

셋톱박스 품질검사를 위한 개선된 지역 방향 패턴 (eLDP) 기반의 비디오 샷 경계 검출

조영탁^{***}, 안기욱^{*}, 김민기^{*}, 이태원^{*}, 송기훈^{***}, 채옥삼^{**},

^{*}(주)이타기술

^{**}경희대학교 컴퓨터공학과

e-mail : {ytcho, kiokahn, mingi, twlee, ghsong}@itatech.co.kr*, oschae@khu.ac.kr**

Video Shot Boundary Detection based on Enhanced Local Directional Pattern(eLDP) for Set-top Box Quality Control

Youngtak Cho^{***}, Kiok Ahn^{*}, Mingi Kim^{*}, Taewon Lee^{*}, Gihun Song^{***}, Oksam Chae^{*}

^{*}ITA Technology Corporation

^{**} Dept. of Computer Science and Engineering, Kyung Hee University

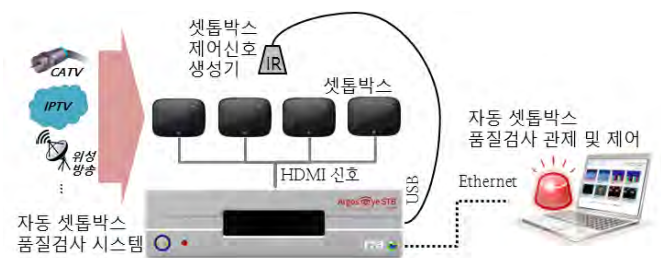
요 약

디지털 비디오의 발전이 가속됨에 따라, 비디오 샷 경계 검출은 비디오 분석 및 카타로깅 등 여러 분야에 있어 필수적인 요소가 되었다. 기존 샷 경계 검출 방법들은 잡음이나 카메라 혹은 물체의 이동, 그리고 색상의 급격한 변화 등에 민감한 성능을 보인다. 본 논문에서는 개선된 지역 방향 패턴 기반(eLDP) 검출 방법을 제안한다. 제안하는 방법은 RGB 색상의 일부와 eLDP의 특징을 결합해 더욱 강인한 샷 경계 검출 성능을 보였다. 또한, 셋톱박스 품질검사 시 필요한 채널 간 동기화의 신뢰성을 높였고, 실시간으로 검사하면서도 안정적인 샷 경계 검출이 가능함을 입증하였다.

1. 서론

디지털 비디오의 발전은 다양한 규격의 비디오를 여러 시장에 서비스로써 진출시켜 시장의 비디오 규격 전환을 활발히 만들고 있다. 이러한 규격 전환은 디스플레이에서의 비디오 규격 지원을 위한 장치를 셋톱박스(STB)로 분리한다. 동시에, 비디오 서비스 사업자는 케이블/인터넷망 및 위성신호 등을 이용한 차별화 전략을 STB로 구현함으로써 STB의 보급을 가속하고 있다. 또한, 비디오 서비스 사업자는 STB에 대한 무결성 검증 이후 양산이라는 작업 흐름을 비용 절감 차원에서 필수 요소로 선택하고 있다. STB 검사를 자동화시키기 위한 노력으로 STB 제어 구성에 대한 스크린샷을 비교 평가하여 검사를 자동화 하는 시스템이 제안되었으나[1], STB 제어에 대한 스크린샷 관리자에 의해 구성되어야 하므로 상당한 사전 작업을 필요로 한다. 게다가, 사용자 인터페이스 재구성에 따른 S/W 업데이트 혹은 새로운 STB에 적용하기 위해 사전 작업을 반복해야 하는 번거로움이 따른다. 이러한 단점을 극복하려는 방법으로 두 대 이상의 STB를 연결하여 사전 작업 없이 STB 제어 구성에 대한 HDMI 출력의 동일성 검사만으로 자동화 가능하도록 시스템을 구성하였다. 그러나, STB는 같은 H/W, S/W 조건에서도 비디오의 동기가 일치하지 않는 상태를 보여, 비디오 샷 경계를 검출하여 HDMI 출력 비디오에 대한 동기를 재구성하고, 이후 동일성 검사를 진행한다. 이러한 방법은 STB가 FHD 해상도의 지원

이 일반화되고, UHD로 전환되는 현시점에 연산을 위한 자원 확보에 유리한 조건을 제공한다. 이처럼, 샷 경계를 검출하는 것은 비디오 분석 및 카타로깅 등 여러 분야에 널리 활용되고 있다.



