

지능형 스마트 TV 응용을 위한 BGOLAM 기반의 성별분류

*오대영, **최지원, ***김창익

한국과학기술원 전기 및 전자공학과

{*Mclaren387, **G1choi}@kaist.ac.kr, ***Changick@kaist.ac.kr

BGOLAM-Based Gender Classification for Intelligent Smart TV Applications

*Oh, Daeyoung, **Choi, Jiwon, ***Kim, Changick

Department of Electrical Engineering, Korea Advanced Institute of Science and Technology(KAIST)

요약

최근 스마트폰, 태블릿 PC와 같은 모바일 스마트 디바이스(mobile smart devices)와 더불어 스마트 TV에 대한 관심이 크게 증가하면서 사용자들의 콘텐츠와 기능에 대한 요구 또한 다양해지고 있다. 스마트 TV가 콘텐츠와 기능적 측면에서 사용자의 편의와 재미, 그리고 유익함을 동시에 만족시키기 위해서는 더욱 지능화된 기능을 탑재할 필요가 있다. 일반적으로 남녀에 따라 TV를 시청하는 경향이 다르기 때문에 현재 TV를 시청하는 사용자의 성별분류(gender classification)를 통해 성별에 따른 다른 채널이나 광고, 응용 프로그램을 달리 제공하는 성별 기반의 스마트 TV 응용을 개발할 수 있게 된다. 본 논문에서는 스마트 TV 응용에 적합한 BGOLAM 기반의 성별분류 방법에 대해 제안하고, 실험을 통해 제안하는 방법의 적절성을 보인다.

1. 서론

최근 스마트 디바이스(smart devices)의 보급과 사용의 증가로 인해 사람들은 시간과 장소에 구애받지 않고 많은 일들을 할 수 있게 되었다. 언제 어디서나 인터넷(internet)에 접속해 오늘의 날씨, 주식, 뉴스, 교통 등 원하는 정보를 쉽게 얻을 수 있으며, TV 광고에서 본 제품에 대한 정보를 곧바로 TV를 통해 인터넷에 접속해 확인할 수도 있다. 지난 몇 년간 스마트 폰과 태블릿 PC로부터 촉발된 스마트 디바이스에 대한 관심이 스마트 TV에서도 이어지면서, 사용자들의 보다 다양한 콘텐츠와 기능에 대한 요구도 증가하고 있다. 스마트 TV가 콘텐츠와 기능적 측면에서 사용자의 편의와 재미, 그리고 유익함을 동시에 만족시키기 위해서는 TV와 사용자간의 상호작용 통해 또는 자동화된 사용자 인식을 통해 사용자가 원하는 바를 스스로 예측하여 그에 따른 적절한 정보를 제공해 주는 것이 매우 중요하다.

일반적으로 남녀에 따라 TV를 보는 경향이 다른 것을 볼 수 있는데, 예를 들어, 남자는 여자보다 스포츠 채널이나 자동차 광고, 혹은 여성이 모델인 광고에 더 관심을 갖는 반면, 여자는 드라마나 화장품 광고에 더 많은 관심을 갖는다. 이와 같은 남녀에 따른 차이로부터 발생하는 선호도에 관한 정보를 수집, 분석하여 다시 사용자에게 제공하기 위해서는 현재 TV를 시청하는 사용자의 성별에 관한 정보가 필요한데, 스마트 TV는 프로세서나 메모리와 같은 리소스 제한적인 환경(resource-limited environment) 이므로, 가능한 단순한 과정을 통해 빠르게 성별을 분류하면서도 성능저하를 최소화 하는 방법이 요구된다.

