디지털 방송 환경에서 비주얼 리듬을 이용한 재색인화

*조용래 **설상훈

고려대학교

*yrcho@mpeg.korea.ac.kr

Indexing method for reusing the existing information with Visual Rhythm in the digital broadcasting

*Cho, Yongrae **Sull, Sanghoon

Korea University

요 약

본 논문은 디지털 방송에서 다양한 부가 정보 제공 및 관련 기기의 기능을 효과적으로 활용하기 위한 연구로서, 방송 시스템에서의 방송이나 편집 등을 고려하여 색인 정보를 재사용하는 알고리즘에 관한 것이다. 이를 위해 본 논문에서는 비주얼 리듬을 이용한 원본 영상과 목표 영상의 매칭을 제안하며, 비주얼 리듬의 히스토그램을 특징 벡터로 사용하여 유사도를 계산한다. 검색 시 목표 영상에 대해 우선 검색 구간을 설정하여 검색 시간을 줄이고자 하였으며, 실제 이 알고리즘을 적용한 결과 약 97%의 정확도의 매칭 결과를 보였다. 또한 결과를 시각적으로 쉽게 알 수 있었기에 오류를 수정하는데 수월하였다. 이를 통해 기존의 색인화 정보를 목표 영상의 복호화 작업 없이 쉽게 재사용 할 수 있어, 불필요한 비용의 증가를 방지하는 효과를 기대할 수 있다.

1. 서론

고화질의 영상과 고음질의 음향 신호는 물론 디지털 위성 방송에서 보는 것과 같이 140여 개의 다양한 채널을 보유한 디지털 방송은 기존의 아날로그 방송 시스템에서 구현하기 힘들었던 다양한 기능들을 제공할 수 있게 하였다. 아날로그 방송의 경우 단순히 방송 프로그램의 영상과 음성 정보만을 제공했던 것에 비하여 디지털 방송의 경우 이전 아날로그 방송과 비교하여 고품질의 음,영상은 물론 전자 홈쇼핑 및 다양한 인터넷 서비스, 기상정보, 교통 정보 및 뉴스와 같은 문자 정보 등 다양한 정보를 서비스 할 수 있다. 이와 같이 많은 부가 가치를 창출할 수 있는 디지털 방송은 큰 잠재적 가치를 가지고 있으며, 이를 활용하기 위한 다양한 기술이 활발하게 연구되고 있다.

디지털 방송을 즐기기 위해서는 영상을 보여줄 디지털 TV와 함께 MPEG-2 형식으로 전송되는 방송 신호를 수신 및 복호화 하여 영상 및 음성 신호와 기타 부가적인 정보를 처리하는 셋톱박스(Set-top Box)가 필요하다. 셋톱 박스는 상기의 핵심적인 기능 외에 기존의 아날로그 TV에서 VTR(Video Tape Recorder) 역할을 하는 PVR(Personal Video Recorder)의 기능을 제공한다. PVR은 기존의 비디오 테이프 대신 내장된 하드디스크를 활용해 영상신호를 저장하는 대용량 저장 장치이며, 최근 본격적인 HD방송과 함께 그 효용성이 크게 증가되어 수요가 빠르게 늘고 있는 장치이다. 이는 기존의 비디오 테이프에서는 미디어의 한계로 불가능했던 랜덤 억세스가 가능하기 때문에 영상을 반영구적으로 저장할 수 있음은 물론 탐색 및 검색, 편집 등의 기능을 활용할 수 있게 한다.

상기된 부가 서비스 및 PVR의 기능을 보다 효과적으로 활용하기 위해서는 기본적으로 동영상이 가지고 있는 영상, 음성 정보 외에 부가적으로 특정 프레임에서의 줄거리 및 해당 프레임에 나타나는 인물, 위치, 상품 정보 등을 사전에 저장하는 색인화 작업이 필요하다. 이는 현재의 기술력으로서 영상 정보만을 통해 해당 프레임에 나타나는 인물이나 장소, 상품을 인식하는데 기술적인 한계가 있기 때문에 사람이 직접 해당 동영상을 살펴보면서 해당 프레임에 대해 색인화를 하는 방법이 대부분이다. 한번의 색인화 작업을 거친 프로그램이 편집되어 재방송 되거나 다른 방송 시스템에서 방송될 경우 기존의 색인화 작업의 결과물은 쓸 수가 없거나 혹은 옳지 못한 정보를 가지고 있을 수 있기 때문에, 이를 방지하기 위해 해당 프로그램에 대해 다시 색인화 작업을 하는 것은 불필요한 비용의 증가로 이어진다.

