

<https://doi.org/10.7236/JIIBC.2022.22.3.1>  
JIIBC 2022-3-1

## 위치 분포 및 그래프 절단에 의한 모발 분류와 영역 분할

### Hair Classification and Region Segmentation by Location Distribution and Graph Cutting

김용길\*, 문경일\*\*

Yong-Gil Kim\*, Kyung-Il Moon\*\*

**요약** 최근 소개된 구글 MediaPipe의 모발 분할 방식은 실시간 모바일 애플리케이션을 위해 특별히 설계된 단일 카메라 입력에서 신경망 기반 모발 분할을 위한 새로운 접근 방식을 제시한다. 상대적으로 작은 신경망으로 가상 머리카락 다시 칠하기와 같은 증강 현실 효과에 매우 적합한 고품질 머리카락 분할 마스크를 생성한다. 그렇지만, 모발 스타일 또는 모발 영역에 잡음이 있는 경우에 모발 분할 정확도가 떨어지는 문제점들이 있다. 이에 본 연구에서는 지정된 라벨에서 모발 위치와 모발 색상 가능성의 추정된 사전 분포에 따라 이미지의 에너지 함수를 구성하고, 이것을 그래프 절단 알고리즘에 따라 최적화시키는 방식으로 초기 모발 영역을 얻는 방식을 도입한다. 그런 다음에, 초기 모발 영역에 클러스터링 알고리즘과 사후 처리 기법을 적용하여 최종 모발 영역을 정밀하게 분할 할 수 있도록 한다. 제안된 방식은 MediaPipe의 모발 분할 파이프라인에 적용된다.

**Abstract** Recently, Google MedeiaPipe presents a novel approach for neural network-based hair segmentation from a single camera input specifically designed for real-time, mobile application. Though neural network related to hair segmentation is relatively small size, it produces a high-quality hair segmentation mask that is well suited for AR effects such as a realistic hair recoloring. However, it has undesirable segmentation effects according to hair styles or in case of containing noises and holes. In this study, the energy function of the test image is constructed according to the estimated prior distributions of hair location and hair color likelihood function. It is further optimized according to graph cuts algorithm and initial hair region is obtained. Finally, clustering algorithm and image post-processing techniques are applied to the initial hair region so that the final hair region can be segmented precisely. The proposed method is applied to MediaPipe hair segmentation pipeline.

**Key Words** : Clustering, Energy function, Hair segmentation, Location distribution, Mediapipe

\*정회원, 조선이공대학교 컴퓨터보안과

\*\*정회원, 호남대학교 컴퓨터공학과(교신저자)

접수일자 2022년 4월 18일, 수정완료 2022년 5월 18일

게재확정일자 2022년 6월 10일

Received: 18 April, 2022 / Revised: 18 May, 2022 /

Accepted: 10 June, 2022

\*Corresponding Author: kimoon@honam.ac.kr

Department of computer engineering, Honam university, korea.

## I. 서론

모발 염색은 최근 몇 년 동안 인기 있는 증강 현실 애플리케이션이 되고 있다. 모발에는 다양한 색상 및 질감 정보가 포함되어 있으며, 이는 모발 스타일과 환경 조명 조건에 따라 크게 달라질 수 있다. 가장 큰 문제점은 현재 사용 가능한 모발 분할 데이터셋이 상대적으로 작다는 점이다. 결과적으로 실제 사용에서 여러 가지 환경 요인으로 인해 모발 분할이 쉽지 않다. 이러한 맥락에서 최근에 소개된 구글 MediaPipe의 모발 분할 방식은 실시간 모바일 애플리케이션을 위해 특별히 설계된 단일 카메라 입력에서 신경망 기반 모발 분할을 위한 새로운 접근 방식을 제시하고 있다. 상대적으로 작은 신경망으로 가상 머리카락 다시 칠하기와 같은 증강 현실 효과에 매우 적합한 고품질 머리카락 분할 마스크를 생성한다. 그렇지만, 모발 스타일 또는 모발 영역에 잡음이 있는 경우에 모발 분할 정확도가 떨어지는 문제점들이 있다.

머리 모양의 사전 탐지와 관련해서 제안된 대부분의 FCNN 모델들은 의미론적 분할 측면에서 객체 모양 제약이 명확하지 않다. 또한 모발 질감 측면에서 야생에서 모발 분할 설정과 관련해서 모발 영역의 크기가 상당히 가변적이라는 문제점이 있다. 이러한 문제점들의 완화와 관련하여 최근 연구는 완전 컨볼루션 신경망(FCNN)의 적용에 앞서 모발 모양을 통합하는 방식으로 진행되고 있다<sup>[1]</sup>. 특정 거리 변환을 기반으로 사전에 모발 모양을 찾기 위해 ASPP(Atrous Spatial Pyramid Pooling)<sup>[2]</sup> 모듈과 함께 완전 컨볼루션 신경망을 활용하고, 다음 단계에서, 모발 모양의 이전과 원본 이미지를 결합하여 대칭 인코더-디코더 완전 컨볼루션 신경망의 입력을 형성하여 최종 모발 분할 출력을 얻는 방식이다. 이와 관련된 정량적 및 정성적 결과는 공개적으로 사용 가능한 LFW-Part 및 Figaro1k 데이터셋에서 우수한 성능을 달성하고 있지만, 여전히 짧은 머리 모양의 애매함과 머리 자세 및 야생 환경에서 모발 영역 분할의 정확도는 만족스럽지 않다.

