

論文2001-38SP-11-13

클러스터링을 이용한 급격한 장면 전환 검출 기법

(Abrupt Shot Change Detection using an Unsupervised Clustering of Multiple Features)

李 熏 哲 *, 高 綸 浩 *, 尹 炳 珠 *, 金 聖 大 **, 俞 相 朝 ***

(Hun Cheol Lee, Yun Ho Ko, Byoung Ju Yun, Seong Dae Kim, and Sang Jo Yoo)

요 약

본 논문에서는 클러스터링을 이용해서 급격한 장면 전환을 찾는 방법을 제안한다. 일반적으로 장면 전환 검출 기법에서 많이 사용되는 특징들은 특별한 상황에서만 잘 적용된다는 단점이 있기 때문에 여러 종류의 특징을 동시에 고려하는 클러스터링 기반의 기법이 많이 사용되고 있다. 하지만 이 경우에는 클러스터의 초기 중심을 정하는 것이 중요한 문제가 된다. 본 논문에서는 k -평균 클러스터링에서의 초기 중심을 적응적으로 바꾸면서 장면 전환 존재 여부를 결정하도록 하였다. 실험 결과 초기 클러스터 중심이 고정된 경우에 비해서 더 좋은 결과를 얻었다.

Abstract

In this paper, we propose an efficient method to detect abrupt shot changes in a video sequence using an unsupervised clustering. Conventional clustering-based shot change detection algorithms use multiple features in order to overcome the shortcomings of a single feature. In such methods it is very important to determine the appropriate initial cluster centers well. In this paper we propose a modified k -means clustering algorithm which estimates the initial cluster center adaptively. Experimental results show that the proposed algorithm works well.

1. 서 론

최근 들어 멀티미디어 데이터의 효과적인 표현 및 검색에 관한 연구가 많이 진행되고 있다. 이들 중에서 비디오와 관련된 연구로는 비디오 인덱싱 및 탐색 등이 있다^[1]. 비디오 데이터는 본질적으로 매우 복잡한

계층 구조를 가지는데 이를 효과적으로 표현하기 위해서는 여러 단계의 계층으로 나눌 필요가 있다. 비디오 데이터의 가장 기본이 되는 단위는 화면(Frame)이며 그 다음 상위 단위는 한 대의 카메라의 연속적인 동작으로 만들어지는 화면들로 구성된 샷(Shot)이다. 화면과 샷은 형태론적 단위(Syntactic units)이다. 그리고 같은 의미를 가지는 샷들로 이루어진 상위 계층의 구조인 씬(Scene)은 의미론적 단위(Semantic units)이다. 연속되는 두 개의 샷 사이의 경계를 찾는 비디오 데이터의 장면 전환 검출은 비디오 인덱싱 및 탐색의 첫 번째 단계로서 매우 중요한 역할을 하며 상위 단지의 비디오 표현을 가능하게 한다.

최근 몇 년 동안 비디오 데이터의 급격한 장면 전환 검출에 관한 연구가 많이 수행되어 왔다. 대부분의 연

* 學生會員, 韓國科學技術院 電氣電子工學科
(Detp. of Electrical Engineering, KAIST)

** 正會員, 韓國科學技術院 電氣 電子工學科
(Detp. of Electrical Engineering, KAIST)

*** 正會員, 仁何大學校 情報通信 專門大學院
(The Graduate School of Information Technology & Telecommunications, Inha University)

接受日字:2001年5月5日, 수정완료일:2001年10月3日

구는 사용하는 특징에 따라서 크게 비압축 영역의 특징을 이용하는 방법과 압축 영역의 특징을 이용하는 방법으로 나눌 수 있다. 그리고 검출 방법에 따라서 크게 문턱 값 기반의 방법과 클러스터링 기반의 방법으로 나눌 수 있다.

우선 가장 많이 사용되는 비 압축 영역의 특징은 픽셀 간 차이 또는 히스토그램에 근거한 화면 간 차이 등이다. 이 경우 화면 간 차이는 전체 영상에 대해서 계산할 수도 있고 부분 영역들에 대해서 계산할 수도 있다. 비압축 영역의 특징을 사용한 기법들로는 three-frames approach^[2], pixel-by-pixel comparison, likelihood ratio, histogram comparison^[5] 등이 있다. 또 압축 영역에서는 주로 MPEG 또는 Motion JPEG 등에 의해 압축된 비디오 데이터로부터 부분적으로 복원시킨 데이터만을 가지고 장면 전환을 찾는다. 여기에는 주로 DCT 계수를 이용한 방법^[8]과 움직임 벡터를 이용한 방법^[7] 등이 대표적으로 사용되었다. 그리고 이런 압축/비압축 영역의 특징들을 이용하는 여러 종류의 거리 척도(distance metric)들이 제안되었다^[6]. 이러한 각각의 특징들은 특별한 상황에만 잘 적용된다는 단점이 있다. 예를 들면 픽셀 간 차이는 움직임이 작은 영상에서는 잘 적용되지만 물체나 카메라의 움직임이 큰 경우에는 성능이 저하된다. 그리고 히스토그램 차이는 빠른 움직임이나 회전에는 강인하지만 장면 전환이 일어난 두 화면이 비슷한 칼라 분포나 밝기 분포를 가지면 성능이 저하되는 단점을 가지고 있다^[9].

