**3. NEXT SECTION etc**

3.1 Unsupervised Clustering을 사용한 Video key Frame Extraction

Key Frame Extraction을 위해 Clustering을 이용한다. Video의 서로 다른 Frame 들을 각 frame들 사이의 공통된 color bins수와 GLCM정보를 활용하여 유사도 측정 서로 다른 M개의 Cluster 로 분류하였다 [1]. 유사도 측정은 다음의 식을 활용하였다.

위의 식에서 는 서로 다른 두 frame 간에 공통된 Color bin의 수를 나타내며 이를 OpenCV의 로 구현하였다. 또한 은 서로 다른 두 frame 간의 GLCM texture 특징의 차이를 나타내며 Python skImage의 greycomatrix와 해당 논문과 같이 Euclidean distance method를 사용하였다.

3.2 object detection을 추가한 key Frame Extraction

Clustering 방식만을 사용하면 key Frame들이 촬영된 시간에 의존한다는 문제가 있다. 이러한 의존도를 낮춰주고 원하는 장면들을 많이 뽑아내기 위해 object detection을 추가하였다.

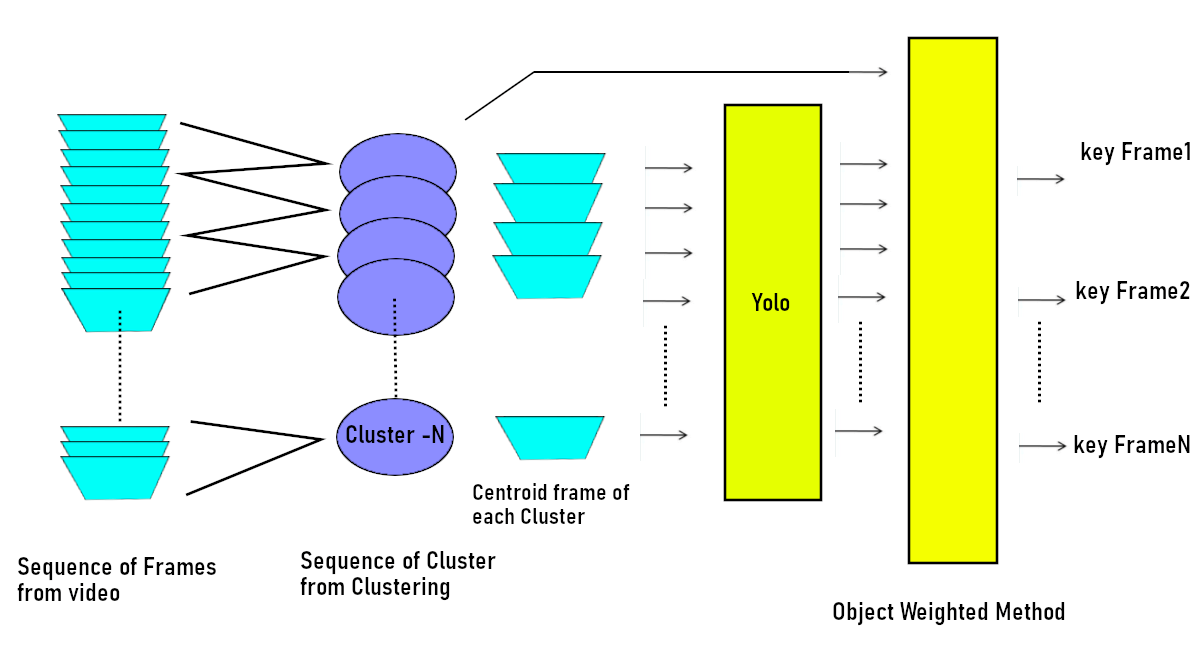
 Figure \*는 object detection이 추가된 key frame extraction을 보여준다. Clustering 방식으로 분류해낸 Frame들에서 각 Cluster의 중심 frame을 Object Detection 모델에 넣어 object label을 뽑아낸다. 이후 뽑아낸 label과 Cluster내부의 Frame 수를 이용하여 각각의 Cluster에서 중심 frame을 Video의 Key frame으로 뽑아낼지 결정하는 구조이다.

FIGURE \* : Object Detection을 적용한 Key Frame Extraction

key Frame을 뽑는 기준에 사용할 식은 다음과 같다.

은 k번째 Cluster 안의 frame수를 나타내며 는 k번째 Cluster의 대표 Frame에 있는 미리 설정된 object 중 번째 object의 가중치를 나타낸다. 따라서 는 k번째 Cluster 안의 Frame의 수와 대표 Frame의 모든 object들의 가중치를 합한 값이다. 같은 object에 대해서는 object detection에 사용한 모델의 특성상 겹쳐 있는 object들을 잘 판별하지 못하는 문제가 있어 한번만 적용하도록 하였다. Object의 가중치 는 과의 합으로 사용되며 key Frame Extraction은 결과에 따른 평가에 객관적인 지표가 없으며 주관적이기 때문에 여러 번의 실험을 통해 적절한 를 찾아야 한다. 본 논문에서는 ‘사람’에 가중치를 두었으며 정도로 진행하였다. 는 Object Detection model을 사용하여 나온 object label을 비교하여 설정한 object가 있으면 가중치가 적용되도록 하였다.

가 특정 임계 값(전체 frame 수 \* )을 넘으면 해당 Cluster의 대표 Frame을 Key Frame으로 뽑아낸다. 는 key frame extraction을 진행하며 실험적으로 적절한 값을 구하여야 한다. 퍼센트 기준 값이 너무 낮으면 중복된 frame이 많이 나오며 퍼센트 기준 값이 너무 높으면 원하는 frame이 나오지 않을 수 있다 [1]. 본 논문에서는 정도로 두었다.

향후 Key Frame Extraction을 Mobile Application에서 구현할 예정이므로 object detection의 속도가 중요한 요소가 된다. Object detection에 R-CNN을 사용해본 결과 오랜 시간이 걸려 Mobile 환경에서 사용자가 사용하기에는 적합하지 않다고 판단했다. 따라서 R-CNN보다 빠른 YOLO(You Only Look Once)모델을 사용했다. [2]YOLO는 Real Time Object Detection을 목표로 만들어졌으며 기존의 R-CNN 또는 Faster R-CNN 보다 빠른 속도를 보여준다. 또한 실생활의 많은 Object를 탐지할 수 있어야 함으로 우리는 COCO(large-scale object detection, segmentation, and captioning dataset)를 이용하여 미리 학습된 모델을 사용하였다.

[1] Janwe, N. J., & Bhoyar, K. K. (2016). Video key-frame extraction using unsupervised clustering and mutual comparison. *International Journal of Image Processing (IJIP)*, *10*(2), 73-84.

[2] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[3] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv:2004.10934*.

Coco <http://cocodataset.org/#home>