# A3C 기반 안저영상 왜곡 보정 기법

천성진*, 추현승**
*성균관대학교 인공지능학과, **성균관대학교 소프트웨어대학

chunsoungjin@skku.edu*, choo@skku.edu**

# A3C-based Fundus Image Distortion Correction Technique

Sungjin Chun*, Hyunseung Choo**
*Dept. of Artificial Intelligence, Sungyunkwan University
**College of Software, Sungkyunkwan University

**요 약**

안저 영상 촬영기술이 발달되며 진단에 사용되는 안저 영상에는 시각적으로 많은 변화가 일어났다. 새로운 촬영 기법인 초광각 안저 영상은 기존 영상에 비해 넓은 범위의 영상을 생성할 수 있다. 촬영 범위가 넓어짐에 따라 이미지에는 왜곡이 발생하고, 이로 인해 안저 영상을 통한 황반 부위 진단에 어려움을 야기하기도 한다. 본 논문에서는 이러한 왜곡을 보정하고 초광각 안저 영상을 기존 안저 영상의 영역으로 변환하는 시스템을 강화학습을 통해 구축한다. 제안하는 방법은 A3C 강화학습법을 사용하며 실험 결과는 제안 방법을 통해 안저 영상을 자동으로 변환할 수 있음을 보여준다.

## 1. 서론

광각 (UWF, ultrawide field) 안저 촬영술이 도입되며 기존 안저 촬영술에 비해 더 넓은 범위의 정보를 이미지에 담을 수 있게 되었다 [1]. 약 200 ° 범위의 정보를 한 장의 이미지에 담으며, 동일한 크기의 이미지에서 황반부 변성 정보가 소실되어 확인이 어려운 문제가 발생한다. 또한 기존에 45 ° 로 촬영된 안저 영상과 비교하였을 때 왜곡이 발생하는 것을 확인할 수 있다.

본 논문에서는 강화학습을 이용하여 초광각 안저 촬영술로 찍힌 안저 영상을 기존 촬영 기법으로 생성된 안저 영상의 영역으로 바꾸는 기술을 제안한다. 강화학습 알고리즘은 Asynchronous Advantage Actor-Critic (A3C)을 사용한다. 제안하는 방법은 안저 영상 외의 다른 domain 의 이미지를 처리할 때도 사용할 수 있을 것으로 기대한다.

## 2. Asynchronous Advantage Actor-Critic

A3C 는 여러 Agent 들이 탐색을 하며 비동기적으로 policy update 를 하는 강화학습법이다 [2]. A3C 는 보상에서 예측한 가치를 차감하는 Advantage 개념을 정의하고 이를 loss 에 사용한다. Reward 를 R, 예측한 가치를 V 라 한다면 Advantage 는 수식 1 과 같이 표현된다.

$$Advantage = R - V \quad (1)$$

Advantage 는 올바른 학습이 진행 중이거나 과소 평가가 발생한 경우 양의 값을, 과대평가가 이뤄진 경우 음의 값을 갖는다. A3C 는 정의된 Advantage 를 사용하여 두 가지 loss 를 정의한다:

$$Value\ loss = Advantage2 \quad (2)$$

Value loss 는 Advantage 의 제곱으로 정의한다. 이는 실제 받는 보상을 과대, 과소평가하지 않게 조절하는 역할을 한다.

Policy loss 는 Softmax Cross Entropy 를 이용하여 정의한다. Softmax 는 어떤 입력 x 에 대해 수식 3 과 같이 정의한다.

$$Softmax(x) = exp(x) / \sum(exp(x)) \quad (3)$$

어떤 입력 x 에 대한 분포를 q(x), 목표하고자 하는 출력의 분포를 p(x)라 하면 수식 4 와 같이 표현된다.

$$Cross\ Entropy(x) = -\sum(p(x) * log(q(x))) \quad (4)$$

Softmax Cross Entropy 는 Cross Entropy 의 q(x)를 Softmax(x)로 대체한다. 즉, 어떤 입력의 분포 q(x)를 어떤 입력이 갖는 상대적 분포로 변환하여 입력으로

해석한다. Softmax Cross Entropy 를 S_CE 로 간략히 줄이면, 수식 5 와 같이 표현된다.

$$S\_CE(p(x), x) = -\sum(p(x) * log(Softmax(x))) \quad (5)$$

Policy loss 는 agent 의 action 을 label 로, Policy 의 output 을 logits 으로 사용한다. Cross Entropy 의 p(x)는 action 에, q(x)는 policy 에 대응한다. 수식 1 과 5 에 따라 Policy loss 를 수식 6 과 같이 정의한다.

