



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0102494
(43) 공개일자 2010년09월24일

(51) Int. Cl.

G06Q 50/00 (2006.01)

(21) 출원번호 10-2009-0020919

(22) 출원일자 2009년03월11일

심사청구일자 2009년03월11일

(71) 출원인

연세대학교 산학협력단

서울 서대문구 신촌동 134 연세대학교

(72) 발명자

이인권

서울특별시 양천구 목6동 한신청구아파트 107동 401호

윤종철

서울특별시 서대문구 창천동 89-26번지 1층

(74) 대리인

특허법인우인

전체 청구항 수 : 총 34 항

(54) 자동 뮤직 비디오 생성 장치와 방법 이를 구현하기 위해 컴퓨터상에서 수행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체

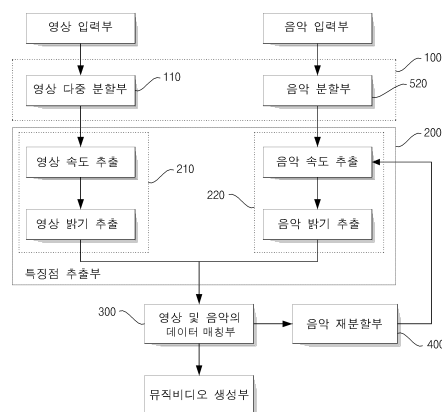
(57) 요약

본 발명은 자동 뮤직 비디오 생성 장치와 방법 이를 구현하기 위해 컴퓨터상에서 수행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체에 관한 것이다.

본 발명에 따른 자동 뮤직 비디오 생성 방법의 일례는 각 이미지 간의 외형 차이를 기준으로 입력된 영상 데이터를 다중 분할하는 영상 분할 단계와 주파수 영역에서의 순간적인 진폭 변화를 기준으로 입력된 음악 데이터를 분할하는 음악 분할 단계를 포함하는 데이터 분할 단계; 분할된 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 영상 분석 단계와 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 음악 분석 단계를 포함하는 데이터 분석 단계; 및 영상 및 음악의 변이점 또는 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 텀이 서로 대응되도록 영상 데이터 조각과 음악 데이터 조각을 매칭하는 데이터 매칭 단계;를 포함한다.

이와 같은 자동 뮤직 비디오 생성 방법의 일례는 뮤직 비디오의 장면 연속성이 최대한 유지되고, 시청자가 집중하는 부분의 동기성을 높여줄 뿐만 아니라 조각 데이터의 매칭을 위한 타임 와핑(Time-warping)을 통해 보다 역동적인 뮤직 비디오의 생성이 가능하도록 하는 효과가 있다.

대표도 - 도1



특허청구의 범위

청구항 1

입력된 영상 데이터의 각 이미지의 외형 매칭(Contour shape matching)을 통한 상기 각 이미지 간의 외형 차이가 미리 결정된 임계값을 초과하는 영상 변이점을 기준으로 상기 입력된 영상 데이터를 분할하는 영상 분할 단계와 입력된 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화가 미리 결정된 임계값을 초과하는 음악 변이점을 기준으로 상기 입력된 음악 데이터를 분할하는 음악 분할 단계를 포함하는 데이터 분할 단계;

상기 영상 분할 단계에 의해 분할된 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 영상 분석 단계와 상기 음악 분할 단계에 의해 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 음악 분석 단계를 포함하는 데이터 분석 단계; 및

상기 영상 및 음악의 변이점 또는 상기 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 팀이 서로 대응되도록 상기 영상 데이터 조각과 음악 데이터 조각을 매칭하는 데이터 매칭 단계;

를 포함하는 자동 뮤직 비디오 생성 방법.

청구항 2

제 1 항에 있어서,

상기 영상 분할 단계에서 상기 각 이미지 간의 외형 차이를 추출하기 위한 전처리 과정으로 영상 데이터의 각 이미지의 외형은 캐니 특징점 추출 기법(Canny edge detector)을 통해 이미지의 특징점으로 표현되는 에지 맵(Edge Map)으로 변환되는 것

을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 3

제 1 항에 있어서,

상기 영상 분할 단계에서

상기 각 이미지의 외형 차이는 다음의 수학식에 따라 두 영상 프레임에서 특징점의 유사성을 측정하여 추출하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

[수학식]

$$I_{i,j} = \sum_{g=1}^7 |1/c_g^i - 1/c_g^j|$$

where

$$c_g^i = \text{sign}(h_g^i) \log_{10} |h_g^i|$$

여기서, i와 j는 영상 데이터의 프레임 번호, h는 Hu-moment로 상기 에지 맵의 특징점으로부터 획득되는 7차원 벡터값, g는 Hu-moment의 개수, $I_{i,j}$ 는 유사도 값을 의미한다.

청구항 4

제 3 항에 있어서,

상기 영상 변이점은

상기 유사도 값으로 이루어진 유사 행렬을 대각선 방향으로 Radial Symmetric gaussian Kernel(RSK) 적용한 다음의 식에 의해 결정되는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

[수학식]

$$EV(i) = \sum_{u=-\delta}^{\delta} \sum_{v=-\delta}^{\delta} RSK(u, v) \cdot I_{i+u, i+v}$$

여기서, $RSK(u, k)$ 는 Radial Symmetric gaussian Kernel, u 와 v 는 커널의 인덱스, δ 는 커널의 크기를 의미한다.

청구항 5

제 4 항에 있어서,

상기 커널의 크기(δ)는 상기 영상 데이터 조각의 길이를 결정하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 6

제 5 항에 있어서,

상기 영상 분할 단계는

사용자에 의해 미리 결정된 복수 개의 상기 커널의 크기에 따라 상기 영상 데이터를 복수 번 분할하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 7

제 1 항에 있어서,

상기 영상 분석 단계에서 상기 영상의 속도 정보는

상기 각 영상 데이터 조각 내에서 각 이미지의 특징점에 대한 속도 정보 또는 상기 각 영상 데이터 조각의 각 이미지에서 관심 영역(Region of interest)에 대한 속도 정보를 포함하는 것

을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 8

제 7 항에 있어서,

상기 영상 분석 단계에서 상기 특징점에 대한 속도 정보는

상기 각 영상 데이터 조각에 내에서 하나의 영상 프레임과 상기 하나의 영상 프레임과 이웃하는 영상 프레임간의 색상 차이 값(D^2)을 최소화하는 다음의 수학식에 의해 추출되는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

[수학식]

$$D^2 = \sum_{p, q \in \phi_{x, y}^i(p, q)} \left(\phi_{x, y}^i(p, q) - \phi_{(x, y) + \text{vec}_{x, y}^i}^{i+1}(p, q) \right)^2$$

여기서, D^2 은 색상 차이 값, ϕ 는 하나의 영상 프레임에서 검출된 상기 캐니 특징점 추출 기법에 의해 구해지는 에지 맵에서 상기 특징점에 대한 윈도우, i 는 상기 하나의 영상 프레임의 인덱스, p 와 q 는 상기 하나의 영상 프레임의 내부 좌표, x 와 y 는 상기 하나의 영상 프레임의 에지 픽셀 포인트, $\text{vec}_{x, y}^i$ 는 영상의 속도 정보이다.

청구항 9

제 7 항에 있어서,

상기 영상 분석 단계에서 상기 관심 영역에 대한 속도 정보는

이미지 공간에서 이웃하는 픽셀들 간의 차이 값인 공간적 집중도와 하나의 영상 프레임에서의 픽셀 위치와 상기 하나의 영상 프레임과 이웃하는 영상 프레임에서의 픽셀의 위치 차이 값인 시간적 집중도를 포함하는 집중도 맵에서 각 픽셀 속도를 합산함으로써 관심 영역에 대한 속도 정보를 추출하는 것

을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 10

제 9 항에 있어서,

상기 관심 영역에 대한 속도 정보는 다음의 수학식에 의해 추출되는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

[수학식]

$$V_{vel}^i = \sum_{x=1}^n \sum_{y=1}^m \|\overline{vec}_{x,y}^i\|$$

여기서, $\overline{vec}_{x,y}^i$ 는 관심 영역 내에서의 상기 $vec_{x,y}^i$, n은 관심 영역의 x 축의 크기, m은 관심 영역의 y 축의 크기이다.

청구항 11

제 1 항에 있어서,

상기 영상 분석 단계에서 상기 영상의 밝기 정보는 다음의 수학식에 의해 추출되는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

$$V_{bri}^i = \sum_{e=1}^{10} B(e)^2 B_{mean_e}$$

여기서, B(e)는 이미지의 밝기에 대한 히스토그램에서 e 레벨에 담겨진 픽셀의 숫자이고, B_{mean_e}는 e 레벨에 들어 있는 픽셀 밝기의 중간값이다.

청구항 12

제 1 항에 있어서,

상기 음악 분할 단계는

상기 순간적인 크기 변화를 추출하기 위한 전처리 과정으로 미리 결정된 단위 샘플 기간의 윈도우로 상기 입력된 음악 데이터를 나누는 단계;와

상기 윈도우로 나누어진 음악 데이터를 각 단위 기간별로 푸리에 변환하여 주파수 영역으로 변환하는 단계;

를 포함하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 13

제 12 항에 있어서,

상기 음악 분할 단계에서 상기 순간적인 진폭 변화는 다음의 수학식에 따라 주파수 영역으로 변환된 음악 데이터의 두 윈도우에서 진폭의 유사성을 측정하는 Novalty Scoring방법을 이용하여 추출하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

[수학식]

$$SM_{i,j} = \frac{A_i \cdot A_j}{||A_i|| ||A_j||}$$

여기서, A_i 는 주파수 영역에서 i 번째 윈도우 신호의 진폭, i 와 j 는 각각 서로 다른 윈도우, $SM_{i,j}$ 는 i 번째 윈도우와 j 번째 윈도우의 유사도 값을 의미한다.

청구항 14

제 13 항에 있어서,

상기 음악 변이점은

상기 유사도 값으로 이루어진 유사 행렬을 대각선 방향으로 Radial Symmetric gaussian Kernel(RSK) 적용하는 다음의 식에 의해 추출되는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

$$EA(i) = \sum_{u=-\delta}^{\delta} \sum_{v=-\delta}^{\delta} RSK(u, v) \cdot SM_{i+u, j+v}$$

여기서, $RSK(u, k)$ 는 Radial Symmetric gaussian Kernel, u 와 v 는 커널의 인덱스, δ 는 커널의 크기를 의미한다.

청구항 15

제 14 항에 있어서,

상기 커널의 크기는 상기 음악 데이터 조각의 길이를 결정하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 16

제 15 항에 있어서,

상기 음악 분할 단계는 사용자에게 의해 미리 결정된 복수 개의 상기 커널의 크기에 따라 상기 음악 데이터를 복수 번 분할하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 17

제 12 항에 있어서,

상기 음악의 속도 정보는

시간 영역에서 상기 음악 데이터 조각에 대한 상기 윈도우에 대한 평균 진폭값 또는 상기 각 윈도우에 대한 평균 진폭값의 변이량을 포함하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 18

제 17 항에 있어서,

상기 각 윈도우에 대한 평균 진폭 값은 다음의 수학식에 의해 추출되는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

[수학식]

$$RMS_i = \frac{1}{U} \sum_{u=1}^U (S_i(u))^2$$

여기서, RMS_i 는 i 번째 윈도우의 평균 진폭 값, U 는 총 샘플링 윈도우 개수, u 는 샘플링 윈도우 인덱스를 의미한다.

