

k -means 클러스터링 기법을 이용한 플리커 저감 비디오 톤 매핑

*허정환 **정제창

한양대학교 전자컴퓨터통신공학과

*hur881122@hanyang.ac.kr **jjeong@hanyang.ac.kr

Flicker Reduction Tone Mapping of Video Using k -means Clustering

*Heo, JeongHwan **Jeong, Jechang

Department of Electronics and Computer Engineering, Hanyang University

요 약

본 논문에서는 넓은 동적 영역 (High Dynamic Range: HDR) 영상의 실시간 처리 방법을 제안한다. HDR 영상은 사람의 눈으로 볼 수 있는 자연 영상과 가깝지만 대부분의 디스플레이 기기들은 좁은 동적 영역 (Standard Dynamic Range: SDR)의 영상 출력만을 지원한다. 이러한 동적 영역의 차이를 사람의 시각 특성을 고려하여 축소하기 위해서는 톤 매핑 기법 (Tone Mapping Operator: TMO)을 수행하여야 한다. 하지만 기존의 이미지 톤 매핑 기법을 실시간 영상에 적용할 경우 영상의 실시간 처리, 플리커링 현상을 해결하는데 어려움이 있다. 본 논문에서는 k -means 클러스터링 기법을 사용하여 플리커 현상에 강인하고 실시간 비디오 톤 매핑이 가능한 알고리즘을 제안한다. 제안하는 알고리즘은 동적 계획법 (Dynamic Programming)을 통해 빠른 전역 해를 찾는 것이 가능하고 세부영상 향상 (Detail Enhancement)의 실시간 처리가 가능함을 보였다.

1. 서론

인지기반 시각 특성 (Human Visual System: HVS)에 따르면, 사람의 눈은 10^{-6} 에서 10^8 cd/m²에 이르는 밝기의 동적 영역 (Dynamic Range)을 볼 수 있다고 한다. 즉 사람이 눈으로 직접 보는 것과 같은 생동감 넘치는 영상을 보기 위해서는, 촬영에서 출력까지의 모든 기기들이 이러한 동적 영역 영상을 지원하여야 한다. 오늘날, 눈으로 직접 보는 것과 같은 생동감 넘치는 동적 영역 영상의 수요가 증가함에 따라, 촬영에서 출력에 이르기까지의 모든 기기들은 넓은 동적 영역 (High Dynamic Range: HDR) 영상 규격을 사용한다. 현재 영상 촬영 기기들은 HDR로 영상을 취득한다. 하지만 아직 대부분의 디스플레이들은 표준 동적 영역 (Standard Dynamic Range: SDR)의 낮은 동적 범위만을 출력 가능하므로 사람이 실제로 보는 것과 유사한 자연스러운 영상을 표현하는 것이 어렵다. 이에 0.0005에서 10,000 니트에 이르는 HDR 영상을 0~255 레벨의 SDR 디스플레이에 출력할 수 있도록 매핑 하는 과정이 필요하다. 이를 톤 매핑 기법 (Tone Mapping Operator: TMO)라고 한다. 톤 매핑 과정은 인간의 비선형적인 인지기반 시각 특성에 의해 수학적으로 완전한 매핑이 정해져 있지 않다. 따라서 기기마다 그 매핑 과정이 다를 수 있다. 하지만 톤 매핑 과정의 역할은 대체적으로 영상의 매우 밝은 영역과 매우 어두운 영역을 알아볼 수 있게 하고 영상의 세부 (Detail) 묘사가 가능하도록 매핑 하는 것이라 할 수 있다. 지난 20년 동안 톤 매핑 문제에 대한 광범위한 연구가 진행되었다.

