

# WebCAM 기반 Video 내에서의 효율적인 이미지 검색 기법

## Efficient Image Retrieval for WebCAM Video

하근희\*, 최정구, 김도년, 조동섭  
이화여자대학교 전자계산학과

Geun-Hee Ha\*, Jeong-Ku Choi, Do-Nyun Kim, Dong-Sub Cho  
Dept. of Computer Science & Engineering, Ewha Womans Univ.

### 요 약

멀티미디어가 발달함에 따라 이를 위한 다양한 검색 기법들이 제안되고 있다. 본 논문에서는 WebCAM 시스템에서 비디오 검색을 하기 위한 텍스트 성분과 영상의 윤곽선 특징을 동시에 이용한 검색 기법을 제안한다. 텍스트 특성 추출에는 8×8 블록 DCT를 기반으로 한 DCT-par와 DCT-energy 기법을 적용하고 윤곽선 추출에는 DCT의 대각선 계수를 이용하는 기법을 이용한다. WebCAM으로 생성된 비디오는 MPEG 표준을 따르는 것으로 가정하고 있으며, 압축된 데이터에 디코딩 과정없이 직접 검색 기법을 적용함으로써 처리 시간을 단축할수 있다는 장점이 있다.

## 1. 서론

### 1.1 WebCAM 시스템

1996년 후반기에 소개된 WebCAM기술은 디지털 카메라를 컴퓨터에 연결하여 일정한 시간 간격을 기본으로 현장 사진을 저장, 압축하여 홈페이지로 전송해주는 기술이다. WebCAM의 장점은 기존의 폐쇄회로 TV가 전용 선로로 연결된 곳에서만 적용될 수 있는 것과는 달리 시간과 공간을 초월하여 세계적으로 인터넷을 통하여 활용, 검색될 수 있다는 것이다. 현재 천여개의 WebCAM 사이트가 운영되고 있으며 교통 감시나 일기, 지진학, 보안과 감시등과 같은 응용분야에서 쓰이고 있다.

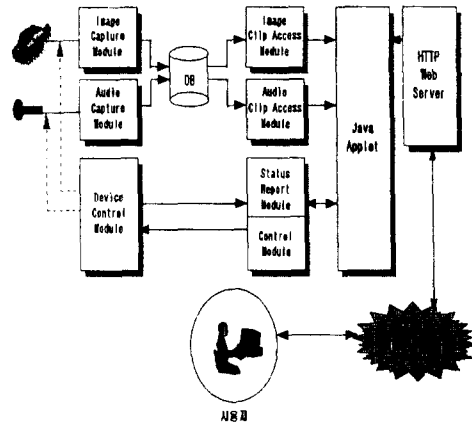


그림 1 WebCAM의 연결 개념도

### 1.2 내용기반 영상/비디오 검색 시스템의 필요성

WebCAM의 역할은 카메라로 보이는 영상을 비디오 매체에 압축한 후 저장하여 서비스 해

본 논문은 정보통신부 97 초고속 정보통신 응용기술 개발 사업의 연구비 지원에 의해 수행된 결과임.

주는 것이다. 사용자는 이 저장된 영상에서 자신이 원하는 정보를 추출해 낼 수 있어야 한다. 현재까지의 데이터베이스 응용분야에서의 검색 방법은 일반적으로 텍스트 기반이었다. 그러나 비디오나 이미지 분야에서 이러한 처리방법을 이용하기는 어렵다. 저장된 데이터베이스에 주석을 달아 검색하려면, 엄청난 양의 이미지 데이터에 따른 주석이 방대해 질 뿐 아니라, 사람이 내용에 따라 일일이 주석을 달기도 힘들다. 따라서, 이미지 데이터의 특징에 따른 내용 정보를 추출하여 저장해 놓고 이를 기반으로 사용자의 질의에 응하는 내용기반 검색에 대한 연구가 필요하다.

## 2. 기존의 내용기반 검색 연구

심춘보 등은 사용자의 다양한 검색 질의를 지원하는 내용-기반 통합 색인 기법을 소개하고 있다[1]. 이 논문에서는 박물관 문화재 중도자기에 대한 이미지와 텍스트를 가지고 통합 색인 기법을 설계하였다. 저장되는 자료에 따라 그에 맞는 색인과 저장 구조를 가진다. 텍스트의 경우에는 자동 색인기를 이용하여 색인어 리스트를 추출한 후 역화일 구조(Inverted file)로 저장하며, 도자기에 관한 색상과 형태 특징은 TV(Telescopic Vector)-트리 색인 기법을 이용하였다.

