

다중 슬라이딩 윈도우들을 이용한 효과적인 샷 경계 검출 방법

*민현석 **진성호 ***노용만

한국정보통신대학교

*min6284@icu.ac.kr

Effective Shot Boundary Detection Using Multiple Sliding Windows

*Hyun-Seok Min **Sung Ho Jin ***Yong Man Ro

Image and Video Systems Lab., Information and Communications University

요약

비디오를 내용 별로 분할하기 위한 최소 단위는 비디오의 샷이다. 따라서 비디오 내용 정보를 분석함에 있어서 비디오의 샷 경계 검출은 필수적인 과정이다. 이러한 샷 전환 과정은 급격한 샷 전환 과정과 점진적인 샷 전환 과정으로 구분할 수 있다. 점진적인 샷 전환 과정의 경우 전환 과정이 여러 프레임들에 걸쳐 발생하는 관계로 점진적인 샷 전환 과정을 검출하기 위하여, 기존 샷 경계 검출 알고리즘은 일정 간격을 슬라이딩한 윈도우 프레임들 간의 차이를 비교하는 방식을 이용하였다. 기존 슬라이딩 윈도우 방법은 점진적인 샷 전환 과정을 검출하기 위하여 고정된 크기의 윈도우 하나만을 이용하였다. 이 경우, 슬라이딩 윈도우의 길이가 점진적인 샷 전환 과정에 비해 짧으면, 샷 전환을 검출하지 못 한다. 슬라이딩 윈도우의 길이가 샷의 길이보다 길면 샷을 점진적인 샷 전환으로 검출하는 오류가 발생된다.

상기 문제점을 개선하기 위하여 본 논문에서는 서로 다른 크기의 다수의 슬라이딩 윈도우들과 적응적 경계치를 적용한 샷 경계 검출 방법을 제안한다.

1. 서론

현재 다양한 멀티미디어 매체의 증가, 디지털 기술의 발전으로 방대한 양의 멀티미디어 데이터가 생성되고 있으며, 이러한 데이터를 효율적으로 이용하고 관리하고자 하는 사용자의 욕구가 증가하고 있다. 대용량의 멀티미디어 데이터의 효율적인 관리를 위하여 비디오에 대한 요약 및 분석이 필요하다. 샷은 동일한 비디오 정보를 담고 있으며 비디오의 내용 기반 요약을 수행하는데 있어서 최소 단위이다. 따라서 비디오 데이터를 분석하기 위해서는 샷 경계 검출은 필수적인 과정이다.

비디오의 샷 경계 검출이란 비디오에서 연속적으로 주사되는 각 프레임들에 담겨있는 정보를 이용하여 샷이 전환되는 지점을 찾아내는 것이다. 그러나 급진적인 샷 전환 과정에 경우, 그 경계가 뚜렷하여 바로 이전 프레임과의 차이를 이용하여 검출할 수 있지만 샷 전환 과정이 여러 프레임들에 걸쳐 나타나는 점진적인 샷 전환 과정은 이런 방식으로는 좋은 검출 성능을 보이지 못 하였다. 이 문제점을 해결하기 위하여 기존 샷 경계 검출 알고리즘은 일정 간격을 띄어서 프레임의 정보를 비교하는 슬라이딩 윈도우 방식을 사용하였다. 기존 슬라이딩 윈도우 방식의 경우, 고정된 윈도우 크기와 경계치를 사용하였다. 그러나 이 경우, 슬라이딩 윈도우의 길이가 점진적인 샷 전환 과정에 비해 짧으면, 샷 전환을 검출하지 못 한다. 슬라이딩 윈도우의 길이가 샷의 길이보다 길면 샷을 점진적인 샷 전환으로 검출하는 오류가 발생된다.