기존의 성별 분류 방법은 대략 생김새 기반(appearance-based) 방법과 특징 기반(feature-based) 방법 두 가지로 나눌 수 있다. 먼저, 생김새 기반 방법은 영상의 픽셀 값을 그대로 특징 벡터(feature vector)와 같이 사용한다. 1990년대 초, Cottrell [1], Golomb 등 [2]에

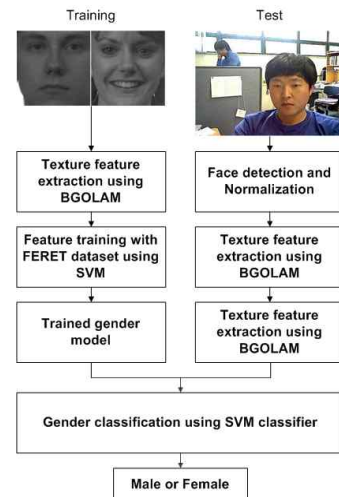


그림 1. 제안하는 방법의 전체 과정

의해 처음 성별분류 방법이 제안 되었는데, 생김새 기반의 방법을 성별 분류에 이용하였다. 생김새 기반의 방법은 영상의 전체적인 얼굴 정보를 표현할 수 있는 장점이 있으나, 원래 픽셀 값을 그대로 이용하기 때문에 얼굴의 위치(location), 표정(expression), 조도(illumination) 변화에 따라 지역적으로 생김새 특징이 달라지는 문제점이 있다. 이러한 문제점을 해결하기 위해 눈, 코, 입, 눈썹 등으로부터 눈썹의 두께, 눈과 눈 사이의 거리와 같은 기하학적 특징(geometric feature)을 이용한 방법이 제안 되었다 [3][4]. 이 기하학적 특징은 조도와 자세(pose) 변화에는 강건하지만, 위치 변화에는 여전히 취약하다. 두 번째, 특징 기반의 방법은 영상의 원래 픽셀 값을 우리가 얻고자 하는 특징을 기술하는 기술자(descriptor)를 이용해 다른 형태의 값으로 변형시켜 특징 벡터로 사용한다. Yang 등은 LBP(Local Binary Pattern)를 이용해 영

상의 텍스처(texture) 특징을 추출하여 성별 분류에 이용하였고 [5], Shobeirinejad 등은 기존의 LBP에서 중심 픽셀과 이웃 픽셀들 간의 미분방향 변화를 이용한 방법으로 수정한 IDP(Interlaced Derivative Pattern)를 특징벡터로 이용하였다 [6].

본 논문에서는 스마트 TV 응용을 위한 학습기반 성별분류 방법에 대하여 제안한다. 먼저, 학습단계에서는 정면얼굴영상(frontal face) 데이터로부터 제안하는 텍스처 특징 추출 방법을 이용해 특징벡터를 추출한 후, 추출한 특징벡터를 SVM(Support Vector Machine)을 통해 트레이닝 하여 성별 모델을 만든다. 테스트단계에서는 입력 비디오가 들어오면 Viola-Jones의 Haar-like 특징 기반 얼굴 검출 방법으로 영상에서 얼굴을 검출하고 [7], 검출된 얼굴 영역에 대해 히스토그램 평활화(histogram equalization) 과정을 거쳐 얼굴 영역을 정규화(normalization)한다. 마지막으로 정규화 된 얼굴 영역으로부터 추출한 텍스처 특징 벡터와 앞서 트레이닝 한 성별 모델을 이용해 최종적으로 남녀를 구별한다. 그림 1은 제안하는 방법의 전체 과정을 보여준다.

본 논문의 구성은 다음과 같다. 2장에서는 얼굴 검출과 검출된 얼굴 영역에 대한 정규화 과정에 대해 다룬다. 3장에서는 제안하는 텍스처 특징 추출 방법에 대하여 설명하고, 4장에서 실험결과를 통해 제안하는 방법의 적절성을 보인다. 마지막으로 5장에서 결론 및 향후 과제를 논의한다.