따라서 본 논문에서는 한번 색인화 작업을 거친 프로그램의 재색인화를 위한 방법으로 비주얼 리듬(Visual Rhythm)을 이용한 패턴 매칭을 통해서 기존에 했었던 색인화 작업의 결과물을 재사용하는 등 다양한 목적으로 활용될 수 있는 방법을 제시한다.

본 논문의 구성은 총 다섯 개의 장으로 이루어졌으며 1장에서는 서론에 대해 다루었고, 2장에서는 기존에 진행된 연구에 대해 언급한다. 3장에서는 비주얼리듬에 대해, 4장에서는 본 논문에서 제안하는 알고리즘에 대해 논의하며, 5장에서는 실험 결과를, 6장에서 결론에 대해 다룬다.

2. 관련된 기존의 연구

본 논문에서 수행하고자 하는 색인화 된 프레임의 정보를 통해 목표 영상에서 해당 영역을 찾는 방법은 3장에서 다루는 비주얼 리듬을

이용한다. 본 연구에서 수행하고자 하는 목표는 이전에 이루어졌던 비디오 검색 알고리즘이나 정지 영상 검색 알고리즘과 유사하다고 볼 수 있지만 이전에 이루어졌던 연구가 주로 장면 전환 혹은 키 프레임 등을 검출 및 추출하고 그 정보를 토대로 검색을 수행한다는 점이 본 논문이 수행하는 방법과 다른 점이다

가. 프레임을 이용한 매칭

두 비디오의 프레임을 비교하여 각각의 프레임의 유사도를 계산, 매칭하는 방법으로 원하는 영역의 유사도를 정확하고 빠르게 계산해 내는 방법이 이와 같은 알고리즘의 관건이다. 이를 이용할 경우 특히 HD급과 같은 고화질 영상을 비교할 때 한 프레임의 특징 벡터를 계산하는 데에 많은 시간이 걸리는 것은 물론, 검색 대상이 되는 영상 구간에 대해서도 같은 작업을 수행해야 하므로 효율이 떨어진다

나. 순차 검색 알고리즘을 이용한 매칭 및 검색

순차 검색은 프레임의 전, 후 특징 벡터들의 연속적인 변화를 파악하여, 이 정보를 토대로 유사도를 계산, 매칭하는 방법이다. 이는 압축 영영에서의 DC 계수와 유클리디안 거리를 활용한 R. Mohan [1] 과 프레임 간 움직임의 크기와 정지 상태를 이용한 V. Vinod [2] 등이 있다. 이 역시 2. 가 에서 언급한 바와 같이 검출 작업을 위해 필요 이상의 계산 및 복호화 과정으로 효율이 떨어지는 단점이 있다.

다. 비주얼 리듬을 이용한 매칭

동영상은 그림.1 에서 보는 것과 같이 3차원의 시공간적인 정보로 구성되며 이를 2차원으로 매핑한 정보가 비주얼 리듬이다. 비주얼 리듬은 미리 정해진 방식으로 각 프레임에서 특정 픽셀을 샘플링하여 그 집합을 이용해서 만들어진다. 현재 전송되는 HDTV 방송 스트림은 MPEG-2 TS(Transport Stream)로 압축되어 전송 되는데 이 경우, 비주얼 리듬을 얻기 위해서는 MPEG의 경우 모든 프레임을 복호화 하지 않고 압축 영역에서 I 프레임의 DC계수와 P 프레임에서 최대 Motion Vector를 활용해 복호화 영역을 최소화 함으로써 계산량을 줄일 수 있다. 일반적으로 전송에러가 발생하기 쉬운 채널을 사용하는 방송에서는 그렇지 않은 채널 환경에 비해 GOP 내 B 프레임의 사용 빈도가 낮기 때문에 빠른 비주얼 리듬의 추출 및 매칭 작업을 위해 B 프레임은 비주얼 리듬 추출에서 고려하지 않았다.