한편 문턱 값 기반의 방법은 결정 함수(discriminant function)를 이용하는 것으로서 현재 화면과 바로 이전 화면 사이의 화면 간 차이를 계산하고 이 값이 결정 함수에 대입될 때 어떤 함수 값이 나오느냐에 따라 장면 전환 존재 여부를 결정하는 것이다. 결정 함수는 선형적일 수도 있고 비 선형적일 수도 있다. 이 방법은 여러 종류의 특징을 이용할 수도 있지만 장면 전환 검출에서는 주로 한 종류의 특징을 사용하는 경우에 사용되어 왔다. 이런 경우에는 결정 함수는 단순한 스칼라 형태의 문턱 값으로 주어지며 화면 간 차이 값이 이 문턱 값보다 더 클 때 현재의 화면을 장면 전환 화면으로 판단하게 된다. 이 방법은 간단한 반면 문턱 값을 효과적으로 결정하는 것이 어렵다. 왜냐하면 장면 전환이 존재하는 화면에서도 화면간 차이 값은 항상 정해진 문턱 값 이상이 나오는 것은 아니기 때문이다. 따라서 이런 단점을 극복하기 위해서 고정된 문턱 값

이 아닌 적응 적으로 변하는 문턱 값을 사용하는 방법들이 제안되었다^{[11][12]}.

한편 클러스터링 기반의 방법은 거리 함수(distance function)을 사용하는 방법이다. 이 방법은 현재 화면을 거리가 가장 가까운 클러스터로 귀속시키는 방법이다. 여기에는 하나의 특징만을 사용해서 클러스터링을 하는 방법^[13] 과 여러 개의 특징을 사용해서 장면 전환을 찾는 방법^{[9][14]} 등이 있다.

한편 이런 클러스터링 기반의 방법의 경우 초기 중심이 클러스터링 결과에 많은 영향을 준다^[10]. 하지만 기존의 클러스터링 기반의 장면 전환 검출 기법에서는 초기 중심을 정하는 방법은 다루지 않았다^{[9][13][14]}. 따라서 본 논문에서는 적응 적으로 클러스터의 초기 중심을 추정하면서 장면 전환을 찾는 방법을 제안한다. 이를 위해서 일반적으로 많이 사용되는 픽셀 간 차이 기법과 히스토그램 차이 기법을 특징으로 사용하였고 클러스터링을 위해서는 k -평균 클러스터링을 사용하였다. 본 논문의 구성은 다음과 같다. 우선 2절에서는 제안된 기법을 설명하고 3절에서는 실험과 결과 분석을 한다. 4절에서는 결론 및 향후과제를 말하고 논문을 맺고자 한다.

II. 제안하는 장면 전환 검출 기법

1. 특징 추출

여러 개의 특징들을 동시에 사용해서 장면 전환 검출을 하기 위해서는 각 특징들은 장면 전환 화면에서 동일한 반응을 보여야 한다. 그리고 서로의 단점을 보완할 수 있어야 한다. 본 논문에서는 장면 전환 검출을 위한 특징으로 히스토그램 차이와 픽셀 간 차이를 사용한다. 그 이유는 다음과 같다. 우선 두 특징은 상호 보완적이다. 일반적으로 장면 전환 검출을 위해 히스토그램이 많이 사용된다. 이것은 서로 이웃하는 두 화면의 히스토그램을 이용해서 bin(bin)별로 차이를 구한 후 이 값들을 선형적 혹은 비 선형적으로 더해서 구한 거리 척도에 의해 장면 전환을 검출한다. 이 방법은 전역적 방법으로서 카메라 움직임이나 물체의 움직임, 회전의 영향을 적게 받는다고 알려져 있다^[9]. 그러나 이 방법은 연속되는 두 샷이 비슷한 칼라 혹은 밝기 분포를 가질 때에는 잘 사용할 수 없다는 단점이 있다. 한편 이에 대응되는 국부적 방법으로 픽셀 간 차이를 사용

하는 방법이 있다. 이 방법은 카메라 움직임이나 물체의 움직임에 매우 민감하다. 즉 움직임이 큰 부분에서는 오 인식(false positive)을 할 확률이 높아진다는 단점이 있다. 하지만 이 방법은 두 화면의 칼라 혹은 밝기 분포가 비슷해도 국부적인 차이만을 고려하기 때문에 장면 전환을 잘 찾을 수 있다.

또 하나의 이유는 서로 다른 특징들을 동시에 사용해서 클러스터링을 할 경우 거리 척도들 사이의 정규화 문제가 발생하는데 히스토그램 차이와 픽셀 간 차이 값들은 최대 값이 정량적으로 나오기 때문에 정규화가 용이하다는 장점이 있다.

이를 위한 식은 다음과 같다. 우선 영상의 사이즈를 $N \times M$, 히스토그램의 bin의 개수를 B , k 번째 화면의 위치에서의 밝기를 $I_k(x, y)$, i 번째 bin에 포함되는 픽셀의 개수를 $H_k(i)$ 라 두자. 그리고 k 번째 화면과 $k+1$ 번째 화면의 히스토그램 차이를 $HD_{(k, k+1)}$, 픽셀 간 차를 $PD_{(k, k+1)}$ 라 두면

$$HD_{(k, k+1)} = \sum_{i=0}^B |H_k(i) - H_{k+1}(i)| \quad (1)$$

$$PD_{(k, k+1)} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} |I_k(x, y) - I_{k+1}(x, y)| \quad (2)$$