다.

청구항 19

제 17 항에 있어서,

상기 각 윈도우에 대한 평균 진폭 값의 변이량은 하나의 윈도우에 대한 평균 진폭 값과 이전 윈도우에 대한 평균 진폭 값의 차이인 다음의 수학식에 의해 추출되는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

[수학식]

$$M_{vel}^i = |RMS_i - RMS_{i-1}|$$

청구항 20

제 1 항에 있어서,

상기 음악의 밝기 정보는 상기 음악 데이터 조각의 주파수 영역에서의 진폭 값에 대한 정보인 스펙트럴 중심 (Spectral centroid) 방법인 다음의 수학식에 의해 추출되는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

[수학식]

$$M_{bri}^i = \frac{\sum p(A_i(p))^2}{\sum (A_i(p))^2}$$

여기서, $A_i(p)$ 는 주파수 영역에서 i 번째 윈도우 신호의 진폭, p 는 주파수 인덱스이다.

청구항 21

제 1 항에 있어서,

상기 데이터 매칭 단계는

상기 각각의 매칭 텀이 서로 대응되도록 하기 위해 상기 영상 및 음악 데이터 조각 각각 길이를 허미트 곡선 (Hermit curve)을 이용하여 정규화하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 22

제 1 항에 있어서,

상기 데이터 매칭 단계는

상기 영상 및 음악 데이터 조각을 매칭하기 위한 기준인 매칭 텀은

상기 영상의 속도 정보를 기반으로 한 영상 데이터 조각의 히스토그램과 상기 음악 데이터 조각에 주파수 영역에서의 진폭값을 기반으로 한 히스토그램을 매칭 기준으로 하는 분위기 매칭 텀; 및

상기 영상 및 음악 데이터 조각의 시간 길이 값을 매칭 기준으로 하는 시간 매칭 텀;

을 더 포함하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 23

제 22 항에 있어서,

상기 분위기 매칭 텀은 다음의 수학식에 의해 매칭값이 결정되는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

$$Hc(y, z) = \sum_{b=1}^K \left(\frac{VH^y(b)}{N_y} - \frac{MH^z(b)}{N_z} \right)^2$$

여기서, $VH^y(b)$ 는 영상의 속도 정보를 기반으로 한 y 번째 영상 데이터 조각의 히스토그램, $MH^z(b)$ 는 주파수 영역

에서 진폭값을 기반으로 한 z 번째 음악 데이터 조각의 히스토그램, y 와 z 는 영상 및 음악 데이터 조각의 인덱스, N_y 와 N_z 는 각 히스토그램을 정규화하기 위한 값으로 각 히스토그램 원소 개수의 합을 의미한다.

청구항 24

제 22 항에 있어서,

상기 데이터 매칭 단계는

상기 각각의 매칭 팀에 대한 가중치를 부여하는 다음의 수학적식을 통해 상기 영상 데이터 조각과 상기 음악 데이터 조각을 매칭하기 위한 최종 매칭값을 구하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

$$Cost_{y,z} = \sum_{i=1}^3 w_i Fc_i(V^y(t), M^z(t)) + w_4 Hc(y, z) + w_5 Dc(y, z)$$

여기서, y 와 z 는 데이터 조각의 인덱스, w_i 는 각 매칭 팀에 대한 가중치 값, $Fc_i(V^y(t), M^z(t))$ 는 각각 상기 영상 및 음악의 변이점을 기준으로 한 매칭 팀, 상기 영상 및 음악 데이터 조각의 속도 정보를 기준으로 한 매칭 팀, 상기 영상 및 음악 데이터 조각의 밝기 정보를 기준으로 한 매칭 팀을 의미하고, $Dc(y, z)$ 는 시간 매칭 팀을 의미한다.

청구항 25

제 1 항에 있어서,

상기 데이터 매칭 단계는

상기 영상 분할 단계에서의 미리 결정된 임계값이 복수 개로 설정되어 상기 영상 데이터를 복수 번 분할한 경우 상기 영상 데이터 조각이 가장 큰 조각부터 작은 조각의 순서로 상기 음악 데이터 조각과 매칭하는 것을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 26

제 25 항에 있어서,

상기 데이터 매칭 단계는

상기 상기 영상 데이터 조각이 가장 큰 것부터 작은 것의 순서로 상기 음악 데이터 조각과 매칭한 최종 매칭 값이 매칭 기준값 미만인 경우

상기 음악 분할 단계의 미리 결정된 임계값을 더 작은 값으로 상기 음악 데이터를 재분할하여 상기 복수 번 분할된 상기 각각의 영상 데이터 조각과 상기 재분할된 음악 데이터 조각과 매칭하는 것

을 특징으로 하는 자동 뮤직 비디오 생성 방법.

청구항 27

입력된 영상 데이터의 각 이미지의 외형 매칭(Contour shape matching)을 통한 상기 각 이미지 간의 외형 차이를 기준으로 상기 입력된 영상 데이터를 복수 번 분할하는 다중 영상 분할 단계와 입력된 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화를 기준으로 상기 입력된 음악 데이터를 분할하는 음악 분할 단계를 포함하는 데이터 분할 단계;

상기 다중 영상 분할 단계에 의해 복수 번 분할된 다중 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 영상 분석 단계와 상기 음악 분할 단계에 의해 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 음악 분석 단계를 포함하는 데이터 분석 단계; 및

상기 영상 및 음악의 변이점 또는 상기 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 팀이 서로 대응되도록 상기 다중 영상 데이터 조각 중 가장 큰 조각부터 작은 조각 순으로 상기 음악 데이터 조각과 매칭하는 데이터 매칭 단계;

를 포함하는 자동 뮤직 비디오 생성 방법.

청구항 28

제 27 항에 있어서,

상기 자동 뮤직 비디오 생성 방법은

상기 데이터 매칭 단계에서 상기 다중 영상 데이터 조각과 상기 음악 데이터 조각을 매칭하기 위한 매칭값이 매칭 기준 값 이하인 경우

상기 음악 데이터를 이전에 분할된 음악 데이터 조각보다 더 작은 데이터 조각으로 분할하는 음악 재분할 단계; 를 더 포함하는 것

청구항 29

제 1 항 내지 제 28 항 중 어느 한 항의 상기 자동 뮤직 비디오 생성 방법을 컴퓨터상에서 수행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체.

청구항 30

입력된 영상 데이터의 각 이미지의 외형 매칭(Contour shape matching)을 통한 상기 각 이미지 간의 외형 차이가 미리 결정된 임계값을 초과하는 영상 변이점을 기준으로 상기 입력된 영상 데이터를 분할하는 영상 분할부와 입력된 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화가 미리 결정된 임계값을 초과하는 음악 변이점을 기준으로 상기 입력된 음악 데이터를 분할하는 음악 분할부를 포함하는 데이터 분할부;

상기 영상 분할부에 의해 분할된 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 영상 분석부와 상기 음악 분할부에 의해 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 음악 분석부를 포함하는 데이터 분석부; 및

상기 영상 및 음악의 변이점 또는 상기 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 텀이 서로 대응되도록 상기 영상 데이터 조각과 음악 데이터 조각을 매칭하는 데이터 매칭부;

를 포함하는 자동 뮤직 비디오 생성 장치.

청구항 31

제 30 항에 있어서,

상기 데이터 매칭부는

상기 영상 및 음악 데이터 조각을 매칭하기 위한 기준인 매칭 텀은

상기 영상의 속도 정보를 기반으로 한 영상 데이터 조각의 히스토그램과 상기 음악 데이터 조각에 주파수 영역에서의 진폭값을 기반으로 한 히스토그램을 매칭 기준으로 하는 분위기 매칭 텀; 및

상기 영상 및 음악 데이터 조각의 시간 길이 값을 매칭 기준으로 하는 시간 매칭 텀;

을 더 포함하는 것을 특징으로 하는 자동 뮤직 비디오 생성 장치.

청구항 32

제 30 항에 있어서,

상기 데이터 매칭부는

상기 영상 분할부에서의 미리 결정된 임계값이 복수 개로 설정되어 상기 영상 데이터를 복수 번 분할한 경우

상기 영상 데이터 조각이 가장 큰 조각부터 작은 조각의 순서로 상기 음악 데이터 조각과 매칭하는 것

을 특징으로 하는 자동 뮤직 비디오 생성 장치.

청구항 33

입력된 영상 데이터의 각 이미지의 외형 매칭(Contour shape matching)을 통한 상기 각 이미지 간의 외형 차이를 기준으로 상기 입력된 영상 데이터를 복수 번 분할하는 다중 영상 분할부와 입력된 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화를 기준으로 상기 입력된 음악 데이터를 분할하는 음악 분할부를 포함하는 데이터 분할부;

상기 다중 영상 분할부에 의해 복수 번 분할된 다중 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 영상 분석부와 상기 음악 분할부에 의해 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 음악 분석부를 포함하는 데이터 분석부; 및

상기 영상 및 음악의 변이점 또는 상기 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 텀이 서로 대응되도록 상기 다중 영상 데이터 조각 중 가장 큰 조각부터 작은 조각 순으로 상기 음악 데이터 조각과 매칭하는 데이터 매칭부;

를 포함하는 자동 뮤직 비디오 생성 장치.

청구항 34

제 33 항에 있어서,

상기 자동 뮤직 비디오 생성 장치는

상기 데이터 매칭부에서 상기 다중 영상 데이터 조각과 상기 음악 데이터 조각을 매칭하기 위한 매칭값이 매칭 기준 값 이하인 경우

상기 음악 데이터를 이전에 분할된 음악 데이터 조각보다 더 작은 데이터 조각으로 분할하는 음악 재분할부;

를 더 포함하는 것을 특징으로 하는 자동 뮤직 비디오 생성 장치.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 자동 뮤직 비디오 생성 장치와 방법 이를 구현하기 위해 컴퓨터상에서 수행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체에 관한 것으로서, 보다 상세하게는, 본 발명은 영상과 음악을 다중 분할 한 후, 영상과 음악의 데이터 조각에 대한 속도, 밝기, 분위기에 대한 분석을 통해 영상 및 음악의 각 데이터 조각을 단계별로 매칭하여 뮤직 비디오를 생성하는 자동 뮤직 비디오 생성 장치와 방법 이를 구현하기 위해 컴퓨터상에서 수행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체에 관한 것이다.

배경 기술

[0002] 최근의 홈 비디오 시장의 확장에 따라, 전문성이 배제된 비디오 편집 기술이 요구되고 있다. 최근의 다양한 편집 킷의 개발을 통해 일반적인 사람들도 비디오 클립의 인위적인 배치를 통한 비디오 편집 기능은 쉽게 사용할 수 있다. 하지만 뮤직 비디오와 같이 영상과 음악의 연계성을 가지는 형태의 편집에 있어선 음악과 영상의 동기화를 만드는 전문적 기술이 필요하다.

