

Intrusion Detection System Based on RGB-D Saliency Object Detection

張少鈞

P76114545

NCKU Robotic Lab

陳子聰

P76115062

NCKU Robotic Lab

楊雲翔

NE6114011

NCKU Robotic Lab

黃子軒

P76114197

NCKU Robotic Lab

November 18, 2022

1 Introduction

1.1 Motivation

在現實生活中，人類會很自然地會將看到的景色自動區分為前景(Foreground)與背景(Background)，前景通常代表具有顯著性的目標物，此目標物通常是人類所感興趣的。背景相對前景來說較為一成不變，相對不存在令人所感興趣的目標物。在演唱會上人們會注意在舞台上的歌手而非人群。顯著性目標檢測(Salient Object Detection, SOD)的目的就是模仿人腦自動判別前景與背景，辨識圖片中最突出、顯著的區域的能力。與傳統的目標檢測(Object Detection)相比，顯著性目標檢測並不如傳統目標檢測辨識物體的分類，而是切分出有興趣的前景，與相對不重要的背景。顯著性目標檢測廣泛應用於不同的領域，如無監督影片目標分割、道路行人重識別等；在機器人領域中：人機交互、目標發現等。

1.2 Objective

這篇文章的目的是將 [RGB-D SOD](#) 技術應用於工廠或住宅的入侵檢測中，利用該技術分析監視器的影像，當有顯著目標出現時，對住戶或工廠管理者發出提醒，並配合可移動的相機來持續追蹤目標。以往此工作需要由專職的保全人員耗費精神力來完成，若使用該技術來進行分析與輔助，則不但可以提高住宅或工廠的安全性，並能進一步減少人力成

本的消耗。

採用RGB-D影像的優勢在於，RGB-D影像相較於普通RGB影像多了深度資訊，當前景與背景顏色相近時，深度資訊能有效幫助我們找出實際的前景，在後續的實驗中，我們也有對比使用RGB影像進行物件偵測，與使用RGB-D影像進行SOD預測，當光線不足時，使用RGB-D影像將具有明顯的優勢，如[Fig.2, Fig.3](#)

2 System framework

整個系統架構預計分為兩部分，第一部分為顯著性目標檢測，第二部分為相機的實時追蹤。整個工作流程為：(1)透過RGB-D相機取得當前的影像資訊，(2)將影像資訊輸入進顯著目標檢測模型來找到欲觀察的目標物，(3)而配合可移動的相機來實現實時追蹤。

2.1 Salient Object Detection

我們採用[SPSN](#)[3]作為顯著目標檢測模型，該模型的輸入為RGB-D影像，輸出為與原圖具有相同大小的二元圖像，每個像素將以1或0分別表示前景與背景。

2.2 Real-Time Tracking

透過顯著目標檢查模型所輸出的二元影像，判斷目標在影像中的位置，藉此來評估目標與相機的相對位置，並進一步去進行相機的移動，使目標始終位於影像中央，以達到目標追蹤的效果。

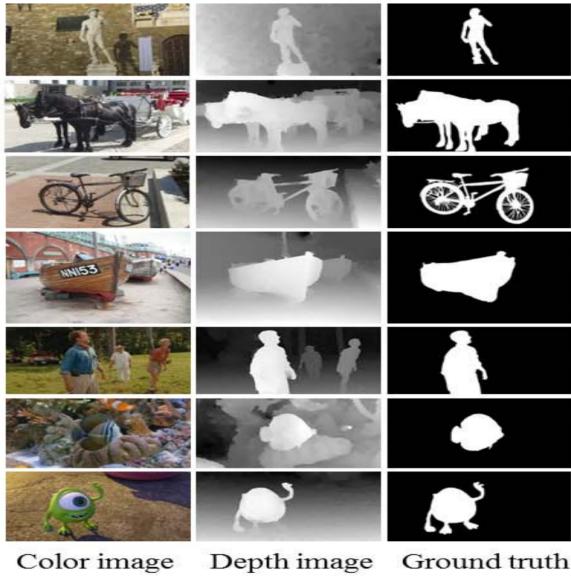


Figure 1: NJU2K

3 Experiments

3.1 Datasets

我們採用了*NJU2K*[2]資料集進行訓練。*NJU2K*是一個包含了1,985組圖像的大型RGB-D資料集，這些圖像是從網際網路與3D電影中收集的，而照片則是由Fuji W3相機所拍攝收集的。每組圖像由兩張RGB影像(分別代表左右鏡頭)與一張depth灰階影像(由左右影像進行深度估算而來)所組成，標籤是一張與原圖具有相同大小的二元圖像，其中每個像素以1或0分別表示前景與背景，Fig.1中展示了該資料集。

