

<https://doi.org/10.7236/JIIBC.2021.21.6.105>
JIIBC 2021-6-15

새로운 Boosted 3-D PCA 기반 Head Pose Estimation 방법

A New Head Pose Estimation Method based on Boosted 3-D PCA

이경민*, 인치호**

Kyung-Min Lee*, Chi-Ho Lin**

요약 본 논문에서는 Boosted 3-D PCA 방법을 데이터 세트로 평가하고 성능을 평가한다. 그런 다음 네트워크의 특징과 성능을 분석하겠습니다. 본 논문에서는 Boosted 3-D PCA 학습방법을 사용하여 300W-LP 데이터 학습을 수행했으며 AFLW2000 데이터 세트를 사용하여 평가를 평가했다. 결과는 이 성능 결과는 기존 랜드마크 대 포즈 방법보다 자유롭게 얼굴 이미지의 데이터 세트를 사용하여 학습할 수 있으므로 실제 상황에서 포즈를 정확하게 예측할 수 있다. 키포인트 세트의 최적화는 독립적이지 않기 때문에, 우리는 계산 시간을 줄일 방법을 확인했다. 이 방법은 Boosted 3-D PCA 성능을 향상시키거나 다양한 애플리케이션 도메인에 적용하는 데 매우 중요한 자원이 될 것으로 예상된다

Abstract In this paper, we evaluate Boosted 3-D PCA as a Dataset and evaluate its performance. After that, we will analyze the network features and performance. In this paper, the learning was performed using the 300W-LP data set using the same learning method as Boosted 3-D PCA, and the evaluation was evaluated using the AFLW2000 data set. The results show that the performance is similar to that of the Boosted 3-D PCA paper. This performance result can be learned using the data set of face images freely than the existing Landmark-to-Pose method, so that the poses can be accurately predicted in real-world situations. Since the optimization of the set of key points is not independent, we confirmed the manual that can reduce the computation time. This analysis is expected to be a very important resource for improving the performance of network boosted 3-D PCA or applying it to various application domains.

Key Words : Explainable AI, Black Box, Boosted 3-D PCA, Head pose, Estimation

*학생회원, 세명대학교 컴퓨터학과
접수일자 2021년 10월 12일, 수정완료 2021년 11월 12일
게재확정일자 2021년 12월 10일

Received: 12 October, 2021 / Revised: 12 November, 2021 /
Accepted: 10 December, 2021

**Corresponding Author: ich410@semyung.ac.kr
School of Computer Science, Semyung University, Korea

I. 서 론

인공지능(AI, Artificial Intelligence) 기술의 핵심은 데이터를 주입하여 이를 학습함으로써 문제를 해결하는데 있다. 특히 자율주행 산업 분야에 많이 적용되는 기술이며, 이 산업이 커짐에 따라 교통안전에 대한 관심이 커지고 있다.^[1-6] 이러한 학습에 의한 방식은 절차를 알려주는 방식에 비해 다수의 변수가 영향을 미쳐서 단순한 수학 함수로 모형화하기 어려운 경우에 더욱 적합하며, 교통안전과 같은 다수의 변수가 존재하는 환경에 적합한 기술이다.

교통환경에 적합한 AI 기술은 수학적 관점에서 인공신경망(Artificial neural network)은 인공 뉴런의 연결 관계에 대응하는 다수의 선형 행렬 함수(Matrix)와 인공 뉴런의 상태 활성화를 모사하는 다양한 비선형 함수의 합성된 형태이다.^[7-9] 이러한 AI 기술은 방대한 규모의 행렬 연산과 합성 함수를 처리하기 위한 대규모 메모리를 처리할 수 있으나, 도출된 결과를 설명할 수 없다.