상기 문제점들을 개선하기 위하여 본 논문에서는 서로 다른 크기를 가지는 다수의 슬라이딩 윈도우들과 적응적 경계치를 적용한 샷 경계 검출 방법을 제안한다. 제안하는 알고리즘은 크게 두개의 과정들로 구분된다. 다수의 슬라이딩 윈도우들을 적용하여 적응적 경계치를 구하고 샷 경계의 후보 영역을 검출하는 과정과 샷 경계 후보 영역을 검

증하는 과정이다. 또한 효율적인 샷 전환 과정 검출을 위해 본 논문에서는 입력 영상을 모두 복호화 하지 않고, DC 값만을 이용한 DC 영상만을 이용한다. 또한 슬라이딩 윈도우의 스택 크기를 한 프레임이 아닌 I 프레임 간격으로 설정하였다.

제안하는 방법의 유효성을 검증하기 위하여 스포츠 및 뉴스 동영상으로 구성된 데이터베이스를 이용하여 실험한다. 실험을 통해 기존의 방법들에 비해 제안하는 방법이 빠른 샷 경계 검출 속도 및 효과적인 검출 성능을 보이는 것을 확인한다.

본 논문은 우선 2장에서 제안하는 샷 경계 검출 알고리즘에 대한 소개를 하고, 3장에서 실험을 수행하고 그 결과를 고찰하였다. 그리고 4장에서 결론을 내리고 향후 연구 과제를 결정하였다.

2. 본론

가. 제안하는 샷 경계 검출 시스템

그림 1은 본 논문에서 제안하는 샷 경계 검출 시스템은 나타내고 있다.

제안하는 샷 경계 검출 시스템은 비디오 영상의 DC 영상에 대하여 HSV 컬러 히스토그램을 구하여 슬라이딩 윈도우 방식을 적용하여 Similarity distance를 구하여 샷 경계 후보군과 적응적 경계치를 설정한다. 그리고 구해진 후보군을 급진적인 샷 전환 후보군과 점진적인 샷 전환 과정 후보군으로 나눈 후, 최종적으로 후보군의 타당성을 검증하여 샷 전환 과정을 포함하고 있는 후보 영역을 찾는 과정으로 이루어진다.

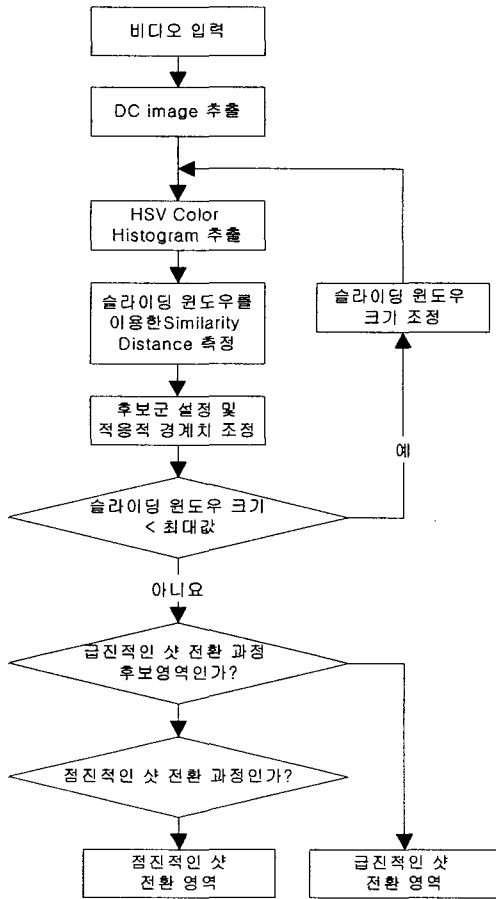


그림 1 샷 경계 검출 시스템

나. HSV 컬러 히스토그램 추출

이미지의 컬러 특징 정보를 샷 전환 검출에 사용하여 할 때, 가장 먼저 고려해야 할 점은 적합한 컬러 모델을 선택하는 것이다. 컬러 모델에는 RGB, CMYK, CIE, HSV등이 있고, 현재 검색에 많이 사용되는 모델은, RGB 컬러 모델과 HSV 컬러 모델이다.