가 된다. 한편 이 두 차이 값들은 서로 다른 단위의 값을 가지므로 동시에 사용하기 위해서는 정규화를 시킬 필요가 있다. 본 논문에서는 임의의 두 연속하는 화면 사이의 차이 값이 0에서 1사이의 값을 가지도록 정규화 시킨다. 우선 히스토그램 차이의 경우 이 값이 가질 수 있는 최대 값은 두 화면의 히스토그램에서 공통적인 bin이 존재하지 않는 경우이므로 $2NM$ 의 값이 되며 픽셀 간 차이의 경우에는 모든 픽셀에서의 밝기 값의 차이가 255일 때(8 bit/pixel인 경우) 최대의 값 $255NM$ 을 가지므로 이 값을 사용해서 정규화를 한다. 따라서 최종적으로 사용할 수 있는 히스토그램 차이와 픽셀 간 차이의 정규화 된 차이 값을 각각 $NHD_{(k, k+1)}$ 와 $NPD_{(k, k+1)}$ 라 두면

$$NHD_{(k, k+1)} = \frac{HD_{(k, k+1)}}{2MN} \quad (3)$$

$$NPD_{(k, k+1)} = \frac{PD_{(k, k+1)}}{255MN} \quad (4)$$

가 된다.

2. 클러스터링을 이용한 장면 전환 검출

앞 절에서 설명한 방법을 이용해서 연속되는 두 장의 화면으로부터 얻은 정규화 된 차이 값들로 구성된 특징 벡터 $x=(NHD, NPD)$ 를 만들 수 있다. 우선 장면 전환이 존재하는 화면들로부터 나온 특징 벡터들은 그림 1(b)와 같이 크기가 일정한 값 이상을 가지면서 임의의 분포를 띤다. 하지만 장면 전환이 존재하지 않는 화면들로부터 나온 특징 벡터들은, 비록 서로 다른 샷에 속한다 할지라도, 그림 1(a)와 같이 비슷한 크기를 가지면서 분포를 이룬다. 또 이와 같이 장면 전환이 없는 화면에서 나오는 특징 벡터들 중 임의의 두 특징 벡터의 거리가 가깝다는 것은 연속되는 두 화면들의 화면 분포의 차이가 비슷하다는 것을 의미한다. 따라서 이런 화면들에서 얻어진 특징 벡터들을 찾기 위해서는 거리가 가까운 것들끼리 모을 필요가 있다. 이런 식으로 클래스 내(intra-class) 거리가 최소가 되도록 특징 벡터들을 모으기 위해서는 클러스터링 기법을 사용해야 한다.

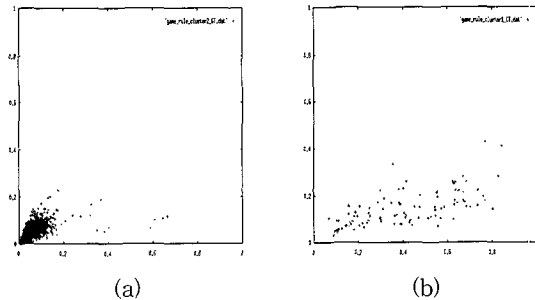


그림 1. 특징 벡터의 분포(x축: NHD, y축: NPD)
(a) 장면 전환이 존재하지 않는 화면에서 구한 특징 벡터(ground truth)
(b) 장면 전환이 존재하는 화면에서 구한 특징 벡터(ground truth)

Fig. 1. Feature vector distribution
(a) Feature vectors calculated from the non-shot-change frames(ground truth)
(b) Feature vectors calculated from the shot change frames(ground truth).

일반적인 클러스터링 기법은 모든 데이터가 미리 주어져 있다고 가정을 한 후 데이터들 사이의 유사 척도(similarity measure)와 클러스터링 기준(clustering criterion)을 정하고 이를 최적화 시키는 방향으로 클러스터링을 진행하게 된다. 하지만 기존의 클러스터링 방법을 장면 전환 검출에 직접 적용하기에는 무리가 있

다. 왜냐하면 임의의 화면에서의 장면 전환 존재 여부를 체크하기 위해서는 모든 화면을 한꺼번에 조사할 필요가 없이 그 화면의 주변 화면들만 보면 되기 때문이다. 이를 위해 본 논문에서는 현재의 화면과 그 주변의 몇 화면만 가지고 클러스터링을 적용하였다.

클러스터링을 위한 대표적인 알고리즘에는 단순 클러스터 찾기(simple cluster-seeking) 알고리즘, 최대 거리(maximum distance) 알고리즘, k -평균 알고리즘 등이 있으나^[10] 본 논문에서는 k -평균 알고리즘을 사용하였다. 그 이유는 다음과 같다. 우선 이 방법은 미리 클러스터의 개수를 지정할 수 있다. 장면 전환 검출에서는 클러스터링 결과 2개의 클러스터만 존재하도록 해야 하는데 최대 거리 알고리즘의 경우에는 2개 이상의 클러스터가 존재할 가능성이 있는 반면 k -평균 알고리즘의 경우에는 미리 클러스터의 개수를 지정할 수 있다는 장점이 있다. 즉 k 는 2의 값을 가지게 된다. 그리고 두 번째 장점으로 단순 클러스터 찾기 알고리즘에서는 지정된 문턱 값에 따라서 클러스터링 결과가 달라지지만 k -평균 알고리즘에서는 클러스터링을 위한 문턱 값을 미리 정할 필요가 없으며 클러스터의 초기 중심만 잘 정해주면 된다^[10]. 클러스터의 초기 중심을 추정하는 방법은 뒤에서 설명한다.

k -평균 클러스터링에 사용된 유사 척도와 클러스터링 기준은 다음과 같다. 우선 유사 척도는 두 벡터 사이의 유클리디언 거리인 $\|x - y\|$ 로 정의하였고 클러스터링 기준(clustering criterion)은 식 5와 같이 정의하였다.