설명 가능한 AI(XAI, Explainable AI)는 딥러닝 기술이 주목받고 이를 통한 다양한 융합 기술이 나타났지만, 모델의 복잡한 구조로 인해 어떤 근거로 해당 결과를 얻었는지 이해하기 어려운 문제 해결하기 위한 기술이다. 머리 자세 추정 및 표정 추적에 관련한 연구는 지속해서 관심을 받고 있으며, 컴퓨터 비전 관점에서 머리 자세 추정이란 입력된 얼굴 영상으로부터 위치와 방향(roll, pitch, yaw)을 추정하는 과정으로 얼굴의 수학적 모델을 사용하는 랜드 마크 기반 접근법^[10,11] 과 매개 변수화된 외형 정보를 바탕으로 분류하는 모델 PAMs (parameterized appearance models) 등 크게 두 가지로 분류된다. 최근에 300W-LP[5] 셋을 사용한 머리 자세 추정에서 효율적인 성능 보이는 Fine-Grained Head Pose Estimation Without Keypoints^[12]의 HopeNet이 발표되었다. 이러한 연구들을 다양한 응용 분야에 활용하기 위해서는 알고리즘의 특징 및 성능을 정확히 분석하는 것이 매우 중요하다. 본 논문에서는 머리 자세 추정에서 효율적인 성능을 보이는 Boosted 3-D PCA를 데이터 셋으로 학습하고 평가하여 네트워크의 특징과 성능 분석하고자 한다^[13-14]. 이러한 분석은 머리 자세 추정 네트워크 Boosted 3-D PCA의 성능을 개선하거나, 다양한 응용 도메인에 적용하기 위한 매우 중요한 자료가 될 것으로 생각한다. 본 논문은 서론, Boosted 3-D PCA 알고리즘에 대한 설명, 머리 자세 추

정 네트워크를 학습하고 평가한 결과를 분석하면서 논문을 마친다.

II. Boosted 3-D PCA

머리 자세 추정 연구 중 300W-LP 데이터 세트 조건에서 좋은 성능을 보이는 Boosted 3-D PCA 은 컨볼루션 신경망을 사용하여 이미지로부터 3D 머리 자세 추정하기 위한 직접적이고 전체적인 접근 방식으로 대부분의 머리 자세 추정 연구 방식인 Key-Point 기반에 비해 뛰어난 정확도를 가진 네트워크이다. Boosted 3-D PCA 은 머리 자세 Euler angle을 예측하는 방법을 제안하였으며, 그림 1과 같이 Boosted 3-D PCA에서 Euler angle은 세 가지 angle(roll, pitch, yaw)마다 하나씩 교차 엔트로피 손실을 줌으로써 세 가지의 신호가 네트워크로 역 전파되어 학습을 향상할 수 있다. 그런 다음 세분된 예측을 개선하기 위해 평균 제곱 오류 손실 방법을 사용하는 회귀 손실을 추가하였다. 각 Euler angle의 최종 손실을 구하는 장치 식은 식(1)과 같다.

$$L = H(y, \hat{y}) + \alpha \cdot MSE(y, \hat{y}) \quad (1)$$

교차로 내 CCTV로부터 얻은 영상에서 정확한 보행자를 검출하기 위해서는 객체의 고성능 특징을 추출할 수 있는 특징 추출기가 필요하다. 특히, 본 논문에서 제안한 시스템은 교차로 환경에서 보행자를 검출하며 이에 대한 특징을 고려한 추출기를 구성하기 위해 PCA를 사용한다. 제안하는 특징 추출기 구조는 2-D PCA 알고리즘과 같이 PCA 연산량을 감소, 정확도 향상 그리고 대용량 입력 이미지를 처리할 수 있는 특징 추출기를 설계하고자 한다. 따라서 제안하는 특징추출 구조는 2-D PCA에서 한 차원 더 PCA를 수행하여 다량의 입력 데이터에 대한 객체 검출 성능을 강화하여 고성능 특징을 추출하는 방법을 제안한다. 제안하는 특징추출 구조는 그림 1과 같이 $W \times H \times D$ 의 이미지 데이터 집합에 대한 3차원 특징을 추출하며, 이를 3-D PCA라고 정의한다. 이 과정에서 Y-Z축에 대한 PCA의 공분산 행렬은 다음과 같이 나타낼 수 있다. 공분산 행렬 C_z 를 계산하기 위해 $z_j^{M \times m_2}$ 와 평균 이미지 \bar{Z} 에 의해 식(2)과 같이 나타낼 수 있다.