$$Policy\ loss = S\_CE(action, policy) * advantage \quad (6)$$

Policy loss 는 정책이 출력되는 확률이 확신을 갖고 학습하되, 학습에서 잘못된 경험이 발생한 것은 advantage 에 따라 역방향 학습이 이뤄질 수 있도록 구성된다. A3C 의 final loss 는 정의된 value loss (수식 2)와 policy loss (수식 6)에 따라 수식 7 로 표현한다.

$$Final\ loss = policy\ loss + 0.5 * value\ loss \quad (7)$$

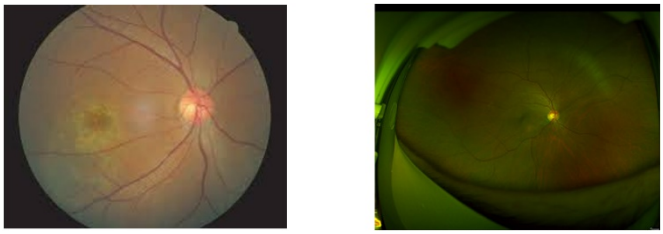
**Algorithm S2** Asynchronous advantage actor-critic - pseudocode for each actor-learner thread.

*// Assume global shared parameter vectors $\theta$ and $\theta_v$ and global shared counter $T = 0$*
*// Assume thread-specific parameter vectors $\theta'$ and $\theta'_v$*
Initialize thread step counter $t \leftarrow 1$
**repeat**
  Reset gradients: $d\theta \leftarrow 0$ and $d\theta_v \leftarrow 0$.
  Synchronize thread-specific parameters $\theta' = \theta$ and $\theta'_v = \theta_v$
  $t_{start} = t$
  Get state $s_t$
  **repeat**
    Perform $a_t$ according to policy $\pi(a_t|s_t;\theta')$
    Receive reward $r_t$ and new state $s_{t+1}$
    $t \leftarrow t+1$
    $T \leftarrow T+1$
  **until** terminal $s_t$ **or** $t - t_{start} == t_{max}$
  $R = \begin{cases} 0 & \text{for terminal } s_t \\ V(s_t, \theta'_v) & \text{for non-terminal } s_t \text{// Bootstrap from last state} \end{cases}$
  **for** $i \in \{t-1, \ldots, t_{start}\}$ **do**
    $R \leftarrow r_i + \gamma R$
    Accumulate gradients wrt $\theta'$: $d\theta \leftarrow d\theta + \nabla_{\theta'} \log \pi(a_i|s_i;\theta')(R - V(s_i;\theta'_v))$
    Accumulate gradients wrt $\theta'_v$: $d\theta_v \leftarrow d\theta_v + \partial (R - V(s_i;\theta'_v))^2 / \partial \theta'_v$
  **end for**
  Perform asynchronous update of $\theta$ using $d\theta$ and of $\theta_v$ using $d\theta_v$.
**until** $T > T_{max}$

(그림 1) A3C 알고리즘 의사 코드 [2]

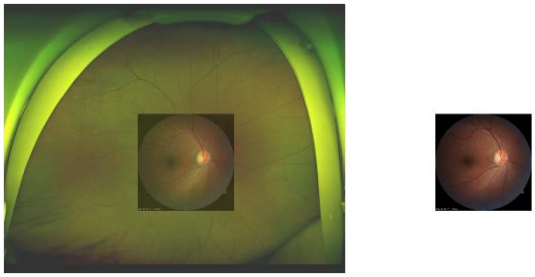
## 3. A3C based Fundus Image Processing System

A3C 를 사용하여 안저영상 변환을 처리하는 모델을 구축한다. 변화하고자 하는 안저 영상은 그림 2 와 같다.



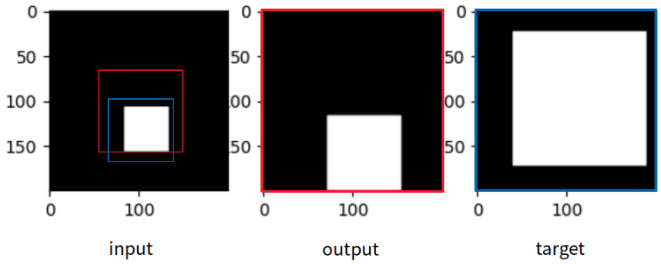
(그림 2) 기존 안저영상(좌), 초광각 안저영상(우)

초광각 안저영상을 사용한 진단은 기존의 안저영상으로 진단할 수 있는 병변을 놓칠 수 있는 가능성이 있다 [3]. 이를 해결하기 위해 촬영기법이 바뀜에 따라 소실되거나 놓칠 수 있는 정보를 초광각 안저영상에 기존의 안저영상 정보를 입힐 수 있는 시스템을 제안한다.



