 속도 변화 등의 요인에 의해 두 비교 시퀀스의 길이가 달라지면 최소 오류 정합 방법을 적용하여 유사성 판단이 가능하도록 하였다. 최소 오류 정합에 따른 계산의 증가는 이분 정합 등의 고속 알고리즘의 도입으로 해결하였다. 또한, 동작의 정확한 순서의 비교를 요하는 경우를 위한 순서 정합 방법과 순서에 상관없이 포괄적 검색을 위한 비순서 정합 방법을 제안하여 실제 응용에 맞추어 다양하게 적용 가능하도록 하였다. 실험을 통하여 연속된 동영상에서 동작 분할 과정 없이 동작 검색이 가능함을 보였으며, 검색 효율에서 비순서 정합은 MPEG-7 모양 변화 기술자를 이용하는 것보다 15% 이상 성능이 향상됨을 확인하였다. 추후 순서 정합을 위한 더욱 최적화된 정합 알고리즘의 적용을 통하여 순서 정합의 검색 성능을 향상시킬 필요가 있을 것이다.

참 고 문 헌

- [1] S.F. Chang, W. Chen, H. Meng, H. Sundaram, and D. Zhong, "A Fully Automated Content-Based Video Search Engine Supporting Multi-Objects Spatio-Temporal Queries," *IEEE Transaction on Circuits and Systems for Video Technology*, Vol.8, No.5, pp. 602-615, 1998.
- [2] Y.P. Tan, S.R. Kulkarni, and P.J. Ramadge, "Rapid Estimation of Camera Motion from Compressed Video with Application to Video Annotation," *IEEE Transaction on Circuits and Systems for Video Technology*, Vol.10, No.1, pp. 133-146, 2000.
- [3] J. Aggarwal and Q. Cai, "Human Motion Analysis: A review," *Computer Vision and Image Understanding*, Vol.73, No.3, pp. 428-440, 1999.
- [4] A.F. Bobick and J.W. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.23, No.3, pp. 257-267, 2001.
- [5] A. Veeraraghavan, A.K. Roy-Chowdhury, and R. Chellappa, "Matching Shape Sequences in Video with Applications in Human Movement Analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.27, No.12, pp. 1896-1909, 2005.
- [6] K. Akita, "Image Sequence Analysis of Real World Human Motion," *Pattern Recognition*, Vol.17, No.1, pp. 73-83, 1984.
- [7] J. Yamato, J. Ohya, and K. Ishii, "Recognizing Human Action in Time Sequential Images Using Hidden Markov Models," *Proceedings of Computer Vision and Pattern Recognition*, pp. 379-385, 1992.
- [8] T. Darrell and A. Pentland, "Space-Time Gestures," *Proceedings of Computer Vision and Pattern Recognition*, pp. 335-340, 1993.
- [9] Y. Cui, D. Swets, and J. Weng, "Learning-Based Hand Sign Recognition Using Shosliffm," *Proceedings of Int'l Conference on Computer Vision*, pp. 631-636, 1995.
- [10] M.S. Choi and W.Y. Kim, "The Description and Retrieval of a Sequence of Moving Objects using Shape Variation Map," *Pattern Recognition Letters*, Vol.25, issue 12, pp. 1369-1375, 2004.
- [11] MPEG-7 Visual Group, "Text of ISO/IEC 15938-3/FDIS Information technology - Multimedia content description interface - Part 3 Visual," *ISO/IEC JTC1/SC29/WG11 N4358*, Sydney, July 2001.
- [12] B. S. Manjunath, Philippe Salembier, and Thomas Sikora, *Introduction to MPEG-7: multimedia content description interface*, John Wiley & Sons, West Sussex, England, 2002.
- [13] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, *Network flows: Theory, Algorithms, and Applications*, Prentice-Hall, 1993.
- [14] C. Papadimitriou and K. Stieglitz, *Combinational Optimization: Algorithms and Complexity*, Prentice-Hall, 1982.
- [15] S. Belongie, J. Malik, and J. Puzicha, "Shape

Matching and Object Recognition Using Shape Contexts," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.24, No.24, pp. 509-522, 2002.

- [16] R. Jonker and A. Volgenant, "A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems," *Computing*, Vol.38, issue 4, pp. 325-340, 1987.
- [17] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, *Introduction to Algorithms - second edition*, The MIT Press, 2001.
- [18] B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada, "Color and Texture Descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.11, pp. 703-715, 2001.



최민석

1992년 3월 ~ 1996년 2월 한양대학교 전자공학과 학사

1996년 9월 ~ 1998년 8월 한양대학교 대학원 전자공학과 석사

1998년 9월 ~ 2004년 8월 한양대학교 대학원 전자통신전파공학과 박사

2002년 3월 ~ 현재 삼육대학교 경영정보학과 겸임교수
2007년 1월 ~ 현재 삼육대학교 정보전략 컨설팅 연구소 책임연구원

2007년 1월 ~ 현재 아이노티 커뮤니케이션 기술이사
관심분야 : 영상처리, 내용기반 영상검색, Computer Vision