HSV 컬러 모델은 색상(Hue), 채도(Saturation), 명암(Value)로 컬러를 나타낸다. 색상(Hue)은 1차색의 혼합으로 생긴 색 자체이며, 0°에서 360°에 걸쳐 표현된다. 채도(Saturation)는 색상의 엷음 정도를 나타내며, 0부터 1까지에 걸쳐 표현된다. 마지막으로 명암(Value)는 밝기의 강도로 0부터 1까지의 값을 가진다. HSV 컬러 모델은 색채, 채도, 명암을 따로 분리해 냄으로써 인간의 시각적 특성을 가장 잘 반영하였으며, 샷 경계 검출에서 이러한 인간의 시각적 특성을 통해 검출의 효율성을 높일 수 있다.

본 논문에서는 각 프레임의 정보 비교의 편의를 위하여 HSV 컬러 정보 히스토그램으로 만들어 사용하였다. 아래 식(1)에 따라 HSV 컬러 모델의 색상(Hue), 채도(Saturation), 명암(Value) 정보를 각각 16개, 4개, 4개의 빈으로 나누었다. 이렇게 나누어진 256개의 빈으로 구성된 히스토그램을 이용하여 각 프레임의 정보를 비교하는데 사용하였다.

$$\begin{aligned} \text{Bin}H &= \left\lfloor \frac{H}{22.5} \right\rfloor \\ \text{Bin}S &= \left\lfloor \frac{S}{0.25} \right\rfloor \\ \text{Bin}V &= \left\lfloor \frac{V}{0.25} \right\rfloor \end{aligned} \quad (1)$$

다. Similarity Distance 측정

샷 전환 과정을 검출하기 위해서는 각 프레임에서 추출한 정보를 비교하여야 한다. 동일 샷 안에 존재하는 프레임은 서로 비슷한 특성을 가지고 있고, 서로 다른 샷에 존재하는 프레임은 서로 다른 특성을 가지고 있다. 그러므로 우리는 각 프레임에서 추출한 컬러 히스토그램 정보를 비교하여 두 프레임이 같은 샷 안에 있는지를 판단할 수 있다. 이렇게 두 프레임간의 정보를 비교하기 위하여, 본 논문에서는 카이제곱 분석을 사용하였다.

카이제곱 검정은 두 범주형 변수가 서로 관계가 있는지 독립인지를 판단하는 통계적 검증 방법으로 두 프레임의 색상 정보가 서로 관계가 있는지 판단할 수 있는 척도가 될 수 있다. 서로 다른 샷에 포함되어 있는 두 프레임들의 정보는 서로 다를 것이기 때문에, 카이제곱 검정 값이 크게 나올 것이고, 아니라면 작게 나올 것이다. 다음 식(2)는 본 논문에서 카이제곱 검정을 위하여 사용한 식이다.

$$\begin{aligned} D_i^k &= d(DC_{F_{(i-\frac{k}{2})}}, DC_{F_{(i+\frac{k}{2})}}) \\ &= \frac{1}{256} \times \sum_{V=0S=0H=0}^3 \sum_{0S=0H=0}^3 \sum_{0H=0}^{15} \\ &\left(\frac{(DC_{F_{(i-\frac{k}{2})}}[V][S][H] - DC_{F_{(i+\frac{k}{2})}}[V][S][H])^2}{\max(DC_{F_{(i-\frac{k}{2})}}[V][S][H], DC_{F_{(i+\frac{k}{2})}}[V][S][H])} \right) \end{aligned} \quad (2)$$

DC_{F_w} : w번째 프레임의 DC 영상의 HSV 히스토그램 정보

라. 후보군 설정 및 경계치 조정

기존 많은 샷 경계 검출 알고리즘에서 고정 경계치를 사용하여 샷 전환 과정을 검출하였다. 그러나 이 경우, 고정 경계치가 너무 낮은 경우 샷 전환 과정이 아님에도 샷 전환 과정으로 검출될 수 있고, 반대로 고정 경계치가 너무 높은 경우 샷 전환 과정임에도 샷 전환 과정으로 검출되지 못할 수 있다.