결과적으로 3차원의 정보를 2차원으로 줄여 간단하게 살펴볼 수 있기 때문에, 매칭 결과 및 정확도를 시각적으로 쉽게 살펴볼 수 있는 장점이 있으며, 데이터 베이스에 저장되는 비디오 검색 및 색인화에 드는 시간적 공간적 비용을 크게 절약할 수 있다.

기존의 색인화된 결과물을 토대로 이루어지는 작업이기 때문에 비주얼 리듬으로 얻어진 두 영상의 변화와 매칭된 영역의 유사도를 간략하게만 살펴보는 것만으로 충분하다.

3. 비주얼 리듬의 정의

가. 비주얼 리듬

본 논문에서 구현하고자 하는 알고리즘을 위한 비주얼 리듬의 정의는 다음과 같다. $f_{DC}(x,y,t)$를 W*H의 크기를 지니는 DC image[3]에

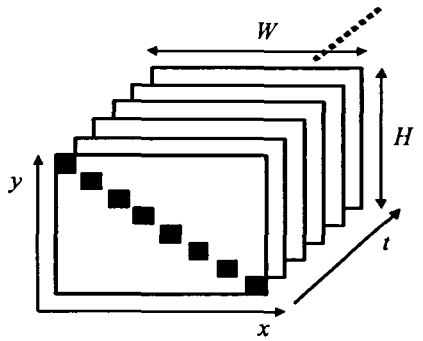

그림. 1 동영상의 3차원 구조

서 본래의 프레임 t 의 시점에서 좌표(x,y)의 픽셀 값이라 정의한다. 그림. 1에서 보는 것과 같이 모든 픽셀들을 사용하지 않고 샘플링 한 픽셀들을 사용하므로, 보다 정확한 실험을 위해 DC 계수로 구성된 DC 영상에 대해 샘플링을 실시한다. 이 때 비주얼 리듬은 다음과 같이 정의 할 수 있다.

$$VR = \{f_{vr}(z,t)\} = \{f_{DC}(x(z),y(z),t)\} \qquad (1)$$

위에서x(z)와 y(z)는 독립 변수 z에 대한 1차원 함수로써, 위 식(1)으로부터 비주얼 리듬은 세로축 z방향으로 시간을 나타내는 가로축 t에서 샘플링 된 픽셀 값들의 누적으로 2차원 평면을 형성하는 것을 알 수 있다. 그림 2.에서 쉽게 알 수 있듯이 비주얼 리듬은 3차원 정보인 DC 영상을 샘플링하여 구성한 2차원 영상이다.

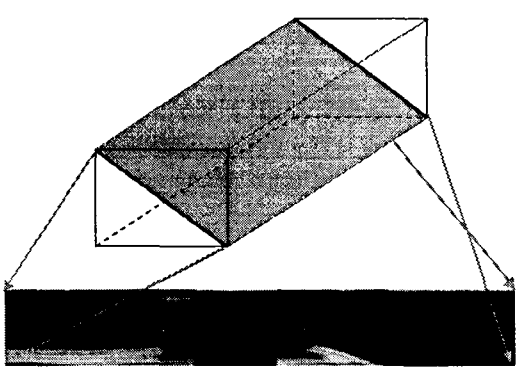

그림. 2 비주얼 리듬의 추출

여기에서 x(z)와 y(z)로 표현되는 샘플링 방식을 바꿔 표현한 H.Kim[3]의 실험에 의하면 일반적으로 수평 방향의 샘플링으로는 수평 방향의 움직임을 검출할 수 없고, 수직 방향의 샘플링으로는 수직 방향의 움직임을 검출할 수 없다. 따라서 본 논문에서는 교차 대각선 샘플링이나 영역 샘플링 방식과 검출에서 유사한 성능을 보이면서 데이터 양을 상대적으로 줄일 수 있는 대각선 방향으로 샘플링 된 비주얼 리듬을 사용하고자 한다.

나. 비주얼 리듬의 고속 추출

많은 영상 압축 표준에서는 이산 코사인 변환(DCT)를 인트라 프레임의 부호화를 위해 사용한다. 이는 비주얼 리듬을 추출하는데 필요한 DC 영상을 얻기 위해 굳이 복호화 할 필요는 없다는 것을 의미한다. DC 계수는 각 인트라 프레임에서 쉽게 얻을 수 있으며, MPEG에

서 사용되는 P- 혹은 B- 프레임 등과 같은 인터 프레임의 DC 영상을 얻기 위한 방법으로 MPEG-1이나 MPEG-2에 대해 Yeo.B,L[4] 과 Song.J[5]에 의해 이미 연구된 바 있다. 그러므로 비주얼 리듬은 적어도 MPEG 등과 같이 DCT를 기반으로 하는 영상 압축 표준에서는 좀 더 빠른 속도로 추출 될 수 있다.