$$J^{(k)} = J_{NSC}^{(k)} + J_{SC}^{(k)} \quad (5)$$

$$J_{NSC}^{(k)} = \sum_{x \in S_{NSC}^{(k)}} \|x - C_{NSC}^{(k)}\|, \quad J_{SC}^{(k)} = \sum_{x \in S_{SC}^{(k)}} \|x - C_{SC}^{(k)}\|$$

$$C_{NSC}^{(k)} = \frac{1}{N_{NSC}^{(k)}} \sum_{x \in S_{NSC}^{(k)}} x, \quad C_{SC}^{(k)} = \frac{1}{N_{SC}^{(k)}} \sum_{x \in S_{SC}^{(k)}} x$$

여기에서 $S_{NSC}^{(k)}$ 은 k 번째 반복 과정 후 장면 전환이 없다고 판단한 특징 벡터들의 집합이고, $S_{SC}^{(k)}$ 은 장면 전환이 있다고 판단한 특징 벡터들의 집합이며 $N_{NSC}^{(k)}$ 와 $N_{SC}^{(k)}$ 는 각각의 집합의 원소의 개수이다. 그리고 $C_{NSC}^{(k)}$ 와 $C_{SC}^{(k)}$ 은 각각에 대한 중심을 의미한다. 식 5를 최소로 만들기 위해서는 매 반복 과정마다 임의의 특징 벡터 x 를 거리가 가장 가까운 클러스터에 귀속시켜야 한다. 이런 반복 과정을 통해 k -평균 알고리즘은 클

러스터의 초기 중심 $C_{NSC}^{(0)}$, $C_{SC}^{(0)}$ 을 정한 후 집합 $S_{NSC}^{(k)}$ 와 $S_{SC}^{(k)}$ 의 원소가 더 이상 변하지 않을 때까지 클러스터링을 수행한다^[10].

3. 초기 클러스터링 과정(Initial clustering stage)

초기 클러스터링 과정은 그림 2와 같이 최초의 장면 전환 화면이 나타날 때를 찾는 과정이며 이 결과는 주 클러스터 과정에서 클러스터들의 초기 중심을 추정하는데 사용된다. 클러스터링 후 장면 전환이 존재한다고 판단된 화면들로 구성된 클러스터를 SC (shot change) 클러스터, 그렇지 않은 클러스터를 NSC (no shot change) 클러스터라고 하자.

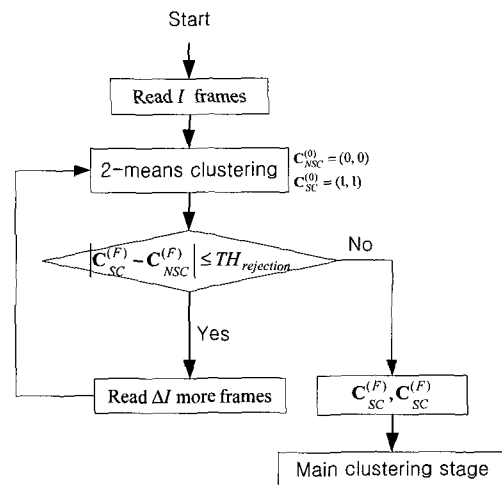


그림 2. 초기 클러스터링 과정
Fig. 2. Initial clustering stage

초기 클러스터링 과정은 다음과 같다. 우선 최초의 I 개의 화면을 이용해서 k -평균 클러스터링을 수행한다. 이 때 각각의 클러스터의 초기 중심은 $C_{NSC}^{(0)} = (0,0)$, $C_{SC}^{(0)} = (1,1)$ 로 둔다. 그리고 나서 클러스터링을 수행한 후 나오는 중심을 $C_{NSC}^{(F)}$ 와 $C_{SC}^{(F)}$ 라 두고 이 중심들 사이의 거리를 계산한다. 이 값이 미리 정해진 문턱 값보다 더 크면 장면 전환이 존재함을 의미하고 그렇지 않으면 장면 전환 화면이 없다는 것을 의미한다^[9]. 만약 I 개의 화면에서 장면 전환이 없다고 판단되면 I 개의 화면의 뒤를 따라 오는 ΔI 개의 화면을 I 개의 화면과 더해서 $I + \Delta I$ 개의 화면을 사용해서 k -평균 클러스터링을 수행한다. 이 과정을 장면 전환 화면이 나올 때까지 되풀이해서 수행하면 최초의 장면 전환이 발생하는 화

면을 찾을 수 있다. 그리고 이 결과 나오는 $C_{NSC}^{(F)}$ 와 $C_{SC}^{(F)}$ 는 주 클러스터링 과정에서 클러스터의 초기 중심을 구하는 데 사용될 수 있다.

4. 주 클러스터링 과정(Main clustering stage)

m 번째 화면에서의 장면 전환 존재 여부를 조사하고 있다고 하자. 이를 위해 우선 이전의 연속적인 N 개의 화면의 특징 벡터들과 현재의 화면의 특징 벡터들로 이루어진 집합 S^m 을 구성한다. 그리고 이 $N+1$ 개의 특징 벡터들로부터 집합 S^m 을 클러스터링 하기 위한 클러스터의 초기 중심을 추정한다. 그리고 나서 집합 S^m 에 대해 k -평균 클러스터링을 수행한다. 클러스터링 결과 현재 조사하는 m 번째 화면은 장면 전환이 존재하는 화면들로 구성된 SC 클러스터(ω_{SC}) 또는 장면 전환이 없는 화면들로 구성된 NSC 클러스터(ω_{NSC}) 둘 중에 하나로 속하게 된다. 이런 식으로 뒤따라오는 화면에 대해서도 같은 방법으로 클러스터링을 수행하면서 장면 전환 존재 여부를 결정한다. 주 클러스터링 과정의 흐름도는 그림 3에 나와 있다.

