An Effective and Efficient Method for Detecting Hands in Egocentric Videos for Rehabilitation Applications

Abstract

Abstract— *Objective:* Individuals with spinal cord injury (SCI) report upper limb function as their top recovery priority. To accurately represent the true impact of new interventions on patient function and independence, evaluation should occur in a natural setting. Wearable cameras can be used to monitor hand function at home, using computer vision to automatically analyze the resulting videos (egocentric video). A key step in this process, hand detection, is difficult to accomplish robustly and reliably, hindering the deployment of a complete monitoring system in the home and community. We propose an accurate and efficient hand detection method that uses a simple combination of existing detection and tracking algorithms. *Methods:* Detection, tracking, and combination algorithms were evaluated on a new hand detection dataset, consisting of 167,622 frames of egocentric videos collected from 17 individuals with SCI performing activities of daily living in a home simulation laboratory. *Results:* The F1-scores for the best detector and tracker alone (SSD and Median Flow) were 0.90±0.07 and 0.42±0.18, respectively. The best combination method, in which a detector was used to initialize and reset a tracker, resulted in an F1-score of 0.87±0.07 while being two times faster than the fastest detector alone. *Conclusion:* The combination of the fastest detector and best tracker improved the accuracy over online trackers while improving the speed over detectors. *Significance:* The method proposed here, in combination with wearable cameras, will help clinicians directly measure hand function in a patient's daily life at home, enabling independence after SCI.

(Bu et al.)이 논문과 가장 비교 할 만한 논문으로

Faster R-CNN detector와 KCF tracker을 결합해 multi-object tracking in 3인칭 영상(third-person video)에 사용함

비교하려는 논문에선 (multi-object detection에서는)simultaneous computation이 필요하지만,

여기 논문에서는 (one type of object)에 대해서는 이러한 복잡성이 필요로 하지 않음을 보임

Main Issue: Object Detection과 Object Tracking model을 결합해 real-time base의 Egocentric Detection을 하겠다! (target for SCI(척수 손상 환자))

이과정의 핵심 단계인 hand detection은 강력하고 안전하게 달성하기가 어려워, home, community에 완벽한 모니터링 시스템을 구축하기 힘듦

Introduction

CERVICAL spinal cord injuries (SCI) significantly reduce the quality of life of the affected individuals and entails an estimated economic cost of $2.7 billion per year in Canada [1]. In particular, the impairment of arm and hand function

동기

improve hand function after SCI are needed. Current assessments of the severity of upper limb impairments are typically performed in clinical settings. To accurately represent the true impact that new interventions have on patient function and independence, evaluation should occur at home. Currently, there are no methods that directly measure and track the effect of therapy on patient hand function in their daily life at home.

상지 손상의 심각성에 대한 판단은 일반적으로 임상환경에서 수행됨

환자의 기능과 독립성에 미치는 실제 영향을 정확하게 나타내려면, 가정에서의 평가를 수행해야한다.

환자의 기능과 독립성에 미치는 실제 영향을 정확하게 나타내려면, 가정에서의 평가를 수행해야한다.

daily activities. Home rehabilitation is of utmost interest as the natural movement information provided by wearable cameras can be used to monitor patient performance and independence in activities of daily living (ADLs), and provide feedback for more effective and more accessible rehabilitation.

wearable camera가 제공하는 자연스러운 움직임 정보를 사용해, 일상 생활활동(ADL)에서 환자의 성과와 독립성을 monitoring하고 보다 효과적으로 접근하기 쉬운 재활을 위한 피드백을 제공할 수 있으므로 home rehabilitation(가정재활)이 가장 중요함

a necessary first step prior to hand function analysis. Robustly and reliably detecting and tracking the hand is affected by factors including partial occlusions, lighting variations, hand articulations, camera motion, and background or objects that are similar in color to the skin.

hand detection 할 때, 고려할 요인 및 해결 과제

손을 강력하고 안정적으로 detection, tracking하는 것은 부분 폐색(partial occlusions), 조명 변화, 손관절, 카메라 동작 및 배경, 피부와 유사한 색상이 유사한 물체를 포함한 요인의 영향을 받음

Related Work

A. Wearable sensors for healthcare purposes

environments and tasks. Egocentric video is appealing in this context because it can capture information not only about the hand itself but also about its interactions with the environment

