

MPEG-7 Visual Identifier

오 원근

한국전자통신연구원 디지털콘텐츠연구본부

e-mail : owg@etri.re.kr;

Weon-Geun Oh

Digital Contents Research Division

Electronics and Telecommunications Research Institute(ETRI)

Abstract

Digital visual information(image and video) plays an important role in our society, and everyday, more and more visual information is available from many sources around the world. And there is an increasing number of cases where the these visual informations are created, stored, retrieved, and re-used by computational systems. But information modification, distortion and compilation occurred through trans-coding or legal/illegal information processing often make these operations difficult. Subordinately, advanced technology or commercialized software for retrieving and identifying the desired information among these massive and diverse ones is strongly required.

In the paper, some recent activities and technical content of MPEG-7 Visual Group are described, especially with regarding to Visual Identifier.

I. 서론

디지털 콘텐츠에 대한 수요가 증가하면서 우리 주변에는 방대한 양의 멀티미디어 콘텐츠(비디오, 음악, 영상 등)가 끈임 없이 생성, 제작, 유통 및 서비스 되고 있다. 그 중에서도 디지털 영상은 고성능, 휴대용 디지털 카메라의 보급과 대용량 저장장치나 휴대용 저장매체의 가격 하락과 함께 가히 폭발적으로 그 수가 증가

하고 있다. 반면에 이러한 영상을 필요시 효율적으로 찾아내고, 원하는 대상을 다른 것과 정확하게 식별하는 제품에 대한 수요도 점점 늘어나고 있는데 그 예가 필요한 영상의 효율적인 검색과 복사(혹은 모사)나 불법적으로 유통되고 있는 영상의 식별 기술이다. 한편, 이들 영상은 인터넷이나 재생장치의 변화 혹은 사용자의 콘텐츠 가공이나 재편집 등에 의해서 본래의 특성(영상의 크기, 형태나 품질, 색상 등)이 변할 경우가 있는데 이 경우에도 변형된 영상을 정확하게 식별하는 기술이 요구된다(그림 1).



그림 1. 영상의 변형

이러한 영상 식별기술(visual identifier)은, 미디어 관리 및 검색, 미디어 모니터링, 콘텐츠 Linking, 콘텐츠 확인 등 매우 광범위한 분야에서 활용이 될 수 있다. MPEG(Moving Picture Experts Group : ISO/IEC JTC 1/SC 29/WG 11)에서는 영상 식별기술의 중요성을 파악하여 수년전부터 기술의 표준화를 추진하고 있으며 정지영상 식별기술에 대해서는 거의 완성단계에 있다.

본 논문에서는, MPEG-7 Visual Group에서 진행중인 영상 식별기술의 표준화에 대한 개요와 기술적인

토의사항, 그리고 최근의 표준기술에 대해 설명한다.

II. 본론

2.1 MPEG-7 VCE 개요

MPEG-7 VCE(Visual Core Experiments)는 그 동안 디지털 정지영상(image)이나 동영상(video)의 관리나 검색기술에 대해 다양한 논의와 표준화를 추진해 왔다. 대표적으로는 유사한 환경(situation)이나 풍경(view) 정보를 기반으로 한 정지영상 클러스터링 기술(VCE-1), 해변/바다/강/실내 등의 카테고리 정보를 기반으로 한 영상 분류기술(VCE-2), 인물정보를 기반으로 한 정지영상의 클러스터링/인덱싱/검색 기술(VCE-3), 얼굴인식 기술(VCE-4), 적외선 영상에서의 얼굴인식 기술(VCE-5) 등이 있다.

이러한 기술들은, 근본적으로 대상 영상들의 유사성을 기반으로 한 방법으로, 영상의 용량이 기하학적으로 늘어나거나 영상의 기하학적인 변형에는 취약한 것이 특징이다. 이러한 기술적 배경으로 MPEG에서는 기하학적/비기하학적으로 변형 혹은 편집이 된 영상을 식별하는 기술에 대한 표준을 진행하게 되었는데 이것이 Image Signature(VCE-6)와 Video Signature(VCE-7)로 표현되는 Visual Identifier 기술이다. 이 기술의 표준은, 2005년 7월(73th MPEG)에 ETRI에 의해서 그 필요성이 제기되었으며, 이후 '06.7.-'06.10.(75th~78th MPEG)에서 실험 및 경쟁조건을 갱신하였으며, '07.1.(79th MPEG)에서는 VCE-7(Video Signatures)이 추가되었고 지난 4월(84th MPEG)에서는 VCE-6의 비기하학적 변형의 정지영상 식별기술에 대해서 FPDAM 단계로 발전이 되었다.

