

## 동영상 검색을 위한 템포럴 텍스처 생성 알고리즘

김도년, 조동섭  
이화여자대학교 컴퓨터학과

### An algorithm for generating temporal texture for video retrieval

DO-NYUN KIM, DONG-SUB CHO

Dept. of Computer Science and Engineering, Ewha womans Univ.

**Abstract** - 텍스처 정보는 정지 영상 뿐 아니라 동영상 분석에서도 많은 정보를 제공한다. 이러한 텍스처 정보를 동영상의 움직임 분류에 사용하여 기준의 색, 색 영역의 배치 정보, 기준 형상, 명도 텍스처 등을 기본 템색 키로 삼는 동영상 검색 시스템에 텍스처 특성을 움직임 정보에 적용하여 저 수준 정보에서 움직임 정보가 직접적으로 추출될 수 있음을 보였다. 이 방법의 장점은 배경 소거, 오브젝트 추출 및 추적, 참조 곡선 탐색 등 많은 계산량을 요구하는 연산들이 없이도 움직임 정보를 압축 동영상에서 추출할 수 있다는 것이다. 또한 동영상은 데이터의 양이 매우 크기 때문에 압축되어 있는 것이 필수인데 본 연구에서는 웨이브릿으로 압축되어 있는 동영상에서 움직임 정보가 고주파 부분에 집중되어 있는 점을 이용하여 역변환을 거치지 않고 직접 템포털 텍스처를 생성하였다. 따라서 계산 속도를 향상시켰으며 계산 과정도 행렬 연산을 기본으로 수행하여 계산 과정을 간단하게 하였다.

### 1. 서 론

이 논문의 목적은 저 수준 움직임 정보가 동영상 분류에 효율적으로 사용될 수 있다는 것을 보이는 것이다. 전통적인 명도 텍스처 분석은 어떤 이미지의 영역 안에 있는 명도 패턴들이 통계적 의미로든 구조적 의미로든 공간적 불변함(spatial invariance)을 확인하는 작업이다.

인간이 어떤 장면에서 물체를 인식할 때 윤곽선 외에 물체 표면의 텍스처도 중요한 변수로 작용한다. 그러나 대부분의 시각 인식 알고리즘이나 영상 처리 알고리즘은 처리하고자 하는 영상의 영역의 명도가 균일하다는 것, 즉, 텍스처가 균일한 명도를 가지고 있다는 것을 가정한다. 그러나 실제 사물의 영상이 항상 균일한 명도를 보이는 것은 아니다. 이러한 텍스처는 무시할 수 없으며 심지어 중요한 정보를 제공하기도 한다.

이 논문의 기본적인 아이디어는 이러한 텍스처 분석 기법을 움직임 정보로 구성된 템포털 텍스처에 직접 적용하여 움직임을 통계적 혹은 구조적인 특징으로 분류하는 것이다. 동영상 분류에 사용되는 중요한 정보 중의 하나로서의 움직임 정보를 객체 추출이나 배경 소거, 색 정보 관리 등 복잡한 과정을 배제하고 압축 영상에서 직접 움직임 정보만을 추출하여 그 정보를 템포털 텍스처로 변환하여 텍스처의 특성을 찾아냄으로써 움직임을 분류한다.

템포털 텍스처를 구분하기 위한 특정 벡터로서 공간적 명도도 의존성을 이용하였으며 운동량과 운동 중심을 계산하여 특정 벡터의 하나로 삼았다. 이 방법의 장점은 여러 단계를 거쳐 만들어지는 동영상 분석 과정 중 움직임의 부류만으로 분류하고자 할 때 객체 분할이나 배경 소거 등의 불필요한 작업을 거치지 않고도 결과를 얻을 수 있으며 또한 동영상의 새로운 특징이라는 독립된 특징으로도 충분한 정보를 제공한다는 것이다.

### 2. 본 론

#### 2.1 템포털 텍스처 생성

##### 2.1.1 용어 정의

템포털 텍스처를 움직임 분류에 필요한 특성으로 사용하기 위해서는 정의를 확장할 필요가 있다. 그 이유는 동영상에서는 Polana[1]가 제안한 것처럼 템포털 텍스처, activity, motion event 등으로 미리 나누어 움직임을 인식하는 것은 세 클래스를 자동적으로 분류해 주는 특정 척도가 제안되지 않은 상태에서는 비현실적이고 Szummer[2]가 제안한 것처럼 단순히 움직이는 텍스처로 의미를 축소하기에는 동영상 분류라는 입장에서는 얻을 수 있는 득이 거의 없으며 수학적으로 강연하다는 Liu[3]의 동질(homogeneous)의 환경이라는 정의는 지나치게 모호하다. 따라서 이러한 연구를 진행하기에 앞서 템포털 텍스처의 정의를 확장하는 것이 필요하다. 본 연구에서 제안한 템포털 텍스처는 다음과 같다.