톤 매핑 기법에 대한 연구들을 간략하게 소개하자면 전역적 톤 매

핑 기법과 지역적 톤 매핑 기법이 있다. 글로벌 톤 매핑 기법은 단일 이미지에 대해 로그기반의 매핑 함수를 사용하여 전역적인 처리를 하는 방법이다. 전역적 톤 매핑 기법은 연산 속도가 빠르지만 이미지의 지역적인 디테일을 보존하기 어려운 단점이 있다[1]. 다음으로 지역적 톤 매핑 기법에는 적응적 로그매핑과 같은 방법이 있다[2]. 전역적 톤 매핑 방법 보다 연산이 복잡하지만 영상의 주변블록을 고려하기 때문에 디테일을 더 잘 보존할 수 있다. 지역적 톤 매핑 기법에는 디테일을 강화 (Detail Enhancement) 하는 과정에서 헤일로 현상 (Halo Artifact)이 발생하는 단점이 있다. 헤일로 현상을 줄이기 위해 다중 스케일 분해 기법을 사용한 톤 매핑이 연구되었다[3][4]. 이러한 톤 매핑 기법들이 연구되어 왔지만 대부분이 단일 이미지에 대한 톤 매핑 기법에 대한 연구이며, 비디오 톤 매핑 문제에 적용 가능한 톤 매핑 기법은 일부이다. 단일 이미지에 대한 톤 매핑 문제를 비디오에 대한 톤 매핑 문제로 확장하려면 시간 영역을 추가로 고려하여야 한다. 따라서 시간 영역을 고려한 이상적인 비디오 톤 매핑 기법의 요구사항을 정리할 필요가 있다. [5]에서는 비디오 톤 매핑 기법의 요구사항을 다음과 같이 서술하고 있다.

- 플리커링, 고스팅 현상이 없는 시간 영역 모델을 사용해야 한다.
- 지역적인 영상의 디테일과 대비를 유지하면서도 모든 상황에서 동적 범위 처리가 가능해야 한다.
- 대량의 효율적인 데이터 처리가 필요하고 처리시간이 가능한 짧아야 한다.
- 파라미터 튜닝이 필요 없어야 함.
- 교정을 위한 알고리즘의 메모리가 최소로 유지되어야 함.

- 매우 다른 특성을 가진 비디오 입력에 대해 고품질의 결과를 생성할 수 있어야 함.
- 노이즈와 영상에 대한 명확한 처리가 있어야 함.

본 논문에서는 k -means 알고리즘과 다층 레이어 분해를 통해 위의 7가지 조건을 모두 만족하는 비디오 톤 매핑 기법을 제안한다. 기존의 k -means 알고리즘을 사용하면 파라미터 튜닝이 필요 없고 다양한 동적 범위의 영상에서도 안정적인 영상을 얻는 것이 가능하다. 하지만 k -means 알고리즘을 통한 톤 매핑은 영상 전체에 적용되는 전역적인 처리이기 때문에 지역인 세부영상 향상 처리가 불가능하다. 따라서 다층 레이어 분해를 통한 세부영상 향상 방법을 추가하여 위의 7가지 조건을 모두 만족하는 톤 매핑 방법을 제안한다.

2. 기존의 알고리즘

2.1 k -means 클러스터링 기법을 통한 HDR 톤 매핑

k -means는 입력 값들의 대표 값을 선정하는 군집화 알고리즘으로 비지도 학습 방법 중 하나이다. k -means는 사람이 직접 입력해 주어야 하는 파라미터가 적기 때문에 HDR 톤 매핑에서의 동적 영역 처리와 잘 부합한다고 할 수 있다. 하지만 연산량이 많고 학습 과정의 중단 조건을 어떻게 설정할 것인지에 대한 문제가 남아 있다. k -means를 톤 매핑 과정에서 사용하기 위해서는 동적 계획법 (Dynamic Programming)을 사용하여 해를 구하는 방법이 있다[6]. 동적 계획법이란 큰 문제를 해결하는데 있어 하위의 작은 문제의 해를 먼저 구하고 이를 재사용하여 효율적으로 최종 해를 구하는 방법이다. 동적 계획법을 1차원 k -means 문제에 적용할 경우 학습의 중단조건을 설정할 필요가 없으며 지역적 최적해 (Local Minima)에 빠지지 않고 항상 전역적 최적해 (Global Minimum)를 찾을 수 있다.