(그림 1) 자동 셋톱박스(STB) 품질검사 시스템

샷은 단일 카메라에서 촬영된 끊어지지 않은 연속적인 영상 시퀀스를 일컫는다. 대부분의 비디오 매체는 영상 시퀀스가 모여 샷을 이루고, 샷이 모여 장면(Scene)을 이룸으로써 하나의 비디오로 구성된다[2]. 샷의 경계는 연속된 영상 사이의 차이를 비교하는 기본적인 연산으로 검출할 수 있다. 그중에서도 가장 간단한 방법은 영상 사이의 각각 같은 위치에 있는 화소들의 밝기 또는 색상 값을 비교하는 것이다. 이는 복잡도가 낮고 빠르다는 장점이 있지만, 전체 화면 영역에서 비디오가 재생되는 경우에는 카메라 또는 물체의 움직임에 굉장히 민감하다. 반대로, 그림 2와 같이 STB 자동 품질 검사 시 나타나는 메뉴에서, 대부분 화면이 변화가 없고 작은 영역에서만 비디오

가 재생되는 경우 샷 경계 검출이 무더지는 단점이 있다. 그 때문에 영상 내에서 이러한 움직임에 강한 특징을 추출하여 비교하는 연구가 이루어졌는데, 검출에 사용되는 특징엔 아래와 같은 종류들이 있다.

- 통계 수치(평균, 표준편차 등)
- 히스토그램(흑백, 색상(RGB, HSV, JND 등))
- 압축 기법(DCT, DFT, Wavelet 등)
- 에지 추적(영상 내 에지 비율의 변화 등)
- 움직임 벡터(Block, Region 단위)



(그림 2) STB 메뉴화면 예시

위 특징을 사용하는 방법 중 가장 일반적인 방법은 색상 히스토그램 기반의 샷 경계 검출 방법이다. 두 영상의 색상 히스토그램을 구하여 두 히스토그램을 비교함으로써 영상 내의 색상 분포 차이를 계산하기 때문에, 단순히 화소 값을 비교하는 것에 비해 카메라 및 물체의 움직임에 강인하다. 하지만 기존의 방법들은 영상 내 형태정보를 포함하지 않기 때문에, 같은 샷에서 영상의 색상 분포가 급격히 변화할 때는 샷 경계로 오검출하는 단점이 있다.

위와 같은 단점들을 극복하기 위해 두 가지 이상의 특징 및 방법을 결합하는 연구가 진행되어왔다. Nagasaka 와 Tanaka[3]는 영상을 16 개의 영역으로 나누어 각 영역에서의 히스토그램을 계산하고, 두 영상으로부터 같은 영역의 히스토그램을 비교해 차이가 가장 큰 8 개의 영역만을 사용함으로써 영상의 움직임과 잡음에 대한 영향을 최소화하고자 하였다. [4]는 RGB 색상 중 채널마다 상위 2 개의 bit 를 사용해 색상으로 총 6 개의 bit 를 특징으로 구성했다. 이를 지역 묘사자인 RILBP(Rotational Invariant Local Binary Pattern) 코드와 결합한 특징으로 히스토그램을 구성함으로써 영상 내의 색상 분포 변화 및 물리 객체의 상태 변화를 극복하고자 하였다. 하지만, 두 방법 모두 같은 샷에서 색상이 급변한 경우에 여전히 샷 경계로 오검출할 가능성이 있고, LBP[5]는 전통적으로 잡음에 민감하고 평평한 지역에서 일관적이지 못한 코드를 생성하는 특징이 있어 안정적이지 못하다.

최근 연구들은 지역적인 정보를 코드화하기 위해 화소 주변의 특정 방향에 대한 에지 크기 값의 변화를 이용하였다. LBP 는 주변 화소와의 밝기 차이를 이용했다면, LDP 는 주변 화소의 에지 크기들을 비교한다. 전통적인 LDP 는 Kirsch 에지 검출기를 이용해 8 개의 방향 에지 값을 얻어 방향 값이 가장 큰 k 개의 bit 를 1 로, 나머지를 0 으로 할당하여 화소마다 8 개의 bit 로 표현되는 LDP 코드를 갖게 된다. [6]에서는 1 로 할당된 bit 중 가장 에지 반응 값이 큰 방향을 most significant bit 로 간주해 코드를 shifting 함으로써

회전 불변의 특성을 갖고자 하였다. 하지만, LDP 도 LBP 와 같이 평평한 지역에서 임의의 코드를 발생시켜 일관된 히스토그램 누적을 방해하여 샷 경계 검출 성능을 저하한다. 또한, STB 자동 품질 검사를 위한 샷 경계 검출은 오검출이 많은 경우 채널 간 동기화에 대한 신뢰성이 낮아지고, 검출 능력이 저하될 경우 동기화에 필요한 시간이 길어져 실시간 처리를 위한 성능 확보에 어려움이 따른다.