Egocentric 비디오는 손 자체 뿐만 아니라, 환경과의 상호작용에 대한 정보를 capture할 수 있기때문에 더 효율적임(appealing)

B. Object or Hand Detection

using convolutional neural networks (CNNs). Existing algorithms can be divided into two categories, region-based and regression-based approaches.

hand detection을 하기위한 방법론이 3가지 존재

Region-based approaches

then perform classification on each proposal. This approach was applied notably in the region-based CNN (R-CNN) but suffered from expensive computational costs as the region proposals must be calculated and classified in every frame

1. region-based CNN은 좋으나 computational cost가 비쌈

R-CNN 제시 but cost가 비쌈

Faster R-CNN 제시

Regression-based approaches

You Only Look Once (YOLO) is one algorithm that uses a single CNN to simultaneously predict bounding boxes and class probabilities, competitively performing with Faster R-CNN, while being significantly faster [22]. Subsequently, the second version of YOLO

2. Regression-based approaches No.1

YOLO -> Faster R-CNN과 유사하게 Performing but more faster

. Another regression-based algorithm that outperformed Faster R-CNN was the Single-Shot Multibox (SSD) Detector

3. Regression-based approaches No.2

Single-Shot Multibox(SSD)

Related Work

C. Object Tracking

Online learning algorithms

[26]. One of the more powerful trackers, the Kernelized Correlation Filter (KCF) tracker, exploits the power of Fourier analysis and circulant matrices by working in the dual space using the kernel trick [27].

trackers offline on a large dataset [29]. However, the top offline algorithms are not feasible to deploy on portable devices due to their slow processing time and report similar accuracy to the KCF tracker [29].

D. Combining Object Detectors and Trackers

may be needed for multi-object detection, we show that a system that focuses on one type of object does not require such complexity.

KCF, MF tracking algorithm이 존재하나 portable device에서 feasible하지 않다.

KCF tracker는 kernel trick을 사용해 이중 공간에서 작업해 푸리에분석 및 순환 행렬의 성능을 활용함

slow processing time 문제로 portable device에는 적합 하지않다.

Offline algorithm은 적합하지 않고, Online learning algorithm이 좀더 적합하다.

비교하려는 논문에선 (multi-object detection에서는)simultaneous computation이 필요하지만,

여기 논문에서는 (one type of object)에 대해서는 이러한 복잡성이 필요로 하지않음을 보임

Method

A. Egocentric Hand Detection Dataset

Therefore, the ANS SCI dataset reflects a range of objects, environments, ADLs, and participants, including different levels of impairment

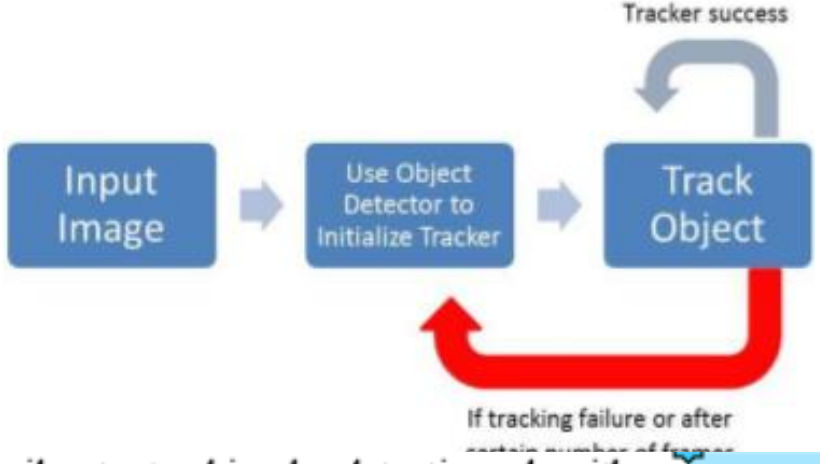
B. Detection and Tracking Only

processing time. Also, the KCF tracker has reported similar accuracy to these offline approaches, while being significantly faster on a CPU [29]. Therefore, we did not implement offline trackers despite their high accuracy, as the proposed combined algorithm would not benefit in efficiency.