본 연구에서 모발 스타일은 긴 머리와 짧은 머리의 두 그룹으로 구분한다. 두 그룹에 대한 모발 위치의 사전 분포 이미지를 각각 추정하며, 훈련 이미지를 통해 모발 색상 가능성 분포 함수를 추정한다. 그런 다음 이미지 임계 기법에 따라 테스트 이미지의 모발 스타일을 추정한다. 또한, 모발 위치 사전 분포와 모발 색상의 사후 분포 함수에 따라 에너지 함수를 구성한다. 에너지 함수 최적화와 관련해서는 그래프 절단 알고리즘을 적

용하고 초기 모발 영역을 추정한다. 마지막으로 초기 모발 영역에 대한 클러스터링 및 사후 처리 방식을 통해 정확한 모발 영역을 추정한 후에 이미지 임계 기법에 따라 테스트 이미지의 모발 스타일을 추정한다. 제안된 기법은 MediaPipe의 모발 분할 파이프라인에 적용했다. 실험 결과 테스트 모발 이미지들에 대한 탐지 정확도는 기존 문헌에서 제시하고 있는 등급을 충족하고 있다. 특히 짧은 모발 영역에 대한 분할 정확도가 향상된 것으로 보이며, 모발 영역에 포함된 잡음 또는 구멍 등을 비교적 잘 처리하고 있는 것으로 확인되었다.

## II. 관련 연구

### 1. 모발 분류 및 분할

모발 영역은 인간의 얼굴 인식 과정에서 단서를 제공하며, 실험에 의하면 친숙한 사람을 인식할 때 머리카락 영역이 가장 중요한 특징임이 입증되었다<sup>[3]</sup>. 또한, 모발은 얼굴 애니메이션에 큰 영향을 미친다고 지적하고 있다<sup>[4]</sup>. 얼굴 인식에도 불구하고 모발 분할은 자동 얼굴 캐리커처 합성 시스템의 핵심 구성 요소이기도 하다. 모발의 특징은 앞서 언급한 몇 가지 장점이 있지만, 현재 대부분의 얼굴 인식 알고리즘은 모발을 고려하지 않고 얼굴 영역의 특징만 사용한다<sup>[5]</sup>. 얼굴 합성의 적용을 위해 일반적으로 밝고 균일한 배경으로 머리카락 영역을 수동 또는 자동으로 추출하는데, 이는 모발 모양이 매우 다양하고 색상도 사람마다 다르기 때문이다. 또한, 복잡한 조명 조건과 배경은 모발 분할 과정을 복잡하게 만든다. 이러한 이유로 영상 기반 모발 분할에 대한 연구는 매우 적은 편이다.

Yacoob과 Davis는 모발 감지 관련 체계적 연구를 수행한 선구자들이다<sup>[6]</sup>. 그들은 간단한 모발 색상 모델을 구성하고 영역 성장 기술을 통해 모발 영역을 감지한다. 얼굴 인식 과정에서 머리카락 길이, 볼륨, 색상, 대칭 및 머리카락 영역의 영향 분석을 목표로 하고 있다. 결과적으로 제안된 모발 검출 알고리즘은 매우 간단하며 배경이 균일한 이미지에만 적용할 수 있다. Lipowezky 등이 제안한 방식은 색상과 질감 정보를 융합하고 가상 모발 재작성 적용을 위한 퍼지 마스크를 통해 모발 영역을 표현한다<sup>[7]</sup>. 그렇지만, 정성적 결과만을 제공하고 정량적 분석이 부족하다. 알고리즘은 복잡한 이미지 매칭 기법을 적용하기 때문에 이미지당 약 5분 정도의 계산 비용도 상당히 높다. Rousset과 Coulon은 색상 정보와

주파수 영역 분석을 통해 초기 모발 영역 바이너리 마스크를 추정한다. 그런 다음 최종 모발 영역 마스크를 얻기 위해 이미지 매트 기술을 적용한다. 200×300 이미지의 평균 처리 시간은 3GHz CPU에서 대략 20초이다. 얼굴 캐리커처 합성 및 얼굴 애니메이션 적용을 위해 Chen 등은 모발 감지를 위한 녹아웃 이미지 매트 기술과 함께 수동으로 레이블이 지정된 헤어 라인을 사용한다<sup>[8]</sup>. Wan 등은 모발 영역 분할을 위해 이미지 임계값과 윤곽 추적 기술을 결합하고 있다<sup>[9]</sup>. 이 방법은 매우 간단하지만 복잡한 배경을 거의 다룰 수 없다.