4. 제안된 알고리즘

본 장에서는, 비주얼 리듬을 활용하여 원하는 프레임을 검색하는 알고리즘에 대하여 설명하고자 한다.

가. 특징 벡터 추출

비주얼 리듬은 각 프레임에서 샘플링 된 픽셀들의 배열의 흐름으로 구성된다. 이는 그림.2 에서 알 수 있듯이 세로축은 각 프레임에서 샘플링 된 픽셀들의 배열을 나타내고 가로축은 시간의 흐름을 나타낸다.

본 논문에서 특징 벡터는 일정한 프레임 구간에 대한 색상 정보들을 이용하여 추출된다. 색상 정보를 표현하는 방법 중, 자주 사용하는 방법인 컬러 히스토그램을 이용하여 특징 벡터를 얻어내는데, 이 정보들을 추출하는데 있어 비주얼 리듬의 R(Red), G(Green), B(Blue) 세 가지 색깔을 모두 사용할 경우 각각 1BYTE씩 3BYTE, 총 2^{24} 개의 색에 대한 히스토그램이 필요하다. 본 논문에서는 검색 속도의 향상을 위해 R, G, B 의 정보 중 상위 비트만을 사용한다. 세 가지 색 중 사람의 눈에 가장 둔감한 B에 대해 2 bit를 할당하고 나머지 R, G에 대해 3 bit를 할당하여 총 256개의 color bin 을 구성하여 이에 대한 히스토그램을 얻는다. 이 히스토그램의 Bin을 유사도 계산을 위한 특징 벡터로 사용하게 되면 특징의 정보량이 줄어들어 계산이 빨라지는 것은 물론 변환 과정을 통한 색상이 변하는 것 역시 양자화를 통해 보정해 줄 수 있는 장점이 있다. 그림.3 는 하나의 픽셀이 RGB24(100,162,84)의 값을 가지고 있을 때, 256개의 color bin 중 하나로 표현되는 예를 보여준다.

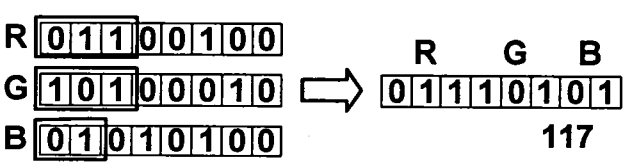

그림. 3 RGB 공간에서의 특징 벡터 추출

나. 검색 구간의 설정

본 논문에서는 검색 정확도를 높이기 위한 방법으로 가장 먼저 검색이 이루어져야 할 구간을 원본 영상의 색인화 된 정보를 바탕으로 정한다. 전혀 다른 영상이 아닌 유사한 비디오 스트림을 가지는 두 영상간의 매칭 작업이므로, 검색하고자 하는 색인화 된 프레임보다 바로 전 단계의 색인화 된 프레임간의 차이를 감안하여 그 구간에서 우선적으로 유사 프레임을 검색한다. 다른 구간보다 먼저 우선 검색 구간에 대한 유사도 계산이 이루어지며, 이 구간에서 검색되지 않았을 경우 그 외 구간에 대해 유사도 계산이 이루어진다. 우선 검색 구간은 프레임 레이트의 변화를 감안하고, 편집을 염두에 두어 전체 영상 길이의 약 5%에 해당하는 길이만큼 허용 오차를 두어 정한다. 즉 프레임 레이트의 변화가 없다고 판단될 경우 전 색인화 프레임과의 차이에서 전체 영상 길이의 5%에 해당하는 곳을 우선 검색 구간으로 설정하고, 그렇지 않을 경우에 프레임 레이트의 변화를 감안한 위치를 중심으로 전체 영상 길이의 5%에 해당하는 구간을 우선 검색 구간으로 설정한다. 이를 그림으로 나타내면 그림. 4와 같다.