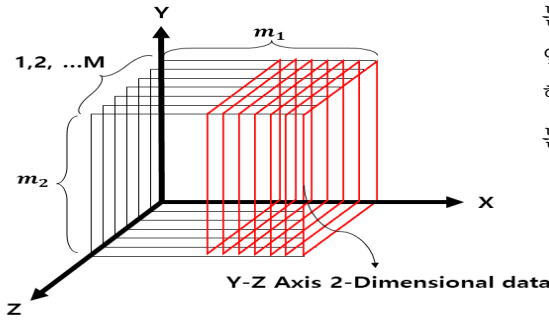


그림 1. Y-Z축 2차원 데이터
 Fig. 1. Y-Z axis 2-Dimensional data

식 (2)는 학습에 필요한 방대한 입력 데이터를 차원 축소를 통해 연산량을 줄일 수 있다. 또한, 차원 축소를 통한 객체의 위치 데이터 손실을 해결하기 위해 X-Y에서 Y-Z 축 변환을 통해 위치 데이터를 얻어 기존 방법 보다는 효과적인 얼굴 검출이 가능하다.

$$C_z = \frac{1}{m_1} \sum_{j=1}^{m_1} (z_j - \bar{Z})(z_j - \bar{Z})^T \quad (2)$$

이렇게 구해진 공분산 행렬 C_z 는 새로운 인식 후보 A가 주어질 때 $Y_z = U_z^T(A - \bar{Z})$ 로 투영된다. 이 과정의 특징은 X-Y 축과 비교하면 배경에 대한 분산이 크다. 따라서 Y-Z 축 데이터를 차원 축소할 경우 효율적인 배경과 전경 분할을 할 수 있는 장점이 있다. 그리고 X-Y축 PCA는 입력 이미지 $x_i^{m_1 \times m_2}$ 에 대한 평균 이미지 \bar{X} 를 통해 공분산 행렬을 계산하며, 공분산 행렬 계산의 복잡성을 개선하기 위해 세 번의 과정을 수행한다. 먼저, 행 방향 PCA는 행 방향 공분산 행렬 C_x 를 이용하여 $x^{m_1 \times dx}$ 만큼 차원을 축소한 특징 행렬 U_x 를 구한다. dx 은 행렬에서 dx 개의 큰 특징 행렬만큼 차원을 줄인 것을 표현하며, 행 방향의 특징 행렬은 $m_1 \times dx$ 의 크기를 갖는다. 이렇게 구해진 행렬 U_x 는 새로운 인식 후보 A가 주어질 때 $Y_x = U_x^T(A - \bar{X})$ 로 투영된다. 그리고 열 방향 PCA는 차원 이미지 열 방향으로 계산된 공분산 행렬 C_y 를 이용하여 $x^{dy \times m_2}$ 만큼 차원 축소한 특징 행렬 V 를 구한다. dy 은 행렬 V 에서 dy 개의 큰 특징 행렬만큼 차원을 줄인 것을 표현하며, 열 방향으로 차원을 축소한 PCA의 특징 행렬은 $dy \times m_2$ 의 크기를 갖는다. 이렇게 구해진 행렬 V

는 $Y_x = (A - \bar{X})V$ 로 투영된다. 마지막으로 행 그리고 열 방향 PCA 결과를 결합하며, 이렇게 생성된 2개의 특징 행렬들을 이용하여 새로운 인식 후보들이 입력되었을 때는 $Y_x = U_x^T(A - \bar{X})V$ 로 투영할 수 있다.