최근 Wang 등은 분할 정복 방식으로 확률적 무발 마스크를 생성하기 위한 구성 모형 기반 모델을 제안했다<sup>[10]</sup>. 굵고 가는 모발 분할 방법과 관련하여 먼저 머리 종자를 얻기 위해 고정 및 평균 이동 기술과 능동 분할을 결합한 다음에 머리카락 가능성 확률을 예측하기 위해 SVM 분류기가 온라인으로 학습되는 방식도 제안되고 있다<sup>[11]</sup>. 제안된 방식은 약속된 결과를 보여주지만, 그 방법은 매우 복잡하고 구현하기가 어렵다. 또한 시간이 많이 소요되는 정교한 훈련 절차가 필요하다. 고품질의 반자동 분할 방식도 점차 등장하고 있는데, Li 등은 관심 영역 내에서 사용자가 제공하는 몇 번의 펜 스트로크를 사용하는 대화형 분할 도구를 제안했으며 해당 영역을 정확하게 추출하기 위해 알고리즘을 반복적으로 적용한 그래프 절단 기술을 제안했다<sup>[12]</sup>. 사실적인 모발 조각의 적용을 위해 Chai 등은 3D 모발 모델을 구성하기 위해 반자동 모발 분할 방법과 가다 추적 기술을 결합하는 것을 제안했다<sup>[13]</sup>. 이 방법은 LI 등이 제안한 방식과 유사하며, 모발의 세분화 단계 중에 모발 영역 안에서 여러 사용자 스트로크가 요구된다.

## 2. Google MediaPipe

MediaPipe는 이미지, 오디오 스트림 및 비디오 스트림과 같은 임의의 감각 데이터에 대한 추론을 수행하는 파이프라인을 구축하기 위한 체계이다<sup>[14]</sup>. MediaPipe를 사용하면 모델 추론, 미디어 처리 및 데이터 변환을 포함한 모듈식 구성 요소의 그래프로 인식 파이프라인을 구축할 수 있다. MediaPipe는 플랫폼 간 지원을 제공하는데, 그래프를 먼저 개발하고 데스크톱에서 테스트한 다음 모바일 장치에서 배포 및 최종 성능 평가를 수행할 수 있다. 중량급의 NN 기반 객체 감지기는 가벼운 템플릿 대응 감지기로 교체될 수 있으며 나머지 그래프는 변경되지 않은 상태로 유지될 수 있다. MediaPipe

를 사용하기 위해서는 몇 가지 기본 개념을 이해할 필요가 있다. 먼저 packet은 기본 데이터 흐름 단위를 나타내고, stream은 타임 스템프가 찍힌 패킷 시퀀스로 예를 들어 카메라의 비디오 스트림이 해당된다. side packet은 타임 스템프가 없는 단일 패킷으로 그것들은 기계 학습 모델, 구성 파일 등과 같은 정적이고 일회성의 입력을 제공하는데 사용될 수 있다. node들은 입력으로 입력 스트림 또는 입력 패킷들을 받아 데이터 변환, 미디어 처리 또는 모델 추론을 행하고 처리된 스트림 또는 패킷들을 출력한다. 각 node는 그것의 기능 코드에 해당하는 하나의 계산기 파일을 갖는다.

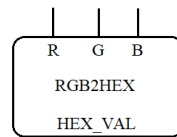


그림 1. RgbToHex  
 Fig. 1. RgbToHex

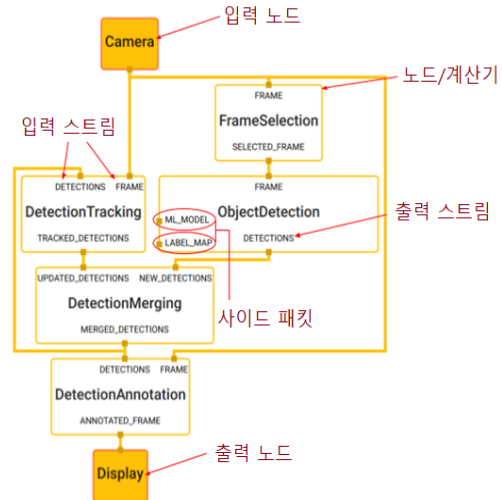


그림 2. 객체 탐지 그래프  
 Fig. 2. object detection graph

그림 1은 RGB 색상 값을 HEX 값으로 변환하는 node의 예이다. 여기서 R, G, B는 node에 대한 입력 패킷이고, HEX\_VAL은 출력 패킷이며, RGB2HEX는 node가 수행하는 계산기 파일에 해당한다. 마지막으로 graph는 node들의 네트워크로 각 node는 입출력으로 서로 연결되며, 연산 또는 작업을 나타낸다. 그림 2는 MediaPipe를 사용한 객체 탐지 그래프의 예이다. Camera 및 Display 상자는 그래프에 대한 외부 입출

력에 해당하고, 나머지 상자는 그래프에서 연산 노드(계산기)들을 나타내고, 노드의 상단으로 들어오는 선과 하단으로 나가는 선은 각각 입출력 스트림을 나타낸다. 파이프라인 저장소 내에서 코드를 수정하기 위해서는 저장소 폴더 내용을 이해할 필요가 있다. 최적화를 위해 mediapipe에서 사용되는 기술로 컴퓨팅 가속을 위한 GPU 활용과 프레임의 병렬 컴퓨팅을 들 수 있다. 입력 노드와 추론 노드 사이의 in-flight 이미지 프레임 수를 1로 제한하면 그 사이의 노드가 들어오는 이미지와 데이터를 과도하게 대기열에 넣는 것을 방지하여 실시간 모바일 애플리케이션에서 원하지 않는 대기 시간과 메모리 사용량을 증가시킬 수 있다. 여러 사물에 대해 하나의 모델 결과를 사용할 수 있는데, 예를 들어, 손 추적의 경우에 초기 프레임에서 hand\_landmark\_detection 모델을 사용하고 이어서 palm\_detection 모델을 사용한다. 추후 프레임에서 palm\_detection 모델은 이전 프레임에서 예측된 랜드마크를 직접 사용할 수 있는데, 랜드마크의 신뢰도가 낮게 예측된 경우에 실행된다.