2. 얼굴 검출 및 얼굴영역 정규화

가. 얼굴 검출

성별 분류를 위한 첫 번째 단계로, 영상에서 얼굴 영역을 검출한다. 기존의 Viola-Jones Haar-like 특징 기반 얼굴 검출 방법을 이용해 영상에서 얼굴을 검출한다. 얼굴의 특징을 추출하기 위한 Haar-like 특징은 그림 2와 같이 영상의 특정 위치에서 두 개의 사각형 각 영역 내에 존재하는 픽셀의 밝기 값을 모두 더한 후 이 두 영역의 전체 밝기 값의 차이를 특징 벡터로 이용한다. 이렇게 구한 특징 벡터는 트레이닝을 거쳐 얼굴 모델을 만들고, 이 얼굴 모델을 이용해 입력 영상에 전체를 체크하면서 얼굴을 찾아낸다. 얼굴의 크기에 관계없이 검출하기 위해 다양한 크기의 검출 창(detection window)을 영상에 적용한다.

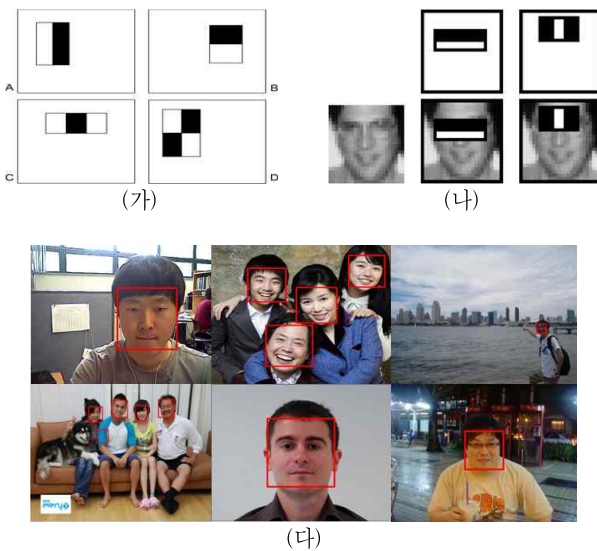


그림 2. 얼굴영역에서의 Haar-like 특징 추출[7] (가) Haar-like 특징 (나) 얼굴영역의 특징 추출 (다) 얼굴검출 결과

나. 얼굴영역 정규화



그림 3. 얼굴영역의 정규화 (가) 그레이 영상 (나) 정규화 영상

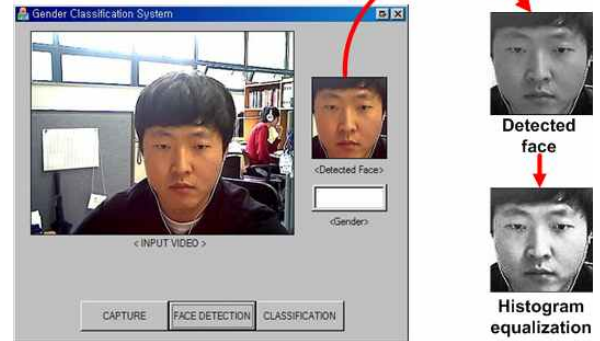


그림 4. 성별분류 시스템에서의 얼굴검출 및 얼굴영역 정규화

검출된 얼굴 영역은 얼굴의 구조적인 특징과 영상이 촬영될 때의 주변 조건에 의해 지역적인 조도 변화가 발생하여 분류성능 저하의 원인이 될 수 있다. 이와 같은 원인에 의한 분류성능 저하를 막기 위해 그림 3과 같이 히스토그램 평활화를 통해 영상의 밝기 분포를 고르게 펼쳐주는 과정을 거친다. 그림 4는 성별분류 시스템에 얼굴을 검출하고 검출된 얼굴 영역에 대한 정규화 과정을 적용한 모습이다.