황본우 등은 뉴스 비디오를 효율적으로 저장, 관리, 검색할 수 있는 기법을 제시하고 있다[2]. 여기서 제안된 뉴스 비디오 파서는 장면 분류, 뉴스 자막의 추출 및 인식, 계층적 뉴스 검색 도구의 구성의 세 부분으로 구성되어있다. 장면 분류에서는 뉴스 비디오를 대표할 키 프레임이나 앵커 프레임을 검출하고 뉴스 자막 추출 및 인식에서는 텍스트 프레임에서의 문자열 추출과 인식을, 계층적 뉴스 검색에서는 효과적인 검색을 위하여 비디오의 내용과 시간상 진행에 따른 관련성 정리를 위하여 키 프레임을 계층적 구조화하였다. 또한 실제로 15분 가량의 뉴

스를 실험한 결과 상당히 좋은 결과를 낸 것으로 나타났다.

영역 지식을 이용하려는 시도에서 파싱 단계에서 필요한 정보들을 추출하여 원래 멀티미디어 데이터와 함께 저장하여 이를 질의 처리에 이용하는 방법이 있는데, 이는 한정된 정보만을 지원할 수 있으며 확장과 변경에 제한을 받게 된다는 문제점이 있다. 반대로, 파싱 단계에서 모든 정보를 추출한다면, 이러한 문제점들이 해결될 수는 있으나 엄청난 파싱 시간이 요구되고 불필요한 데이터가 생긴다는 문제점이 있다. 이병구 등은 이의 절충점을 찾아 파싱 단계에서 이미지 내에 존재하는 객체에 대한 특징을 추출하고, 주요 객체에 대한 특징들의 조합으로 이루어지는 사건은 질의 처리 시점에서 영역 지식을 이용하여 사건을 판별하도록 하고 있다 [3].

Pass와 Zabih가 제안한 CCV에는 공간 정보가 없다는 문제점이 있는데, 장석환 등은 이러한 인식 한계의 문제점을 극복하기 위해 공간 정보와 모양 정보를 정의한 RV(region vector)를 이용하고 있다[4]. 이 알고리즘은 인지적 인식에 큰 역할을 하는 색깔의 뭉친 정보에 민감하게 반응하므로 사람이 유사한 것으로 인식하는 이미지들을 동일한 이미지로 분류한다.

Su-Shing Chen은 다양한 종류의 정보로 이루어진 디지털 도서관에서 이용될 검색 기법을 위한 내용-기반의 BC 트리 인덱싱을 정점과 영역들에 대한 의미와 결합하는 인덱싱 기법을 제안하고 있다[5].

## 3. DCT를 이용한 검색 방법

### 3.1 기존 DCT 이용 검색 방법

일반적으로 비디오 데이터는 용량이 매우 크기 때문에 데이터베이스의 효율적인 관리를 위해서 압축된 형태로 저장, 관리된다. 압축 비디오에 비압축 비디오에 대한 분할 기술을 적용하기 위해서는 우선 복호화(decompression) 작

업이 필요하다. 복호화는 단순한 작업이기는 하지만, 처리 시간이 오래 걸리기 때문에 매우 비효율적인 방법이다. 따라서 최근에는 압축 비디오에서 직접적으로 셋의 경계 프레임을 추출하는 방법이 연구되고 있다. 압축 비디오의 분할 방법은 사용되는 특징에 따라서 DCT 계수를 사용하는 방법과 이동 벡터를 사용한 방법으로 나누어진다. DCT 계수를 사용한 방법은 연속된 프레임의 DCT 계수의 차가 임계값 이상이 되는 프레임을 장면의 경계 프레임으로 추출하는 방법으로써, 대부분의 압축 비디오 분할 방법이 DCT 계수를 사용하고 있다. 이 방법은 MPEG 형식으로 압축된 비디오에 적용할 경우 처리 시간이 빠른 반면, 점진적인 장면 전환은 검출하지 못한다는 문제점을 가지고 있다 [6].

Tao와 Dickinson은 블록 DCT 압축 영상에 적용될 수 있는 두가지 텍스처 분류 기법, 즉 DCT<sub>par</sub>와 DCT<sub>energy</sub>를 제안하였다. DCT<sub>par</sub>는 parametric 접근 방법으로서 텍스처를 고정된 Gaussian process로 모델화하고 DCT의 대각선화 성질을 이용한다. DCT<sub>energy</sub>는 DCT 영역에서의 파워 스펙트럼의 개념을 이용하는 것이다. 이 두가지 기법은 디코딩없이 압축된 데이터에서 행해지며 양자화 잡음에 강하게 설계되었다[7].