### III. 모발의 분류 및 영역 분할

본 연구에서 RGB 색상 공간 대신에 YCbCr 색상 공간을 사용한다. 따라서 이미지에서 픽셀  $x=[u, v]^T$ 의 색상 값은  $I(x)=[Y(x), Cb(x), Cr(x)]^T$ 로 나타낸다. 픽셀  $x$ 에서 라벨은  $f_x \in \{0, 1\}$ 로 정의한다. 여기서 0은 모발이 없는 영역, 1은 모발 영역이다. 제안되는 알고리즘은 얼굴이 포함된 이미지에서 최적의 라벨링  $f_x$ 을 얻는 데에 있다.  $P(f_x)$ 는 라벨 위치의 사전 분포이고,  $P(I(x)|f_x)$ 는 라벨 색상 관련 사후 분포를 나타낸다.

#### 1. 모발의 분류

훈련 이미지에 관해 얼굴 감지 알고리즘을 사용하여 이미지에서 얼굴 영역을 감지한 후에 얼굴 영역을 확장하여 감지된 영역에 모발 영역을 포함하고 이미지 해상도를 정규화한다. 정규화된 각 훈련 이미지에 대해 모발 영역에 수동으로 라벨을 지정한다. 모든 픽셀  $x$ 에서  $P(f_x)$ 를 추정하는 데에 있어서 픽셀  $x$ 가 모발 영역에 속하는 이미지의 수와 훈련 이미지의 전체 수 사이의 비율을 설정할 수 있다. 그렇지만, 머리 스타일은 사람마다 크게 다르다. 따라서 본 연구에서는 긴 머리와 짧은 머리 스타일에 각각  $P(f_x)$ 와  $P(f_x)$ 를 대응시킨다. 머리카락의 아래쪽 테두리가 턱 아래에 있는 경우 긴 머리 스타일에 속한다고 정의하고, 그렇지 않으면 짧은 머리 스타일에 속하는 것으로 한다. 모발 색상 사후 분포  $P(I(x)|f_x=1)$ 는 라벨이 지정된 모발 영역에 속하는 YCbCr 색상 공간 정보로 설명된다. Y, Cb 및 Cr 채널 각각을 64개 레벨로 다운 샘플링한 후에 각 차원에 64개의 계급이 있는 3차원 색상 히스토그램을 구성할 수 있다. 모발 색상 사후 분포를 근사화하기 위해 색 히스토그램을 정규화한다. 모발 색상은 일반적으로 색상 공간의 조밀한 영역에 집중되므로 구현 중 메모리 공간 소비를 줄이기 위해 색상 히스토그램에 0이 아닌 계급을 저장할 수 있다. 모발이 없는 경우에 색상 사후 분포  $P(I(x)|f_x=0)$ 는 균등 분포로 근사시킨다.  $P(f_x)$ 와  $P(I(x))$ 의 계산을 위해서는 모든 테스트 이미지의 모발 스타일을 추정해야 한다. 테스트 이미지에서 모발 영역을 취한 다음에 머리 이미지를 정규화시킨다. 모발 스타일 분류 단계는 다음과 같이 작성할 수 있다. [단계 1] 아래 그림 3과 같이 세 개의 특징점  $F_1=[x_1, y_1]^T$ ,  $F_2=[x_2, y_2]^T$ ,  $F_3=[x_3, y_3]^T$ 을 탐지한다.

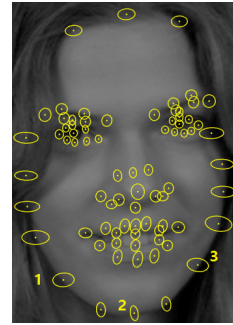


그림 3. 3개의 특징점 탐지  
 Fig. 3. 3 feature point detection

[단계 2] 색상 이미지를 명암 이미지로 변환하고 해당 히스토그램  $H$ 를 구성한다. 다음 방정식에 따라 분할 임계값  $t$ 를 설정한다.

$$t = \operatorname{argmax}_k \sum_{i=0}^k H(i) < \alpha M \quad (0 \leq k \leq 90) \quad (1)$$

여기서,  $H(i)$ 는 휘도 레벨  $i$ 의 픽셀 수이고,  $M$ 은 얼굴 이미지의 픽셀 수이다. 실험에 따라  $\alpha$ 값을 설정할 수 있는데, 여기에서는  $\alpha=0.2$ 로 설정한다. 또한 모발 영역이 작을 때 임계값이 너무 크지 않도록 90이하로 설정한다. [단계 3] 에 따라 얼굴 이진 이미지  $I_b$ 를 생성한다. 즉, 밝기의 값이  $t$ 이하이면  $I_b$  값을 0으로 설정한다는 것을 의미한다. 그렇지 않으면 1로 설정된다. 형태학적 필터링 기법을 사용하여 이진 이미지에서 잡음을 제거한다.

