레터논문 (Letter Paper)
방송공학회논문지 제20권 제4호, 2015년 7월 (JBE Vol. 20, No. 4, July 2015)
http://dx.doi.org/10.5909/JBE.2015.20.4.641
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

이진 프레임 기술자를 이용한 유사중복 동영상 프레임 단위 정합

김 경 래[a], 이 준 태[a], 장 원 동[a], 김 창 수[a)‡]

Frame-level Matching for Near Duplicate Videos Using Binary Frame Descriptor

Kyung-Rae Kim[a], Jun-Tae Lee[a], Won-Dong Jang[a], and Chang-Su Kim[a)‡]

요 약

본 논문에서는 이진 프레임 기술자와 이를 이용한 프레임 단위 유사중복 동영상 정합 알고리즘을 제안한다. 우선 동영상으로부터 취득한 프레임을 패치(patch)단위로 나누고 패치간의 관계를 이진으로 나타낸다. 그리고 두 동영상 프레임 간의 정합비용과 보상비용으로 비용 함수를 표현한다. 초기 정합과 반복적인 정합 갱신을 통해 비용 함수를 최소화한다. 실험을 통해 제안하는 이진 프레임 기술자의 적합성과 프레임 단위 정합 알고리즘 성능이 우수함을 확인한다.

Abstract

In this paper, we propose a precise frame-level near-duplicate video matching algorithm. First, a binary frame descriptor for near-duplicate video matching is proposed. The binary frame descriptor divides a frame into patches and represent the relations between patches in bits. Seconds, we formulate a cost function for the matching, composed of matching costs and compensatory costs. Then, we roughly determine initial matchings and refine the matchings iteratively to minimize the cost function. Experimental results demonstrate that the proposed algorithm provides efficient performance for frame-level near duplicate video matching.

Keyword : Near-duplicate video, frame-level video matching, and binary frame descriptor.

Ⅰ. 서 론

최근 YouTube같은 동영상 공유 사이트가 대중화 되면서 온라인상의 동영상 수가 급격히 증가하고 있다. 특히 자막 삽입, 밝기 및 해상도 변화 등의 재생산을 통해 공유되는 다수의 유사중복 동영상들은 잉여 정보를 제공하여 동영상 검색 및 관리의 효율성을 저하시킨다. 동영상 자체를 기술자로 나타내어 유사중복 동영상 검출을 하는 시도들이 있다[1,2]. 하지만 동영상 단위 기술자는 동영상의 시간적 관계를 무시하기 때문에 여러 단편으로 엮인 동영상에 대해 취약점이 있다. 따라서 동영상의 중복 여부를 보다 정밀하게 판별하기 위해 실제로 중복된 구간을 검출할 필요가 있다.

본 논문에서는 유사중복 동영상 검출에 적합한 이진 프

a) 고려대학교 전기전자공학부(School of Electrical Korea University)
‡ Corresponding Author : 김창수(Chang-Su Kim)
E-mail: changsukim@korea.ac.kr
Tel: +82-2-3290-3806
ORCID: http://orcid.org/0000-0002-4276-1831
※이 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2009-0083495).
· Manuscript received May 22, 2015; Revised July 14, 2015; Accepted July 14, 2015.

레임 기술자와 반복적인 정합 갱신을 통한 유사중복 동영상의 프레임 단위 정합 알고리즘을 제안한다.

Ⅱ. 관련 연구

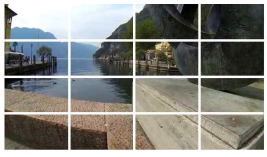
본 장에서는 대표적인 기존 기법에 대해 간단히 설명한다. Chen 등[3]은 추출된 키 프레임들과 그것들의 유사성을 기반으로 그래프를 구성하고 서로 강하게 연결된 부분그래프를 추출함으로써 유사중복 구간을 찾아낸다. Chiu 등[4]은 두 동영상의 프레임들 간의 유사행렬에 허프 변환을 적용하여 유사중복 부분 시퀀스를 검출한다. 그러나 이 기법들[3,4]은 오직 유사중복 구간의 유무만을 결정한다. 반면에 다이나믹 프로그래밍(dynamic programming)을 기반으로 최적화 문제를 해결함으로써 더욱 정밀한 프레임 단위의 정합을 가능하게 하는 기법들도 연구되고 있다[5-7]. 하지만 높은 연산량과 많은 메모리 공간을 요구하는 단점이 있다.