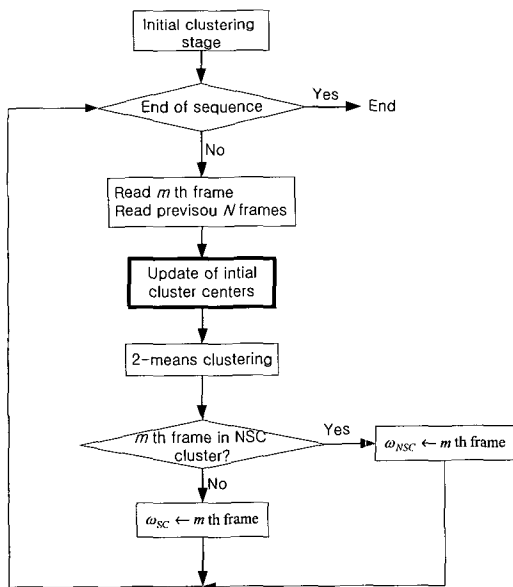


그림 3. 주 클러스터링 과정
Fig. 3. Main clustering stage

5. 초기 클러스터 중심의 추정(Estimation of initial cluster centers)

이제 초기 클러스터 중심을 추정하는 과정을 설명한

다. 이 과정의 흐름도는 그림 4에 나와 있다. $C_{SC,m}^{(0)}$ 와 $C_{NSC,m}^{(0)}$ 를 각각 m 번째 화면의 장면 전환 존재 여부를 결정하기 위하여 S^m 안에 있는 $N+1$ 개의 특징 벡터들을 클러스터링 할 때 사용하는 SC 클러스터와 NSC 클러스터의 초기 중심이라고 하자. 그리고 이 특징 벡터들을 클러스터링 한 후 나오는 SC 클러스터와 NSC 클러스터의 중심을 각각 $C_{SC,m}^{(F)}$ 와 $C_{NSC,m}^{(F)}$ 라고 하자. 그림 1에서는 NSC 클러스터에 속하는 화면들은 거의 편차가 없이 일정한 값 부근에 몰려 있는 반면 SC 클러스터에 속하는 화면들은 장면 전환 화면들에서 나타나는 화면간 차이 값들의 크기가 일정하지 않기 때문에 특징 값들의 편차가 아주 큼을 알 수 있다. 이런 사실을 이용해서 클러스터의 초기 중심을 추정하는 과정은 다음과 같다.

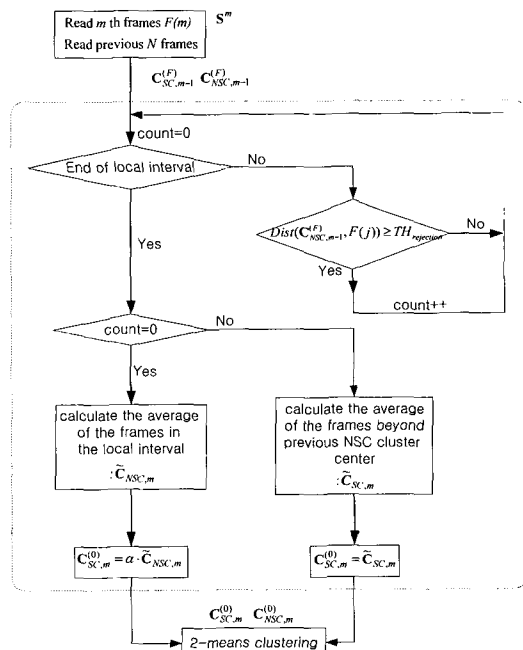


그림 4. 초기 클러스터 중심($C_{SC,m}^{(0)}$ 과 $C_{NSC,m}^{(0)}$)의 추정 과정(점선 부분)
Fig. 4. Estimation of initial cluster centers

우선 NSC 클러스터의 경우에는 클러스터 중심이 거의 변하지 않기 때문에 $C_{NSC,m}^{(0)}$ 는 집합 S^{m-1} 의 클러스터링 결과 나온 $C_{NSC,m-1}^{(F)}$ 의 값을 그대로 이용한다. 그리고 SC 클러스터의 초기 중심($C_{SC,m}^{(0)}$)은 주어진 데이터에 따라서 항상 변하기 때문에 이 값은 주어진 데

이터를 이용해서 추정해야 한다. 이 방법의 기본 개념은 S^m 에 포함된 화면들 중 잠재적으로 장면 전환 화면이 될 가능성이 있는 화면들을 미리 선택하고 이로부터 SC 클러스터의 초기 중심 $C_{SC,m}^{(0)}$ 을 추정한다는 것이다. 우선 $N+1$ 개의 특징 벡터 중에서 $C_{NSC,m-1}^{(F)}$ 에서 일정 값($TH_{rejection}$) 이상 떨어진 특징 벡터들을 선택한다. 이런 특징 벡터를 만드는 화면들은 잠재적으로 장면 전환이 존재하는 화면이 될 가능성이 높은 화면들이다. 이렇게 선택한 특징 벡터들을 평균함으로써 $C_{SC,m}^{(0)}$ 를 추정한다. 만약 S^m 에 그런 특징 벡터가 없으면 S^m 에 포함된 $N+1$ 개의 특징 벡터들이 모두 NSC 클러스터에 속할 확률이 높다는 의미이다. 이 경우에는 이 특징 벡터의 평균값에 적당한 상수 α 를 곱해서 $C_{SC,m}^{(0)}$ 를 추정한다. 왜냐하면 SC 클러스터의 초기 중심은 NSC 클러스터의 초기 중심보다 적당한 크기의 상수 배만큼 멀리 떨어져 있기 때문이다. 이런 방법으로 추정하는 클러스터의 초기 중심은 클러스터링 될 화면에 따라서 적응적으로 달라진다.