##### 용어 정의1 템포털 텍스처 (temporal texture)

임의의 움직임으로부터 생성되었으며 지역적으로는 통계적 유사성을 보이는 패턴

템포털 텍스처를 웨이브릿으로 압축된 동영상으로부터 생성하기 위하여 움직임 정보를 포함한 웨이브릿 계수 열을 선택하여야 한다. 이때 사용되는 웨이브릿 계수 열을 M 요소라고 하고 이 M 요소는 계수 열 중에서 저주파 정보를 포함하는 S 요소에서 선택하게 된다.

##### 용어 정의2 M 요소 (motion component)

웨이브릿 변환을 한 결과로서 생성된 계수 열들 중 S 요소의 고주파 정보

웨이브릿 변환을 하게 되면 서로 다른 단계의 다중 해상도를 가진 부대역을 얻을 수 있다. 2차원 영상의 경우 부대역은 저주파 성분(LL)과 고주파 성분(LH,HL,HH)으로 나뉘게 되며, LL 대역의 영상은 이전 영상의  $\frac{1}{2}$ 의 해상도를 가진 닳은 형태가 된다(그림1 참조). 이를 매끈한 성분(smooth component), 즉 S 요소라고 부른다. 또한, LH,HL,HH 대역은 각각 수직, 수평, 대각선 방향의 고주파 성분을 가진다. 이를 세부 성분(detail component), D 요소라고 부른다. 웨이브릿으로 변환된 각 대역은 변환 단계 L과 방향(LL,LH,HL,HH)을 나타내는 O를 이용하여 (L,O)의 형태로 표현될 수 있다. 각 대역은 원 영상과 닳은 형태를 갖게 된다. 즉, S 요소에 영상의 중요한 정보가 거의 들어가게 된다. 웨이브릿 변환된 동영상에서 영상의 중요한 정보가 포함된 S 요소를 선택하면 그 S 요소의 D 요소에 움직임 정보가 있다. 이러한 S 요소의 D 요소를 움직임 성분이라는 의미로 M 요소(motion component)라는 용어를 정의하

였다.

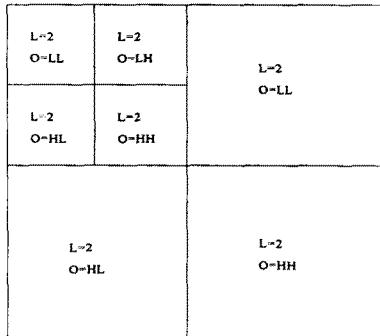


그림 1 2차원 영상의 웨이브릿 계수

### 2.1.2 템포럴 텍스처 생성

동영상의 템포럴 텍스처 특성을 알아내기 위해서는 대상 동영상의 템포럴 텍스처를 만들어야 한다. 움직임 특성을 분석하기 위한 템포럴 텍스처는 분석 대상이 된 동영상의 움직임 정보를 표현해야 한다. 현재 사용되는 영상 데이터는 수 년 전의 영상 데이터보다 훨씬 데이터 양이 크기 때문에 압축 영상에서 직접 만들어 내려 하였다. 대상으로 한 압축 방법은 웨이브릿 변환을 이용한 압축 방법이며 temporal한 특성을 압축하므로 3차원 변환과는 의미 상 차이를 보인다.

### 알고리즘1 temporal 텍스처 생성

Step1 : 동영상에 대하여 웨이브릿 변환하고, 혹은 이미 변환된 결과에 대해 S 요소를 선택한다.

Step2 : division rule에 따라 S 요소를 선택하여 이 중에서 고주파 성분, 즉 M 성분을 선택한다.

Step3 : step2에서 구해진 성분을 컨벌루션하여 한 장의 영상으로 만든다.

웨이브릿 변환의 결과로 만들어진 이미지 열에서 동영상을 분류하고 템포럴 텍스처를 만들 때는 움직임 특성이 많이 반영된 부분을 선택하는 것이 중요하다. 따라서 이러한 부분을 선택하는 규칙이 필요하게 된다. 이 규칙을 본 연구에서는 division rule이라 하고 이 division rule은 움직임 특성을 많이 포함한 부분을 추출해 내는 우선 순위에 관한 규칙이다. 동영상에서 일정한 간격으로 n번 샘플링하여 n개의 영상을 추출한다. 이렇게 추출된 n의 영상을 웨이브릿 변환하여 그림2와 같은 웨이브릿 계수 열을 구한다. 동영상의 해상도가  $i \times i$  일 때 웨이브릿 계수 열은 동영상의 각 프레임의 크기와 같은  $i \times i$ 인 배열 n개로 구성된다. 여기서 움직임 정보가 모여있는 M 요소를 선택하는 방법이 알고리즘2이다.