[0003] 종래의 뮤직 비디오 제작 방식은 미리 만들어진 음악에 동기화되기 위한 영상 촬영을 위하여 전문적인 촬영 기술을 요구하였다. 음악과 영상의 분리된 작업환경의 제약 때문에 수많은 시행 착오를 요구하였고, 일반적인 홈 비디오 사용자에겐 동기화된 뮤직 비디오의 제작이 쉽지 않은 일이다.

[0004] 따라서, 종래의 영상과 음악의 동기화에 대한 연구는 대부분 주어진 음악에 맞도록 영상 데이터를 수정 또는 재조합하는 방법으로 진행되었다.

[0005] Foote는 음악의 반복적인 특성을 이용하여, 음악의 유사행렬을 계산하고 여러 조각으로 나눴다.

[0006] 또한 비디오의 밝기 변화와 카메라 움직임을 분석하여 비디오도 여러 조각으로 나눈 후, 각 조각의 변환점을 맞

추는 방법을 소개하였다.

- [0007] 비디오 조각 각각의 특징과 음악의 특징을 고려하는 방법은 Hua에 의해 시도되었다. Hua는 일반인이 찍은 홈비디오의 경우 화면의 질이 낮고 필요 없는 부분들이 많을 것이라는 전제를 바탕으로 물체의 동작, 카메라의 동작, 오디오 등을 토대로 비디오샷마다 집중도를 계산하여 중요한 샷만을 요약하는 방법을 소개하였다. 선택된 비디오 샷들은 박자의 세기를 기준으로 나누어진 음악 조각들의 빠르기에 맞도록 대응시킴으로써 동기화를 시도하였다.
- [0008] Mulhem은 비디오 편집 전문가들이 통상적으로 사용하는 몇 가지 미적인 규칙들을 사용하여 음악의 변화에 적절한 내용의 비디오 조각을 붙여 나가는 방법을 제시하였다.
- [0009] 이와 같은 종래의 방법들은 비디오 조각의 재조합을 통한 동기화를 시도하였다면, Jehan은 주어진 비디오의 재생 속도를 부분적으로 조정하면서 특징점의 동기화를 이루었다.
- [0010] 한편, 이와 같은 종래의 뮤직 비디오 생성 방법들은 비디오의 연속성을 유지하기 어렵고, 시청자가 집중하는 부분에 대한 동기화가 낮고, 역동적인 뮤직 비디오의 생성이 어려운 문제점이 있었다.

발명의 내용

해결 하고자하는 과제

- [0011] 본 발명은 자동 뮤직 비디오 생성을 위해 음악 데이터를 분할하고 영상 데이터를 다양한 크기 레벨의 영상 데이터 조각으로 복수 번 분할하고 영상과 음악 데이터 조각으로부터 각 조각의 길이, 속도, 밝기, 분위기, 변이점을 추출하여 영상 및 음악의 각 데이터 조각을 단계별로 매칭함으로써 뮤직 비디오의 장면 연속성이 최대한 유지되고, 시청자가 집중하는 부분의 동기성을 높여줄 뿐만 아니라 조각 데이터의 매칭을 위한 타임 와핑(Time-warping)을 통해 보다 역동적인 뮤직 비디오의 생성이 가능한 자동 뮤직 비디오 생성 장치와 방법 이를 구현하기 위해 컴퓨터상에서 수행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체를 제공하는데 그 목적이 있다.
- [0012] 본 발명의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있으며, 본 발명의 실시예에 의해 보다 분명하게 알게 될 것이다. 또한, 본 발명의 목적 및 장점들은 특허 청구 범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제 해결수단

- [0013] 본 발명에 따른 자동 뮤직 비디오 생성 방법의 일례는 입력된 영상 데이터의 각 이미지의 외형 매칭(Contour shape matching)을 통한 각 이미지 간의 외형 차이가 미리 결정된 임계값을 초과하는 영상 변이점을 기준으로 입력된 영상 데이터를 분할하는 영상 분할 단계와 입력된 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화가 미리 결정된 임계값을 초과하는 음악 변이점을 기준으로 입력된 음악 데이터를 분할하는 음악 분할 단계를 포함하는 데이터 분할 단계; 영상 분할 단계에 의해 분할된 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 영상 분석 단계와 음악 분할 단계에 의해 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 음악 분석 단계를 포함하는 데이터 분석 단계; 및 영상 및 음악의 변이점 또는 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 팀이 서로 대응되도록 영상 데이터 조각과 음악 데이터 조각을 매칭하는 데이터 매칭 단계;를 포함한다.
- [0014] 여기서, 영상 분할 단계에서 각 이미지 간의 외형 차이를 추출하기 위한 전처리 과정으로 영상 데이터의 각 이미지의 외형은 캐니 특징점 추출 기법(Canny edge detector)을 통해 이미지의 특징점으로 표현되는 에지 맵(Edge Map)으로 변환할 수 있다.
- [0015] 여기서, 각 이미지의 외형 차이는 다음의 수학식에 따라 두 영상 프레임에서 특징점의 유사성을 측정하여 추출할 수 있다.

[0016] [수학식]

$$I_{i,j} = \sum_{g=1}^7 |1/c_g^i - 1/c_g^j|$$

where

$$c_g^i = \text{sign}(h_g^i) \log_{10} |h_g^i|$$

[0017]

[0018] 여기서, i와 j는 영상 데이터의 프레임 번호, h는 Hu-moment로 에지 맵의 특징점으로부터 획득되는 7차원 벡터 값, g는 Hu-moment의 개수, I_{i,j}는 유사도 값을 의미한다.

[0019] 또한, 영상 변이점은 유사도 값으로 이루어진 유사 행렬을 대각선 방향으로 Radial Symmetric gaussian Kernel(RSK) 적용한 다음의 식에 의해 결정할 수 있다.

[0020] [수학식]

$$EV(i) = \sum_{u=-\delta}^{\delta} \sum_{v=-\delta}^{\delta} \text{RSK}(u, v) \cdot I_{i+u, i+v}$$

[0021]

[0022] 여기서, RSK(u,k)는 Radial Symmetric gaussian Kernel, u와 v는 커널의 인덱스, δ 는 커널의 크기를 의미한다.

[0023] 여기서, 커널의 크기(δ)는 영상 데이터 조각의 길이를 결정할 수 있다.

[0024] 또한, 영상 분할 단계는 사용자에게 의해 미리 결정된 복수 개의 커널의 크기에 따라 영상 데이터를 복수 번 분할할 수 있다.

[0025] 또한, 영상 분석 단계에서 영상의 속도 정보는 각 영상 데이터 조각 내에서 각 이미지의 특징점에 대한 속도 정보 또는 각 영상 데이터 조각의 각 이미지에서 관심 영역(Region of interest)에 대한 속도 정보를 포함할 수 있다.

[0026] 또한, 영상 분석 단계에서 특징점에 대한 속도 정보는 각 영상 데이터 조각에 내에서 하나의 영상 프레임과 하나의 영상 프레임과 이웃하는 영상 프레임간의 색상 차이 값(D^2)을 최소화하는 다음의 수학적식에 의해 추출할 수 있다.

[0027] [수학식]

$$D^2 = \sum_{p,q \in \phi_{x,y}^i(p,q)} \left(\phi_{x,y}^i(p,q) - \phi_{(x,y)+vec_{x,y}^i}^{i+1}(p,q) \right)^2$$

[0028]

[0029] 여기서, D^2 은 색상 차이 값, ϕ 는 하나의 영상 프레임에서 검출된 캐니 특징점 추출 기법에 의해 구해지는 에지 맵에서 특징점에 대한 윈도우, i는 하나의 영상 프레임의 인덱스, p와 q는 하나의 영상 프레임의 내부 좌표, x와 y는 하나의 영상 프레임의 에지 픽셀 포인트, $vec_{x,y}^i$ 는 영상의 속도 정보이다.

[0030] 또한, 영상 분석 단계에서 관심 영역에 대한 속도 정보는 이미지 공간에서 이웃하는 픽셀들 간의 차이 값인 공간적 집중도와 하나의 영상 프레임에서의 픽셀 위치와 하나의 영상 프레임과 이웃하는 영상 프레임에서의 픽셀의 위치 차이 값인 시간적 집중도를 포함하는 집중도 맵에서 각 픽셀 속도를 합산함으로써 관심 영역에 대한 속도 정보를 추출할 수 있다.

[0031] 또한, 관심 영역에 대한 속도 정보는 다음의 수학적식에 의해 추출할 수 있다.

[0032] [수학식]

$$V_{vel}^i = \sum_{x=1}^n \sum_{y=1}^m \|vec_{x,y}^i\|$$

[0033]

[0034] 여기서, $\overline{vec}_{x,y}^i$ 는 관심 영역 내에서의 $vec_{x,y}^i$, n은 관심 영역의 x 축의 크기, m은 관심 영역의 y 축의 크기이다.

[0035] 또한, 영상 분석 단계에서 영상의 밝기 정보는 다음의 수학적식에 의해 추출할 수 있다.

$$V_{bri}^i = \sum_{e=1}^{10} B(e)^2 B_{mean_e}$$

[0037] 여기서, B(e)는 이미지의 밝기에 대한 히스토그램에서 e 레벨에 담겨진 픽셀의 숫자이고, B_{mean_e}는 e 레벨에 들어 있는 픽셀 밝기의 중간값이다.

[0038] 또한, 음악 분할 단계는 순간적인 크기 변화를 추출하기 위한 전처리 과정으로 미리 결정된 단위 샘플 기간의 윈도우로 입력된 음악 데이터를 나누는 단계;와 윈도우로 나누어진 음악 데이터를 각 단위 기간별로 푸리에 변환하여 주파수 영역으로 변환하는 단계;를 포함할 수 있다.

[0039] 여기서, 음악 분할 단계에서 순간적인 진폭 변화는 다음의 수학적식에 따라 주파수 영역으로 변환된 음악 데이터의 두 윈도우에서 진폭의 유사성을 측정하는 Novalty Scoring방법을 이용하여 추출할 수 있다.

[0040] [수학적식]

$$SM_{i,j} = \frac{A_i \cdot A_j}{||A_i|| ||A_j||}$$

[0042] 여기서, A_i는 주파수 영역에서 i번째 윈도우 신호의 진폭, i와 j는 각각 서로 다른 윈도우, SM_{i,j}는 i번째 윈도우와 j번째 윈도우의 유사도 값을 의미한다.

[0043] 또한, 음악 변이점은 유사도 값으로 이루어진 유사 행렬을 대각선 방향으로 Radial Symmetric gaussian Kernel(RSK) 적용하는 다음의 식에 의해 추출할 수 있다.

$$EA(i) = \sum_{u=-\delta}^{\delta} \sum_{v=-\delta}^{\delta} RSK(u, v) \cdot SM_{i+u, j+v}$$

[0045] 여기서, RSK(u,k)는 Radial Symmetric gaussian Kernel, u와 v는 커널의 인덱스, δ 는 커널의 크기를 의미한다.

[0046] 또한, 커널의 크기는 음악 데이터 조각의 길이를 결정할 수 있다.

[0047] 또한, 음악 분할 단계는 사용자에게 의해 미리 결정된 복수 개의 커널의 크기에 따라 음악 데이터를 복수 번 분할할 수 있다.