이에 본 논문에서는 적응적 경계치와 여러 슬라이딩 윈도우를 사용한다. 슬라이딩 윈도우는 크기가 작은 윈도우부터 사용한다. 이는 점진적인 샷 전환 과정이 작은 슬라이딩 윈도우보다 더 긴 경우, 작은 슬라이딩 윈도우를 이용하여 구한 Similarity distance가 큰 슬라이딩 윈도우를 이용하여 구한 Similarity distance보다 작기 때문에 점진적인 샷 전환 과정에 경우 작은 슬라이딩에서 적용시킨 경계치는 큰 슬라이딩 윈도우에서 구한 Similarity distance보다 작아 점진적인 샷 전환 과정을 놓치지 않기 때문이다.

또한 경계치는 슬라이딩 윈도우에 따라 적응적으로 변하기 때문에 객체의 움직임 등에 의해 변화가 일어난 경우, 후보 영역은 여러 슬라이딩 윈도우를 거치면서 줄어들게 된다.

다음은 샷 전환 후보 영역의 설정 조건과 각 경우의 적응적 경계치 설정 방법을 나타낸 것이다. 초기 영상 정보를 가지고 있지 않기에 본 논문에서는 L이라는 최소 경계치를 사용하였고, α 는 경계치의 적응적 변화 속도를 결정하는 변수이다. 이전에 구해진 경계치가 다음 슬라이딩 윈도우를 이용하여 구한 값에 얼마나 빠르게 적응 할지를 결정한다.

표 1. 후보 영역 설정 조건 및 적응적 경계치 설정

1. 현재 프레임이 샷 전환 후보 영역이 아닐 경우 if $D_i^k > Threshold$, - 샷 전환 후보 영역 시작된다.

$Threshold_i = Threshold_{i-1} \times \alpha + (1 - \alpha) \times D_i^k$
 else if $D_i^k > D_{i-k}^k \times L$ - 샷 전환 후보 영역이 시작된다.
 $Threshold_i = D_{i-k}^k \times L \times \alpha + (1 - \alpha) \times D_i^k$
 else - 샷 전환 후보 영역에 포함되지 않는다.
 $Threshold_i = Threshold_i \times \alpha + (1 - \alpha) \times D_i^k \times L$
 2. 현재 프레임이 샷 전환 후보 영역일 경우
 if $D_i^k < Threshold_i$ - 샷 전환 후보 영역 종료된다.
 $Threshold_i = D_i^k \times L$
 else if $D_i^k \times L < D_{i-k}^k$ - 샷 전환 후보 영역이 종료된다.
 $Threshold_i = D_i^k \times L$
 else - 샷 전환 후보 영역에 포함된다.
 $Threshold_i = Threshold_i \times \alpha + (1 - \alpha) \times D_i^k$
 k : 슬라이딩 윈도우 크기
 i : 현재 프레임 번호
 L : 샷 전환 과정 검출을 위한 최소 경계치
 α : 경계치의 적응적 변화 속도를 결정하는 변수

마. 후보 영역 구별

후보군은 크게 급진적인 샷 전환 과정이 포함된 후보 영역과 점진적인 샷 전환 과정이 포함된 후보 영역으로 나눌 수 있다. 급진적인 샷 전환 과정의 경우 샷 전환 과정이 연속된 두 프레임 사이에서 일어나므로 후보 영역 안에서 연속된 두 프레임의 Similarity Distance가 후보 영역의 경계치 보다 크다면 그 후보 영역은 급진적인 샷 전환 과정을 포함하고 있다고 볼 수 있다. 또한 점진적인 샷 전환 과정의 경우에는 여러 프레임에 걸쳐 일어나고 적응적 경계치는 큰 슬라이딩 윈도우에 적용되어 있어 두 프레임의 차이보다는 크기 때문에 급진적인 샷 전환 과정을 포함한 후보 영역으로 검출되지 않는다.