Ⅲ. 이진 프레임 기술자

본 연구에서는 프레임 단위 유사중복 동영상 정합에 효율적인 프레임 기술자를 제안한다.

빈번하게 일어나는 유사중복 유형 중 하나인 좌우 반전 동영상에 강인하게하기 위해 전처리 과정을 적용한다. 프레임을 수직으로 이등분하여 왼쪽과 오른쪽 영역의 평균 휘도를 η_L와 η_R으로 나타낸다. 만약 η_R이 η_L보다 큰 값을 가질 경우, 프레임을 좌우로 반전하여 나타낸다.

전처리과정 후, 그림 1와 같이 프레임을 4×4 패치로 나누고 각 패치들의 평균 휘도를 계산한다. η_i가 패치i의 평균 휘도 값을 나타낼 때, 비트(bit) 값 B_{ij}는 다음과 같다.



204	203	49	38
81	95	66	36
117	110	142	102
65	104	157	131

그림 1. 프레임 4x4 패치단위 분할과 그 패치들의 평균 휘도 값
Fig. 1. division of a frame into 4x4 patches and the average luminances of those patches

$$B_{ij}=\begin{cases}0 & \text{if } \eta_i-\eta_j<0,\\ 1 & \text{if } \eta_i-\eta_j\geq 0\end{cases} \quad (1)$$

가능한 모든 패치 쌍 $N=\{(1,2),(1,3),...,(15,16)\}$에 대해 비트 값 B_{ij}을 계산하여 $120={}_{16}C_2$차원을 갖는 이진 프레임 기술자 $\mathbf{B}=\{B_{ij}:(i,j)\in N\}$을 계산한다.

두 프레임 x, y의 이진 기술자를 $\mathbf{B}(x)$와 $\mathbf{B}(y)$로 나타냈을 때, x와 y 간의 유사성은 해밍거리를 계산함으로써 $d_H(\mathbf{B}(x),\mathbf{B}(y))=\sum_{(i,j)\in N}B_{ij}(x)\otimes B_{ij}(y)$로 나타낸다. 여기에서 $\otimes$는 배타적 논리합(exclusive or) 연산자를 나타낸다.

제안하는 이진 프레임 기술자는 패치의 휘도 값인 지역 정보로부터 추출된 전역 프레임 기술자로서, 특히 검출하기 어려운 톤 및 밝기, 색상 변화 등의 다양한 유사중복 유형에 강인하다.

Ⅳ. 프레임 정합

본 연구에서는 두 유사중복 동영상에서 추출한 프레임 시퀀스 간의 정밀한 프레임 단위 정합을 한다. 연산량을 줄이기 위해 각 동영상으로부터 균일하게 초당 1개의 프레임을 추출한다. 동영상 X와 Y에서 추출된 프레임 집합을 각각 $X=\{x_1,x_2,...,x_M\}$와 $Y=\{y_1,y_2,...,y_N\}$으로 나타낸다.

제안하는 기법은 최적화 문제를 정의하고 풀어냄으로써 두 시퀀스 간의 대응하는 정합을 구할 수 있다. 다이나믹 프로그래밍을 이용한 기법들[5-7]은 전역 최적화 해결방법이지만 많은 계산량과 메모리를 요구한다. 본 연구에선 초기 정합을 실시한 후 반복적인 정합 갱신을 통해 연산량을 줄이면서 효과적으로 정합비용 함수를 최소화한다.