3.2 Implementation Details

我們將*NJU2K*以8:1:1的比例切割成訓練集、驗證集與測試集，輸入圖像統一縮放至 352×352 ，採用了Adam optimizer，learning rate為 $8e-5$ ， $\beta_1 = 0.9$ ， $\beta_2 = 0.999$ ，並在NVIDIA RTX 3080 Ti GPU上總共訓練了200 epoch。

3.3 Compare With Yolov7 in a dark environment

為了證實採用RGB-D影像的優勢，我們使用了Real Sense Camera在室內、三種不同強弱的燈光下進行資料的收集。該相機是採用雙鏡頭進行深度估算，以產生深度圖。我們根據收集到的影像設計了兩組實驗進行對比，第一組實驗為使用RGB-D影像與SPSN[3]，結果如Fig.2，第二組實驗為使用RGB影像與yolov7[4]，結果如Fig.3。透過觀察兩組結果可以發現，在亮度最低(Dark)的影像中，yolov7無法正常偵測出人影，而在亮度適中(Low Light)的影像中，yolov7預測出的人影信賴度為0.33，反觀SPSN，在各種亮度下，都能預測出大致上的人體輪廓。

References

- [1] Wang, Wenguan, et al. "Salient object detection in the deep learning era: An in-depth survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.6 (2021): 3239-3259.
- [2] R. Ju, L. Ge, W. Geng, T. Ren and G. Wu, "Depth saliency based on anisotropic center-surround difference," 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 1115-1119, doi: 10.1109/ICIP.2014.7025222.
- [3] Lee, Minhyeok, et al. "SPSN: Superpixel Prototype Sampling Network for RGB-D Salient Object Detection." European Conference on Computer Vision. Springer, Cham, 2022.
- [4] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." arXiv preprint arXiv:2207.02696 (2022).

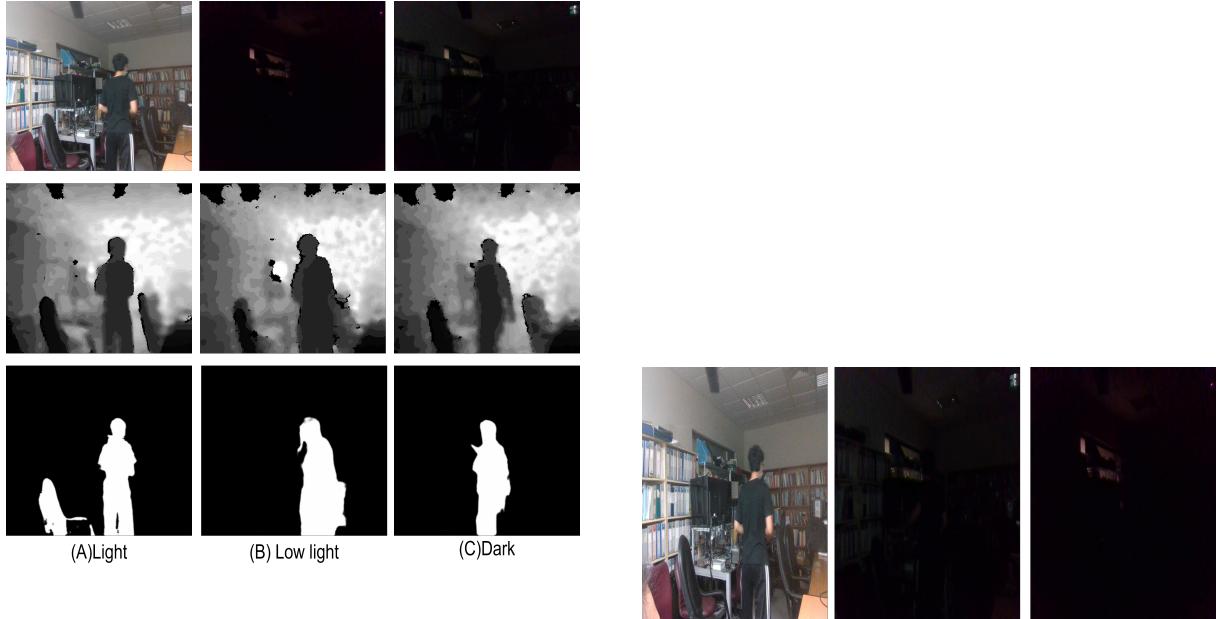


Figure 2: **RGB-D SOD Result:** 第一列為原始RGB影像，第二列為影像深度圖，第三列為SPSN模型偵測出的遮罩。(A),(B),(C)分別為三組不同強弱燈光下的圖像。



Figure 3: **Yolov7 Result:** 第一列為原始RGB影像，第二列為yolov7模型的預測結果。(A),(B),(C)分別為三組不同強弱燈光下的圖像。