HDR 톤 매핑 문제에서 k -means를 적용하려면 넓은 동적 영역 이미지의 다양한 밝기 값들 중 대표 밝기를 뽑아 SDR 이미지를 만드는 방법으로 적용할 수 있다. 입력 밝기의 종류가 $\{u_1, u_2, \dots, u_N\}$ 가 지닌 HDR 그레이 이미지를 $I(x, y)$ 라 하고 $\{c_1, c_2, \dots, c_K\}$ 의 256가지 밝기를 갖는 SDR 그레이 이미지를 $\hat{I}(x, y)$ 라 하자.

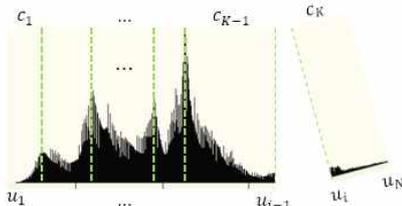


그림 1 재귀식으로 표현한 1차원 히스토그램 군집

그림 1과 같이 HDR 그레이 이미지의 밝기는 1차원이기 때문에 이를 오름차순으로 정렬하면 k -means의 최종 해는 최 우측 군집 c_K 와 이를 제외한 나머지 값들의 군집화 결과로 분리하여 생각할 수 있다. n 가지의 밝기 입력과 k 가지의 출력에 대한 비용함수 D 는 다음과 같은 재귀식으로 표현할 수 있다.

$$D(n, k) = \min_{k \leq i \leq n} (D(i-1, k-1) + f(u_i, \dots, u_n)) \quad (1)$$

f 는 $\{u_i, \dots, u_n\}$ 가 속한 군집의 비용함수이다. 여기서 최종 군집의 크기를 알 수 없으므로 i 값을 특정할 수 없다. 따라서 가능한 모든 i 값에 대해 계산 값을 저장해 두어야 한다. $D(n, k)$ 의 모든 결과 값을 $N \times K$ 매트릭스에 저장한다. $D(1, 1)$ 를 푸는 것을 시작으로 $D(N, 1)$ 까지의 모든 해를 구하면 이를 통해 k 가 2인 비용함수를 이어서 계산할 수 있다. 이후 (N, K) 지점에서 최종 해를 구할 수 있다. 이렇게 구해진 SDR 이미지는 밝기가 치우쳐진 동적 영역 영상에서도 안정적인 결과물을 보여준다.

2.2 컬러 영역으로의 톤 매핑

앞서 2.1 장에서 k -means 클러스터링을 통해 HDR 그레이 이미지를 SDR로 톤 매핑 하는 과정에 대해 논하였다. 이번 장에서는 완성된 그레이 매핑을 통해 매핑을 각 RGB의 채널로 효율적으로 확장하는 과정을 설명한다. 입력 $\{u_1, u_2, \dots, u_N\}$ 를 통해 구한 $\{c_1, c_2, \dots, c_K\}$ 를 최근접 이웃법으로 매핑 하는 과정을 $F(s)$ 라 하면,

$$F(s) = \underset{l}{\operatorname{argmin}} |c_l - s| \quad (2)$$

로 표현할 수 있다. 이를 이용하면 다음과 같이

$$I_{out}^{ch} = F(I_{log}^{ch}) \text{ 와 } I_{out}^{gr} = \frac{I^{ch}}{I_{log}^{gr}} \cdot F(I_{log}^{gr}) \quad (3)$$

의 두 가지 방법을 고려할 수 있다. 여기서 I^{ch} 은 그레이 이미지, I^{gr} 은 3차원 컬러 이미지, I_{log} 는 로그스케일 이미지를 뜻한다. 첫 번째 방법은 그레이 이미지로 구한 매핑함수 $F(s)$ 를 각 채널에 적용한 결과이며, 두 번째 방법은 그레이 이미지의 매핑 결과를 컬러성분의 가중에 따라 적용하는 방법이다. 두 번째 방법의 경우 일부 영상에서 컬러 포화가 발생하기 쉬우므로 [7]에서는 식 (3)을 0.7과 0.3의 선형 결합으로 사용하였고 안정적인 결과를 보였다.