[단계 4] 다음 식에 따라 모발 스타일  $h_s$ 을 분류한다. 모발의 아래쪽 테두리가 턱 아래에 있으면 모발이 긴 스타일에 속한다고 정의했기 때문에, 결과적으로 목과 얼굴의 교차점에서 턱 아래 위치까지 연속적인 모발 픽셀이 있는지 감지하여 모발 스타일을 결정한다. 탐지 정확도와 견고성 간의 균형을 위해 모수  $\beta$ 를 사용한다. 실험에 따라  $\beta$ 를 설정하는데 여기에서는 0.2를 사용한다. 여기서  $h$ 는 이미지의 높이이고,  $H$ 는 수직 방향에서 히스토그램이다.

$$h_s = \begin{cases} long, & H_v(i) > 0, (y_1 + y_3)/2 \leq i \leq (1 - \beta)y_2 + \beta h \\ short, & otherwise \end{cases} \quad (2)$$

## 2. 모발 영역 분할

모발 영역의 분할을 위해 에너지 함수를 구성하고 그래프 절단 알고리즘으로 최적화한다. 에너지 함수는 관측값과 라벨 간에 가능성  $E_d(f_x)$ 과 인접 라벨과의 완만성  $E_s(f_x)$ 로 나타낼 수 있다.

$$E(f_x) = E_d(f_x) + E_s(f_x). \quad (3)$$

그렇지만, 이러한 에너지 함수는 너무 많은 변수를 포함하고 있어서 기존의 기울기 기반 최적화 알고리즘을 적용하기가 어렵다<sup>[15]</sup>. 에너지 함수의 최적화와 관련해서 그래프 절단 알고리즘을 적용할 수 있다<sup>[16]</sup>. 그래프 절단 알고리즘은 에너지 함수를 특정한 그래프 데이터 구조로 변환하고, 에너지 함수 최적화 문제를 그래프의 최소 절단 집합을 얻는 문제로 변환한다. 이 문제는 최대 유량 방식에 따라 효율적으로 풀 수 있다. 그래프 절단 알고리즘은 지금까지 컴퓨터 비전 문제에서 많은 성공적인 응용 프로그램을 얻고 있다<sup>[17]</sup>. 그래프 절단 알고리즘을 사용한 근사 최적 해는 다음과 같다.

$$\hat{f}_x = \arg \min_f E(f_x) \quad (4)$$

그래프 절단 알고리즘의 계산 복잡도는 입력 이미지의 해상도와 밀접한 관련이 있다<sup>[18]</sup>. 따라서 입력 이미지의 해상도를 적절한 크기로 down-sampling하여 대략적인 최적의 모발 분할 결과를 추정한다. 그런 다음 모발 영역을 원래 해상도로 up-sampling하면 된다. 그렇지만, 그래프 절단 알고리즘에 의한 근사 최적화 결과는 보편적으로 모발 영역이 배경 및 얼굴 영역 픽셀로 오염된다. 그 이유는 에너지 함수에  $E_s$ 의 항이 포함되어 있어 모발을 매끄럽게 만드는 경향이 있기 때문이다. 또한 더 낮은 해상도에서 초기 모발 영역 추정에 따른 오차가 있다. 이러한 문제점은 K-means 클러스터링 알고리즘을 적용하여 모발 영역을 개선할 수 있다<sup>[19]</sup>. [단계 1]

일반적으로 모발 영역은 인접한 배경 및 얼굴 영역의 영향을 받는다. 배경, 얼굴 및 모발 영역의 색상 정보는 비교적 균일하다. 따라서 모발 영역의 모든 픽셀에 대해 K-means 클러스터링 알고리즘을 적용하여 3개 그룹으로 구분한다. 3개 클러스터 중심에 관한 RGB 색상이 각각  $C_1, C_2, C_3$ 이라 할 때, 해당 밝기의 값  $Y_1, Y_2, Y_3$ 을 계산하면 각 3개의 클러스터에 해당하는 영역을 얻을 수 있다. [단계 2] 모발 영역은 일반적으로 밝기가 낮으므로  $k = \arg \min_i Y_i (i = 1, 2, 3)$ 의 클러스터에 속하는 것으로 가정할 수 있다. 이제 나머지 2개 클러스터에 속하는 픽셀들이 모발 영역에 속하는지 또는 그렇지 않은 것인지 결정한다. 모발 영역에서 밝기와 색상과 관련하여 어느 정도의 불균일성을 허용하는 다음 식이 충족되면, 클러스터  $C_i$ 에 속하는 픽셀들은 모발 영역에 해당한다.

$$Y_i < m Y_k \cdot |R_i - B_i| < 3\sigma_{rb} \cdot |G_i - B_i| < 3\sigma_{gb}, i \neq k \quad (5)$$

위 식에서  $m$ 은 모발 영역에서 밝기의 불균일성을 제어하는데, 훈련 이미지에서 모발 감지 정확도를 최대화하여 결정할 수 있다. 또한, 훈련 이미지에서 모발 영역의  $|R-B|$ 와  $|G-B|$ 의 히스토그램을 얻을 수 있다. 여기서  $3\sigma_{rb}$ 와  $3\sigma_{gb}$ 는 가우스 분포의  $3\sigma$  규칙에 해당한다. 즉, 95% 확률로 현재 구간에 속하기 위한 것이다. 이 단계에서 얻은 모발 영역에 잡음 또는 구멍이 있을 수 있는데, 이와 관련해서는 모발 영역의 연결된 노드에 라벨을 지정하고 해당 영역의 면적을 계산한다. 면적이 임계값보다 작은 영역은 모두 제거한다.