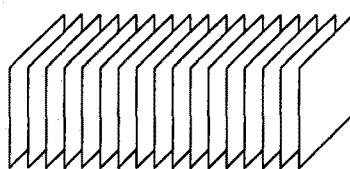


그림 2 웨이브릿 계수 열

### 알고리즘2 division rule

$C_k$  는  $i \times i$ 인 배열,  $k=1 \dots n$

$n = 2^m$ ,  $m$ 은 음이 아닌 정수

$$S_N = \{k | \{\forall C_k\}, k = 1 \dots 2^{N+2}\}$$

$$M_N = \{k | \{\forall C_k\}, k = 2^{N+1} + 1 \dots 2^{N+2}\}$$

여기서  $N=0 \dots l$

$$l은 2^{l+2} = n$$

$n = 2^m$ 인 이유는 2의 몇수일 때 웨이브릿 변환이 효율적이기 때문이다. 따라서 1도 2의 몇수일 때 효율적이 된다. 이 알고리즘에 의하면  $S_0$ 의 경우에는  $C_1, C_2, C_3, C_4$ 가 선택되며 이중 고주파 성분, 즉 D 요소인  $C_3, C_4$ 가  $M_0$ 가 된다.

그림3은 실험에 사용한 동영상이며 그림4에서는 division rule에 의해 S 요소와 M 요소를 선택할 때  $N=0$ 인 경우의 결과 계수를 역변환하여 보인 것이며 그림5는  $N=1$ 인 경우이다.

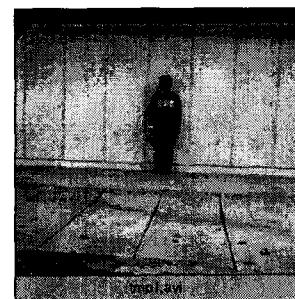


그림 3 실험 동영상의 예

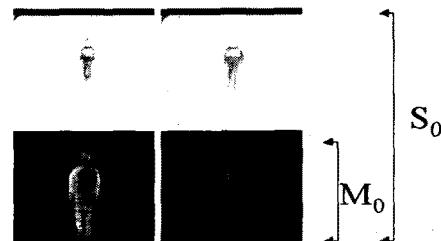


그림 4 각 요소 선택 ( $N=1$ 의 경우)

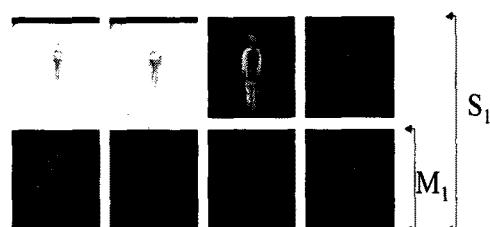


그림 5 각 요소 선택 ( $N=1$ 의 경우)

그림 5에서 선택한 S 요소 중 고주파 부분에 움직임 정보가 모여 있으므로 고주파 부분을 추출하여 필요한 경우에는 convolution 연산을 거쳐 텍스처 영상으로 변환한다. 그림6에서는 해당 픽셀에서 가장 큰 값을 가진 픽셀로 치환하는 방법을 사용하였다.

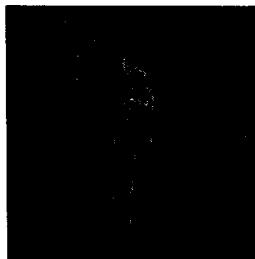


그림 6 그림3의 동영상에 대한 템포럴 텍스처

## 2.2 구현 및 실험

### 2.2.1 구현 환경

avi 파일 정보 실험에 사용한 동영상을 일상 생활에서 직접 촬영한 데이터도 대상으로 하였으며 캡쳐 시스템 사양은 CPU는 Pentium II 450MHz (RAM: 256MB) 사용한 캡쳐 카드는 Matrox Marvel G200 AGP이고 프로그램은 Presto! VideoWorks을 사용하였으며 카메라 FUJIX FOTOVISION FV7을 사용하였다. 또한 참고 문헌 [7][8]에서 제공한 동영상도 실험 대상으로 하였다. 실험실에서 직접 촬영한 경우의 avi 파일의 사양은 칼라 정보는 RGB (8-bit)이고 프레임 수는 30 frame/sec, 촬영 시간은 5초이다. 템포럴 텍스처를 생성하는 프로그램과 웨이브릿 변환 프로그램, 특성 벡터 계산 모듈 등은 C 언어로 작성하였다.