또한 이 과정에서 급진적인 샷 전환 후보 영역의 경우에는 경계치 보다 큰 차이를 보인 두 연속된 프레임들이 샷 전환이 일어난 정확한 지점이라고 할 수 있다.

표2는 여러 후보 영역 중 급진적인 샷 전환 과정을 구별하는 조건이다.

표 2, 급진적인 샷 전환 과정 구별 조건

$Distance_i > Threshold''$ (any $i \in B_s'' \dots B_e'' - 1$)
 B_s'' : n 번째 후보 영역의 시작 프레임 위치
 B_e'' : n 번째 후보 영역의 끝 프레임 위치
 $Distance_i$: i번째 프레임과 i+1번째 프레임과의 차이
 $Threshold''$: n 번째 후보 영역의 경계치

바. 후보 영역 검증 과정

적응적 경계치를 사용하여 슬라이딩 윈도우를 적용하였지만 모든 후보 영역이 샷 전환 과정을 포함하고 있다고 할 수 없다. 카메라의 움직임, 객체의 움직임, 조명의 변화 등으로 인한 변화도 가지고 있기 때문에 후보 영역에 대한 검증 과정이 필요하다.

본 논문에서는 다음과 같은 검증 조건을 추가 하였다.

표 3. 후보 영역 검증 조건

Condition 1. $PD'' < Threshold''$
 Condition 2. $PD'' < L$
 Condition 3. $PD_{HSV}'' > PD_{HS}'' \times L$
 $PD'' = PD_{HS}'' = D_{Mean}^{Length}$
 $Threshold'' = Threshold^{Mean}$
 $PD_{HS}'' = D_{Mean}^{Length}$ when (all $V=0$)

PD'' 는 n번째 후보 영역의 양 끝 프레임을 비교하여 얻은 Similarity Distance이다. 점진적인 샷 전환 과정에 경우 변화가 시작하는 프레임부터 변화가 끝나는 프레임까지 전환 과정으로 정의한다. 그러므로 두 구간 사이의 차이는 알고리즘 과정에서 구한 적응적 경계치 이상의 차이를 보여야 하며, 그 차이는 알고리즘에서 정한 최소 차이 이상이어야 한다. (검증 조건 1,2)

샷 전환은 조명의 변화만으로는 일어나지 않는다. 그러므로 조명의 변화가 다른 변화와 비슷하여야 한다. 그래서 본 논문에서는 HSV 중 명도(Value) 값 중 조명에 민감한 명도(Value) 값을 0으로 가정한 Similarity Distance인 PD_{HS}'' 와 기존의 Similarity Distance인 PD_{HSV}'' 를 비교하여 조명의 변화에 의해 후보 영역으로 선택된 후보 영역을 검증하였다. (검증 조건 3) PD_{HSV}'' 는 지금까지와 같이 구한 Similarity distance이다. PD_{HS}'' 는 비교할 두 프레임의 V값을 모두 0으로 놓고 비교한 Similarity Distance이다.

3. 실험 및 결과

실험에 사용한 다큐멘터리 동영상은 61개의 급진적인 샷 전환 과정과 디졸브와 페이드 인/아웃으로 구성된 20개의 점진적인 샷 전환 과정을 가지고 있으며, 총 10000 프레임으로 구성되어 있다. 또한 스포츠 동영상은 다큐멘터리 동영상과 같이 10000 프레임으로 구성되어 있지만, 29개의 급진적인 샷 전환 과정만을 포함하고 있다.

본 실험에서는 4개의 슬라이딩 윈도우를 사용했으며, 슬라이딩 윈도우의 스텝 크기는 영상의 I 프레임 간격인 15 프레임으로 설정하였다. 최소 슬라이딩 윈도우 크기를 I 프레임 간격인 15 프레임으로 설정하였고, 나머지 3개의 슬라이딩 윈도우의 크기는 30,45,60 프레임으로 각각 설정하였다. 또한 적응적 경계치 변화 속도를 가름하는 α 로 0.8을 사용하였고, 본 논문에서 사용한 최소 경계치 L의 값으로 1.5로 사용하였다.

$$Recall = \frac{N_{Correct}}{N_{Correct} + N_{Miss}} \quad (4)$$

$$Precision = \frac{N_{Correct}}{N_{Correct} + N_{False}} \quad (5)$$

$N_{Correct}$ 은 샷 전환 과정을 올바르게 찾은 개수를 나타내고, N_{miss} 는 샷 전환 과정인데 검출하지 못한 개수를 나타낸다. N_{False} 는 샷 전환 과정임에도 샷으로 검출하지 못한 개수를 나타낸다.