3. 텍스처 특징 추출

검출된 얼굴영상으로부터 성별에 따른 텍스처 특징을 추출한다. 우리는 관찰을 통해 남녀에 따라 픽셀 값의 변화량이 다르게 나타나는 특징과, 얼굴의 구조적인 특징에 의해 텍스처 분포가 다르게 나타나는 사실을 근거로 텍스처 특징을 남녀를 구분하는 특징벡터로 이용한다. 기존의 BGLAM(Basic Gray Level Aura Matrix)은 아우라 집합(aura set)이라 하는 집합론적 개념과 이 집합의 원소의 개수를 셀으로써 영상의 텍스처 정보를 추출하는 방법이다 [8]. 여기서 아우라는 영상에서 우리가 찾고자 하는 텍스처의 분포를 나타낸다. 영상에서 256개의 그레이 레벨(gray level) 값들을 하나의 집합으로 보았을 때, 그 집합은 여러 개의 부분집합, 즉 다양한 텍스처의 집합들로 구성되어 있다. 무수히 많은 텍스처의 집합들에 대한 분포를 측정하면 텍스처 분포의 통계적 특성을 알 수 있고, 이 통계적 특성을 이용하여 행렬을 만든 것이 BGLAM이다. BGLAM은 영상의 텍스처 특징을 잘 표현하는 장점이 있어 텍스처 합성, 모델링 등의 응용에 주로 사용되어 왔다. 하지만 아우라 행렬(aura matrix)은 그레이 레벨의 개수에 비례하여 행렬이 생성되며, 따라서 256 그레이 레벨에 대한 아우라 행렬은 256x256 크기를 갖게 된다. 행렬의 크기는 곧 특징벡터의 차원의 개수를 의미하는데, 이 경우 BGLAM은 특징벡터의 차원이 너무 커 빠른 처리가 요구되는 스마트 TV 응용에 대한 특징벡터로 부적절하다. 그래서 우리는 픽셀의 그레이 레벨 값 대신 픽셀의 미분방향값(gradient orientation)을 이용하여 그레이 레벨 값을 양자화(quantization)하는 방법을 제안한다. 픽셀의 미분방향값을 이용하면 주변픽셀 간의 관계에 관한 정보까지 포함하므로 판별력(discriminant power)이 증가하고, 영상구조를 보다 잘 표현할 수 있는 장점과 함께 256 그레이 레벨을 10개의 레벨로 양자화 함으로써 특징벡터의 차원을 줄일 수 있다. 그림 5는 제안하는 BGOLAM의 생성 과정을 보여준다. 먼저, 입력 영상에서 픽셀의 그레이 레벨 값에 대한 미분방향 값을 구하고 각 픽셀이 갖는 각도를 10개로 나누어 양자화 한다. 그리고 기존의 BGLAM을 구하는 방법과 동일하게 아우라 행렬을 구하게 되는데, 여기서 아우라 행렬의 각 원소 값을 행렬의 전체 원소의 합으로 나누어 정규화 한 후 특징벡터로 사용한다.

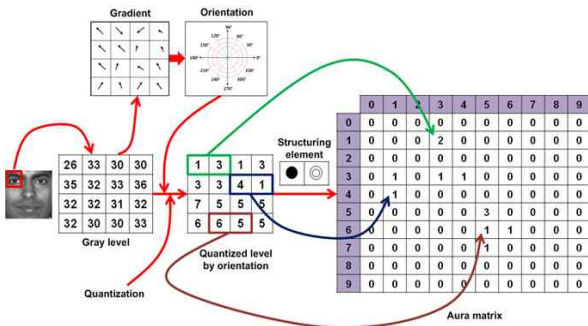


그림 5. BGOLAM 생성 과정

4. 실험결과

실험은 Intel(R) Core(TM)2 Duo CPU E8400 @ 3.00GHz, RAM 3.25GB PC의 Visual Studio 2010(C++) MFC 환경에서 이루어졌으며, 얼굴검출 및 얼굴영역 정규화 과정은 OpenCV 라이브러리를 사용하였다. 영상 데이터는 그림 6과 같은 FERET 데이터베이스로부터 임의로 남녀 각각 250장씩 총 500장의 영상에서 얼굴 영역만을 추출해 실험을 진행하였다. 그리고 신뢰성 검증을 위해 5-집단 교차검증 (5-fold cross validation)을 수행하였다.