[0048] 또한, 음악의 속도 정보는 시간 영역에서 음악 데이터 조각에 대한 윈도우에 대한 평균 진폭값 또는 각 윈도우에 대한 평균 진폭값의 변이량을 포함할 수 있다.

[0049] 또한, 각 윈도우에 대한 평균 진폭 값은 다음의 수학적식에 의해 추출할 수 있다.

[0050] [수학적식]

$$RMS_i = \frac{1}{U} \sum_{u=1}^U (S_i(u))^2$$

[0052] 여기서, RMS_i는 i번째 윈도우의 평균 진폭 값, U는 총 샘플링 윈도우 개수, u는 샘플링 윈도우 인덱스를 의미한다.

[0053] 또한, 각 윈도우에 대한 평균 진폭 값의 변이량은 하나의 윈도우에 대한 평균 진폭 값과 이전 윈도우에 대한 평균 진폭 값의 차이인 다음의 수학적식에 의해 추출할 수 있다.

[0054] [수학식]

[0055]
$$M_{vel}^i = |RMS_i - RMS_{i-1}|$$

[0056] 또한, 음악의 밝기 정보는 음악 데이터 조각의 주파수 영역에서의 진폭 값에 대한 정보인 스펙트럴 중심 (Spectral centroid) 방법인 다음의 수학식에 의해 추출할 수 있다.

[0057] [수학식]

[0058]
$$M_{bri}^i = \frac{\sum p(A_i(p))^2}{\sum (A_i(p))^2}$$

[0059] 여기서, $A_i(p)$ 는 주파수 영역에서 i 번째 윈도우 신호의 진폭, p 는 주파수 인덱스이다.

[0060] 또한, 데이터 매칭 단계는 각각의 매칭 팀이 서로 대응되도록 하기 위해 영상 및 음악 데이터 조각 각각 길이를 허미트 곡선(Hermit curve)을 이용하여 정규화할 수 있다.

[0061] 또한, 데이터 매칭 단계는 영상 및 음악 데이터 조각을 매칭하기 위한 기준인 매칭 팀은 영상의 속도 정보를 기반으로 한 영상 데이터 조각의 히스토그램과 음악 데이터 조각에 주파수 영역에서의 진폭값을 기반으로 한 히스토그램을 매칭 기준으로 하는 분위기 매칭 팀; 및 영상 및 음악 데이터 조각의 시간 길이 값을 매칭 기준으로 하는 시간 매칭 팀;을 더 포함할 수 있다.

[0062] 또한, 분위기 매칭 팀은 다음의 수학식에 의해 매칭값이 결정할 수 있다.

[0063]
$$Hc(y, z) = \sum_{b=1}^K \left(\frac{VH^y(b)}{N_y} - \frac{MH^z(b)}{N_z} \right)^2$$

[0064] 여기서, $VH^y(b)$ 는 영상의 속도 정보를 기반으로 한 y 번째 영상 데이터 조각의 히스토그램, $MH^z(b)$ 는 주파수 영역에서 진폭값을 기반으로 한 z 번째 음악 데이터 조각의 히스토그램, y 와 z 는 영상 및 음악 데이터 조각의 인덱스, N_y 와 N_z 는 각 히스토그램을 정규화하기 위한 값으로 각 히스토그램 원소 개수의 합을 의미한다.

[0065] 또한, 데이터 매칭 단계는 각각의 매칭 팀에 대한 가중치를 부여하는 다음의 수학식을 통해 영상 데이터 조각과 음악 데이터 조각을 매칭하기 위한 최종 매칭값을 구할 수 있다.

[0066]
$$Cost_{y,z} = \sum_{i=1}^3 w_i Fc_i(V^y(t), M^z(t)) + w_4 Hc(y, z) + w_5 Dc(y, z)$$

[0067] 여기서, y 와 z 는 데이터 조각의 인덱스, w_i 는 각 매칭 팀에 대한 가중치 값, $Fc_i(V^y(t), M^z(t))$ 는 각각 영상 및 음악의 변이점을 기준으로 한 매칭 팀, 영상 및 음악 데이터 조각의 속도 정보를 기준으로 한 매칭 팀, 영상 및 음악 데이터 조각의 밝기 정보를 기준으로 한 매칭 팀을 의미하고, $Dc(y, z)$ 는 시간 매칭 팀을 의미한다.

[0068] 또한, 데이터 매칭 단계는 영상 분할 단계에서의 미리 결정된 임계값이 복수 개로 설정되어 영상 데이터를 복수 번 분할한 경우 영상 데이터 조각이 가장 큰 조각부터 작은 조각의 순서로 음악 데이터 조각과 매칭할 수 있다.

[0069] 또한, 데이터 매칭 단계는 영상 데이터 조각이 가장 큰 것부터 작은 것의 순서로 음악 데이터 조각과 매칭한 최종 매칭 값이 매칭 기준값 미만인 경우 음악 분할 단계의 미리 결정된 임계값을 더 작은 값으로 음악 데이터를 재분할하여 복수 번 분할된 각각의 영상 데이터 조각과 재분할된 음악 데이터 조각과 매칭할 수 있다.

[0070] 또한, 본 발명에 따른 자동 뮤직 비디오 생성 방법의 다른 일례는 입력된 영상 데이터의 각 이미지의 외형 매칭 (Contour shape matching)을 통한 각 이미지 간의 외형 차이를 기준으로 입력된 영상 데이터를 복수 번 분할하는 다중 영상 분할 단계와 입력된 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화를 기준으로 입력된 음악 데이터를 분할하는 음악 분할 단계를 포함하는 데이터 분할 단계; 다중 영상 분할 단계에 의해 복수 번 분할된 다중 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 영상 분석 단계와 음악 분할 단계에 의해 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 음악 분석 단계를 포함하는 데이터 분석 단계; 및 영상 및 음악의 변이점 또는 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 팀이 서로 대응되도록 다중 영상 데이터 조각 중 가장 큰 조각부터 작은 조각 순으로

음악 데이터 조각과 매칭하는 데이터 매칭 단계;를 포함한다.

- [0071] 여기서, 자동 뮤직 비디오 생성 방법은 데이터 매칭 단계에서 다중 영상 데이터 조각과 음악 데이터 조각을 매칭하기 위한 매칭값이 매칭 기준 값 이하인 경우 음악 데이터를 이전에 분할된 음악 데이터 조각보다 더 작은 데이터 조각으로 분할하는 음악 재분할 단계;를 더 포함할 수 있다.
- [0072] 또한, 본 발명에 따른 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체는 전술한 자동 뮤직 비디오 생성 방법을 컴퓨터상에서 수행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체를 포함한다.
- [0073] 또한, 본 발명에 따른 자동 뮤직 비디오 생성 장치의 일례는 입력된 영상 데이터의 각 이미지의 외형 매칭 (Contour shape matching)을 통한 각 이미지 간의 외형 차이가 미리 결정된 임계값을 초과하는 영상 변이점을 기준으로 입력된 영상 데이터를 분할하는 영상 분할부와 입력된 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화가 미리 결정된 임계값을 초과하는 음악 변이점을 기준으로 입력된 음악 데이터를 분할하는 음악 분할부를 포함하는 데이터 분할부; 영상 분할부에 의해 분할된 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 영상 분석부와 음악 분할부에 의해 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 음악 분석부를 포함하는 데이터 분석부; 및 영상 및 음악의 변이점 또는 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 텀이 서로 대응되도록 영상 데이터 조각과 음악 데이터 조각을 매칭하는 데이터 매칭부;를 포함한다.
- [0074] 여기서, 데이터 매칭부는 영상 및 음악 데이터 조각을 매칭하기 위한 기준인 매칭 텀은 영상의 속도 정보를 기반으로 한 영상 데이터 조각의 히스토그램과 음악 데이터 조각에 주파수 영역에서의 진폭값을 기반으로 한 히스토그램을 매칭 기준으로 하는 분위기 매칭 텀; 및 영상 및 음악 데이터 조각의 시간 길이 값을 매칭 기준으로 하는 시간 매칭 텀을 더 포함할 수 있다.
- [0075] 또한, 데이터 매칭부는 영상 분할부에서의 미리 결정된 임계값이 복수 개로 설정되어 영상 데이터를 복수 번 분할한 경우 영상 데이터 조각이 가장 큰 조각부터 작은 조각의 순서로 음악 데이터 조각과 매칭할 수 있다.
- [0076] 또한, 본 발명에 따른 자동 뮤직 비디오 생성 장치의 다른 일례는 입력된 영상 데이터의 각 이미지의 외형 매칭 (Contour shape matching)을 통한 각 이미지 간의 외형 차이를 기준으로 입력된 영상 데이터를 복수 번 분할하는 다중 영상 분할부와 입력된 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화를 기준으로 입력된 음악 데이터를 분할하는 음악 분할부를 포함하는 데이터 분할부; 다중 영상 분할부에 의해 복수 번 분할된 다중 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 영상 분석부와 음악 분할부에 의해 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 음악 분석부를 포함하는 데이터 분석부; 및 영상 및 음악의 변이점 또는 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 텀이 서로 대응되도록 다중 영상 데이터 조각 중 가장 큰 조각부터 작은 조각 순으로 음악 데이터 조각과 매칭하는 데이터 매칭부;를 포함한다.
- [0077] 여기서, 자동 뮤직 비디오 생성 장치는 데이터 매칭부에서 다중 영상 데이터 조각과 음악 데이터 조각을 매칭하기 위한 매칭값이 매칭 기준 값 이하인 경우 음악 데이터를 이전에 분할된 음악 데이터 조각보다 더 작은 데이터 조각으로 분할하는 음악 재분할부;를 더 포함할 수 있다.

효 과

- [0078] 본 발명에 따른 자동 뮤직 비디오 생성 장치와 방법 이를 구현하기 위해 컴퓨터상에서 수행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체는 자동 뮤직 비디오 생성을 위해 음악 데이터를 분할하고 영상 데이터를 다양한 크기 레벨의 영상 데이터 조각으로 복수 번 분할하고 영상과 음악 데이터 조각으로부터 각 조각의 길이, 속도, 밝기, 분위기, 변이점을 추출하여 영상 및 음악의 각 데이터 조각을 단계별로 매칭함으로써 뮤직 비디오의 장면 연속성이 최대한 유지되고, 시청자가 집중하는 부분의 동기성을 높여줄 뿐만 아니라 조각 데이터의 매칭을 위한 타임 와핑(Time-warping)을 통해 보다 역동적인 뮤직 비디오의 생성이 가능하도록 하는 효과가 있다.