식 (3)의 Recall은 샷 경계 검출에서 샷 전환 과정을 놓치지 않고 찾아낸 확률을 나타내고, 식 (4)의 Precision은 샷 전환 과정으로 찾은 샷 전환 과정이 실제로 샷 전환 과정인 확률을 나타낸다. 제안된 알고리즘의 성능을 알아보기 위하여 위의 두 식으로 제안된 알고리즘을 평가하였다.

표 4,5 같은 영상을 가지고 기존 알고리즘[5]을 가지고 실험한 결과를 나타낸 표이다. 기존 알고리즘은 하나의 슬라이딩 윈도우를 이용하였고, 슬라이딩 윈도우 양 끝 프레임 정보 차이가 일정 값 이상의 차

이를 보일 때, 샷 전환이라고 정의하였다. 또한 기존 하나의 슬라이딩 윈도우를 쓰는 경우 점진적인 샷 전환과 급진적인 샷 전환 과정을 동일하게 검출하여 표4,5에서는 급진적인 샷 전환 과정과 점진적인 샷 전환 과정을 결과의 총계만을 나타내었다.

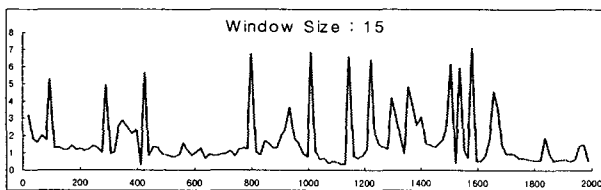
표 4. 제안하는 알고리즘을 이용한 샷 검출 결과(다큐멘터리)

	샷 검출				정확도	
	Ori.	Cor.	Miss	False	Recall	Precision
총계	81	75	6	20	0.9259	0.7895

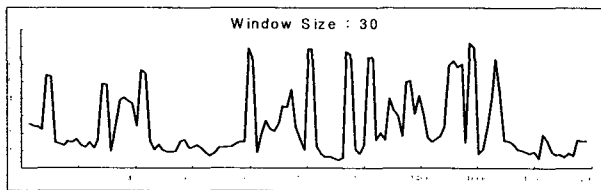
표 5. 제안하는 알고리즘을 이용한 샷 검출 결과(스포츠)

	샷 검출				정확도	
	Ori.	Cor.	Miss	False	Recall	Precision
총계	29	28	1	17	0.9655	0.6222

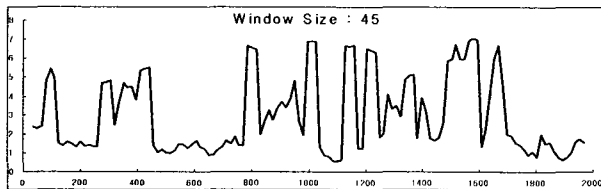
다음은 각 슬라이딩 윈도우를 이용한 Similarity Distance 결과들과 Similarity distance를 이용하여 검출한 샷 전환 과정 영역을 나타낸 표이다.