2.3 비디오 영역으로의 톤 매핑

이번 장에서는 매핑을 시간영역으로 확장하는 방법을 설명한다. 일반적인 비디오의 경우 시간축으로 인접한 영상들간의 밝기 분포는 비슷한 경향을 보인다. 즉 시퀀스 전체에 동적 계획법을 실행하는 것보다 밝기의 분포가 극적으로 변하는 프레임, 혹은 동일한 프레임 간격으로 동적 계획법을 실행하는 것이 효율적이다. 이렇게 동적 계획법이 실행된 프레임을 키 프레임이라 하면 키 프레임 사이의 프레임은 양쪽 키 프레임의 결과를 선형으로 보간하여 사용한다. 현재 프레임 $F_t(s)$ 의 양방향으로 존재하는 키 프레임이 k 와 $k+1$ 이라고 할 때,

$$F_t(s) = w_k F_k(s) + w_{k+1} F_{k+1}(s) \quad (4)$$

밝기의 분포가 변하는 지점에 대한 척도는 Earthmover's distance (EMD) 등을 사용하기도 하지만 임계치에 대한 파라미터를 결정해야 하는 문제가 존재한다. 실시간 처리를 수행 할 때 계산량이 균일해야 함을 고려 할 때 일정한 간격마다 동적 계획법을 실행하는 방법을 사용할 수 있다. 일정한 간격의 키 프레임을 사용하여도 비디오 톤 매핑의 품질에 큰 문제가 없었다.

2.4 다층 레이어 분해를 통한 톤 매핑

다층 레이어 분해를 통한 톤 매핑 기법은 지역적 톤 매핑의 대표적인 기법이라 할 수 있다. 다층 레이어 분해란 HDR 영상을 베이스 레이어와 디테일 레이어로 구분하여 처리하는 기법을 말하는데 베이스 레이어는 주로 영상의 뼈대에 해당하는 구조 정보를 담고 있으며 디테일 레이어는 영상의 세부묘사에 해당하는 명암, 모서리 정보를 담고 있다고 할 수 있다[3][4]. 디테일 레이어와 베이스 레이어를 나누는 데는 모서리 보존 필터 (Edge Preserving Filter)같은 비선형 필터를 사용한다. 주로 원본 이미지에 모서리 보존 필터링을 거쳐 베이스 레이어를 만들고 원본 이미지에서 베이스 레이어를 뺀 이미지를 디테일 레이어를 만든다. 이후 디테일 레이어는 스케일 조정을 통해 디테일을 강화 한다. 최근에는 모서리 보존 필터의 성능에 따라 후광 현상이 적게 발생한다는 점이 밝혀져 양방향 필터 (Bilateral Filter [8]), 가중 최소 제곱 필터 (Weighted Least Square Filter [9]) 유도 영상 필터 (Guided Filter: GIF[10]) 등 다양한 모서리 보존 필터를 사용한 톤 매핑 연구가 이루어지고 있다.