## 3. 분류 및 영역 분할의 MediaPipe 적용

TensorFlow Lite를 사용하여 모발 분할을 수행하는 MediaPipe의 기본 모발 분할 그래프의 노드는 그래프에 들어오고 나가는 GPU 상의 이미지 그래프 관련 노드 및 selfie-mode 테스트를 위한 노드로 시작하여 나머지 5개의 노드로 구성된다. 첫 번째 그래프 노드는 GPU 상의 입력 이미지를 512x512 이미지로 변환시킨다. 이미지 신축과 관련해서 기본적으로 전체 이미지를 전체 변환된 이미지로 대응시키기 위해 stretch 신축 방식을 사용하기 때문에 이미지 종횡비가 변경되어 이미지 내에 물체가 변형된다. 그렇지만 그래프에서 사용되는 모발 분할 모델은 그러한 변형을 인지하지 못한다. 두 번째 그래프 노드는 모발 영역의 전달 관련 패킷을 생성하는데, 알파 채널이 사용된다. 세 번째 그래프 노드는 이미지 텐서를 취해 모발 분할을 나타내는 텐서를

출력하는 TensorFlow Lite 모델을 수행한다. 네 번째 그래프 노드는 TensorFlow Lite 모델에 의해 생성된 분할 텐서를 0과 1 사이의 값으로 변환시키고, 마지막 그래프 노드는 지정된 색상으로 모발 영역을 나타낸다. 본 연구에서 제안된 방식과 관련하여 모발 스타일 분류는 두 번째 그래프 노드 상에서 처리되고, 모발 영역 분할은 세 번째 그래프 노드 상에서 진행되고, 모발 영역의 잡음 처리는 마지막 그래프 노드 상에서 처리된다(그림 4 참조).

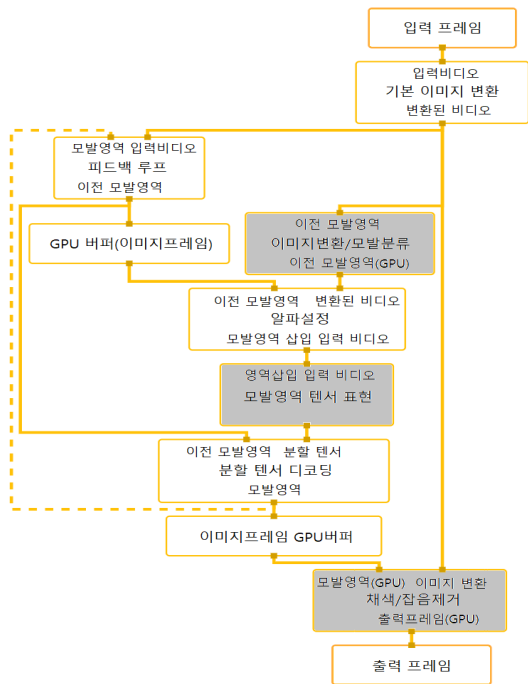


그림 4. MediaPipe 그래프 노드 수정  
Fig. 4. Modify MediaPipe Graph Nodes

의 분할 오차가 발생하는데, 긴 머리의 경우에 6개, 짧은 머리의 경우에 8개의 분할 오차가 발생하는 것으로 확인되었다. 결과적으로 모발 분할 과정에서 모발 스타일 분류가 중요하다. 모발 스타일 분류와 관련하여 히스토그램 상한값 관련  $\alpha$  값으로는 0.2를 사용하였고, 탐지 정확도와 관련된  $\beta$  값은 0.2를 사용하였다. 또한, 모발 분할 다듬기와 관련해서  $\sigma_{rb}$  및  $\sigma_{gb}$  값으로는 14와 10을 각각 사용하였다. 밝기의 값 계산과 관련해서는  $Y=0.299R+0.587G+0.114B$ 의 공식을 사용하였다.

```
<html>
<head>
<title>Hair Segmentation Graph Runner</title>
<link href="default_demo.css" rel="stylesheet">
<meta charset="utf-8">
<script>const templUnused = WebAssembly.Module;</script>
<script src="hair_segmentation_bin.js"></script>
<script src="hair_segmentation_packed_assets_loader.js"></script>
<script src="serializer.js"></script>
<script src="video_processor.js"></script>
<script src="default_demo_api.js"></script>
<script src="worker_api_wrapper.js"></script>
<script src="passthrough_renderer.js"></script>
</head>
<body>
<!-- VIDEO INTERFACE -->
<div id="video_interface" class="fill abs interface" style="display:
<video class="abs video" crossOrigin="anonymous" playsinline loop s
display: none; muted="muted">
Loading embedded camera/video failed.
</video>
<div class="video-display">
<div>
<canvas id="output" class="displayCanvas"></canvas>
</div>
<div id="message" class="abs message message-bgfill">Initializing
<div style="display: flex; align-items: center; bottom: 0; right:
<span>Hair Segmentation</span>
</div>
<div class="controls" style="display:none">
<select class="demo-selector">
</select>
</div>

<div class="help-page">
<iframe/>
</div>
</div>
</div>
</body>
</html>
```