발명의 실시를 위한 구체적인 내용

- [0079] 이하, 본 발명의 바람직한 실시예를 첨부된 도면들을 참조하여 상세히 설명한다.
- [0080] 우선 각 도면의 구성 요소들에 참조 부호를 부가함에 있어서, 동일한 구성 요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다. 또한, 이하에서 본 발명의 바람직한 실시예를 설명할 것이나, 본 발명의 기술적 사상은 이에 한정하거나 제한되지 않고 당업자에 의해 변형되어 다양하게 실시될 수 있음은 물론이다.
- [0081] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세히 설명하고자 한다.
- [0082] 도 1은 본 발명에 따른 자동 뮤직 비디오 생성 장치의 일례를 설명하기 위한 도이다.
- [0083] 도시된 바와 같이, 자동 뮤직 비디오 생성 장치의 일례는 데이터 분할부(100), 데이터 분석부(200), 데이터 매칭부(300)를 포함하고, 음악 재분할부(400)를 더 포함할 수 있다.
- [0084] 데이터 분할부(100)는 입력된 영상 데이터를 다중 분할하는 영상 다중 분할부(110)와 입력된 음악 데이터를 분할하는 음악 분할부(120)를 포함한다.
- [0085] 데이터 분석부(200)는 분할된 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 분석하는 영상 분석부(210)와 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 분석하는 음악 분석부(220)를 포함한다.
- [0086] 데이터 매칭부(300)는 영상 데이터 조각과 음악 데이터 조각의 매칭을 위하여 추출된 영상과 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 팀이 서로 대응되도록 영상 데이터 조각과 음악 데이터 조각을 서로 매칭한다.
- [0087] 이하에서는, 전술한 데이터 분할부(100), 데이터 분석부(200), 데이터 매칭부(300)를 보다 상세히 설명한다.
- [0088] 영상 다중 분할부(110)는 영상 데이터의 각 이미지의 외형 매칭(Contour shape matching)을 통한 각 이미지 간의 외형 차이가 미리 결정된 임계값을 초과하는 영상 변이점을 기준으로 입력된 영상 데이터를 분할하는 기능을 한다.
- [0089] 이와 같이 영상 다중 분할부(110)가 영상 데이터를 분할하는 것은 프레임간의 유사성을 기준으로 한다. 여기서, 프레임간의 유사성이란 두 개의 이미지가 얼마나 비슷한 색을 가지고 있는냐로 구분된다. 하지만 움직이고 있는 영상에서는 단순히 같은 좌표의 색상차이만으로는 유사성을 따지기 힘들다.
- [0090] 따라서, 본 발명에서는 각 이미지의 외형 매칭(Contour shape matching)을 통한 유사성 측정을 통해 영상 데이터를 분할하는 것이다. 임의의 N개의 이미지로 이루어진 영상 조각 $V_i (i=1, \dots, N)$ 로 이루어졌을 때, 이와 같이 영상 다중 분할부(110)에서 각 이미지 간의 외형 차이를 추출하기 위한 전처리 과정으로 영상 데이터의 각 이미지의 외형은 캐니 특징점 추출 기법(Canny edge detector)을 통해 이미지의 특징점인 F_i 로 표현되는 에지 맵(Edge Map)으로 변환될 수 있다. 여기서, 노이즈로 작용할 수 있는 작은 외곽선을 방지하기 위해 영상 클립을 미리 가우시안 필터링 할 수 있다.
- [0091] 이미지의 특징점인 F_i 는 이미지의 외형을 이루는 픽셀 단위의 특징점들이 모인 외곽선으로 이루어져 있다. 외곽선으로부터 얻어지는 7개의 Hu-moment를 $h_u^i (u=1, \dots, 7)$ 이라 했을 때, 각 이미지의 외형 차이는 다음의 수학적식 1에 따라 두 영상 프레임에서 특징점의 유사성을 측정하여 추출할 수 있다.

수학적식 1

$$I_{i,j} = \sum_{g=1}^7 |1/c_g^i - 1/c_g^j|$$

where

$$c_g^i = \text{sign}(h_g^i) \log_{10} |h_g^i|$$

[0092]

[0093] 여기서, i와 j는 영상 데이터의 프레임 번호, h는 Hu-moment로 에지 맵의 특징점으로부터 획득되는 7차원 벡터

값, g 는 Hu-moment의 개수, $I_{i,j}$ 는 유사도 값을 의미한다.

[0094] 이와 같은 수학적 식 1에서 Hu-moment는 이동, 회전, 그리고 크기에 독립적이기 때문에 움직이는 영상의 흐름이 끊기지 않는다면 각 프레임에서 이미지의 외형은 서로 유사하다고 판정을 한다.

[0095] 이와 같은 각 이미지의 외형 차이값들의 집합으로 이루어지는 유사행렬 $I_{i,j}$ 는 도 2a의 (a)와 같이 표현될 수 있다. 이와 같은 도 2의 (a)에서 색이 동일한 경우 외형 차이가 없는 것으로 표현되고, 색이 변하는 지점이 외형 차이가 발생하는 경계로 표현된다.

[0096] 따라서, 유사행렬 $I_{i,j}$ 의 대각선 방향은 동일한 두 프레임의 외형 차이를 측정한 것이므로 항상 동일한 색을 가지게 된다.

[0097] 이와 같이 유사도 값으로 이루어진 유사 행렬을 대각선 방향으로 Radial Symmetric gaussian Kernel(RSK) 적용한 다음의 수학적 식 2에 의해 영상의 변이점이 결정된다.

수학적 식 2

$$EV(i) = \sum_{u=-\delta}^{\delta} \sum_{v=-\delta}^{\delta} RSK(u, v) \cdot I_{i+u, i+v}$$

[0098] 여기서, $RSK(u, k)$ 는 도 2b와 같은 형태의 Radial Symmetric gaussian Kernel(RSK), u 와 v 는 커널의 인덱스,

δ 는 커널의 크기를 의미한다.

[0100] 이와 같은 수학적 식 2에 의해 얻어진 영상 변이점 $EV(i)$ 가 미리 결정된 임계값을 초과하는 경우, 영상 변이점이 되는 프레임의 전후로 이미지 간의 외형 차이가 있어 영상의 흐름이 끊기는 것을 의미하고, 이와 같은 영상 변이점이 영상 데이터의 분할 기준이 되는 것이다.

[0101] 이와 같은 영상 변이점을 결정할 때 결정의 기준이 되는 값은 커널의 크기(δ)값이다. 따라서 이와 같은 커널(δ)값을 변화시키면 분할의 정도를 조절할 수 있어 영상 데이터 조각의 길이를 결정할 수 있는 것이다.

[0102] 보다 상세하게는, 만약 큰 커널(δ)값을 사용하면 짧은 변화는 영상 데이터로부터 무시되어 긴 길이를 가지는 영상 데이터 조각을 분할/생성할 수 있고, 상대적으로 작은 커널(δ)값을 사용하면 짧은 변화에 민감하게 작용하므로 영상 데이터로부터 상대적으로 짧은 길이를 가지는 영상 데이터 조각을 분할/생성할 수 있는 것이다.

[0103] 예를 들어 커널의 크기 값에 따라 영상 데이터를 다중 분할한 결과의 일례가 도시된 도 2a의 (b)를 참조하면, 1 번째 레벨의 커널(δ)값은 128인 경우이고, 2 번째 레벨의 커널(δ)값은 64, 3 번째 레벨의 커널(δ)값은 32이다.

[0104] 여기서, 커널(δ)값이 128인 1 번째 레벨의 경우 128보다 작은 영상 변이점은 무시되어 가장 큰 영상 데이터 조각으로 분할되는 것이고, 커널(δ)값이 64인 2 번째 레벨의 경우 64보다 작은 영상 변이점만 무시되므로 중간 크기의 영상 데이터 조각으로 분할되는 것이고, 커널(δ)값이 32인 3 번째 레벨의 경우 32보다 작은 영상 변이점만 무시되므로 가장 작은 영상 데이터 조각으로 분할되는 것이다.

[0105] 여기서, 영상 다중 분할부(110)는 전술한 도 2a의 (b)와 같은 예처럼 입력된 영상 데이터를 분할하는 기준이 되는 임계값을 복수 개 설정된 경우, 각 임계값의 크기 레벨 별로 입력된 영상 데이터를 복수 번 분할할 수 있는 것이다. 이는 보다 다양한 크기 레벨의 영상 데이터 조각을 얻을 수 있도록 하고 데이터 매칭 단계에서 보다 정밀한 매칭이 가능하게 하여, 뮤직 비디오의 영상 연속성을 매우 양호하게 유지할 수 있는 효과가 있는 것이다.

[0106] 다음, 영상 분석부(210)는 분할된 영상 데이터 조각으로부터 영상의 속도 정보 또는 밝기 정보를 추출하는 기능을 한다.

[0107] 영상의 속도 정보는 각 영상 데이터 조각 내에서 각 이미지의 특징점에 대한 속도 정보 또는 각 영상 데이터 조각의 각 이미지에서 관심 영역(Region of interest)에 대한 속도 정보를 포함할 수 있다.

[0108] 영상 분석부(210)에서 특징점에 대한 속도 정보를 추출하기 위해 특징점 위의 한 점에 대하여 w 의 크기를 가지

는 윈도우 $\phi_{x,y}(p,q)$ (p,q 는 윈도우 내에서의 좌표)를 선언해 준다. 이후, 특징점에 대한 속도 정보는 각 영상 데이터 조각에 내에서 하나의 영상 프레임과 하나의 영상 프레임과 이웃하는 영상 프레임간의 색상 차이 값(D^2)을 최소화하는 다음의 수학적식에 의해 추출될 수 있다.

수학적식 3

$$D^2 = \sum_{p,q \in \phi_{x,y}^i(p,q)} \left(\phi_{x,y}^i(p,q) - \phi_{(x,y)+vec_{x,y}^i}^{i+1}(p,q) \right)^2$$

[0109]

[0110] 여기서, D^2 은 색상 차이 값, Φ 는 하나의 영상 프레임에서 검출된 캐니 특징점 추출 기법에 의해 구해지는 에지 맵에서 특징점에 대한 윈도우, i 는 하나의 영상 프레임의 인덱스, p 와 q 는 하나의 영상 프레임의 내부 좌표, x 와 y 는 하나의 영상 프레임의 에지 픽셀 포인트, $vec_{x,y}^i$ 는 영상의 속도 정보이다.

[0111] 이와 같은 수학적식 3에서 i 번째 프레임과 $i+1$ 번째 프레임의 색상차이인 D^2 을 최소화하는 벡터 $vec_{x,y}^i$ 가 특징점 위의 한 점(x, y)에 대한 움직임 벡터가 된다. 윤곽선이 아닌 부분의 움직임은 무시하기 위해 $F_i(x,y)=0$ 인 픽셀에서는 벡터를 (0,0)으로 고정해 준다. 보다 안정적인 결과를 위하여 지역적인 Lucas-Kanade 기법을 적용하여 결과를 보완하였다.

[0112] 특징점 이미지 F_i 를 사용한 속도 측정의 장점은 특징점 위의 점끼리만 비교하기 때문에 측정에 있어서 노이즈의 영향을 작게 받는다는 점이다. 하지만, 시청자가 쉽게 집중하지 않는 배경에서도 특징점이 생성될 수 있고, 특징점의 속도 측정에 영향을 줌으로써 분석의 정확성이 떨어질 위험이 있다.

[0113] 따라서, 본 발명에서는 집중도 맵(Saliency Map)기반의 이미지 집중적 탐색 기법을 통해 관심 영역에 대한 속도 정보를 강조한다.

[0114] 영상의 집중도 맵을 구성하기 위해서 두 가지 측면을 고려한다. 첫째는 하나의 이미지 상의 이웃점과의 차이를 고려한 공간적 집중도이고, 다른 하나는 프레임간의 집중도를 고려한 시간적 집중도이다.