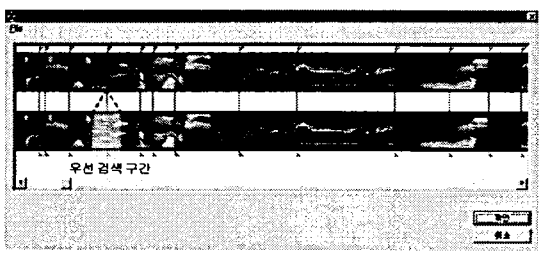

그림. 4 우선 검색 구간의 설정

다. 유사도 계산 알고리즘

본 논문에서는 질의 영상을 찾기 위한 방법으로 비주얼 리듬으로 장면 검출을 수행하기 위한 M.Chung[4]의 알고리즘을 응용하였다

질의 영상의 50 frame 에 해당하는 비주얼 리듬과 재색인화 하고자 하는 영상의 비주얼 리듬과의 SAD를 계산하여 질의 영상에 해당하는 구간을 찾아낸다. 이와 같은 방식으로 검색 구간 내에 검출된 영상이 존재한다면 기존에 색인화 했던 정보를 해당 구간에 적용시킬 수 있다. 만약 프레임 레이트의 변화가 있다고 예상될 경우 그에 비례하여 검색할 구간의 frame 수를 조절한다. 프레임 레이트의 변화는 [3]에서 이루어진 연구를 바탕으로 영상의 시작 위치를 대략적으로 맞춘 이후 shot이 나타나는 간격을 측정하여 이를 토대로 결정한다.

상기 알고리즘을 적용하기 위해 우선적으로 첫 색인화 된 프레임을 검색하는 작업이 이루어져야 한다. 이 경우 앞에서 논의한 전 단계의 색인화된 프레임 정보가 없기 때문에 제안된 알고리즘을 적용할 수 없다. 이미 영상의 프레임이나 시작 위치에 대한 정보가 존재할 경우 그 것을 활용해 찾을 수 있겠으나, 비슷한 두 영상인 것과, 데이터베이스가 구성되지 않아도 수행 될 수 있도록 프레임 레이트의 변화가 없을 때를 기준으로 75 frame에 해당하는 비주얼 리듬 정보를 활용하여 첫 유사 프레임 검색이 이루어진다. 첫 유사 프레임을 찾았을 경우 이 후엔 전 단계의 색인화된 프레임의 위치와 질의 프레임과 이후 49frame, 총 50frame 의 특징 벡터를 활용해 유사도 계산 작업이 반복적으로 이루어진다. 유사도 계산은 히스토그램 간의 유클리디안 거리 값들로 이루어지며, 이는 다음과 같이 나타낼 수 있다.

$$D(z,t) = \sum \left| F_{VR1}(z,t) - F_{VR2}(z,t) \right| \qquad (2)$$

여기에서 F_{VR1} 과 F_{VR2}는 각각 원본 영상과 목표 영상의 비주얼 리듬을 토대로한 히스토그램 값을 의미한다. 히스토그램간의 거리값이 어떤 임계치 이하로 나타날 경우, 해당 구간은 유사 구간으로 판단하여 계산을 시작한 프레임간의 매칭을 수행한다.

위와 같은 우선 검색 구간에서 실제로 탐색을 수행한 결과는 그림. 5 와 같고, 검색 구간 중 가장 적은 오차를 보이는 프레임을 우선적

으로 선택한다. 일반적으로 대부분의 우선 검색 구간에서 그림.5 와 같은 유사도 분포를 보인다. 우선 검색 구간 내에서 만약 같은 거리값을 가지는 프레임이 검색될 경우는 예상 위치와 가장 가까운 프레임을 선택하여 매칭한다.

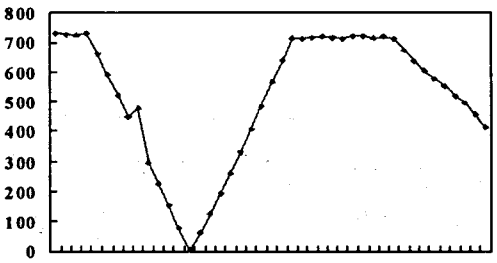

그림. 5 우선 검색 구간 내 유사도 분포

5. 실험 결과

가. 실험 개요

이 실험에서 사용한 동영상은 실제로 색인화 작업이 이루어진 본 방송 이후 재방송을 한 6개의 영상과 본 방송만 방영하는 임의 편집된 실험 영상 21개로 총 1260분의 분량에 해당하며, 실험에 사용된 PC는 P4-1.4GHz, 512M RAM의 사양이다.