그림 5. 모발 분할 그래프 러너  
Fig. 5. hair split graph runner

#### IV. 실험 결과

제안된 방법을 사용한 모발 분할 결과는 80개 이미지에 관해 (각각 40개의 짧은 머리 그룹과 긴 머리 그룹) 7개 모발 이미지 분할이 잘못된 것으로 나타났다. 세부적으로 짧은 머리 스타일의 경우가 5개, 긴 머리 스타일의 경우가 2개로 파악되었다. 이는 기존 연구 문헌에서 소개되고 있는 분류 오류 10%대에 근접하는 것으로 제안된 방법이 비교적 우수한 모발 분할 성능을 갖음이 확인되었다. 특히, 모발 스타일을 분류하지 않으면, 14개



그림 6. 모발 영역분할 결과  
Fig. 6. hair split areas result

MediaPipe에서 Tensorflow 모델 사용과 관련하여 TFLite 모델을 다운로드하고 필요한 종속성을 설치한 다음에 확장자가 ptxt인 그래프 파일을 mediapipe/graphs/portrait\_segmentation 폴더에 생성한다. MediaPipe의 hair\_segmentation 파이프라인을 사용하기 때문에, 본 연구에서 제안된 방식은 ptxt 파일을 복사하여 제안된 방식으로 수정하였다(그림 4 참조). 그래프용 BUILD 파일 생성과 관련해서는 그래프 생성 폴더에 확장자 없이 "BUILD"라는 파일을 만들어 hair\_segmentation의 BUILD 파일의 내용을 복사하였다. 이 파일에는 그래프 파이프라인에서 사용되는 계산기 종속성이 포함되어 있으며, txt 그래프(.ptxt)를 이진 그래프(binarypb)로 변환한다. MediaPipe Visualizer를 사용한 그래프 노드 작성은 그래프에 있는 계산기만 사용해야 하고, 그래프 실행에 있어서 단일 스레드 사용 문제, GPU의 텐서플로 추론이 지원되지 않는 등 몇 가지 제한 사항이 있다. 따라서, 일반 웹 브라우저상에서 모발 분할 작업이 용이하도록 그래프 노드들을 ptxt 파일 및 자바스크립트 파일들로 처리하였다. 이와 관련하여 그림 5는 모발 분할 그래프 수행 관련 index 파일이다. 그림 6의 우측(b) 이미지는 제안된 방식을 MediaPipe 그래프 노드에 적용한 결과이다. 그림 6의 좌측(a) 이미지는 기존 MediaPipe 모발 분할 알고리즘을 적용한 것으로 모발 스타일이 짧은 경우에는 분할 정도가 다소간 떨어지는 것이 확인되었으며, 특히, 모발 영역에 잡음이 있거나 모발이 단정하지 않으면 분할 정확도가 떨어지는 것으로 확인되었다.

## V. 결 론

GPU와 같은 가속 추론을 활용하는 복잡한 지각 파이프라인을 위해 특별히 설계된 개방형 소스 체제인 MediaPipe는 이러한 작업에 빠르고 비교적 정확한 솔루션을 제공한다고 할 수 있다. 사실 이러한 모든 것을 실시간으로 의미론적으로 일관되게 중단 간 솔루션으로 결합하는 것은 여러 종속된 신경망의 동시 추론을 요구하는 어려운 문제이다. 본 연구에서는 MediaPipe를 활용한 모발 분할 그래프 노드들을 소개하고, 개선점에서 모발 스타일 추가 및 모발 분할 다듬기와 관련하여 그래프 절단 및 클러스터링 알고리즘 적용, 모발 영역의 잡음 처리를 소개하였다. 이러한 개선과 관련하여 정량적 실험을 위한 이미지 데이터 개수가 다소간 충분하지

못해 제안된 모발 분할 방식이 강력하다고 할 수는 없지만, 기존 연구 문헌에서 제시하고 있는 분할 정확도를 충족하고 있음이 확인되었다. 끝으로 MediaPipe의 적용과 관련하여 MediaPipe Visualizer를 사용한 그래프 노드 작성은 몇 가지 제한 사항들로 인해 그래프 노드들을 작성하기가 쉽지 않다. 이에 일반 브라우저상에서 쉽게 사용할 수 있도록 자바스크립트로 노드들을 나타냈다.