본 논문에서는 전체 화면 영역에서 비디오가 재생되는 경우뿐만 아니라, STB 메뉴 화면과 같이 작은 영역에서 비디오가 재생되더라도 오검출률이 낮고, 안정적인 샷 경계 검출이 가능한 개선된 지역 방향 패턴 기반의 비디오 샷 경계 검출 방법을 제안한다. 기존 방법에서 발생했던 평평한 지역과 잡음에 민감한 LDP 를 개선하여 카메라의 움직임이나 물체의 이동, 회전뿐만 아니라 평평한 지역, 그리고 잡음에도 더욱 강한 샷 경계 검출 성능을 기대할 수 있다. 끝으로 RGB 색상의 일정 bit 와 eLDP 코드를 결합해 최종적인 샷 경계를 검출하는 방법을 제안한다.

2. 제안하는 방법

기존의 히스토그램 기반 비디오 샷 경계 검출 방법을 STB 품질검사에 적용했을 때 발생하는 단점을 보완하기 위해, 본 논문에서는 개선된 지역 방향 패턴과 주요 RGB 색상 bit 를 결합한 특징 히스토그램을 기반으로 하는 검출 방법을 제안한다. 제안하는 방법에서는 기존의 LDP 에서 평평한 지역에서 일관적이지 못한 코드를 발생시키는 문제를 default 코드를 적용함으로써 해결하고자 한다. 또한, 기존 논문에서 선별된 색상 bit 중 주요한 bit 를 ITU-R 규격에 따라 새롭게 결정하였다.

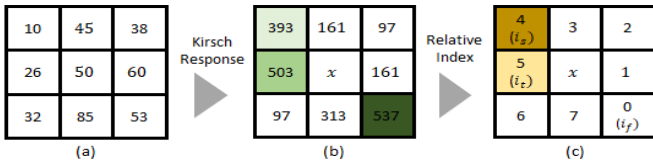
1) 개선된 지역 방향 패턴 (enhanced Local Directional Pattern)

Enhanced Local Directional Pattern(eLDP) 코드는 영상의 형태 정보를 표현함과 동시에 평평한 지역에서는 default 코드를 취함으로써 평평하거나 그렇지 않은 지역에 대한 텍스처 패턴의 구분이 가능하다.

각 화소에 대한 형태 정보를 코드화하기 위해, Kirsch 에지 검출기를 이용하여 에지 반응 값을 구한다[6]. 계산된 Kirsch 에지 반응 값을 기반으로, 기본적인 eLDP 코드 생성에 대한 수식은 아래와 같이 나타낼 수 있다.

$$eLDP(x) = \begin{cases} 2^3 * i_s(x) + i_t(x), & m_f(x) \geq T_m \\ 0, & m_f(x) < T_m \end{cases} \quad (1)$$

위 수식에서 x 는 화소의 위치를, i_s, i_t 는 가장 Kirsch 에지 반응 값이 큰 방향과 두 번째, 세 번째로 반응 값이 큰 방향 사이의 상대적인 인덱스(1~7 중 하나의 정수)를 나타낸다. $m_f(x)$ 는 현재 화소에 대한 Kirsch 에지 반응 값이 가장 큰 방향의 반응 값을, T_m 은 평평한 지역을 구분하기 위한 임계치를 나타낸다. 그림 3은 i_s, i_t 를 구하는 예를 보여준다.