Shen과 Sethi는 압축된 영상에 디코딩없이 컨벌루션 연산을 직접 수행하여 에지를 추출하는 기법을 제안하였다. 이러한 디코딩없이 압축된 데이터에서 직접적으로 처리하는 다양한 방법들의 결과들은 주로 비디오 조작과 합성 작업, 또는 비디오 분할 작업에 초점이 맞추어지고 있다. 에지 추출과 같은 기하학적 특징 추출에 연관된 작업들은 디지털 비디오의 오브젝트 추출 혹은 추적과 같은 많은 응용분야에 매우 중요하다. Arman et al.은 에지가 존재하는지 여부를 결정하기 위하여 DCT 블록에서 0이 아닌 계수들을 이용하였고 Shen과 Sethi는 방향과 강도 같은 에지 파라미터들을 결정하기 위

하여 DCT 계수들의 패턴을 이용하였다[8].

정화자는 DCT를 이용하여 윤곽선을 추출하는 방법을 제안하였는데 이들 계수중 윤곽선에 가장 민감한 주파수 성분을 가진 계수를 여러 가지 실험을 통하여 선택하였다. 이 방법은 잘 설계된 윤곽선 검출 방법에 비하여 품질 면에서 약간 떨어질 수도 있으나, 윤곽선을 추출하는 과정이 정지 영상의 압축 표준인 JPEG을 탑재한 프로세서를 사용하는 경우 양자화 테이블의 일부 계수를 조정해 주기만 함으로써 윤곽선 영상을 얻을 수 있다. 이 경우 기존에 사용 중인 하드웨어를 그대로 살릴 수 있어 고속의 윤곽선 검출이 필요한 응용 분야에 잘 이용될 수 있으며, 전송이 필요한 경우 저전송율로 구현 가능하다고 주장하였다[9].

### 3.2 제안한 방법

영상의 내용에 의해 유사영상 검색이 가능한 내용 기반 영상검색은 텍스처 패턴, 색, 모양, 공간들과 같은 특징집합의 추출, 인덱싱, 유사도 매칭 및 검색이 요구되며 영상처리 기법을 이용하여 영상의 특징 프리미티브를 얻어낼 수 있어야 한다. 기존의 키워드 검색은 데이터베이스에 영상을 인덱싱하는데 사용될 수 있으나 이는 가시적인 데이터의 텍스트 묘사를 완벽하고 일관되게 할 수 없으며, 사람은 각 영상에 대한 텍스트 정보로 영상을 확인하고 해석하기 때문에 인덱싱을 자동화시킬 수 없다. 따라서 내용기반 접근 방법을 통해 영상 데이터의 가시적인 특징에 기반하여 검색하는 방법들이 연구되어 왔다. 이러한 방법은 인간의 가시적이고 직관적인 개념과 일치하는 영상 데이터의 특징이 영상검색을 위한 인덱스가 되며 이는 키워드 인덱싱에 사용되는 것과 동일한 방법으로 인덱싱된 칼라키나 텍스처키, 모양키를 사용하여 유사도 검색이 수행되도록 하고 있다[10]. 기존 연구들으로써는 윤곽선 특징, 유한 요소 모델, 얼굴특징, 다각형 근사, 칼라 히스토그램, 텍스처 등을 이용한 영상검색 연구가 있었다.

이러한 연구들의 문제점을 살펴보면, 첫 번째로 칼라 히스토그램이나, 텍스처는 저수준의 특징으로써 세부영역 지정이나 문맥정보가 없기 때문에 잘못 매칭될 가능성이 많으며 두 번째로 윤곽특징, 사각형 모양, 다각형 근사, 유한요소 모델과 같은 모양 특징들은 저수준 특징과 고수준 특징의 중간단계의 표현으로서, 물체의 위치에 매우 민감하며, 물체나 영역의 경계선 검출을 위해 강력한 영역분할 알고리즘이 필요하다. 그러나 영역을 완전자동으로 분할하는 기법이 없으며 현재 대부분 반자동 영역분할 기법을 제공하고 있다. 얼굴 특징이나 인공위성 사진 같은 고수준 특징들은 세부영역지정이나 문맥정보가 필요하고 영상의 특징정보를 추출하고 검출하기 위해서 지능화된 영상 분석 도구가 필요하다.