나. 실험 결과

우선 기존의 색인화 정보를 활용하기 위한 목표 영상의 비주얼 리듬을 측정하는데 걸리는 시간은 1시간 분량의 경우 약 3분의 시간이 소요된다. 실제로 각종 부가 정보를 색인화한 드라마와 쇼 프로그램 영상 6개의 경우 색인화 정보가 존재하는 프레임의 수는 아래의 표.1 과 같다. 색인화 정보는 장면 전환에 따른 줄거리 및 인물 정보와 음악 정보, 장소 정보에 대하여 작업이 이루어 졌다.

분야	평균 장면 전환 수	기타 색인화 위치
드라마	24.0	93.5
연예/오락	22.4	14.8
기타	8.6	11.2

표 1. 프로그램 별 줄거리 정보 및 부가 정보 분포

하나의 프로그램에 존재하는 부가 정보의 양은 프로그램마다 다르겠지만 평균적으로 위와 같은 양을 가지고 있었으며, 본 논문에서는 매칭이 얼마나 정확하게 이루어지는가를 측정하기 위해 상기 영상들 외 기타 영상들에 대해서는 측정된 색인화 프레임의 수 정도를 가지는 임의 위치에서 색인화를 가정하고 실험하였다.

원 영상의 색인화 정보와 그의 비주얼 리듬을 사용해 목표 영상의 재색인화를 실행하였을 때, 제안된 알고리즘에 의한 매칭 결과는 아래 표.2 와 같다.

	색인화 프레임 수	missing	false
개수	407	5	7

표 2. 실험 결과

표.2 에서 볼 수 있듯이 매칭 결과는 만족할 만한 수준이었으며, 매칭 작업에 걸리는 시간 역시 한 시간 분량의 영상을 기준으로 평균 10초 내외로 단시간에 검색이 가능했다. 오차가 발생한 부분도 그림. 6 에서 보는 것과 같이 쉽게 확인할 수 있었기 때문에 수정 작업도 쉽게 이뤄질 수 있다.

6. 결 론

제 5 장의 표에서 보는 것과 같이 하나의 방송 프로그램에는 많은 부가 정보가 존재한다. 기존의 색인화 정보를 이용하여 재색인화 작업을 할 경우 만족할만한 정확도를 얻을 수 있었으며, 비주얼 리듬이 가지는 특성으로 오류 역시 한 눈에 확인할 수 있었다. 비단 특정 프로그램의 재색인화 외에도 비슷한 순서를 지니며 일부 광고가 추가 및 삭제된 스트림의 매칭 작업이나 이와 유사한 영상 스트림에도 본 논문의 알고리즘을 적용할 수 있을 것이라 생각된다.

단 프레임 별 특징 벡터를 추출하여 비교하는 알고리즘이기 때문에 프레임 레이트가 크게 줄었을 경우, 비주얼 리듬의 영상 정보만으론 프레임 레이트의 변화까지는 파악할 수 없어서 매칭 작업의 정확도가 떨어진다. 이를 보정하기 위해 프레임 레이트 변화에 따른 유사도 변화에 대한 더 많은 연구가 필요하다.

그림. 6 miss 발생 프레임의 예

참고문헌

[1] Mohan, R, "Video sequence matching," Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal processing, vol. 6, pp. 3697 - 3700, May 1998.

[2] Vinod, V.V. "Activity based video shot retrieval and ranking," Proceedings. Fourteenth International Conference on Pattern Recognition, vol 1, pp. 682 - 684, Aug. 1998

[3] Hyeokman Kim, Jinho Lee and Moon-Ho Song, S. "An efficient graphical shot verifier incorporating visual rhythm," IEEE International Conference on Multimedia Computing and Systems, Vol. 1, pp. 827 - 834, 1999

[4] Yeo, B. L and Liu, B, "Rapid Scene Analysis on Compressed Video," IEEE Transactions on Circuit and Systems for Video Technology, vol. 5. pp. 533-544, 1995

[5] Song, J and Yeo, B.L, "Spatially Reduced Image Extraction from MPEG-2 Video," Fast Algorithms and Application. Proceedings of SPIE Storage and Retrieval for Image and Video Database VI, Vol. 3312. pp. 92-107, 1998