III. 실험 및 분석

본 논문에서는 Boosted 3-D PCA의 성능을 다시 확인하기 위해 동일 데이터 세트인 300W-LP 데이터 세트를 사용한다. 300W-LP 데이터 세트는 랜드 마크 검출 모델을 학습시키기 위한 데이터 세트로, 학습 데이터 세트는 얼굴 모델이 원본 이미지와 왜곡된 이미지로 구성되어 있으며, 왜곡된 이미지는 yaw의 angle이 변경되어 여러 yaw angle에서 머리 자세 추정 학습이 가능한 데이터 세트다. 또한, 다른 네트워크와 성능 평가 비교를 위해 사용되었던 AFLW2000 데이터 세트를 통해 성능 평가한다.

표 1. AFLW2000 데이터 세트 실험
 Table 1. AFLW2000 dataset experiment

Method	Yaw	Pitch	Roll	MAE
Multi-Loss ResNet50($\alpha = 2$)	6.430	6.555	5.433	6.149
Multi-Loss ResNet50($\alpha = 1$)	6.917	6.614	5.660	6.396
Multi-Loss ResNet50($\alpha = 2$) test	6.464	6.551	5.429	6.150
ours	6.910	6.632	5.668	6.405

boosted 3-D PCA의 학습 및 평가 결과는 표 1과 같이 비슷한 결과를 확인하였다. 표 1은 사람의 얼굴의 방향 추정을 위해 범용적으로 성능이 우수한 ResNet50을 사용한 결과와 본 논문에서 제한된 방법을 비교한 결과 표이다. 이 성능 결과는 기존 Landmark-to-Pose 방법보다 자유로운 얼굴 사진의 데이터 세트를 사용한 학습이 가능하여 실제 상황에서 정확하게 포즈를 예측이 가능하였고, 헤드 모델 선택, 랜드 마크 감지 방법, 헤드 모델 정렬에 사용되는 Keypoint의 집합에 대한 최적화가 독립적이지 않아 연산 시간을 단축할 수 있는 장점을 확인하였고, 사람의 바라보는 방향을 그림 2와 같이 추정할 수 있다.

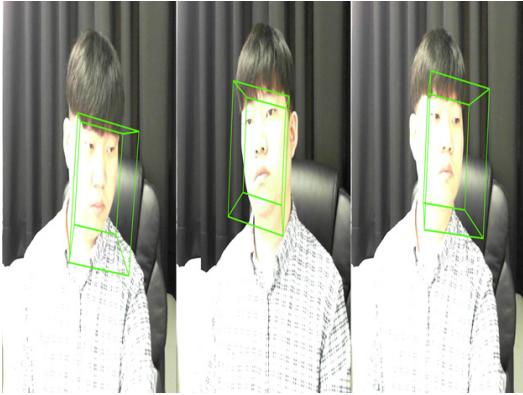


그림 2. Boosted 3-D PCA를 통한 머리 자세 추정 결과
Fig. 2. Results of head posture estimation through Boosted 3-D PCA

IV. 결 론

본 논문에서는 머리 자세 추정에서 좋은 성능을 보이는 boosted 3-D PCA의 성능을 개선하거나 다양한 응용 도메인에 적용하기 위해 boosted 3-D PCA를 300W-LP 데이터 셋으로 학습하고 AFLW2000 데이터 셋으로 평가하여 네트워크의 성능 및 특성을 분석하였다. 분석 결과 boosted 3-D PCA는 데이터 셋의 머리 자세 추정에 좋은 성능을 가졌으며, 학습 방식에 대해 다른 네트워크에 비해 간단하지만 효율적인 방법으로 연산 시간 및 크기를 줄이면서 좋은 성능을 가지는 것에 큰 의미가 있다. 이러한 분석자료는 boosted 3-D PCA를 개선하는데 많은 도움을 줄 수 있을 것으로 기대하며, 향후 다양한 응용 도메인에 적용한 가능한 개선된 boosted 3-D PCA 연구에 매우 중요한 자료가 것으로 생각한다.