따라서 본 논문에서는 영상의 칼라 히스토그램이나 텍스처 성분을 추출하였고, 영상의 윤곽선 특징을 사용하였다. 이러한 방식은 특징 추출과정에 웨이브렛같은 압축 기법이 많이 도입되었으나 본 논문에서는 특히, 정지영상이나 동영상의 압축 표준으로 되어 있는 JPEG이나 MPEG 표준에서 사용되는 DCT를 기반으로 적용하였다. 이러한 기법은 압축되어 있는 영상을 디코딩하지 않고 특징을 추출하여 영상을 검색할 수 있다는 가능성을 보여준다는 것이 큰 의미가 있다.

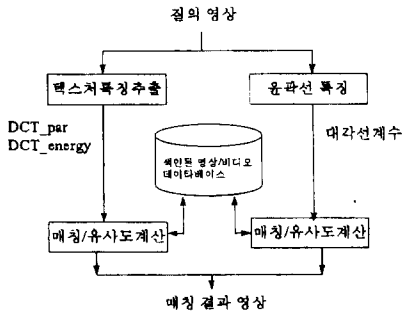


그림 2 시스템 구성도

텍스처 특징 추출에 사용되는 Tao와

Dickinson[7]이 제안한 블록 DCT 압축 영상에 적용할 수 있는, DCT-par와 DCT-energy라 불리는 두가지 텍스처 분류 기법을 사용하였다. 첫 번째는 데이터의 블록에서 다변량의 가우시안 분포를 가정하는 파라메트릭 (parametric) 접근 방법인 반면에 두 번째 방법은 푸리에 영역 파워 스펙트럼 기법(Fourier domain power spectrum method)이다. 비록 이 알고리즘들을 다른 블록 크기들, 혹은 좋은 대각선화 성질을 가진 다른 단일한 변환들로 확장하기 위해서는 별도의 연구가 필요하지만 JPEG 표준에 따르기 위하여  $8 \times 8$  블록 DCT를 가정하였다.

● A Parametric Approach : DCT-par

텍스처는 식(1)과 같은 random field model로 표현될수 있다.

$$y(s) = \sum_{r \in N} \theta_r y(s \oplus r) + e(s) \quad (1)$$

여기서  $(y(s), s \in \Omega)$ 는 영상이며  $\Omega = \{s = (i, j) : 0 \leq i, j \leq M-1\}$ 이고 N은 (0, 0)을 포함하지 않는 이웃 집합이다. 프로세스  $(e(s))$ 는 지역적 성질을 가지고 있는 가우시안 (Gaussian)으로 가정된다. 식(1)은 행렬 형태로 다음과 같이 다시 쓸수 있다.

$$B_\theta y = e \quad (2)$$

여기서  $B_\theta$ 는 변환행렬이며 Gaussian vector의 선형변환은 여전히 Gaussian이기 때문에 영상  $(y(s))$ 는 Gaussian process이다.

2차원 DCT는 선형변환이므로 DCT 영역 텍스처는 또한 Gaussian process로 표현될 수 있다. 텍스처의 DCT 계수들이 각각 독립적이라 가정할 수 있으며 블록간의 분산은 같다고

가정할 수 있고 이것은 식(1)에 내재된 가정에 의해 증명된다. 요약하여 주어진 특정 텍스처에서 어떤 DCT 계수들의 8×8 블록이나 다음에 따라 분산되어 있다고 가정할 수 있다.

$$p(x_1, \dots, x_{64}) = \prod_{i=1}^{64} \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \quad (3)$$

각 텍스처 유형 벡터들  $\{\mu_i, i=1, \dots, 64\}$ 와  $\{\sigma_i, i=1, \dots, 64\}$ 의 추정에 의하여, 다른 텍스처들은 Bayesian decision rule등을 이용하여 구분될 수 있으며, 특정 텍스처 타입이 영상 안에 있는지 결정하기 위한 신뢰성 검사가 이루어질 수 있다. 이러한 작업에서 텍스처 분류에서 실험 결과를 볼 수 있다.

실제적인 압축 과정(JPEG과 같은)에서의 고려사항은, 조잡하게 양자화된 높은 빈도수의 구성요소들은 종종 0인 계수들을 발생시키며 실재하는 분별력을 제공하지 않으므로 실제 구현에서는 64 계수들의 1/4만 사용한다. 개별적인 분별력에 기반하여 16개의 계수들이 최적화되어 선택되고 따라서 여기서도 저주파 성분이 16 계수만 사용하였다. 주어진 임의의 텍스처  $\{y(s)\}$ 는 크기가 8K \* 8L 이고 클래스 i에 대한 함수는 다음과 같이 주어진다. 여기서  $x$ 는 영상  $y$ 의 블록 DCT 변환된 값이다.