III. 실험 결과 및 분석

제안하는 알고리즘의 분석을 위해 104개의 급격한 장면 전환이 존재하는 뮤직 비디오를 사용해서 실험을 하였다. 그리고 성능 분석을 위해 일반적으로 장면 전환 검출에서 많이 사용되는 소환 비율(recall rate)과 정확 비율(precision rate)을 측정하였다. 여기에서 소환 비율이란 실제로 존재하는 장면 전환 화면의 개수와 정확하게 찾은 장면 전환 화면의 개수 사이의 비율이고 정확 비율이란 장면 전환이 있다고 판단한 화면의 개수와 정확하게 찾은 장면 전환 화면의 개수 사이의

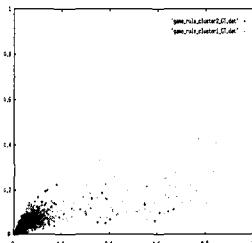


그림 5. 테스트 영상의 화면간 차이 값의 2차원 분포(x축: NHD, y축: NPD)
Fig. 5. Feature vector distribution of test sequence(x-axis: NHD, y-axis: NPD)

비율을 의미한다. 그림 5는 특징 벡터를 2차원 평면에 도시한 것이다. $I=10$, $\Delta I=5$, $TH_{rejection}=0.2$, $N=15$, $\alpha=4.0$ 으로 두고 실험을 하였다.

● 문턱 값 기반의 방법과 클러스터링 기반의 방법의 성능 비교

첫째 실험에서는 문턱 값 기반의 방법과 클러스터링 기반의 방법의 성능을 비교하였다. 일반적으로 문턱 값 기반의 방법에서 고정된 문턱 값을 사용하면 문턱 값의 크기에 따라 성능이 달라지므로 최적의 문턱 값을 정하기가 힘이 든다. 따라서 여기에서는 적응적으로 변하는 문턱 값으로 장면 전환을 결정하는 방법을 사용하였다^[11]. 이 방법은 현재 화면에서의 화면 간 차이 값이 sliding window 구간 안에 있는 샘플들 중에서 최대이고 이 값이 sliding window 안에 있는 두 번째로 큰 값보다 n 배 이상 클 때 장면 전환이 있다고 판단한다. 표 1에 있는 위의 두 개의 행은 각각 히스토그램 차이(HD)와 픽셀 차이(PD)를 사용해서 실험한 결과이고 셋째 행은 제안하는 방법에 의한 실험 결과이다. 실험 결과 제안하는 방법의 성능이 문턱 값 기반의 방법보다 더 좋게 나옴을 알 수 있었다. 왜냐하면 문턱 값 기반의 방법은 한 종류의 특징을 사용하기 때문에 앞 절에서 기술한 특징의 단점으로 인해 성능의 저하가 일어나기 때문이다. 특히 픽셀 간 차이 값은 움직임이 큰 영역에서는 오 인식(false positive)을 하게 되므로 정확 비율에서 성능이 더 떨어짐을 알 수 있다.

● k -평균 클러스터링의 초기 중심이 적응적으로 변하는 경우와 고정된 경우의 성능 비교

표 1. 문턱 값 기반의 방법과 제안하는 클러스터링 방법의 성능 비교

Table 1. Performance comparison between thresholding-based method and proposed clustering-based method

특징과 검출 방법	성능	
	소환 비율(%)	정확 비율(%)
Adaptive HD Thresholding	100/104(96.15%)	100/108(92.59%)
Adaptive PD Thresholding	95/104(91.34%)	95/109(87.15%)
Proposed HD+PD Clustering	102/104(98.07%)	102/112(91.07%)

둘째 실험에서는 k -평균 클러스터링을 이용한 장면 전환 검출에서 초기 중심이 변하지 않는 경우^[9]와 제안하는 방법의 성능을 비교하였다. 표 2에 있는 위의 5개의 행은 SC 클러스터와 NSC 클러스터의 초기 중심이 고정된 경우의 결과이며 마지막 행은 초기 중심이 적응적으로 변하는 경우의 실험 결과를 의미한다. 우선 초기 중심이 적응적으로 변하는 경우 소환 비율은 가장 좋고 정확 비율도 초기 중심이 고정된 경우에 나오는 최고의 성능만큼 좋아짐을 알 수 있다.

우선 클러스터의 초기 중심이 고정된 경우에 위와 같은 결과가 나오는 이유는 다음과 같다. 일반적으로 특징 벡터들이 원점으로부터 멀어질수록 장면 전환 화면일 확률이 커지는 경향이 있다. 그리고 SC 클러스터의 초기 중심 부근과 그 이상에 있는 화면들이 장면 전환 화면으로 클러스터링 될 가능성이 높아진다. SC 클러스터의 초기 중심을 고정시키고 클러스터링을 하는 경우에는 이 초기 중심이 원점으로부터 멀어진다면 그 클러스터 중심 부근과 그 이상에 있는 화면들이 장면 전환 화면이 될 확률은 높아지며 따라서 정확 비율은 증가한다. 하지만 이 경우 초기 중심 ($C_{SC}^{(0)}$) 부근과 그 이상에 있는 화면들만 SC 클러스터로 선택하기 때문에 소환 비율은 낮아진다. 이런 현상은 단순히 하나의 문턱 값을 가지고 장면 전환 검출을 할 때 문턱 값을 높이는 경우와 동일한 것이다. 즉 고정된 문턱 값 기반의 장면 전환 검출 기법에서 문턱 값을 크게 하면 소환 비율은 감소하는 반면 정확 비율은 증가하는 것과도 같은 현상이다.