3. 제안하는 알고리즘

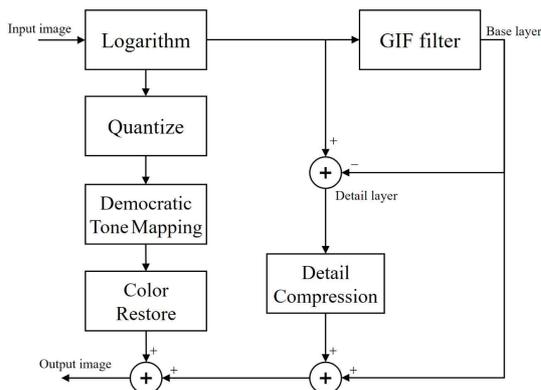


그림 2 제안하는 톤 매핑 방법의 흐름도

기존의 동적 계획법을 통한 톤 매핑은 전역적 톤 매핑 과정이기 때문에 지역적인 영상 디테일이 향상되기 어렵다. 따라서 그림 2와 같이 동적 계획법을 통한 글로벌 톤 매핑과정에 추가로 베이스 레이어와 디테일 레이어를 나누어 디테일을 향상시킬 수 있도록 구성한다. Democratic Tone Mapping 단계는 동적 계획법을 통한 히스토그램 영역의 k -means 톤 매핑 단계이다. 우측에서는 GIF를 사용하여 베이스 레이어를 구성하였으며 원본이미지에서 베이스 레이어를 빼는 방법으로 디테일 레이어를 만들었다. 해당 흐름도를 사용 할 경우 GIF의 비선형적인 필터링으로 시각적으로 모서리 부분을 경계로 영상을 분해해 낼 수 있다. 분해된 디테일 레이어는 디테일 압축 단계를 통해 정

규화 되어 디테일 향상효과를 낸다. 디테일 압축 단계에서는 시그모이드 함수를 사용하였다. GIF 필터 대신 양방향 필터, 가중 최소 제곱 필터, 등의 다른 모서리 보존 필터를 고려해본 결과 유도 영상 필터가 계산 복잡도가 낮으면서도 좋은 성능을 보였다. Color Restore 단계에서는 식 (3)을 모두 사용하는 방법으로 구현 하였고 동적 계획법을 통한 톤 매핑 과정은 매 20 프레임 마다 수행하며 레이어 분해를 통한 디테일 향상 과정은 매 프레임마다 수행하였다.

4. 실험 및 결과

제안하는 방법을 평가하기 위해 기존의 동적 계획법을 사용한 이미지 톤 매핑결과와 제안하는 톤 매핑 방법의 주관적 화질을 비교하였다. 또한 실시간 영상의 톤 매핑 결과에 플리커링이 발생하는지 확인하였다. 실험환경은 Intel Core i5-3570@3.40Ghz 의 PC에서 MATLAB R2018a 버전을 사용하였다.



그림 3 memorial 영상의 톤 매핑 방법 비교
(a) 동적 계획법을 사용한 톤 매핑 방법 (b) 제안하는 톤 매핑 방법

그림 3은 memorial 영상의 실험 결과이다. (a)의 결과는 동적 계획법을 통한 k -means 글로벌 톤 매핑만 수행된 결과로 (b) 영상보다 지역적인 영상 강화가 약함을 알 수 있다. 또한 (b)영상에서 모서리 부분에서 후광효과 등의 결함이나 대비가 높은 영역의 선명도 문제가 없음을 알 수 있다.



그림 4 KitchenWindow 영상의 톤 매핑 방법 비교
(a) 동적 계획법을 사용한 톤 매핑 방법 (b) 제안하는 톤 매핑 방법

그림 4는 KitchenWindow 영상의 실험 결과이다. 극단적으로 밝은 영역을 확대한 영상이며 HDR 영상의 지역적인 디테일 정보가 효과적으로 강화