$$T_i(x) = - \sum_{l=0}^{K-1} \sum_{m=0}^{L-1} \sum_{n=0}^{63} \left[ \frac{(x_{l,m,n} - \mu_{i,m,n})^2}{2\sigma_{i,m,n}^2} + \log(\sqrt{2\pi\sigma_{i,m,n}}) \right] \quad (4)$$

● DCT Power Spectrum Approach  
: DCT-energy

DCT 파워 스펙트럼 방법을 간단히 요약하면 DCT 영역에서 다음과 같은 에너지 분포를 계산할 수 있고 여기서  $p(i,j), 0 \leq i,j \leq 3$ 은 특성벡터로 사용된다.

$$p(i,j) = \frac{[DCT(i,j)]^2}{\sum_{i=0}^3 \sum_{j=0}^3 [DCT(i,j)]^2} \quad (5)$$

4. 결론 및 향후 연구

본 연구에서는 WebCAM에 기반한 내용기반 비디오 검색 기법에 관하여 살펴보았다. WebCAM 시스템은 영상 및 비디오 정보가 실시간으로 탐색되어야 하는 특징이 있으며 이러한 특징에 기반하여 우선, 현재까지의 내용 기반 검색 기법들을 살펴 보았다. 기존 연구들은 윤곽선 특징, 유한 요소 모델, 얼굴특징 등을 이용하였으나 본 논문에서는 영상의 텍스처 성분을 추출하고 윤곽선 특징을 이용한 검색 기법을 제안하였다. 실시간 탐색과 더불어 중요한 고려사항은 정지영상 표준, 동영상 표준인 JPEG과 MPEG에서 사용되는 DCT를 기반으로 하여 특징을 추출하였다는 것이다. 텍스처 분류 기법인 DCT-par, DCT-energy와 DCT의 대각선 계수를 영상 및 비디오 특징 추출에 사용하였다. 따라서 제안한 시스템의 특징은 윤곽선 추출이나 텍스처 특징 추출에 있어서 성능은 약간 떨어지더라도 방대한 데이터를 처리함에 있어서 디코딩을 수행하지 않고도 특징을 추출할 수 있다는 점과 정지영상, 동영상 표준에 포함될 수 있는 구조로 되어 있다는 점이다.

향후 연구 계획으로는 현재 제안한 알고리즘을 실제 구현하여 WebCAM에서의 적용이 남아있다 하겠다. 또한 영상 데이터베이스에서 질의 영상에 대한 유사도를 검사하기 위한 효율적인 척도를 개발해야 한다.

참고문헌

[1] 심춘보 외, "내용-기반 멀티미디어 정보 검색을 위한 효율적인 색인 기법의 설계", 한국정보과학회 가을 학술발표논문집, pp.187-190, 1997.  
[2] 이미숙 외, "내용 기반 색인 및 검색을 위

- 한 실시간 뉴스 비디오 파서의 설계 및 구현”, 한국정보과학회 봄 학술발표논문집, pp.365-368, 1997.
- [3] 이병구 외, “멀티미디어 데이터베이스를 위한 영역지식을 이용한 내용기반 검색”, 한국정보과학회 가을 학술발표논문집, pp.147-150, 1997.
- [4] 장석환 외, “칼라 이미지 데이터베이스에서 Region Vector를 이용한 내용기반 이미지 검색”, 한국 정보과학회 가을 학술발표 논문집, pp.143-146, 1997.
- [5] Su-Shing Chen, “Content-Based Indexing of Spatial Objects in Digital Libraries”, Visual Communications and Image Representation, Vol.7, No.1, March, pp.16-17, 1996.
- [6] 이미숙 외, “내용 기반 영상 및 비디오 검색 기술의 연구 현황”, 정보과학회지, 제 15권 제 9호, pp.10-19, 1997.
- [7] Bo Tao, Bradley Dickinson, “Texture Classification on Block-Transformed Data”, Proc. of Visual Communications and Image Processing'97, San Jose, California, Feb. pp. 964-970, 1997.
- [8] B.Shen, I.K.Sethi, “Convolution-Based Edge Detection for Image/Video in Block DCT Domain”, Visual Communication and Image Representation, Vol.7, No.4, December, pp.411-423, 1996.
- [9] 정화자, “DCT를 이용한 윤곽선 추출”, 정보과학회논문지(C), 제3권 제2호, pp.90-97, 1997.
- [10] 최현섭 외, “텍스처패턴과 윤곽점 기울기 성분을 이용한 내용기반 화상 검색시스템의 설계 및 구현”, 한국정보처리학회 논문지 제4권 제1호, pp. 54-66, 1997.