### 2.2.2 템포럴 텍스처 생성

템포럴 텍스처 생성의 예로서 원경으로 보이는 장면이 이동하는 동영상에서  $16(2^4)$ 장을 샘플링하여(그림7) 웨이브릿 변환하였다. 이 결과  $C_1 \dots C_{16}$ 인 계수 열이 생성되며 이 결과를 각각 역변환하여 표시한 결과가 그림8에 제시되었다. 이 계수 열에 division rule을 적용하여 S 요소를 선택하고 여기에서 M 요소를 결정하여 컨벌루션하여 템포럴 텍스처를 만들어낸다(그림9).

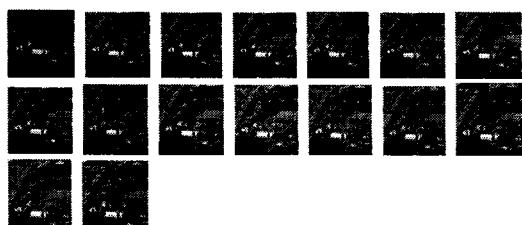


그림 7 동영상에서 샘플링한 영상

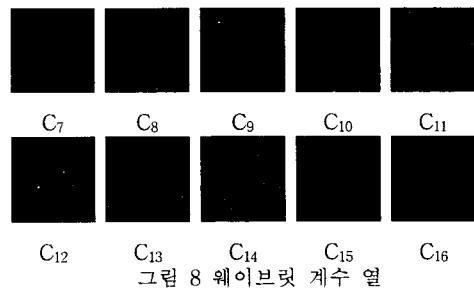
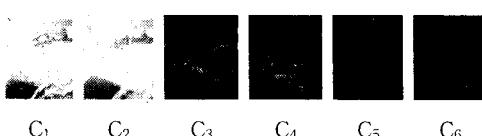


그림 8 웨이브릿 계수 열

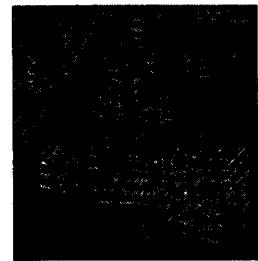


그림 9 생성된 템포럴 텍스처

## 3. 결 론

텍스처 정보는 영상을 분석, 이해할 때 많은 정보를 제공한다. 그러나 대부분의 인식 시스템에서는 이러한 텍스처 정보를 무시하거나 동일한 명도 정보를 가진 영역으로 가정하여 처리하였다. 실제 시스템에서는 텍스처 정보를 이러한 형태로 임의로 처리할 수 없고 따라서 실제로는 여러 가지 오류의 원인이 되기도 한다. 이러한 텍스처 정보를 동영상의 움직임 분류에 사용하여 기존의 색, 색 영역의 배치 정보, 기준 형상, 명도 텍스처 등을 기본 탐색 키로 삼는 동영상 검색 시스템에 움직임 정보가 텍스처 특성을 움직임 정보에 적용하여 저 수준 정보에서 직접적으로 추출될 수 있음을 보였다. 이 방법의 장점은 배경 소거, 오브젝트 추출 및 추적, 참조 곡선 탈색 등 많은 계산량을 요구하는 연산들이 없이도 움직임 정보를 동영상에서 추출할 수 있다는 것이다.

## (참 고 문 헌)

- [1] Polana, Ramprasad B., "Temporal texture and activity recognition," Ph.D. Thesis, University of Rochester, 1994.
- [2] Szummer, Marcin Olof, "Temporal texture modeling," Master Thesis, MIT, 1995.
- [3] Liu, Fang, "Modeling spatial and temporal textures," Ph.D. Thesis, MIT, 1997.
- [4] Richard W. Conners, Charles A. Harlow, "A Theoretical comparison of texture algorithms," IEEE trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No.3, pp.204-222, 1980.
- [5] 김교식, 한준희, "Texture 의 시각적 분류기준," 한국정보과학회 학술발표논문집, Vol.20, No.2, pp.337-340, 1993.
- [6] 김희승, 영상인식, pp.175-199, 생능, 1993.
- [7] <http://sipi.usc.edu/services/database/Database.html>
- [8] <http://iris.usc.edu/Image-database/Motion.html>
- [9] Ismail Haritaoglu, David Harwood, Larry S. Davis, "W4: Who? When? Where? What? A real time system for detecting and tracking people," Int. conf. on face and gesture recognition, 1998.