[0115] 공간적 집중도는 가우시안 거리(Gaussian distance)를 기반으로 계산한다. 이와 같은 가우시안 거리의 공식은 다음의 수학적식 4와 같이 나타낼 수 있다.

수학적식 4

$$G_{l,\eta}^i(x,y) = G_l^i(x,y) - G_{l+\eta}^i(x,y)$$

[0116]

[0117] 여기서, G_l^i 는 i 번째 프레임에서 가우시안 피라미드 상의 l 번째 레벨을 의미한다. 이와 같은 이미지로부터 얻어진 가우시안 피라미드에서 단계별 차이를 통해 이미지에서 집중되는 부분이 어디인지 검출해 낼 수 있는 것이다. 즉 낮은 주기 데이터와 높은 주기 데이터의 차이를 통해 상대적으로 변이가 많은 부분을 찾을 수 있다.

[0118] 집중 정도는 밝기에 가장 영향을 많이 받으므로 각각의 프레임 V_i 의 색상을 YUV 공간으로 변환 한 후 수학적식 4에 대입한다.

[0119] 시간의 집중도는 특정 위치에서 픽셀의 움직임이 앞뒤 프레임과 얼마나 차이가 나는지의 고려를 통해 추출할 수

있다.

[0120] 이와 같은 시간의 집중도는 수학식 3에서 추출한 특징점에 대한 영상의 속도 정보 $vec_{x,y}^i$ 를 통해 다음의 수학식 5와 같이 표현할 수 있다.

수학식 5

$$T^i(x, y) = N \left(\|vec_{x,y}^i - vec_{x,y}^{i+1}\| \right)$$

[0121]

[0122] 여기서, N은 데이터 정규화를 의미하고, i 번째 프레임에서 (x,y) 위치의 픽셀의 가속도를 뜻한다. 가속도의 크기가 클수록 i 번째 프레임의 집중성이 강해진다고 가정한다.

[0123] 여기서, 더 고려해야 할 점은 카메라의 움직임 정보이다. 일반적인 촬영에 있어서 카메라를 움직인다는 것은 주요 물체를 추적한다는 뜻이 강하다.

[0124] 예를 들어 정지된 카메라의 경우는 가속도가 클수록 그것이 집중도가 강한 부분이라 볼 수 있지만, 카메라가 움직일 경우는 오히려 가속도가 작은 부분, 즉 카메라가 모션을 쫓아가는 부분이 더 큰 집중도를 가진다고 볼 수 있다. 따라서 본 발명은 ITM 방법을 통해 카메라의 움직임 양을 측정한다.

[0125] 영상의 떨림을 고려하여 4픽셀 이상의 카메라의 움직임이 있을 때에는 $T_s^i(x,y)$ 가 아닌 $1-T_s^i(x,y)$ 를 적용한다.

[0126] 이와 같이 수학식 4와 5를 통해 얻어진 두 개의 집중도를 통해 사용자의 주의를 끄는 관심 영역을 추출할 수 있는 것이다.

[0127] 이와 같이 얻어진 집중도 정도를 기준으로 다음의 수학식 6과 같은 식을 통해 i 번째 프레임에서의 포커스 x_f^i , y_f^i 를 찾을 수가 있는 것이다.

수학식 6

$$x_f^i = \frac{1}{CM} \sum_{x=1}^n \sum_{y=1}^m G^i(x, y) T^i(x, y) x$$

$$y_f^i = \frac{1}{CM} \sum_{x=1}^n \sum_{y=1}^m G^i(x, y) T^i(x, y) y$$

[0128]

[0129] 우리는 포커스를 기준으로 전체 이미지 크기의 1/4 크기만큼의 시야를 관심 영역으로 정의하여, 관심 영역 안에 들어오는 벡터 $vec_{x,y}^i$ 만 $\overline{vec}_{x,y}^i$ 에 저장한다.

[0130] 이와 같은 수학식 4 내지 6을 통해 구해진 이미지 공간에서 이웃하는 픽셀들 간의 차이 값인 공간적 집중도와 하나의 영상 프레임에서의 픽셀 위치와 하나의 영상 프레임과 이웃하는 영상 프레임에서의 픽셀의 위치 차이 값인 시간적 집중도를 포함하는 집중도 맵에서 각 픽셀 속도를 합산함으로써 관심 영역에 대한 속도 정보를 추출하게 된다.

[0131] 이와 같은 관심 영역에 대한 속도 정보는 다음의 수학식 7에 의해 추출된다.

수학식 7

$$V_{vel}^i = \sum_{x=1}^n \sum_{y=1}^m \|\overline{vec}_{x,y}^i\|$$

[0132]

[0133] 여기서, $\overline{vec}_{x,y}^i$ 는 관심 영역 내에서의 $vec_{x,y}^i$, n은 관심 영역의 x 축의 크기, m은 관심 영역의 y 축의 크기이

다.

[0134] 도 3은 특징점 기반의 비디오 속도 분석의 일례이다. 도 3의 (a)는 분할된 영상 데이터 조각이고, (b)는 분할된 영상 데이터 조각의 이미지에 대한 특징점 추출 결과, (c)는 움직임 벡터의 크기를 나타내는 맵, (d)는 집중도 기반의 관심 영역에 대한 일례, (e)는 관심 영역 외부의 움직임 벡터가 제거된 움직임 벡터 맵이다.

[0135] 다음, 영상의 밝기 정보 추출부는 이미지의 밝기에 대한 히스토그램을 이용하여 영상의 밝기 정보를 추출한다. 이미지의 밝기 정보는 프레임을 이루고 있는 전체 픽셀의 밝기 분포를 통해 얻어진다. 본 발명은 이를 위해 히스토그램 분석 방법을 사용하여 영상의 밝기 정보를 추출한다.

[0136] 이를 위해 우선 V_i 를 흑백 레벨로 변환한 뒤 밝기 히스토그램을 분석한다. 일례로 히스토그램을 10개의 레벨로 나누었을 때 프레임을 대표하는 밝기 정보는 수학적 식 8에 의해 추출된다.

수학적 식 8

$$V_{bri}^i = \sum_{e=1}^{10} B(e)^2 B_{mean_e}$$

[0137] 여기서, $B(e)$ 는 이미지의 밝기에 대한 히스토그램에서 e 레벨에 담겨진 픽셀의 숫자이고, B_{mean_e} 는 e 레벨에 들어 있는 픽셀 밝기의 중간값이다.

[0139] 실제 이미지에서 전체적으로 중간 밝기를 가지고 있는 이미지보다 한 쪽은 밝고 다른 한쪽은 어두운 이미지가 더 밝은 이미지를 준다. 따라서 $B(u)$ 의 제곱항을 두어, 하나의 이미지에서 히스토그램의 레벨 차이가 크게 날수록 더 큰 밝기 값을 가지도록 한 것이다.

[0140] 분석된 영상과의 매칭을 위해서는 음악 역시 데이터 조각 단위로 잘라서 분석을 해야한다. 그러나 웨이브 형식에서 얻을 수 있는 정보는 시간에 대한 음악 데이터의 진폭 정보만 얻을 수 있다. 따라서, 주어진 진폭에 대한 신호 분석을 통해 음악을 분할하고 속도와 밝기를 측정해야 한다.

[0141] 이를 위해 음악 분할부(120)는 입력된 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화가 미리 결정된 임계값을 초과하는 음악 변이점을 기준으로 입력된 음악 데이터를 분할하는 기능을 한다.

[0142] 이와 같이 음악 분할부(120)가 순간적인 크기 변화량을 추출하기 위해서 전처리 과정으로 미리 결정된 단위 샘플 기간의 윈도우로 입력된 음악 데이터를 나누는 단계와 윈도우로 나누어진 음악 데이터를 각 단위 기간별로 푸리에 변환하여 주파수 영역으로 변환하는 단계를 포함한다.

[0143] 그리고, 음악 분할 단계에서 순간적인 진폭 변화는 다음의 수학적 식 9에 따라 주파수 영역으로 변환된 음악 데이터의 두 윈도우에서 진폭의 유사성을 측정하는 Novalty Scoring 방법을 이용하여 추출한다.

수학적 식 9

$$SM_{i,j} = \frac{A_i \cdot A_j}{||A_i|| ||A_j||}$$

[0144] 여기서, A_i 는 주파수 영역에서 i 번째 윈도우 신호의 진폭, i 와 j 는 각각 서로 다른 윈도우, $SM_{i,j}$ 는 i 번째 윈도우와 j 번째 윈도우의 유사도 값을 의미한다.

[0146] 그리고, 음악 변이점은 수학적 식 9에 의해 구해진 유사도 값으로 이루어진 유사 행렬을 대각선 방향으로 Radial Symmetric gaussian Kernel(RSK) 적용하는 다음의 수학적 식 10에 의해 추출된다.

수학식 10

$$EA(i) = \sum_{u=-\delta}^{\delta} \sum_{v=-\delta}^{\delta} RSK(u, v) \cdot SM_{i+u, j+v}$$

[0147]

[0148] 여기서, $RSK(u, k)$ 는 Radial Symmetric gaussian Kernel, u 와 v 는 커널의 인덱스, δ 는 커널의 크기를 의미한다. 이와 같은 커널은 영상 분할에서와 같이 크기를 조절하여 음악 데이터 조각의 길이를 결정할 수 있다.

[0149] 이와 같이 음악 분할 단계는 사용자에게 의해 미리 결정된 복수 개의 커널의 크기에 따라 음악 데이터를 복수 번 분할할 수 있는 것이다.

[0150] 도 4는 주파수 영역에서 유사도 분석을 통한 Novalty Scoring 생성 방법의 일례를 보여준다.

[0151] (a)는 주파수 영역으로 변환된 음악 데이터의 유사도 행렬의 일례이고, (b)는 서로 다른 크기의 커널을 가진 RSK를 사용하여 Novalty Score를 분석한 결과의 일례이다.

[0152] 다음, 음악 분석부(220)는 분할된 음악 데이터 조각으로부터 음악의 속도 정보 또는 밝기 정보를 추출하는 기능을 한다.

[0153] Novalty Score가 음악 변이점을 나타내긴 하지만, 음악의 전환점에서만 큰 값을 가지므로 음악의 속도를 표현한다고 보기는 어려우므로 웨이브의 진폭을 사용하여 음악의 속도 정보를 분석한다.

[0154] 이와 같은 음악의 속도 정보는 시간 영역에서 음악 데이터 조각에 대한 각 윈도우에 대한 평균 진폭값 또는 각 윈도우에 대한 평균 진폭값의 변이량을 포함할 수 있다.

[0155] 일반적으로 음악에서의 속도는 비트를 기반으로 하나, 음악의 속도 정보를 추출하기 위해서는 정확한 비트를 추적하는 것보다는 전체적인 음악의 빠르기, 즉 속도를 알아야 한다. 따라서 신호의 진폭을 사용한 분석을 통해 음악의 속도 정보를 추출한다.

[0156] 음악 데이터 조각에서 시간 t 에 대한 음악 데이터 신호의 진폭을 $S_i(u)$ 라고 했을 때, 음악 데이터를 영상 데이터의 프레임과 동일한 샘플 단위인 1/30초 단위의 윈도우로 분리한다.

[0157] 분리된 음악 데이터 조각에 대해서 각 윈도우에 대한 평균 진폭 값은 다음의 수학식 11에 의해 추출할 수 있다.