먼저, 기존 방법들과의 분류 시간에 관한 비교 결과는 그림 5와 같다. 양자화 레벨의 크기를 다양하게 실험한 결과 가장 높은 분류정확도를 보이는 레벨이 10이었으며, 비교 결과는 BGOLAM의 양자화 레벨을 10으로 한 것을 기준으로 하였다. 더 큰 레벨로 양자화할 때 정확도가 떨어지는 원인은 미세하게 양자화할수록 남녀의 특징벡터의 겹침(overlap)이 발생하여 SVM에서 특징벡터를 구분하는 경계(boundary)가 모호해지기 때문이다. 그래프를 통해 제안하는 방법이 기존의 방법들보다 빠른 시간 안에 분류를 수행하는 것을 알 수 있었다.



그림 6. FERET 얼굴 데이터 (가) 원본 영상 (나) 검출된 얼굴영역

표 1. 성별 분류 시스템 수행 시간

단 계	수행시간(msec)	%
1. 얼굴검출	84.09	39.42
2. 얼굴영역 정규화	0.02	0.00009
3. 특징벡터 추출	40.85	19.15
4. 성별분류	88.42	41.45
총 계	213.3	100

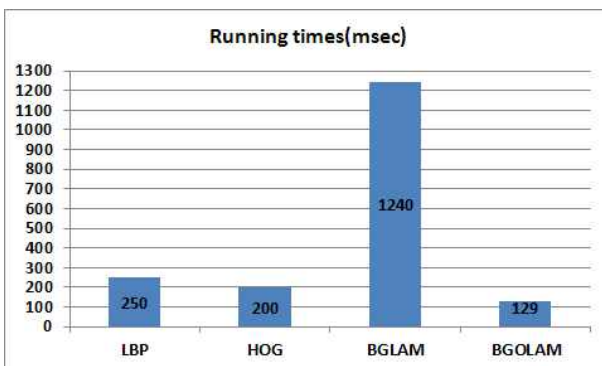
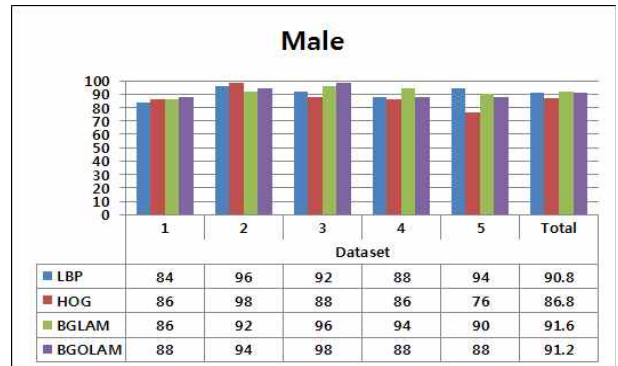
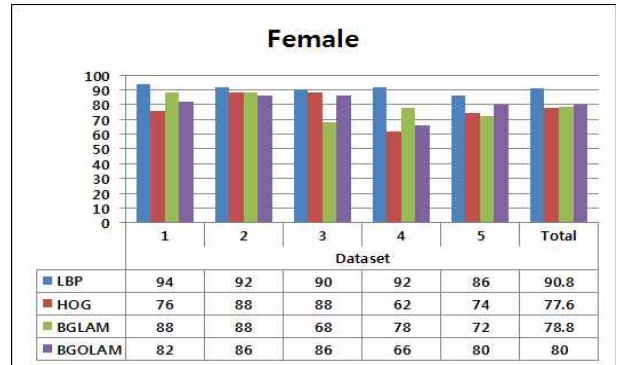


그림 6. 기존 방법들과의 분류 시간 비교



(가)



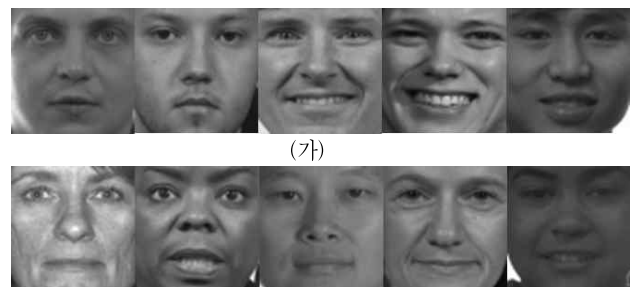
(나)



(다)