(그림 3) 상대적인 인덱스 i_r, i_t 생성의 예

그림 3의 (a)는 원본 영상을, (b)는 Kirsch 마스크를 통해 구한 (a) 영상의 Kirsch 에지 반응 값을, (c)는 에지 반응 값이 가장 큰 화소를 기준(0)으로 하여 반시계방향으로 인덱싱을 한 결과를 보여준다.

eLDP는 전통적인 LDP와는 다르게 8bit 대신 6bit로 표현할 수 있다. 기존 방법에서는 에지 반응 값이 가장 큰 세 개의 방향에 대해 코드를 1로 할당하여 총 8bit로 표현했다면, 제안하는 방법에서는 에지 반응 값이 두 번째, 세 번째로 큰 방향의 인덱스를 에지 반응 값이 첫 번째로 큰 방향을 기준으로 설정하여 각각 i_s, i_t 에 할당한다. 이는 i_f 를 기준으로 나머지 두 방향에 대한 상대적인 방향의 위치 차이를 코드화함으로써 회전 불변의 특성을 갖게 하며, 이는 수식 (1)과 같이 6bit로 표현할 수 있다. 또한, $m_f(x)$ 의 크기를 임계치 T_m 과 비교함으로써 만약 $m_f(x)$ 이 T_m 보다 작다면 해당 화소 x 의 코드로 0을 할당함으로써 default 코드로 표현한다. 이러한 방법은 평평한 지역에 대한 구분력을 추가함으로써 더욱 안정적인 코드 생성을 가능하게 한다. 임계치 T_m 보다 $m_f(x)$ 가 크면 상대적인 방향의 차이가 0이 나올 수 없으므로, bit 수를 늘리지 않고서도 default 코드를 부여할 수 있다.

2) RGB 색상 및 eLDP 코드를 결합한 특징 생성

방법 [7]에서는 24bit의 RGB 중 각 채널 당 상위 4개의 bit를 사용해 총 12개의 bit만을 사용하였고, [4] 방법에서는 12개의 bit를 반으로 줄인 뒤 LBP와 결합하여 특징을 생성하였다. 하지만 위 방법들은 각 R, G, B 채널의 가중치를 고려하지 않고 균일하게 bit를 선정하였다. 이에 제안하는 방법에서는 ITU-R 규격에 따라 각 채널에서 선정되는 bit 수를 다르게 결정한다. 먼저, ITU-R BT. 601 정의[8]에 따르면 RGB 영상을 휘도 신호로 나타내는 수식은 아래와 같다.

$$Y = 0.299R + 0.587G + 0.114B \quad (2)$$

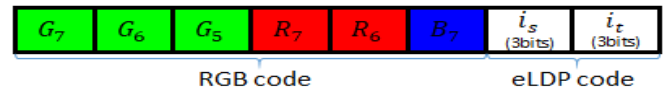
이는 사람의 눈에 더 민감한 색상에 큰 가중치를 주는, 과장의 비시감도(relative visibility)에 따라 계산된 수식이다. 이에, 제안하는 방법에서는 각 채널의 중요도에 따라 색상 bit를 추출하였다. 제안하는 방법에서 고려하는 색상 bit는 그림 4와 같다.

R_7	R_6	R_5	R_4	R_3	R_2	R_1	R_0
G_7	G_6	G_5	G_4	G_3	G_2	G_1	G_0
B_7	B_6	B_5	B_4	B_3	B_2	B_1	B_0

(그림 4) 제안하는 방법에서 고려하는 색상 bit

그림 4와같이, G 채널에서는 3개의 bit, R 채널에서는 2개의 bit, B 채널에서는 1bit를 사용한다. 이렇게 고려된 색상 bit와 eLDP 코드를 결합하면 그림 5와 같

은 최종 12bit의 특징이 생성된다.



(그림 5) 색상 정보와 eLDP 코드를 결합한 최종 특징

위와 같이 최종적으로 추출된 특징은 $2^{12} = 4096$ 개의 종류로 구성되며, 이는 즉 영상을 4,096개 bin으로 구성된 히스토그램으로 재구성할 수 있음을 의미한다.

3) 히스토그램 비교를 통한 샷 경계 검출

제안하는 방법에서는 색상 정보와 eLDP 코드가 결합된 특징(이하 CeLDP라 칭함)의 빈도를 누적함으로써 히스토그램을 구성한다. 즉, 모든 화소에 대해 CeLDP를 계산하여 구해진 k 번째 영상의 히스토그램 $H_k(\tau)$ 는 아래 수식과 같이 표현된다.

$$H_k(\tau) = \sum_{x=1}^x f(CeLDP(x), \tau) \quad (3)$$

$$f(a, \tau) = \begin{cases} 1, & a = \tau \\ 0, & otherwise \end{cases} \quad (4)$$

수식 (3)과 (4)에서 τ 는 CeLDP 코드이다.