## References

- [1] Y. Yongzhe, B. Anthony, D. Stefan, N. Xavier, G. Christophe, C. Thierry, "Human Hair Segmentation In The Wild Using Deep Shape Prior", 2020. DOI: <https://doi.org/10.1016/j.patrec.2020.06.014>.
- [2] V. Badrinarayanan, A. Kendall, R. C. Segnet, "A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", *IEEE Trans. Pattern Anal. Mach. Intell.* (12), 2481-2495, 2017. DOI: <https://doi.org/10.1109/TPAMI.2016.2644615>
- [3] P. Sinha and T. Poggio, "'United' we stand", *Perception*, vol. 31, no. 1, p. 133, 2002. DOI: <https://doi.org/10.1068/p3101no>
- [4] G. Davis, H. Ellis, and J. Shepherd, Eds., *Perceiving and Remembering Faces*, Academic Press, London, UK, 1981.
- [5] U. Lipowezky, O. Mamo, and A. Cohen, "Using integrated color and texture features for automatic hair detection", in *Proceedings of the IEEE 25th Convention of Electrical and Electronics Engineers in Israel*, pp. 51-55, December 2008. DOI: <https://doi.org/10.1109/EEEI.2008.4736632>
- [6] B. Klare, M. Burge, J. Klontz, R. Bruegge, and A. Jain, "Face recognition performance: role of demographic information", *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1789-1801, 2012. DOI: <https://doi.org/10.1109/TIFS.2012.2214212>
- [7] Y. Yacoob and L. S. Davis, "Detection and analysis of hair", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1164-1169, 2006. DOI: <https://doi.org/10.1109/TPAMI.2006.139>
- [8] H. Chen, Z. Liu, C. Rose, Y. Xu, H. Y. Shum, and D. Salesin, "Example-based composite sketching of human portraits", in *Proceedings of the 3rd International Symposium on Non-Photorealistic Animation and Rendering (NPAR '04)*, pp. 95-102, June 2004. DOI: <https://doi.org/10.1145/987657>
- [9] L. Wan, L. Dong, Y. Shen, "Creating a system with

realism of portrait and cartoon”, Computer Simulation, vol. 24, no. 5, pp. 169-172, 2007 (Chinese).  
DOI:https://doi.org/10.1155/2014/748634

- [10] N. Wang, H. Ai, and S. Lao, “A compositional exemplar-based model for hair segmentation”, in Proceedings of the Asian Conference on Computer Vision, pp. 171-184, 2010.  
DOI:https://doi.org/10.1007/978-3-642-19318-7\_14
- [11] D. Wang, X. Chai, H. Zhang, H. Chang, W. Zeng, and S. Shan, “A novel coarse-to-fine hair segmentation method”, in Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, pp. 233-238, March 2011.  
DOI: https://doi.org/10.1109/FG.2011.5771403
- [12] Y. Li, J. Sun, C. Tang, and H. Shum, “Lazy snapping”, ACM Transactions on Graphics, vol. 23, no. 3, pp. 303-308, 2004.  
DOI : https://doi.org/10.1145/1015706.1015719
- [13] P. Viola and M. J. Jones, “Robust real-time face detection”, International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.  
DOI:https://doi.org/10.1023/B:VISI.0000013087.49260.fb
- [14] V. Kolmogorov and R. Zabih, “What energy functions can be minimized via graph cuts?”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 2, pp. 147-159, 2004.  
DOI: https://doi.org/10.1109/TPAMI.2004.1262177
- [15] Y.G Kim, “Image Reconstruction Using Poisson Model Screened from Image Gradient,” The Journal of The Institute of Internet, Broadcasting and Communication(JIIBC) , Vol.18, No 2, pp. 117-123, 2018.  
DOI:https://doi.org/10.7236/JIIBC.2018.18.2.117
- [16] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. L. Chang, M. G. Yong, and J. Lee, Mediapipe: A framework for building perception pipelines. 2019.
- [17] A. Delong, Advances in graph-cut optimization: multi-surface models, label costs, and hierarchical costs [Ph.D. thesis], The University of Western Ontario, Ontario, Canada, 2011.
- [18] Y.G Kim, K.I Moon, “Image Restoration Based on Inverse Filtering Order and Power Spectrum Density,” The Journal of The Institute of Internet, Broadcasting and Communication(JIIBC) Vol.16 No. 2, pp. 113-122, 2016.  
DOI:http://dx.doi.org/10.7236/JIIBC.2016.16.2.113.
- [19] S. Lloyd, “Least squares quantization in PCM”, IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129-137, 1982.  
DOI: https://doi.org/10.1109/TIT.1982.1056489

## 저 자 소 개

### 김 용 길(정회원)



- 1990년 : 호남대학교 전산통계학과 졸업(이학사)
- 1992년 : 광주대학교 대학원 컴퓨터학과 졸업(공학석사)
- 2021년 : 호남대학교 대학원 컴퓨터공학과 졸업(공학박사)
- 2014년 ~ 현재 : 조선이공대학교 컴퓨터보안과 부교수

•관심분야 : 네트워크보안, 지능시스템, 정보보호

### 문 경 일(정회원)



- 1982년 : 서울대학교 계산통계학과 졸업(이학사)
- 1988년 : 서울대학교 대학원 계산통계학과 졸업(이학석사)
- 1991년 : 서울대학교 대학원 계산통계학과 졸업(이학박사)
- 1987년 ~ 현재 : 호남대학교 컴퓨터공학과 교수

•관심분야 : 지능 시스템, 복잡성 과학