수학식 11

$$RMS_i = \frac{1}{U} \sum_{u=1}^U (S_i(u))^2$$

[0158]

[0159] 여기서, RMS_i 는 i 번째 윈도우의 평균 진폭 값, U 는 총 샘플링 윈도우 개수, u 는 샘플링 윈도우 인덱스를 의미한다.

[0160] 만약, 악기가 빠른 속도로 연주가 될 경우에는 진폭값의 변이량이 커지고, 느린 속도로 연주가 될 경우에는 진폭값의 변이량이 작아지게 된다. 따라서 음악의 속도를 추출하기 위해서는 RMS의 변이량을 사용하여 하나의 윈도우에 대한 평균 진폭 값과 이전 윈도우에 대한 평균 진폭 값의 차이인 다음의 수학식 12에 의해 추출될 수 있다.

수학식 12

$$M_{vel}^i = |RMS_i - RMS_{i-1}|$$

[0161]

[0162] 이와 같이 얻어진 음악 데이터 조각의 속도 정보는 도 5와 같이 표현될 수 있다.

[0163] 도 5에서 (a)는 Novalty Scored의 일례이고 (b)는 음악의 속도를 추출하기 위해서는 RMS의 변이량의 일례이다.

[0164] 도식된 바와 같이, 음악이 빠른 비트를 가지는 영역에서는 Novalty Score와는 달리 RMS 변이량이 커지고, 늦은

비트를 가지는 영역에서는 RMS 변이량이 작아진다.

[0165] 이런 현상의 원인은 대부분의 음악이 타악기가 중심이 된 비트를 가지고 있고, 비트가 빠를수록 타악기에서 생성되는 음악의 진폭 변이량이 커지기 때문이다.

[0166] 다음, 음악 분석부(220)에서 음악의 밝기 정보를 얻기 위해 스펙트럴 중심(Spectral centroid) 방법을 사용한다. 스펙트럴 중심 값은 일반적으로 사운드의 밝기를 측정하는데 사용된다.

[0167] 음악에서의 밝기란 음색과 밀접한 관련성을 가지게 된다. 예를 들어 바이올린과 같은 고음을 내는 악기의 경우 밝은 소리를 내게 되고, 베이스와 같이 저음을 내는 악기의 경우 어두운 소리를 내게 된다.

[0168] 이와 같은 음악의 밝기 정보는 음악 데이터 조각의 주파수 영역에서의 진폭 값에 대한 정보인 스펙트럴 중심(Spectral centroid) 방법인 다음의 수학적 식 13에 의해 추출된다.

수학적 식 13

$$M_{bri}^i = \frac{\sum p(A_i(p))^2}{\sum (A_i(p))^2}$$

[0169]

[0170] 여기서, $A_i(p)$ 는 주파수 영역에서 i 번째 윈도우 신호의 진폭, p 는 주파수 인덱스이다.

[0171] 지금까지는 영상과 음악의 매칭을 위한 분할과 각각의 조각 데이터의 속도 정보와 밝기 정보를 추출하였다.

[0172] 이와 같이 얻어낸 정보를 바탕으로 데이터 매칭부(300)에서는 영상 및 음악의 변이점 또는 영상 및 음악 데이터 조각의 속도 정보 또는 밝기 정보를 기준으로 한 매칭 팀이 서로 대응되도록 영상 데이터 조각과 음악 데이터 조각을 매칭하는 기능을 한다.

[0173] 여기서, 영상 및 음악 데이터 조각을 매칭하기 위한 기준인 매칭 팀은 영상의 속도 정보를 기반으로 한 영상 데이터 조각의 히스토그램과 음악 데이터 조각에 주파수 영역에서의 진폭값을 기반으로 한 히스토그램을 매칭 기준으로 하는 분위기 매칭 팀 및 영상 및 음악 데이터 조각의 시간 길이 값을 매칭 기준으로 하는 시간 매칭 팀을 더 포함할 수 있다.

[0174] 여기서, 영상 및 음악의 변이점을 기준으로 한 매칭 팀은 영상의 경우 외형 차이가 급격한 영상 변이점과 음악 데이터의 주파수 영역에서의 순간적인 진폭 변화가 급격한 음악 변이점과의 매칭이 하나의 매칭 쌍을 이룬다.

[0175] 또한, 영상 및 음악 데이터 조각의 속도 정보를 기준으로 한 매칭 팀은 영상과 음악에서 각각 구해진 $vec_{x,y}^i$ 값과 M_{vel}^i 가 하나의 매칭 쌍을 이룬다.

[0176] 또한, 영상 및 음악 데이터 조각의 밝기 정보를 기준으로 한 매칭 팀은 영상과 음악에서 각각 구해진 V_{bri}^i 값과 M_{bri}^i 값이 하나의 매칭 쌍을 이룬다.

[0177] 또한, 영상 및 음악 데이터 조각의 분위기를 기준으로 한 매칭 팀은 영상의 속도 정보를 기반으로 한 영상 데이터 조각의 히스토그램과 음악 데이터 조각에 주파수 영역에서의 진폭값을 기반으로 한 히스토그램이 하나의 매칭 쌍을 이룬다.

[0178] 분위기 매칭 팀은 수학적 식 14에 의해 매칭값이 결정된다.

수학적 식 14

$$Hc(y, z) = \sum_{b=1}^K \left(\frac{VH^y(b)}{N_y} - \frac{MH^z(b)}{N_z} \right)^2$$

[0179]

[0180] 여기서, $VH^y(b)$ 는 영상의 속도 정보를 기반으로 한 y 번째 영상 데이터 조각의 히스토그램, $MH^z(b)$ 는 주파수 영역에서 진폭값을 기반으로 한 z 번째 음악 데이터 조각의 히스토그램, y 와 z 는 영상 및 음악 데이터 조각의 인덱스, N_y 와 N_z 는 각 히스토그램을 정규화하기 위한 값으로 각 히스토그램 원소 개수의 합을 의미한다.

[0181] 분위기 동기화의 경우, 기본적으로 작은 움직임이 많은 영상의 경우, 저음과 매칭되는 것이 적절하다고 가정하고, 반대로 큰 움직임이 많은 복잡한 장면의 경우 고음과 매칭이 되는 것이 적절하다고 가정할 수 있다.

[0182] 그리고, 시간을 기준으로 한 매칭 팀은 영상 및 음악 데이터 조각의 각 시간 길이 값이 하나의 매칭 쌍을 이룬다.

[0183] 이와 같이 각각의 매칭 팀이 서로 대응되도록 하기 위해 영상 및 음악 데이터 조각 각각 길이를 허미트 곡선(Hermit curve)을 이용하여 정규화하는 과정을 거친다.

[0184] 또한, 데이터 매칭 단계는 각각의 매칭 팀에 대한 가중치를 부여하는 다음의 수학적 식 15를 통해 영상 데이터 조각과 음악 데이터 조각을 매칭하기 위한 최종 매칭값을 구한다.

수학적 식 15

$$Cost_{y,z} = \sum_{i=1}^3 w_i Fc_i(V^y(t), M^z(t)) + w_4 Hc(y, z) + w_5 Dc(y, z)$$

[0186] 여기서, y 와 z 는 데이터 조각의 인덱스, w_i 는 각 매칭 팀에 대한 가중치 값, $Fc_i(V^y(t), M^z(t))$ 는 각각 영상 및 음악의 변이점을 기준으로 한 매칭 팀, 영상 및 음악 데이터 조각의 속도 정보를 기준으로 한 매칭 팀, 영상 및 음악 데이터 조각의 밝기 정보를 기준으로 한 매칭 팀을 의미하고, $Dc(y, z)$ 는 시간 매칭 팀을 의미한다.

[0187] 이와 같이 사용자는 각각의 매칭 팀에 대해 가중치를 할당함으로써 뮤직 비디오의 결과를 조절할 수 있다. 예를 들어 시간 동기화의 가중치를 줄이게 되면 서로 다른 길이를 가지는 음악과 영상이 매칭될 확률이 커지게 되므로 역동적인 뮤직 비디오가 만들어질 가능성 커진다.

[0188] 만약 급격한 변이점에 대한 가중치를 크게 한다면 음악과 영상의 특징점에서의 매칭 확률이 상승하게 된다.

[0189] 또한, 데이터 매칭부(300)는 영상 분할 단계에서의 미리 결정된 임계값이 복수 개로 설정되어 영상 데이터를 복수 번 분할한 경우 영상 데이터 조각이 가장 큰 조각부터 작은 조각의 순서로 음악 데이터 조각과 매칭할 수 있다.

[0190] 이와 같이 함으로써 매칭함으로써 영상과 음악 데이터 조각의 유사도가 상대적으로 더 높게 매칭할 수 있어 뮤직 비디오의 영상 연속성이 최대한 유지될 수 있는 것이다.

[0191] 또한, 영상 데이터 조각이 가장 큰 것부터 작은 것의 순서로 음악 데이터 조각과 매칭한 최종 매칭 값이 매칭 기준값 미만인 경우 음악 분할 단계의 미리 결정된 임계값을 더 작은 값으로 음악 데이터를 재분할하여 복수 번 분할된 각각의 영상 데이터 조각과 재분할된 음악 데이터 조각과 매칭할 수도 있다.

[0192] 이를 위해 데이터 매칭부(300)에서 다중 영상 데이터 조각과 음악 데이터 조각을 매칭하기 위한 매칭값이 매칭 기준값 이하인 경우 음악 데이터를 이전에 분할된 음악 데이터 조각보다 더 작은 데이터 조각으로 분할하는 음악 재분할부(400)를 둘 수도 있다.

[0193] 이와 같은 본 발명의 실시를 위해 총 30분 분량의 비디오를 촬영하여 100초 가량의 뮤직 비디오의 자동 생성을 시도하였다. 비디오는 총 142개의 데이터 조각들로 나누어졌고, 음악의 경우 최초로 11개의 조각들로 나누어졌으나 다중 레벨 매칭에 의해 최종적으로 16개의 조각들로 나누어져 매칭이 이루어졌다.

[0194] 다음의 표 1은 음악과 영상의 특징 분석을 위한 전처리 계산 시간을 나타낸다.