샷 경계는 위와 같이 구해진 히스토그램을 이용하여 인접한 영상 사이의 히스토그램 차이 값을 특정 임계치와 비교하여 임계치보다 높으면 검출하게 된다. 현재 영상과 이전 영상 사이의 히스토그램 차이를 계산하는 수식은 아래와 같다.

$$D(i) = \sum_{b=0}^{B-1} |H_k(b) - H_{k-1}(b)| \quad (5)$$

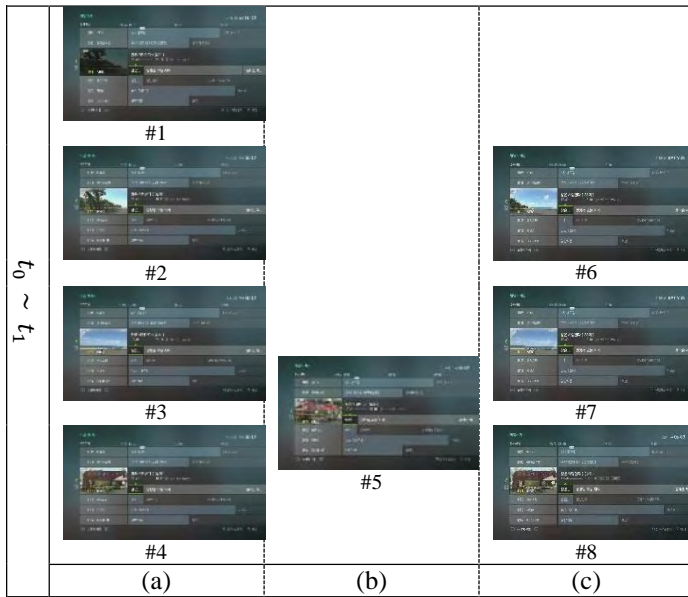
여기서 B 는 히스토그램 bin의 개수를 의미하며 제안하는 논문에서는 4,096이 된다.

$D(i)$ 는 하나의 1차원 신호로 표현할 수 있다. 이 신호는 smoothing을 통해 히스토그램 차이 계산으로 인해 발생한 작은 변화를 제거할 수 있다. Smoothing된 신호를 $D_{smooth}(i)$ 라 하면, 해당 신호를 통해 같은 샷 내에서는 작은 값을, 샷의 경계에서는 값이 커지게 되어 잡음과 잘 구분되어 나타난다. 따라서, 샷 경계를 구분하는 임계치는 $D_{smooth}(i)$ 신호의 최댓값의 일정 퍼센티지를 취함으로써 설정한다(예: 50%). 이러한 방식을 통해 비디오 샷은 영상 시퀀스에서 CeLDP 히스토그램 변화가 갑작스럽게 일어났을 때만 검출하는 것이 가능하다.

3. 실험 결과

제안하는 샷 경계 검출 방법의 성능평가를 위해 RGB 히스토그램 기반[7], RGB+RILBP 히스토그램 기반[4], 그리고 제안하는 방법을 이용한 비디오 샷 경계 검출을 수행하였다. 평가하기 위한 검출 환경으로는 Intel i7 CPU, 16GByte RAM, Quadro K2200d GPGPU가 내장된 KT Skylife STB를 활용하였으며, 24시간동안 비디오 샷 경계 검출을 수행하여 그 결과를 분석하였다. 그림 6은 전체 비디오 중 세로축의 시간

$[t_0 \sim t_1]$ 의 흐름에 따라 각 방법 (a)~(c)에서 검출한 영상들(#1~#7)을 나타낸다.



(그림 6) 샷 경계 검출 결과 (a) RGB histogram[7], (b) RGB+RILBP histogram[4], (c) 제안하는 방법

먼저 RGB[7]는 네 개의 영상(#1~#4)을 검출하였는데, 그림 6의 영상 #1과 #2를 비교해 보면 메뉴 화면에서 Fade In 효과가 일어난 것을 볼 수 있다. 어두운 화면에서 점점 밝아지면서 영상의 색상 분포에 큰 변화가 일어나, 이를 감지하여 샷 경계로 오검출하는 결과를 보였다. 다음으로 RGB+RILBP[4]는 한 개의 영상(#5)만을 검출하였다. 이는 영상 내에 샷 경계가 존재하지만 이를 검출하지 못했음을 의미한다. 마지막으로 제안하는 방법은 세 개의 영상(#6~#8)을 검출하였는데, 색상의 변화에도 오검출하지 않으면서 요구하는 샷 경계를 적절히 검출했음을 나타낸다.