KCF tracker가 offline tracker과 비슷한 성능을 냄

offline tracker가 정확성이 더 높으나, portable device에 적합하지 않아, 실제로 효율성이 없다.

C. Combining Object Detectors and Trackers



Similar to tracking-by-detection algorithms, we proposed the use of an object detector to automatically initialize and reinitialize an object tracker upon failure or after a certain number of frames (Fig. 2). This method was proposed since

고장이 발생하거나, 특정 프레임 수 후에 object tracker를 자동으로 초기화하고, 이를 위해 다시 초기화하기 위해 object tracker를 사용하는 방법 제안

Method

C. Combining Object Detectors and Trackers

the main problem with tracking algorithms is the inability to recover from occlusions or lost objects, thus making it difficult to perform adequately after failure. Therefore, we aid successful recovery from occlusions and quick motions by using a detector. Further, since online trackers require manual initialization, the process is only semi-automatic.

tracking problem

교합(occlusion) 또는 손실된 객체를 복구 할 수 없기 때문에 실패 후 적절히 수행하기가 어려움

KCF tracker가 offline tracker과 비슷한 성능을 냄

Another problem many online trackers face is tracker drift. Using a detector to reset the tracker after a certain number of frames minimizes the effect of tracker drift, thus avoiding the propagation of errors and improving performance. We refer to this proposed method as "Detector-Assisted Tracking" (DAT).

online tracker의 문제점: tracker drift

However, they performed detection and tracking simultaneously in every frame and then compared the state of each to obtain the correct location of the object. In contrast, we only use the detector to initialize the tracker at the beginning of a video and to reinitialize the tracker when it fails or after a certain number of frames. Therefore, either the

해결방법으로, 기존의 방법(Bu et al.)과 달리 이 논문에서는 detector를 video 시작 할 때와, 실패하거나 특정 frame 수를 초과했을 때만 초기화 시켜준다.

Method

D. Evaluation Method

To account for participants' functional capabilities, ADLs, environments, and variability, the dataset was split into 3 groups to generate balanced training and testing sets for a cross-validation process. The split was based on participants

참가자의 기능적 기능, ADL, 환경 및 가변성을 설명하기 위한 dataset을 교차검증 프로세스(cross-validation process)를 위한 균형 훈련 및 테스트 세트를 생성하기 위해 3개의 그룹으로 나누었다.

Conclusion

Based on the results obtained from detectors and trackers alone, we expected a combination between YOLOv2 and MF or KCF to perform the best. Even though KCF performed poorly on its own, it had the potential to perform well upon reset due to its high precision (Table 3). After evaluation, we

KCF만 사용하는 것은 performance가 좋지않으나 잠재력이 있어서 사용가치가 크다

more accurate than MF alone. Therefore, combining detection and tracking algorithms resulted in successful recovery from occlusions and quick motions while improving the speed over detectors alone.

Detection과 tracking algorithm을 조합하는 것이 detector만 사용하는 것 보다 효율성이 높음

FPS respectively, which would meet our goal. This was a limitation of our study as we were unable to force these trackers to only use the CPU on CPU-i7. However, even on a mid-range CPU-i5, we see a significant increase in speed compared to detectors alone.

cpu를 i-7에서 사용하는 것이, 이 논문에서의 목적을 충족시킬 결과인데, cpu-i7을 사용 강요할 수가 없다.

하지만 CPU i5에서도 여전히 현저한 속도 향상을 보여준다

정리

- 가정에서도 사용하기 편하게 portable device에 대한 연구이기에 offline tracker보다 online tracker를 사용하였고, 동시에 detector와 tracker를 사용하는 것이 아닌 특정 부분에서 detector를 사용하고 특정부분에서 tracker를 사용하여 효율성을 높이자 이다.
- online tracker에서 가장 효율성 있는 것이 KCF tracker이 offline과 비슷함
- 하지만 tracking algorithm에서 occlusion, lost objects 될 경우 오류가 발생하거나 tracker drift가 일어 남.
  - 방법론제안 (Detector-Assisted-Tracking(DAT))
- online tracker에서 가장 효율성 있는 것이 KCF tracker이 offline과 비슷함
- YOLO_KCF, YOLO_MF 등의 실험결과를 통해 YOLO(v2)_KCF가 가장 효율적 이였다.