표 1

미디어	길이	분할	속도	밝기
영상	30분	2.5시간	4.5시간	5분
음악	100초	8.3초	4.8초	2.5초

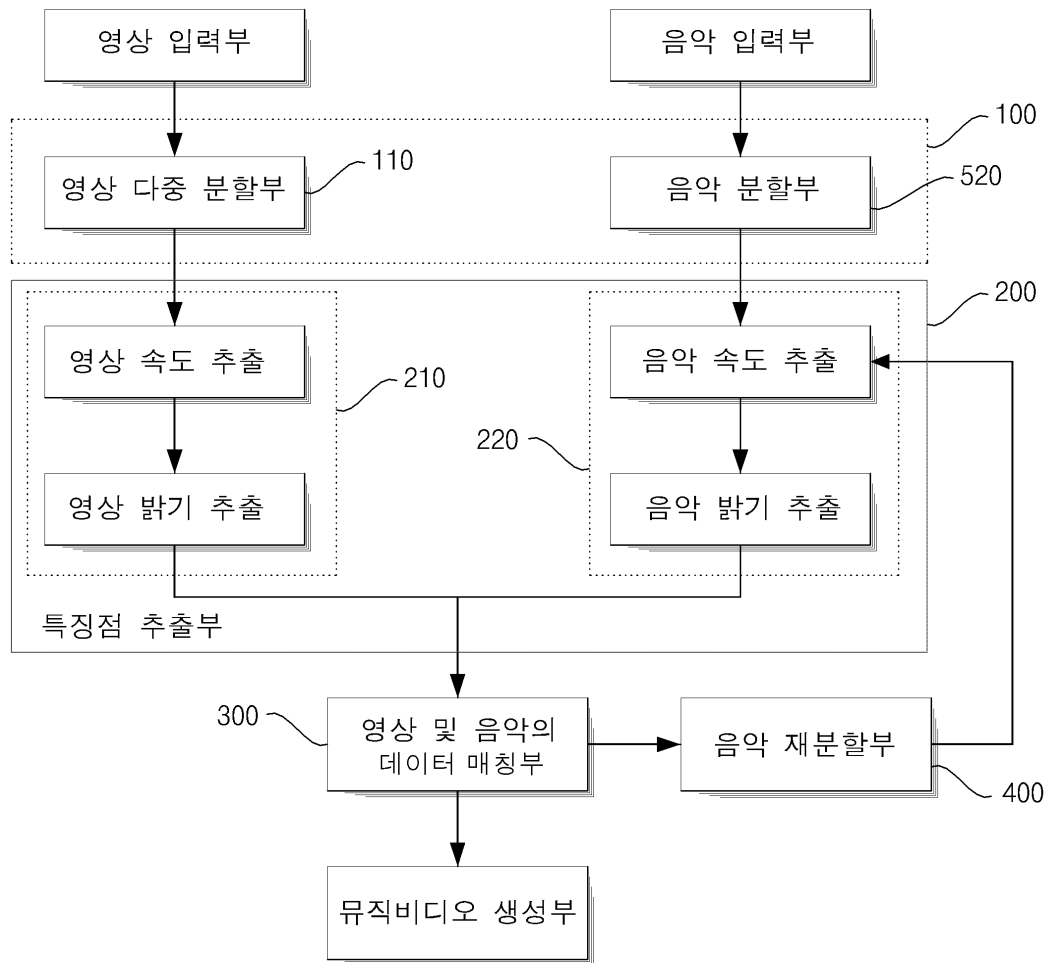
- [0196] 음악의 경우 1차원 시그널 분석이기 때문에 오랜 시간이 걸리지 않았지만 비디오의 경우 2차원 데이터이기 때문에 상대적으로 많은 시간이 걸리게 된다.
- [0197] 따라서, 비디오 데이터의 경우 전처리로 추출된 영상의 속도 정보와 밝기 정보를 미리 데이터 베이스에 저장해둠으로써 실제 뮤직 비디오 생성 시간을 절약할 수 있다.
- [0198] 한편, 상술한 본 발명의 실시예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터로 읽을 수 있는 기록매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다. 상기 컴퓨터로 읽을 수 있는 기록매체는 마그네틱 저장매체(예를 들면, ROM, 플로피 디스크, 하드 디스크, 자기 테이프 등), 광학적 판독 매체(예를 들면, CD-ROM, DVD, 광데이터 저장장치 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장매체를 포함한다.
- [0199] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에서 다양한 수정, 변경 및 치환이 가능할 것이다. 따라서, 본 발명에 개시된 실시예 및 첨부된 도면들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예 및 첨부된 도면에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면의 간단한 설명

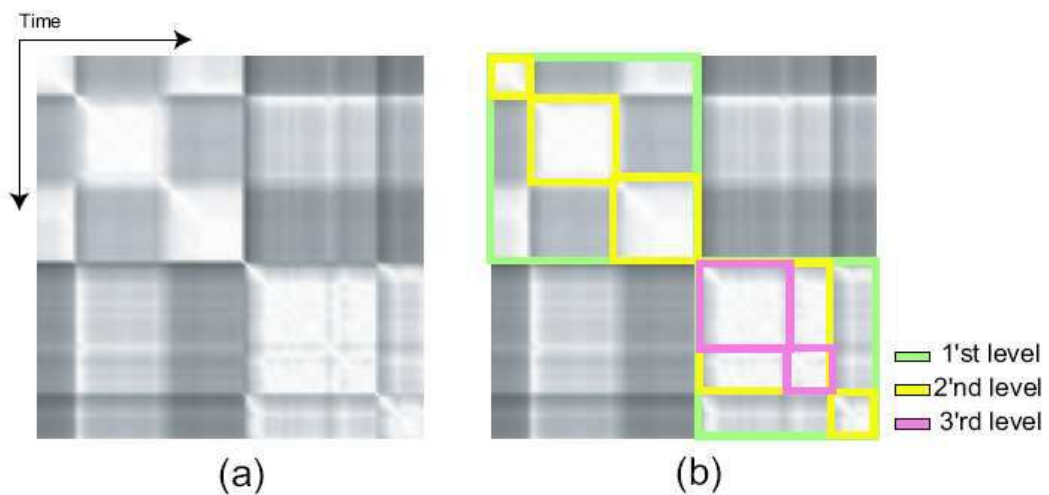
- [0200] 도 1은 본 발명에 따른 자동 뮤직 비디오 생성 장치의 일례를 설명하기 위한 도.
- [0201] 도 2a는 자동 뮤직 비디오 생성 장치의 일례에서 영상 다중 분할부에서 영상 분할하는 방법의 일례 설명하기 위한 도.
- [0202] 도 2b는 Radial Symmetric gaussian Kernel(RSK)의 형태
- [0203] 도 3은 특징점 기반의 비디오 속도 분석의 일례.
- [0204] 도 4는 주파수 영역에서 유사도 분석을 통한 Novalty Scoring 생성 방법의 일례.
- [0205] 도 5는 음악 조각 데이터의 Novalty Score와 RMS 변이량 비교.

도면

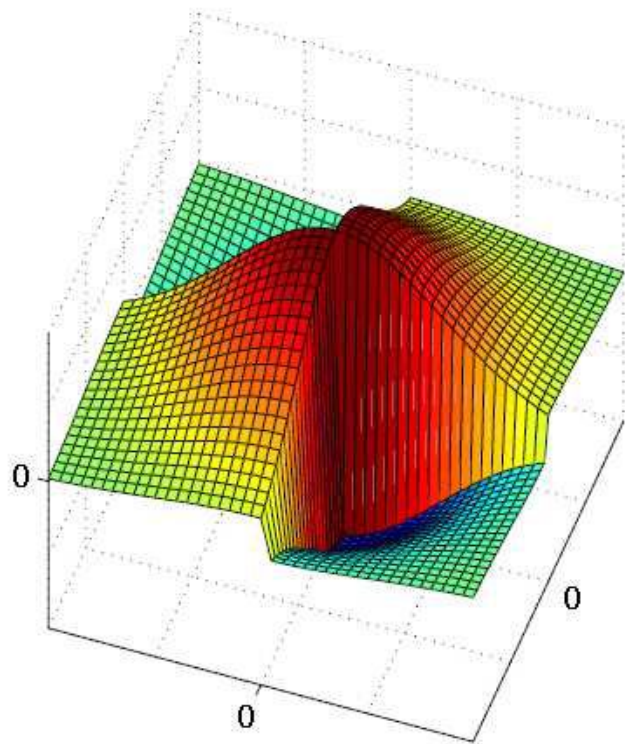
도면1



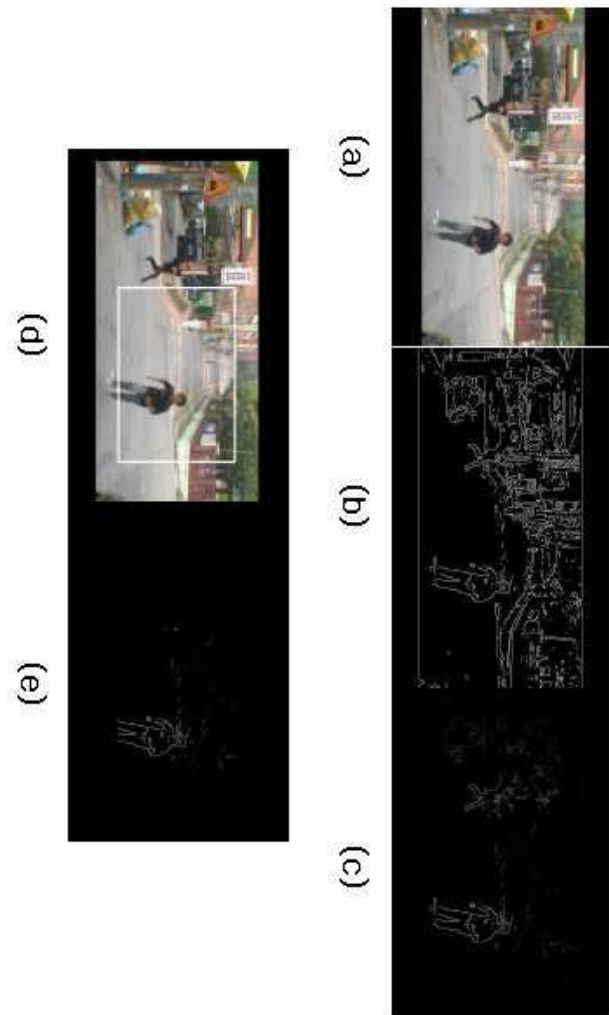
도면2a



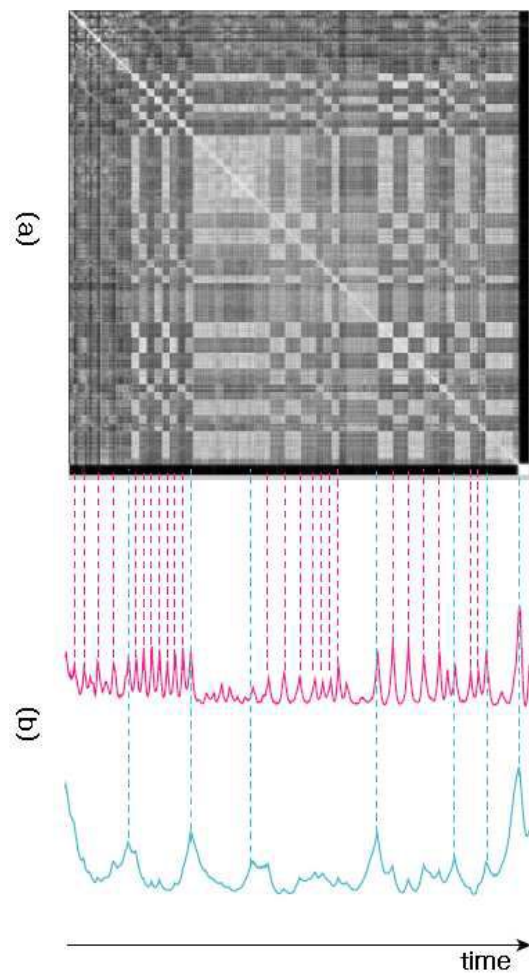
도면2b



도면3



도면4



도면5