각 방법을 통해 검출한 샷 경계 검출 성능의 정량적인 평가를 위해 제안하는 논문에서는 False Discovery Rate(FDR)을 이용하였다. FDR은 $FDR = \frac{FP}{TP+FP} * 100$ 과 같이 계산되며, TP(True Positive)는 검출한 샷 경계 중 올바르게 검출한 샷 경계의 수를, FP(False Positive)는 오검출한 샷 경계의 수를 의미한다. 표 1은 각 샷 경계 검출 방법에 대한 TP, FP, FDR을 나타낸다.

<표 1> 각 샷 경계 검출 방법에 대한 TP, FP, FDR

	RGB[7]	RGB+RILBP[4]	Proposed
TP	2894	1842	2589
FP	623	249	287
FDR	17.71%	11.91%	9.98%

FDR은 그 값이 낮을수록 샷 경계 검출 성능이 뛰어나다. RGB[7]는 검출한 영상이 가장 많으면서도 FDR이 가장 높아 다른 방법들에 비해 검출 성능이 낮고, RGB+RILBP[4]는 다른 방법과의 TP 수를 비교했을 때 올바른 샷 경계를 검출하는 빈도가 낮은 것을 알 수 있다. 제안하는 방법은 FDR의 값을 통해 검출한 샷 중 잘못된 샷 경계를 검출하는 비율이 10%로, 오

검출률이 가장 낮은 것을 알 수 있다. 즉, 제안하는 방법은 기존 방법들보다 샷 경계 검출 성능이 뛰어나고 입증되었다.

4. 결론

본 논문에서는 셋톱박스 품질검사를 위한 개선된 지역 방향 패턴(eLDP) 기반의 비디오 샷 경계 검출 방법을 제안하였다. 제안하는 방법은 카메라의 움직임이나 물체의 이동, 회전뿐만 아니라 평평한 지역 및 잡음에도 더욱 강인한 검출 성능을 보였다. RGB 색상의 일정 bit와 eLDP 코드를 결합함으로써 특징을 더욱 강인하게 하였고, 이를 이용해 전체 화면 영역에서 비디오가 재생되는 경우뿐만 아니라 STB 메뉴 화면과 같이 작은 영역에서 비디오가 재생되더라도 오검출률이 낮은 안정적인 샷 경계 검출이 가능했다. 또한, 제안된 샷 경계 검출 방법을 STB 자동 품질 검사 시스템에 적용하여 채널 간 동기화에 대한 신뢰성을 높였고, 실시간 처리가 가능한 안정적인 샷 경계 검출률을 확보하였다.

※ 본 연구는 중소기업청의 기술혁신개발사업의 일환으로 수행하였음. [S2309071, IPTV/DCATV/위성방송의 품질검증을 위한 자동 콘텐츠 분석 및 에러 탐지기술 기반의 실시간 품질 모니터링 시스템 개발]

참고문헌

- [1] Stb-tester.com. (2017). *stb-tester : Automated Testing for Set-Top Boxes and Smart TVs*. [online] Available at: <http://Stb-tester.com> [Accessed 18 Sep. 2017].
- [2] Boreczky, J. S., & Rowe, L. A. (1996). Comparison of video shot boundary detection techniques. *Journal of Electronic Imaging*, 5(2), 122-128.
- [3] Nagasaka, A., & Tanaka, Y. (1992). Automatic video indexing and full-video search for object appearances. *Journal of Information Processing*, 15(2), 316.
- [4] 안기욱, et al. "원격 스마트 협업을 위한 컬러와 LBP 코드 기반 비디오 샷 경계 검출." *대한전자공학회 학술대회* (2015): 593-596.
- [5] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12), 2037-2041.
- [6] Jabid, T., Kabir, M. H., & Chae, O. (2010, August). Local directional pattern (LDP)—A robust image descriptor for object recognition. *In Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on* (pp. 482-487). IEEE.
- [7] Mas, J., & Fernandez, G. (2003). Video shot boundary detection based on color histogram. *Notebook Papers TRECVID2003, Gaithersburg, Maryland, NIST*.
- [8] Series, B. T. (2011). Studio encoding parameters of digital television for standard 4: 3 and wide-screen 16: 9 aspect ratios.
- [9] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.