하지만 제안된 방법은 소환 비율도 가장 높았고 정확 비율도 다른 방법 못지 않게 높게 나오음을 알 수 있다. 즉 같은 소환 비율을 가지는 경우와 비교해서 정확 비율은 훨씬 높아짐을 알 수 있다. 그 이유는 다음과 같다. 그림 1의 실제 결과(ground truth)에서 알 수 있듯이 비록 장면 전환이 존재하는 화면들의 특징 벡터들이라 할지라도 장면 전환이 존재하지 않는 특징 벡터보다 더 원점에 가까울 수 있다. 따라서 클러스터의 초기 중심을 고정시키면 이런 경우에는 제대로 검출하지 못함을 알 수 있다. 반면 제안하는 방법은 장면 전환의 가능성이 있는 화면들만 먼저 선택해서 SC 클러스터의 초기 중심을 추정하기 때문에 어느 정도 특징 값이 적은 장면 전환 화면들도 잘 검출할 수 있다. 이

것은 문턱 값이 고정되지 않고 적응적으로 변하면서 장면 전환을 찾는 기법과 개념적으로 동일하다고 볼 수 있다. 그림 6에서는 클러스터링 된 결과를 보여준다. 클러스터링 결과 그림 1의 실제 결과(ground truth)와 마찬가지로 SC 클러스터에 속하는 화면 중에는 특징 벡터의 크기가 NSC 클러스터에 속하는 화면의 특징 벡터보다 더 작은 것도 있음을 알 수 있다.

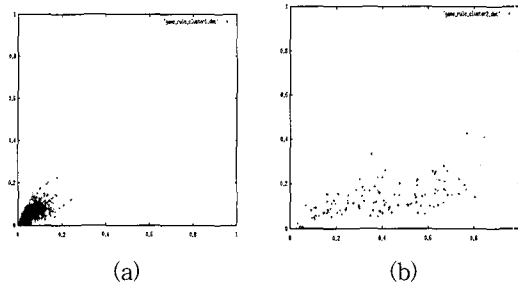


그림 6. 클러스터링 결과
(a) 장면 전환이 없다고 판단한 클러스터
(b) 장면 전환이 있다고 판단한 클러스터

Fig. 6. Clustering results
(a) NSC cluster (b) SC cluster

표 2. 클러스터의 초기 중심이 고정된 경우와 적응적으로 변하는 경우의 성능 비교

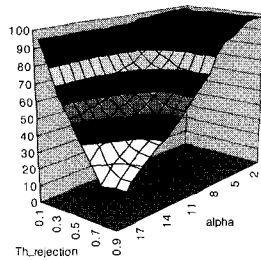
Table. 2. Performance comparison between fixed vs. adaptive initial cluster centers

성능	소환 비율(%)	정확 비율(%)
NSC & SC 클러스터의 초기 중심		
(0.0, 0.0) & (0.1, 0.1)	102/104(98.07%)	102/809(12.6%)
(0.0, 0.0) & (0.3, 0.3)	85/104(81.73%)	85/103(82.52%)
(0.0, 0.0) & (0.5, 0.5)	64/104(61.53%)	64/71(90.14%)
(0.0, 0.0) & (0.7, 0.7)	34/104(32.69%)	34/38(89.47%)
(0.0, 0.0) & (0.9, 0.9)	14/104(13.46%)	12/15(93.33%)
Adaptive Cluster Center	102/104(98.07%)	102/112(91.07%)

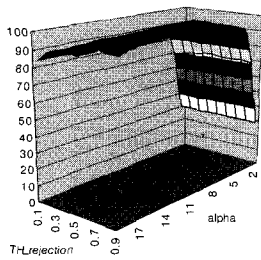
● 파라미터 값의 변화에 따르는 성능 분석

일반적으로 k -평균 클러스터링을 이용하는 방법은 문턱 값 기반의 검출 방법과는 달리 장면 전환을 결정

하기 위해서 문턱 값을 정할 필요가 없다는 장점이 있다. 다만 제안하는 기법에서는 SC 클러스터의 초기 중심을 추정하기 위해 두 종류의 파라미터 값($TH_{rejection}$, α)이 필요하다. 여기에서는 이 값들이 장면 전환 검출 성능에 끼치는 영향을 분석하였다.



(a)



(b)

그림 7. α 와 $TH_{rejection}$ 의 변화에 따르는 장면 전환 검출 성능

(a) 소환 비율 (b) 정확 비율

Fig. 7. Performance of shot change detection as a function of α and $TH_{rejection}$

(a) Recall rate (b) Precision rate

실험 결과는 그림 7에 나와 있다. 이 그림에서 $TH_{rejection}$ 과 α 의 값이 증가함에 따라 소환 비율은 감소하고 정확 비율은 증가함을 알 수 있다. 그 이유는 다음과 같다. SC 클러스터의 초기 중심 추정에서 $TH_{rejection}$ 의 값이 감소하면 장면 전환 화면이 될 가능성이 있는 화면들의 개수가 증가하게 된다. 즉 SC 클러스터의 초기 중심이 원점에 더 가깝게 된다. 이런 현상은 α 의 값이 감소해도 동일하게 발생한다. 따라서 $TH_{rejection}$ 과 α 의 값이 증가함에 따라 소환 비율은 감소하고 정확 비율은 증가하게 된다. $TH_{rejection}$ 과 α 는 문턱 값을 적응적으로 바꾸면서 장면 전환을 찾는 방법^[11]에서 사용된 n 과 동일한 역할을 한다.