2.2 MPEG-7 VCE-6(Image Signature)

Image Signature 기술은, 정지영상을 식별하는 것이 주목적이고, 기술적으로는 식별의 정확성과 안정성에 초점을 두고 있으며, 대표적인 응용 예는 정지영상의 모사품(혹은 복제품)을 검출하는 하는 것이다. 이 기술의 표준을 진행하는데 있어서 기술을 평가하는 기준은 크게, 독립성 테스트(independence test)와 강인성 테스트(robustness test)가 있다.

2.2.1 실험 영상 데이터베이스

실험 영상 데이터베이스는, Nova사의 상용 데이터베이스인 "Art Explosion 800000"이며, 독립성 테스트를 위해 이중 135K의 서로 다른 정지영상을 선정하였다.

2.2.2 독립성 테스트(independence test)

이 영상들은, 제안된 기술의 독립성에 대한 임계값(threshold), 즉 각 기술의 최대 false positive rate를 결정하는데 사용되며, VCE-6에서는 이 값을 다음과 같이 설정하였으며, 이 값은 다음의 강인성 테스트의 최저 임계값(threshold)으로 사용된다..

- For basic modifications only: $\leq 0.05\text{ppm}$ (parts per million)
- For all modifications: $\leq 10.0\text{ ppm}$ (parts per million)

여기서 변형(modification)의 종류는 다음 절(2.2.3)의 표 1과 같으며, 제안된 기술은 135K의 영상의 모든 쌍(pair)의 갯수($N*(N-1)/2$)에 대해 비교를 하여야 한다.

2.2.3 강인성 테스트(robustness test)

강인성 테스트를 위해서 각 기술들은 표 1과 같은 다양한 영상의 변형에 대해 안정성을 실험하고, 그 결과(success ratio)는 표 1의 카테고리별로 작성된다. 이 실험을 위해서 2차적으로 사용된 영상 데이터베이스는, M(=10K)개의 원 영상을 선정하고, 각 원영상은 표 1에 있는 44종류의 변형을 가한 후에 실험영상으로 활용된다.

표 1. 정지영상 변형의 종류 및 레벨

Modifications	Levels	Heavy	Medium	Light
Brightness change		+25%	+20%	+10%
Color to monochrome conversion		N/A	N/A	$I = 0.299 \times R + 0.587 \times G + 0.114 \times B$
JPEG compression with varying Quality factors		QF 30	QF 60	QF 80
Color reduction			GIF 8bit version	16bit: RGB(565)
Gaussian Noise		12.0	8.0	4.0
Histogram equalization		N/A	N/A	YES
Image enhancement via Auto-level		N/A	N/A	YES
Blur		mask size: 7x7	mask size: 5x5	mask size: 3x3
	Rotation	270°	180°	90°
	Rotation	45°	25°	10°
	Scaling (width-height ratio)	25% (50%)	49% (70%)	81% (90%)
	Translation	49% (30%)	64% (20%)	81% (10%)
	Flip	N/A	N/A	left-right
	Aspect ratio change	N/A	Width x 1.5 (e.g. 4:3 → 6:3)	Width x 1.33 (e.g. 4:3 → 16:9)
	Crop	36% (60%)	60% (77%)	81% (90%)
	Skew	+10°	+6°	+4°
	Perspective (Focal length=500)	+10°	+6°	+4°
	Combined Crop, Translation & Scaling	C=70% (84%), S=50%	C=81% (90%), S=65%	C=90% (95%), S=80%

2.2.3 성능평가(performance measure)

성능평가의 기준은 크게 계산속도와 알고리즘에 사용되는 특징의 크기 및 정합율(SR : Success Ratio)로 정하였다. 성능평가를 위해서는 우선, 원영상과 표 1에 있는 각각의 변형 카테고리상의 변형된 영상에 대해 제안된 기술을 통해 비교하여 정확하게 정합된 비율(SR)을 다음의 식에 의해 계산한다.

$$SR = K/M$$

여기서 SR은 정합율, M은 실험용 데이터베이스에서 선정된 영상의 개수(본 실험에서는 M=10K로 정합)이고, K는 변형된 총 영상(10K*44=440K)중 정확하게 원 영상과 정합된 경우의 수를 나타낸다.

계산 속도는 영상의 비교 횟수를 나타내는데, 빠른 알고리즘과 느은 알고리즘에 대해 두 가지로 나누었다. 전자는 비기하학적 변형 영상에 대해서, 후자는 전체(비기하학적/기하학적 변형 영상)에 대한 처리속도를 나타낸다. VCE-6에서는 PC-class computer (CPU<=3.4GHz)에서의 계산 속도를 기준으로 하였으며, 최하기준을 각각 2,000,000회/초와 100,000회/초로 정하였다.

특징의 크기에 대한 평가는, 알고리즘에서 사용한 특징을 bit수로 평가하였는데, 빠른 알고리즘은 1Kbit를, 느은 알고리즘은 1KByte를 최고 임계값으로 정하였다.