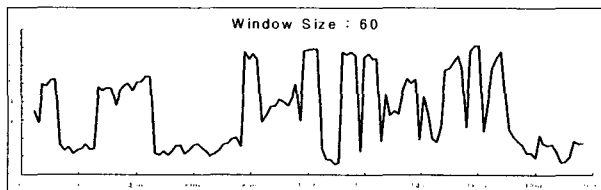
(a) Similarity Distance (윈도우 크기:15)



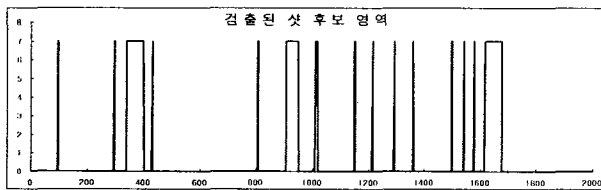
(b) Similarity Distance (윈도우 크기:30)



(c) Similarity Distance (윈도우 크기:45)



(d) Similarity Distance (윈도우 크기:60)



(e) 검출된 샷 전환 영역

그림 2. 샷 경계 검출 결과

다음 표6, 표7는 본 논문에서 제안하는 시스템을 이용한 실험 결과를 나타낸 것이다.

표 6. 제안하는 알고리즘을 이용한 샷 검출 결과(다큐멘터리)

	샷 검출				정확도	
	Ori.	Cor.	Miss	False	Recall	Precision
급진적인 샷 전환	61	57	4	2	0.9385	0.9683
점진적인 샷 전환	20	18	2	1	0.9091	0.9524
총계	81	75	6	3	0.9259	0.9615

표 7. 제안하는 알고리즘을 이용한 샷 검출 결과(스포츠)

	샷 검출				정확도	
	Ori.	Cor.	Miss	False	Recall	Precision
급진적인 샷 전환	29	28	1	3	0.9655	0.9032

위에 나타난 결과로 기존 슬라이딩 윈도우를 이용한 샷 검출 알고리즘에 비해 본 논문에서 제안하는 알고리즘이 정확도에서 높은 성능을 보이는 것을 확인할 수 있다. 기존 알고리즘에 비해 특히 카메라와 객체의 움직임이 많은 스포츠 영상에서 좋은 성능을 보였다. 이 실험을 통해 제안하는 샷 경계 검출 시스템이 기존 슬라이딩 윈도우를 이용하는 샷 검출 알고리즘에 비해 카메라와 객체의 움직임에 대해 민감하지 않고 강한 성능을 보이는 것으로 알 수 있다.

4. 결론 및 향후 연구 과제

본 논문에서는 다중 슬라이딩 윈도우와 적응적 경계치를 이용하여 샷 경계를 검출하는 시스템을 제안하였다. 급진적인 샷 전환과정과 샷 전환 과정이 여러 프레임에 걸쳐 일어나는 점진적인 샷 전환 과정 모두를 효과적으로 찾기 위하여 다중 슬라이딩 윈도우와 적응적 경계치를 사용하였다.

제안하는 알고리즘은 각 프레임의 정보를 분석하여 효과적으로 샷 경계를 검출하였다. 이러한 정확한 샷 경계 검출은 비디오 데이터의 요약 및 관리에 효율적으로 이용될 수 있다. 향후 본 논문에서 제안한 알고리즘을 통해 검출한 샷 정보를 이용하여 사용자에게 동영상 정보를 계층적으로 분석하여 제공하는 연구를 진행하겠다.

Acknowledgement

본 논문은 에서는 문화콘텐츠진흥원의 '방송 콘텐츠 분석을 통한 부가 콘텐츠 제작 시스템 개발'에 관한 연구 과제 수행의 일환으로 얻어진 연구 결과입니다.

참고 문헌

- [1] Alan Hanjalic, "Shot-Boundary Detection: Unraveled and Resolved?" IEEE Transactions on circuits and systems for video technology, VOL. 12, NO. 2, FEBRUARY 2002
- [2] Robert A. Joyce, Bede Liu, "Temporal Segmentation of Video Using Frame and Histogram Space", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 8, NO. 1, FEBRUARY 2006
- [3] Sang Heun Shim, Seung Ji Yang, Jeong Hyun Yoon, Ki Hyun Kim and Yong Man Ro, "Real-time Shot boundary detection for digital video camera using the MPEG-7 descriptor," Real-Time Imaging VI, Proceeding on SPIE Electronic Imaging, Vol. 4666, pp.161-171, 2002-01