1. 최적화 문제

본 연구에서는 최적화 문제를 정의한다. 프레임 집합 X에 속한 프레임 x_m에 대응하는 프레임이 $y_n\in Y$ 일 때 정합 함수는 $\lambda(m)$는 n의 값을 갖는다. 즉, 다음과 같은 관계식을 갖는다.

$$\lambda(m)=\begin{cases}n & \text{if } x_m \text{ is matched to } y_n,\\ 0 & \text{if } x_m \text{ is unmatched.}\end{cases} \quad (2)$$

최적화된 정합 함수 λ^*는 아래의 비용 함수를 최소화함으로써 구할 수 있다.

$$\lambda^*=\arg\min_{\lambda}\left(\sum_{m:\lambda(m)>0}d(m,\lambda(m))+\sum_{m:\lambda(m)=0}d_c(m)\right) \quad (3)$$

여기서 $d(m,\lambda(m))$은 정합비용, $d_c(m)$은 보상비용을 나타낸다. 정합비용은 프레임 x_m과 $y_{\lambda(m)}$의 이진 기술자 간의 해밍거리 $d_H(\mathbf{B}(x_m),\mathbf{B}(y_{\lambda(m)}))$을 통해 계산한다.

정합비용 합의 최소화만을 고려하면 전부 정합이 되지 않도록 최적화 되므로 정합되지 않는 프레임에 보상비용 $d_c(m)$을 부여한다. 최적화된 정합 함수 λ^*에서 정합비용이 보상비용보다 작은 프레임쌍만 정합한다. 유사중복 유형에 따라 프레임 정합비용이 상이하기 때문에 다른 값을 갖도록 적응적으로 보상비용을 결정한다. 프레임 x_m와 Y의 모든 프레임간의 정합비용 집합 $\mathbf{D}_\mathrm{m}=\{\mathrm{d(m,1)},\ldots,\mathrm{d(m,N)}\}$을 계산한다. $\mathbf{D}_m$을 오름차순으로 정렬하고 정렬된 색인을 $a_1,\ldots,a_N$으로 나타낸다. 정렬된 비용 간 차이 값이 가장 큰 위치의 색인 $\delta^*=\mathrm{argmax}_{\delta\in\{2,\ldots,N\}}(d(m,a_\delta)-d(m,a_{\delta-1}))$을 찾는다. 차이 값이 가장 큰 두 정합비용의 평균이 프레임 x_m의 보상비용 $d_c(m)$이 된다.

$$d_c(m)=\frac{d(m,a_{\delta^*})+d(m,a_{\delta^*-1})}{2} \tag{4}$$

2. 프레임 단위 정합 알고리즘

초기 정합을 통해 대표 정합 벡터 α^*를 결정하고 이를 기준으로 반복적인 정합 갱신을 통해 최적화 문제를 해결한다. 입력 받은 두 동영상의 프레임 간의 모든 정합비용을 계산한다. 대응되는 프레임 간의 정합비용이 다른 프레임과의 정합비용보다 작은 값을 가질 확률이 높으므로 각 프레임 마다 최저 정합비용을 갖는 프레임과 정합하여 다음과 같이 초기 정합 함수를 얻는다.

$$\lambda_{\mathrm{Init}}(m)=\mathrm{argmin}_{1\le n\le N}d(m,n) \tag{5}$$

λ_{Init}로부터 정합 벡터 $\alpha_m=\lambda_{\mathrm{Init}}(m)-m$을 계산한다. 정합 벡터에 대한 히스토그램이 $h(\alpha_m)$일 때, 가장 빈번한 정합 벡터를 대표 정합 벡터 $\alpha^*=\mathrm{argmax}_{\alpha_m}h(\alpha_m)$로 나타낸다. 아래와 같이 정합 벡터 α_m가 α^*-1보다 작거나 α^*+1보다 큰 경우 잡음으로 판단하여 정합을 제거함으로써 초기 정합 함수 λ'_{Init}를 완성한다.

$$\lambda'_{\mathrm{Init}}(m)=\begin{cases}0, & \alpha_m<\alpha^*-1 \text{ and } \alpha_m>\alpha^*+1\\ \mathrm{argmin}_{1\le n\le N}d(m,n), & \text{otherwise}\end{cases} \tag{6}$$