2.2.4 평가 절차

평가 절차는, 표 1의 변형 카테고리상의 light 레벨에서 각각 SR이 75%가 넘을 경우에 대해서 제안된 기술중에서 우선적으로 SR이 높은 기술이 선정된다. SR이 특별히 차이가 없을 경우에는 계산 속도를 비교해서 선정한다.

2.3 MPEG-7 VCE-7(Video Signature)

Video Signature, 동영상(video clip)을 식별하는 것이 주목적이고, 기술적으로도 Image Signature 와 같이 식별의 정확성과 안정성에 초점을 두고 있으며, 대표적인 응용 예는 동영상의 모사품(혹은 복제품)을 검출하는 하는 것이다. 이 기술의 표준을 진행하는데 있어서 기술을 평가하는 기준도 독립성 테스트(independence test)와 강인성 테스트(robustness test)가 있다.

2.3.1 실험 영상 데이터베이스

실험 영상 데이터베이스는, VCE-6에 참가하는 여러 기관(예, KBS, NTT, NEC, 인하대 등)에서 수집한 3분 clip을 기준으로 200시간 분량의 다양한 장르(뉴스, 스포츠, 드라마, 자연 풍경, 만화, 홈비디오 등)의 동영상을 활용하였다. 선정된 원동영상 clip은 두 가지 방식에 의해서 query clip과 비교과정을 거치게 되는데, 하나는 3분 clip에서 잘라낸 2초, 5초, 10초의 부분영상을 query로 이용하는 direct content 정합방식이고, 다른 하나는 잘라낸 2초, 5초, 10초의 부분영상과 다른

영상을 조합해 3분 clip을 query로 이용하는 partial content 정합방식이다(그림 2).

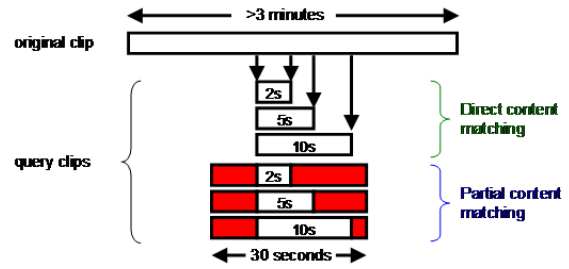


그림 2. 6종류의 Query 동영상

2.3.2 독립성 테스트(independence test)

독립성 테스트를 위해서, 실험 영상 데이터베이스에서 선정된 4,083개의 동영상 clip은, 6종류의 query에 대해 모든 쌍(6*4,083*(4,082-1)/2=50,000,418회)에 대해 비교를 한다. VCE-7에서는, 각각의 query에 대해 1ppm을, 전체(6개) query에 대해서는 50ppm을 적용하였다. 이 임계값(threshold)은 다음의 강인성 테스트에 사용된다.

2.3.3 강인성 테스트(robustness test)

강인성 테스트를 위해서 각 기술들은 표 2와 같은 다양한 동영상의 변형에 대해 안정성을 테스트하고, 그 결과는(success ratio)는 표 2의 카테고리별로 작성된다. 이 실험을 위해서 3분 이상의 1,000개 원 동영상 clip(필름, 뉴스, 도큐멘터리, 만화, 스포츠, 홈비디오 등)이 선정되고, 각 clip에 대해서 3개의 query(2, 5, 10초 동영상 clip)과 이들을 30초 이상으로 타 영상과 조합한 query를 사용해 이들을 원 동영상 clip과 비교하여, 강인성을 테스트한다.

표 2. 동영상 변형의 종류 및 레벨

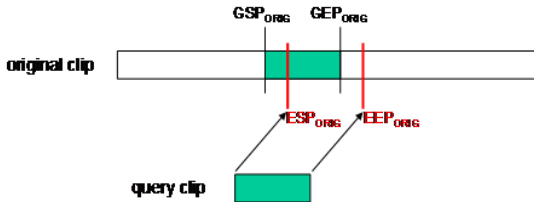
Modifications	Levels	Coding	Heavy	Medium	Light
Text/logo overlay		MPEG-2	30%	20%	10%
Severe compression (at CIF resolution)		AVC	64kbps	56kbps	512kbps
Resolution reduction (from SD)		MPEG-2	N/A	QCIF	CIF
Frame-rate reduction (from 30 fps)		AVC	4fps	5fps	15fps
Capturing on camera (at SD resolution)		MPEG-2	10%	5%	0%
Analog VCR recording & recapturing (100% of image captured)		MPEG-2	3 times	2 times	1 time
Color to monochrome conversion		MPEG-2	N/A	N/A	$I = 0.299 \times R + 0.587 \times G + 0.114 \times B$
Brightness change		MPEG-2	+36	-18	+9
Interlaced / Progressive conversion		MPEG-2	N/A	N/A	$P \rightarrow I \rightarrow P$ $I \rightarrow P$