References

- [1] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23 No. 6 pp. 681-685, 2001.
DOI: <https://doi.org/10.1109/34.927467>
- [2] I. Matthews and S. Baker, "Active Appearance Models Revisited", *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 135- 164, 2004.
DOI: <https://doi.org/10.1023/B:VISI.0000029666.37597.d3>
- [3] Xiangyu Zhu et al, "Face Alignment in Full Pose Range: A 3D Total Solution Supplemental Material", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 1, pp. 78-92, 2019.
DOI: <https://doi.org/10.1109/TPAMI.2017.2778152>
- [4] Jong Gun Song, Young Sil Lee, Won Tae Jang, HoonJae Lee and Tae Yong Kim, "Face Recognition Authentication Scheme for Mobile Banking System", *The Journal of The Institute of Internet, Broadcasting and Communication(JIIBC)*, Vol. 8, No. 2 pp.38-42, 2016.
<https://www.dbpia.co.kr/Article/NODE10252650>
- [5] Seung Ho Lee, "A Method for Determining Face Recognition Suitability of Face Image", *Journal of Korea Academia-Industrial cooperation Society*, Vol. 19 No. 11, pp. 295-302
<http://www.dbpia.co.kr/Article/NODE07578395>
- [6] Lee Hui Kueh, John Tark Lee, Kwon Soon Lee, "Face Recognition Using Newly Regularized LDA", *Journal of Korean Institute of Information Technology (JKIIT)*, Vol. 8, No 6, pp. 59-68, 2010.
<http://www.dbpia.co.kr/Article/NODE01464640>
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", *Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.
DOI: <https://doi.org/10.1109/CVPR.2016.90>
- [8] R. Girshick, "Fast R-CNN," *Proc of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440-1448, 2015.
DOI: <https://doi.org/10.1109/ICCV.2015.169>
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60, pp. 84-90, 2017.
DOI:<https://doi.org/10.1145/3065386>
- [10] J. Redmon, S. Divvala, R. Girshick, and A. F. "You Only Look Once: Unified, Real-Time Object Detection", *Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016
DOI: <https://doi.org/10.1109/CVPR.2016.91>
- [11] J. M. Saragih, S. Lucey, and J. F. Cohn. "Deformable model fitting by regularized landmark mean-shift", *International Journal of Computer Vision*, Vol. 91, No. 2, pp. 200-215, 2011.
DOI:<https://doi.org/10.1007/s11263-010-0380-4>
- [12] Xiangxin Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild", *In Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 2879-2886, 2012.
DOI: <https://doi.org/10.1109/CVPR.2012.6248014>
- [13] Nataniel Ruiz et al, "Fine-Grained Head Pose Estimation Without Keypoints", *The IEEE Conference on Computer Vision and Pattern Recognition Workshops(CVPRW)*, pp. 2074-2083, 2018.
DOI: <https://doi.org/10.1109/CVPRW.2018.00281>
- [14] G. Ma, A. Kummert, S.B. Park, S.M. Schneiders, and A.

Loffe, "A Symmetry Search and Filtering Algorithm for Vision Based Pedestrian Detection System," SAE Technical Paper, 2008..

DOI: <https://doi.org/10.4271/2008-01-1252>

저 자 소 개

이 경 민(학생회원)



- 현재 : 세명대학교 일반대학원 박사과정(컴퓨터학 전공)
- e-mail : leegis0904@semyung.ac.kr
- 주요관심분야 : 스마트응용, IoT, 빅데이터, 모바일컴퓨팅, 임베디드 시스템, Cad 알고리즘, SoC 설계, 자율주행시스템

인 치 호(정회원)



- 2020년 6월호 참조

※ 본 논문은 2021년 교육부 대학혁신지원사업에 의해 지원되었음