$\lambda'_{\mathrm{Init}}(m)$가 0의 값을 갖는 프레임을 기준으로 정합 갱신을 실시한다. α^*-1보다 같거나 크고 α^*+1보다 같거나 작은 정합 벡터를 갖는 정합 중에 정합비용 $d(m,n)$을 최소화하는 프레임 y_{n^*}을 아래 식과 같이 찾는다.

$$n^*=\mathrm{argmin}_{m+\alpha^*-1\le n\le m+\alpha^*+1}d(m,n) \tag{7}$$

앞서 언급했듯이, 식(3)의 비용 함수를 최소화하기 위해 해당 되는 정합비용 $d(m,n^*)$이 보상비용 $d_c(m)$보다 클 경우 정합하지 않는다. 정합 갱신은 다음과 같은 관계식을 갖는다.

$$\lambda^*(m)=\begin{cases}n^*, & \text{if } d(m,n^*)<d_c(m)\\ 0, & \text{otherwise}\end{cases} \tag{8}$$

정합 함수가 0의 값을 갖는 모든 프레임에 대해 정합 갱신을 실시한다. 새로운 정합이 생기지 않을 때까지 정합 갱신을 반복적으로 수행하여 최적화된 정합 함수 λ^*를 결정한다.

V. 실험 결과

본 논문에서는 제안하는 이진 프레임 기술자가 유사중복 동영상 정합에 적합한지 확인하고 프레임 단위 정합 알고리즘의 성능 확인을 위한 실험 결과를 제시한다. 실험의 신뢰도를 높이기 위해 자막 및 로고 삽입, 색상 및 밝기 변화, 화질 변화 등 여러 종류의 유사중복 동영상을 동영상 공유 웹사이트인 YouTube, Vimeo, Soku로부터 직접 수집했다. 실험 데이터 동영상들의 평균 길이는 약 3분이고, 320×240에서 640×480까지 다양한 해상도로 구성되어있다.

모든 동영상에 전처리 과정으로 흔한 유사중복 유형 중 하나인 보더를 제거한다. 동영상 전반에 걸쳐 값의 변화가 거의 없는 화소는 보더에 속한다고 판단하고 제거한다.

1. 이진 프레임 기술자

제안하는 이진 프레임 기술자 평가를 위해 테스트 셋 100개를 구성했다. 한 개의 테스트 셋은 유사중복 동영상 5개로부터 추출된 100개의 프레임으로 구성되어 있어, 총 10,000개의 프레임을 실험에 사용했다. 문턱 값을 조정해 가며 프레임 간의 정합을 통해 정확도와 재현율을 계산한다. 프레임 간의 정합 비용이 문턱 값보다 낮으면 유사중복

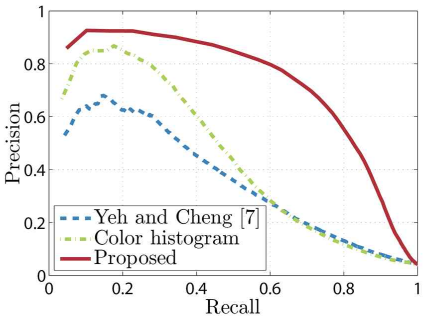


그림 2. 문턱값 변화에 따른 정확도-재현율 커브
Fig 2. Precision-recall curves with threshold adjustments

프레임으로 판단한다. 제안하는 기술자와 Yeh와 Cheng[7]의 전역 기술자와 색 히스토그램을 비교하여 성능을 평가하였다. RGB 각 채널을 16단계로 양자화하고 일렬로 연결하여 48차원의 색 히스토그램을 생성한다.

그림 2는 정확도-재현율 커브를 통해 세 개의 기술자를 비교하고 있다. 제안 기법이 기존 기법들보다 유사중복 프레임을 구별하는데 있어서 확연히 좋은 성능을 보인다.

2. 프레임 단위 정합 성능

본 장에서는 프레임 단위 정합 성능을 측정하기 위해 500쌍의 유사중복 동영상을 실험 데이터로 사용했다. 두 동영상 간의 실제 중복 구간과 제안하는 알고리즘 결과를 비교하여 정확도와 재현율을 측정한다. 또한 500쌍 동영상 정합 평균 수행 시간을 측정하여 계산 복잡도를 측정한다.