2.2.3 성능평가(performance measure)

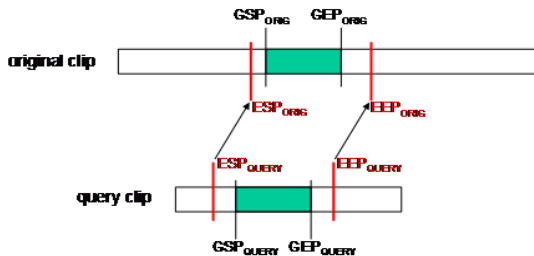
성능평가의 기준은 역시 계산속도와 알고리즘에 사용되는 특징의 크기 및 정합율(SR : Success Ratio)로

정하였다.

Direct content 정합방식의 경우의 정합의 조건은, 그림 3(a)과 같이 원 동영상 clip의 크기와 query clip의 크기가 같아야 하고, 50% 이상 겹쳐야 하고, partial content 정합의 경우는 그림 3(b)과 같이 query clip이 원 동영상 clip내에 포함되어야 한다.



3(a) Direct content 정합



3(b) Partial content 정합

그림 3. 정합의 조건

성능평가를 위해서는 우선, 원 동영상 clip과 표 2에 있는 각각의 변형 카테고리상의 변형된 query clip에 대해 제안된 기술을 통해 비교하여 정확하게 정합된 비율(SR)을 다음의 식에 의해 계산한다.

$$SR = K/M$$

정합율(SR)은, 6개의 query를 기준으로 모든 변형의 종류와 레벨(총 22 카테고리)에 대해서 계산한다. 여기서 M은 132,000(1,000*22*6)이 되고 K는 정합조건을 만족하는 query의 개수를 나타낸다.

VCE-7에서의 계산 속도의 임계값은, 1,000 clip의 쌍에 대한 전체 계산시간을 1초 이내로 정하였다. 또, 특징의 크기에 대한 평가는, 알고리즘에서 사용한 특징을 bit수로 평가하였는데, 1초의 content에 해당하는 특징의 크기가 30kbit를 넘지 않는 것으로 정하였다.

2.2.4 평가 절차

평가 절차는, 2초 query clip에 대해 direct/partial

content 정합에 대해 표 2의 변형 카테고리상에 있는 9개의 light 레벨에서 각각 SR이 75%가 넘을 경우를 만족해야 하고, 이 조건을 만족하지 못할 경우에는 정합속도를 낮추어서 비교를 한다.

III. 결론

현재 MPEG-7 VCE-6는, Mitsubishi UK의 “Trace Transform과 Harris Corner 기반의 알고리즘”과 ETRI/인하대의 “Concentric Circle과 Modified Generalized Symmetry Transform 기반의 알고리즘”이 정합을 벌이고 있으며, 2008년 10월(86th MPEG)을 목표로 표준을 진행하고 있으며, VCE-7은 2008년 7월(85th MPEG)에 CFP를, 2009년 1월(87th MPEG)부터는 제안된 기술에 대한 평가를 시작하기로 예정되어 있다. 특히, MPEG-7 VCE-7은 기술의 중요성과 산업에 대한 파급성이 큰 것으로 인해 기존의 ETRI, Mitsubishi UK, NEC 등은 물론, Fraunhofer HHI(독일), Joost Technologies(미국), Institute for Infocomm research(싱가포르) 등의 새로운 기관이 추가로 참가하고 있어 향후 치열한 경쟁이 예상된다.

MPEG-7 Visual Identifier 기술은, 2010년경에 테라바이트(1조 Gbyte)로 예상되는 초대용량 콘텐츠 시대에 콘텐츠의 관리 및 검색에 필수적인 기술로 활용될 것이며, 특히 콘텐츠의 내용과 특성의 다양성이 그 특징인 UCC(User Created Content)에 대해서는 그 필요성이 더욱 크게 부각될 것이다.

참고문헌

- [1] Video subgroup, “Description of MPEG-7 Visual Core Experiments”, MPEG Doc. No. N767, April 2008.
- [2] Paul Brasnett, Miroslaw Bober, “the Trace Transform Image Signature to Complex Conditions”, MPEG Doc. NO. m15172, January 2008.
- [4] Weon-Geun Oh, Won-Keun Yang, Ayoung Cho, Dong-Seok Jeong, “The Extra Experiment Result to Verify the Method of Performance Measure on MPEG-7 VCE-6”, MPEG Doc. NO. m15131, January 2008.
- [5] Paul Brasnett, Miroslaw Bober, “Proposal on Image Signature for Complex Conditions”, MPEG Doc. NO. m15449, April 2008.