(그림 3) 기존 안저영상과 초광각 안저영상의 촬영 영역

그림 3 에서 볼 수 있듯이, 기존 안저영상으로부터 추가될 정보들은 초광각 안저영상의 특정 영역에 mapping 된다. 강화학습 agent 는 이러한 영역을 자동으로 탐색하고 영상에서 자동으로 cropping 하는 action 을 하도록 디자인한다. 초광각 안저사진에서 cropping 될 목표 영역과 그 외의 영역을 추상화하여 agent 의 입력으로 사용한다. 목표 영역을 흰색, 그 외의 영역을 검은색으로 추상화하고 agent 는 cropping window 의 좌표 값을 action 출력 값으로 갖는다. Agent 가 출력한 좌표 값을 따라 cropping window 를 설정하고 cropping window 를 따라 이미지를 잘랐을 때 결과 이미지가 목표 영역과 일치하는 정도에 따라 reward 를 설정한다.

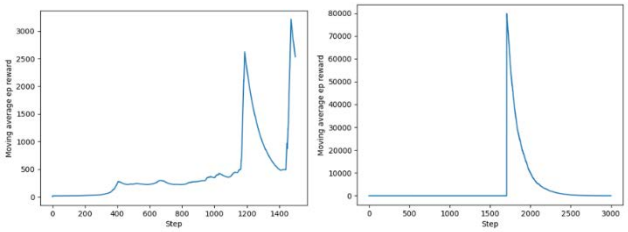



(그림 4) target area (blue), action window (red)

그림 4 의 red window 은 agent 의 action output 을 따라 설정된다. 설정된 red window 을 따라 최종 output 영상이 설정되고, 환경은 target area 와 최종 output 영상의 차를 이용하여 reward 를 계산한다. 이러한 작업은 A3C 의 각 agent 별로 반복되며 탐색하고 학습된다.

## 4. 실험 환경 및 결과

Reward 는 target area 좌표와 red window 좌표의 distance 를 기반으로 구성한다. 추가적으로 Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), 그리고 RMSE (pixel by pixel)를 사용하였다. RMSE (pixel by pixel)은 penalty reward 로 사용한다. 실험을 진행하며 평균적인 reward 그래프를 출력하며 학습의 정도를 비교하고 분석한다. CartPole 문제를 A3C 로 해결하고 도출된 reward 그래프를 대조군 그래프로 사용한다. 제안한 방법의 학습 정도를 reward 그래프로 표현하고 두 그래프를 비교 분석한다.




(그림 5) 수렴 완료된 CartPole reward 그래프(좌), 제안하는 방법의 reard 그래프(우)

A3C 는 병렬 탐색 특성과 주기적인 optimal policy 를 master agent 로부터 동기화 받는 특성에 의하여 학습 진행에 따른 평균 reward 그래프가 튀는 형태를 띈다. 그림 5 의 왼쪽 그래프는 학습이 비교적 쉽게 이뤄지는 CartPole 문제를 해결하며 그려지는 A3C 의 평균 reward 그래프다. 그림 5 의 오른쪽 그래프는 제안하는 방법이 학습 중에 만드는 그래프이며, 그래프 양상을 비교하면 탐색 횟수, reward 의 spike 정도는 다르지만 동일하게 학습이 진행됨을 확인할 수 있다.

## 5. 결론 및 향후 연구 계획

제안하는 방법을 통해 안저 이미지 처리 과정을 확인할 수 있다. 이러한 시스템을 기반으로 이미지 처리를 강화학습을 통해 처리하고 출력된 이미지를 기반으로 warping domain 을 맞출 수 있다. 학습된 모델은 이미지를 입력을 받아 목표하는 영역으로 자동으로 처리하는 모델이다. 진행된 연구 결과를 기반으로 3000 x 3000 초광각 안저 영상 입력을 자동으로 alignment 하는 시스템 연구를 계속할 예정이다.

Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 글로벌핵심인재양성지원사업 (2019-0-01579), ICT 명품인재양성 사업(IITP-2020-0-01821), 인공지능대학원 (No.2019-0-00421)의 지원을 받아 수행된 연구임

## 참고문헌

[1] Takao Hirano, et al. "Assessment of diabetic retinopathy using two ultra-wide-field fundus imaging systems, the Clarus® and Optos™ systems," BMC ophthalmology 18.1 (2018): 1-7.

[2] Mnih Volodymyr, et al. "Asynchronous methods for deep reinforcement learning," International conference on machine learning, PMLR, 2016.

[3] Lee, Dong Hyun, et al. "Identifiable peripheral retinal lesions using ultra-widefield scanning laser ophthalmoscope and its usefulness in myopic patients." J Korean Ophthalmol Soc 55.12 (2014): 1814-1820.