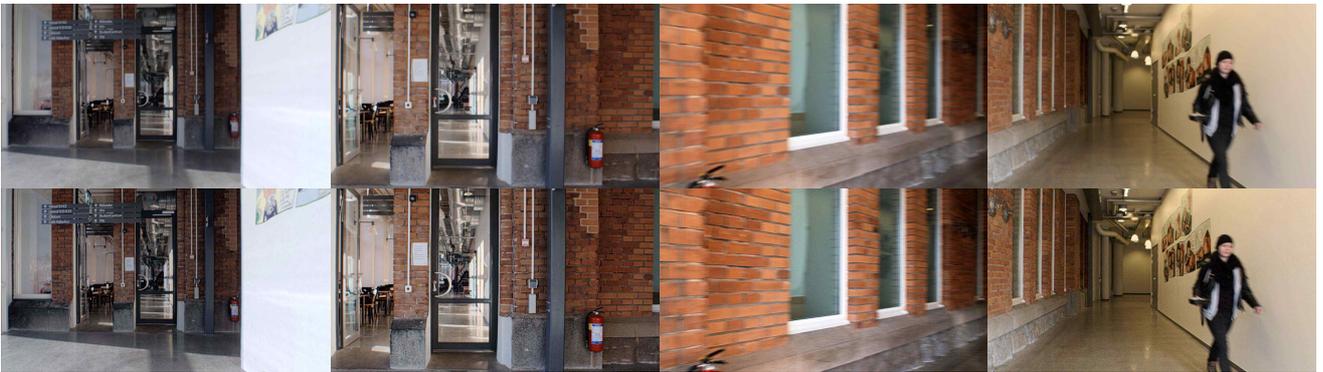


그림 5 hallway 영상의 각 1,100,200,300 번째 프레임을 비교한 결과 (위: 동적 계획법을 사용한 톤 매핑 방법, 아래: 제안하는 톤 매핑 방법)

되어 (b) 영상에서 강화 효과가 두드러져 선명함을 알 수 있다.

그림 5는 hallway 영상의 실험 결과이다. 플리커링 발생 여부를 보기 위해 연속적인 장면을 확인하였으며 플리커링이나 기타 다른 결함이 보이지 않았다.

5. 결론

본 논문에서는 동적 계획법을 통한 비디오 톤 매핑 방법을 제안하였다. 기존의 동적 계획법은 안정적인 영상 출력을 보여주지만 히스토그램을 통한 전역적 톤 매핑 방법이었기 때문에 디테일 향상이 이루어지지 않던 문제가 있었다. 이를 다층 레이어 분해를 통해 디테일 향상을 이룰 수 있었다. 제안하는 방법은 이전의 알고리즘에 비해 세부 영상 향상을 이루었고 극단적으로 동적 범위가 넓은 영상에서도 처리가 가능함을 보였다. 또한 실시간 비디오 처리에 문제가 없음을 확인하였다.

감사의 글

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음.[2014-0-00670, ICT 장비용 SW 플랫폼 구축]

참 고 문 헌

- [1] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph. Forum.*, vol. 22, no. 3, pp. 419-426, Sep. 2003.
- [2] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman, "Spatially Nonuniform Scaling Functions for High Contrast Images," in *Proc. of Graph. Interf.* 93, pp. 245-253, May 1993.
- [3] J. Tumblin, and G. Turk, "LCIS: A boundary hierarchy for detail-preserving contrast reduction." in *Proc. of the Computer graphics and interactive techniques*, pp. 83-90, July. 1999.
- [4] B. Gu, W. Li, M. Zhu and M. Wang, "Local Edge-Preserving Multiscale Decomposition for High Dynamic Range Image Tone Mapping," in *IEEE*

Transactions on Image Processing, vol. 22, no. 1, pp. 70-79, Jan. 2013.

- [5] G. Eilertsen, R. Wanat, R. Mantiuk and J. Unger, "Evaluation of tone mapping operators for hdr-video," in *Comput. Gr. Forum* Vol. 32, No. 7, pp. 275 - 284, Nov. 2013.
- [6] J. Kleinberg and É. Tardos, "Algorithm Design," Addison-Wesley, Boston (2005)
- [7] M. Oskarsson, "Temporally Consistent Tone Mapping of Images and Video Using Optimal K-means Clustering," in *J. Math Imaging Vis.*, Vol. 57, No. 2, pp. 225-238, Feb. 2017.
- [8] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high dynamic range images," *ACM Trans. Graphics*, vol. 21, no. 3, pp. 257-266, Jul. 2002.
- [9] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graphics*, vol. 27, no. 3, pp. 1-10, Aug. 2008.
- [10] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Learn.*, vol. 35, no. 6, pp. 1397-1409, Jun. 2013.