제안하는 알고리즘과 다이나믹 프로그래밍을 통한 전역 최적화 문제 해결 방법[5]을 비교한다. 표 1은 정확도와 재현율, 평균 수행 시간을 비교하고 있다. 제안하는 알고리즘이 다이나믹 프로그래밍과 비교하여 확연히 적은 계산량으로 대등한 성능을 보여주고 있다. 표 2는 장르별 프레임 정합 성능을 나타내고 있다. 제안하는 알고리즘이 다양한 장르의 동영상에서 골고루 좋은 성능을 내고 있으며, 비교 알고리즘보다 대등하거나 더 좋은 성능을 나타내고 있다.

표 1. 프레임 단위 정합 성능 비교
Table 1. Comparison of frame matching performance

	Precision	Recall	Time (s.)
Dynamic prog. [5]	0.95	0.97	0.105
Proposed algorithm	0.98	0.96	0.0034

표 2. 장르별 정합 성능 비교
Table 2. Comparison of matching performance by genre

장르	비율	Proposed algorithm		Dynamic prog. [5]	
		Precision	Recall	Precision	Recall
UCC/광고	20%	0.98	0.93	0.94	0.96
뉴스/스포츠	3%	0.96	0.98	0.88	0.95
뮤직비디오	25%	0.99	0.97	0.96	0.98
애니메이션	21%	0.99	0.97	0.97	0.98
영화 예고편	7%	0.97	0.97	0.93	0.98
영화 클립	13%	0.96	0.91	0.9	0.93
예능/음악쇼	11%	0.99	0.97	0.95	0.98

Ⅵ. 결 론

본 논문에서는 유사중복 동영상 간의 프레임 단위 정합을 통해 정밀한 중복구간 검출 기법을 제안하였다. 동영상으로부터 추출된 프레임을 이진 프레임 기술자로 표현하고 프레임 시퀀스 간의 관계를 비용 함수로 표현한다. 초기 정합을 실시하고 반복적인 정합 갱신을 통해 최적화 문제를 해결한다. 실험 결과에서는 이진 프레임 기술자가 기존 방법[7]보다 유사중복 동영상 정합에 효과적인 것을 보였고, 제안하는 정합 알고리즘이 전역 최적화 기법과 비교하여 낮은 계산량을 요구하면서 대등한 성능을 나타냈다.

참 고 문 헌 (References)

[1] S. Hu, “Efficient video retrieval by locality sensitive hashing,” in Proc. IEEE ICASSP, 2005, vol. 2, pp. 449-452.
[2] H. J´egou, M. Douze, and C. Schmid, “Improving bag-offeatures for large scale image search,” Int. J. Comput. Vis., vol. 87, no. 3, pp. 316-336, 2010.
[3] T. Chen, S. Jiang, L. Chu, and Q. Huang, “Detection and location of near-duplicate video sub-clips by finding dense subgraphs,” in Proc. ACM Multimedia, Nov. 2011, pp. 1173-1176.
[4] C.-Y. Chiu, T.-H. Tsai, Y.-C. Liou, G.-W. Han, and H.-S. Chang, “Near-duplicate subsequence matching between the continuous stream and large video dataset,” IEEE Trans. Multimedia, vol. 16, no. 7, pp. 1952-1962, Nov. 2014.
[5] Y.-Y. Lee, C.-S. Kim, and S.-U. Lee, “Video frame-matching algorithm using dynamic programming,” J. Electron. Imaging, vol. 18, no. 1, pp. 1-3, Mar. 2009.
[6] M.-C. Yeh and K.-T. Cheng, “Video copy detection by fast sequence matching,” in Proc. ACM CIVR, July 2009, pp. 45:1-45:7.
[7] M.-C. Yeh and K.-T. Cheng, “A compact, effective descriptor for video copy detection,” in Proc. ACM Multimedia, Oct. 2009, pp. 633-636.