그림 7 기존 방법들과의 분류정확도 비교 (가) 남자 (나) 여자 (다) 전체 분류 성능

특히, BGLAM과 비교하여 제안하는 방법이 약 9.61배 빠른 분류 시간을 보였다. 그림 7은 분류정확도에 대한 비교 결과이다. 남자의 분류정확도(classification accuracy)는 조금씩 차이는 있지만 제안하는 방법이 높은 정확도를 보였다. 하지만 여자의 경우는 제안하는 BGOLAM이 HOG와 BGLAM보다는 높지만 LBP보다는 낮은 분류정확도를 보였다. 그림 7에서 보듯이 잘못 분류된 영상들은 육안으로 보기에도 구분이 어려운 경우가 많았으며, 특히 여자의 영상 데이터 셋에



(가)

(나)

그림 7. 잘못 분류된 영상의 예

남자 같은 여자의 영상 비율이 더 높았는데 이것이 분류정확도에서 여자의 분류정확도가 남자에 비해 낮은 결과를 초래한 원인 중의 하나로 보인다. 앞서 언급했듯이, 남녀에 따라 다르게 나타나는 픽셀의 변화량과 얼굴의 구조적인 특징이 만들어내는 텍스처 특징이 여자 같은 남자와 남자 같은 여자에서 반대로 나타나 잘못 분류된 결과를 초래한 것이다. 그런 측면에서 보았을 때, 그림 7 (나)의 여자 얼굴영상 같은 경우의 잘못 분류된 결과가 오히려 적절한 분류 결과 일수도 있을 것이라 생각된다.

5. 결론

본 논문에서는 스마트 TV 응용을 위한 성별분류 방법을 제안하였다. 간결한 과정을 통해 분류가 이루어지기 때문에 빠른 수행시간과 함께 기존 방법과 비교하여 성능저하를 최소화 할 수 있었다. 따라서, 스마트 TV뿐만 아니라, 서비스 로봇과 같은 제한적인 리소스를 갖는 실시간 응용에도 적용이 가능할 것으로 보인다. 그러나 가능한 간결하고 빠른 방법의 연구에 초점을 맞추어 개발하여 정확성 측면에서는 상대적으로 부족한 성능을 보인다. 따라서 향후 현재의 시스템 안에서 분류의 정확성을 향상 시킬 수 있는 방법과 남녀의 얼굴에서 성별을 구분짓는 주요 특징이 무엇인지에 대해 정량적으로 분석하고자 한다.

감사의 글

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음. (NIPA-2011-(C1090-1111-0003))

참고문헌

- [1] G. Cottrell, and J. Metcalfe, "Empath: Face, emotion, and gender recognition using holons," *Neural Information Processing Systems*, vol. 3, pp. 564-57, 1991.
- [2] B. Golomb, D. Lawrence, and T. Sejnowski, "Sexnet: A neural network identifies sex from human faces," *Advances in neural information processing systems*, vol.3,1991,pp.572-577.
- [3] S. Mozaffari, H. Behravan, and R. Akbari, "Gender classification using single frontal image per person: combination of appearance and geometric based features," *IEEE International Conference on Pattern Recognition*, pp. 1192-1195, 2010.
- [4] A. Samal, V. Subramani, and D. Marx, "Analysis of sexual dimorphism in human face," *Journal of Visual Communication and Image Representation*, vol. 18, issue 6, pp.453-463, Dec. 2007.
- [5] Z. Yang, and H. Ai, "Demographic Classification with Local Binary Patterns," *LNCS4642*, pp. 464-473, 2007.
- [6] A. Shobeirinejad, and Y. Gao, "Gender classification using interlaced derivative patterns," *IEEE International Conference on Pattern Recognition*, pp.1509-1512, 2010.
- [7] P. Viola, and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *IEEE International Conference on Computer Vision and Pattern Recognition*, vol.1, pp.1-9, 2001.
- [8] X. Qin, and Y.-H. Yang, "Basic Gray Level Aura Matrices: Theory and its Application to Texture Synthesis," *IEEE International Conference on Computer Vision*, pp. 1-8, 2005.