IV. 결론 및 향후 과제

본 논문에서는 한 종류의 특징만을 사용하지 않고 여러 종류의 특징을 동시에 사용하는 클러스터링 방법을 이용하여 비디오 영상에서 급격한 장면 전환을 검출하는 방법을 제안하였다. 이를 위해 초기 클러스터 중심을 적응적으로 변화시키면서 k -평균 클러스터링을 하는 방법을 제안하였다. 앞으로는 여러 종류의 특징들의 정량적 분석을 통해 더 많은 종류의 특징을 동시에 고려할 수 있는 방법에 대한 연구가 필요하다. 그리고 압축 영역에서의 특징들을 이용해서 클러스터링을 하는 방법에 대한 연구도 함께 진행되어야 한다.

참고 문헌

- [1] F. Idris and S. Panchanathan, "Review of Image and Video Indexing Techniques", *Journal of Visual Communication and Image Representation*, Vol. 8, No. 2, June, pp.146-166, 1997.
- [2] I. K. Sethi and N. Patel, "A statistical approach to scene change detection", *Proceedings of SPIE Storage and Retrieval for Image and Video Databases III*, 1995, Vol. 2420.
- [3] 이훈철, 이청우, 김성대, "클러스터링을 이용한 비디오 영상에서의 급격한 장면 전환 검출 기법", 제 12회 신호처리 합동 학술 대회 논문집, Oct.1999
- [4] Huncheol Lee, CheongWoo Lee and SeongDae Kim, "Abrupt shot change detection using an unsupervised clustering of multiple features", *Proceeding of ICASSP 2000*, Vol. 4, pp.2015~2018.
- [5] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video", *Multimedia Systems I*, 1993, pp. 10-28.
- [6] R. Brunelli, O. Mich, and C. M. Modena, "A Survey of the Automatic Indexing of Video Data", *Journal of Visual Communication and Image Representation*, Vol. 10, pp.78-112, 1999.

- [7] Jianhao Meng, Yujen Juan, and Shih Fu Chang, "Scene Change Detection in an MPEG Compressed Video Sequence", *Digital Video Compression : Algorithms Tech*, Vol. 2419, Feb. 1995, pp.14-25.
- [8] F. Arman, A. Hsu, and M. Chiu, "Feature management for large video database", *Proceedings of SPIE Storage and Retrieval for Image and Video Database*, 1993, pp.2-12.
- [9] A. Mufit Ferman and A. Murat Tekalp, "Efficient Filtering and Clustering Methods for Temporal Video Segmentation and Visual Summarization", *Journal of Visual Communication and Image Representation*, Vol.9, No.4, December, pp. 336-351, 1998.
- [10] Julius T. Tou and Rafael C. Gonzales, "Pattern Recognition Principles", Addison-Wesley Publishing Company, 1974.
- [11] B. L. Yeo and B. Liu, "Rapid Scene Analysis on Compressed Domain", *IEEE Tran. on Circuits and Systems for Video Technology*, Vol-5, No-6, pp. 533-544, 1995.
- [12] Y. Yusoff, W. Christmas and J. Kittler, "Video Shot Cut Detection using Adaptive Thresholding", *Proc. of British Machine Vision Conference*, 2000.
- [13] B. Günsel, et. al, "Video Indexing through Integration of Syntactic and Semantic Features", *Proc. Workshop on Applications of Computer Vision*, pp. 90 - 95, 1996.
- [14] M. R. Naphade, et. al, "A High Performance Shot Boundary Detection Algorithm using Multiple Cues", *Proc. of ICIP*, Vol-3, pp.884-887, 1998.

 저 자 소 개

李 熏 哲 (學生會員)

1995년 KAIST 전기 및 전자공학과 학사. 1997년 KAIST 전기 및 전자공학과 석사. 1997년~현재 KAIST 전기 및 전자공학과 박사 과정, 주관심 분야는 영상 처리, 컴퓨터 비전, 비디오 검색 등임.

高 綸 浩 (學生會員)

1995년 부산대학교 전자공학과 학사. 1997년 KAIST 전기 및 전자공학과 석사. 1997년~현재 KAIST 전기 및 전자공학과 박사 과정, 주관심 분야는 영상 및 멀티미디어 통신, 워터마킹 등임.

金 聖 大 (正會員) 論文 第 38卷 SP編 第 2號 參照

尹 炳 珠 (學生會員)

1993년 경북대학교 전자공학과 학사. 1996년 KAIST 전기 및 전자공학과 석사. 1996년~현재 KAIST 전기 및 전자공학과 박사과정. 주관심 분야는 영상 코딩, 영상 처리, 컴퓨터 비전 등임.

俞 相 朝 (正會員)

1988년 한양대학교 전자통신공학과 학사. 1990년 KAIST 전기 및 전자공학과 석사. 2000년 KAIST 전기 및 전자공학과 공학 박사. 1990년 3월~2001년 2월 한국 통신 연구 개발 본부. 2001년 3월~현재 인하대학교 정보 통신 전문 대학원 조교수. 주관심 분야는 멀티미디어 통신, 트래픽 모델링 및 큐잉 해석, 멀티미디어 QoS 제어 등임.