Department of Electrical Engineering, National Tsing Hua University
Special Topic on Implementation Research Report

On-The-Go Photographer
隨行攝影師

Major Category: 資工領域

Group Number: B260

Advisor: 黃朝宗 教授

Member: 陳均豪、施品瑄

Research Period: From 2022/02/15 to 2022/12/05

Abstract

In today's society, smartphones have become a must-have item for everyone, bringing a lot of convenience to everyone's life. In various programs, the camera function is often used. We can record the scenes around us and turn them into memories to cherish. However, how to take good-looking photos has always been a question that everyone has been thinking about. In addition to manufacturers' continuous enhancement of camera specifications, having a good composition is the most adjustable direction for users. We can see various composition skills teachings on the Internet, but there are too many small details to pay attention to when shooting, which often makes people not know what to take. We aim to solve users' troubles when taking pictures because of how to compose. Thus, we want to design a smartphone camera APP. According to the user's initial shooting screen, the APP will provide real-time and clear guidance on adjusting the position and angle. Achieve the best composition without greatly changing the original picture content.

This camera APP mainly uses the technologies of object detection and pose estimation. The former will first find out the person's position in the picture, and the latter will locate the 17 key points of the person. We also add the Kalman filter to optimize the stability so that the guidance prompt will not cause large jumps due to short-term detection errors. Our main adjustment direction is to achieve the rule of thirds, which is the most widely used. When the user is shooting, an arrow is displayed on the screen in time to provide position guidance, and there use a gyroscope in the upper half of the screen to provide angle guidance. This APP will help the user achieve the best composition picture by changing these two parts. When the user adjusts to a suitable position, a prompt will appear to inform that the current view is the best, and it is recommended to shoot. We have also completed many basic functions of the camera, such as adjustment of focus, click to focus, and automatic flash. Let users use our APP more smoothly.

What we have done so far is help users adjust to achieve the rule of thirds. In the future, the user might fine-tune the recommended composition parameters and set their favorite shooting style or further use the analysis of the photographer’s photo collection to achieve the effect of shooting style transfer. The practicability of this APP would be greatly improved, and everyone can have more choices when composing pictures.

摘要

在現今的社會中，智慧型手機成了大家必備的物品，帶給大家生活上許多的方便，在各種程式裡，相機功能更是常被使用的，隨手就可以將身邊的景象記錄下來，變成回憶好好珍藏，但要如何拍出好看的照片，一直都是大家不斷思考的問題，除了廠商持續增強相機的規格外，有好的構圖畫面則是我們使用者最能夠調整的方向，在網路上也能看到大家推出的各種構圖技巧教學，但實際在拍攝時要注意的小細節實在太多，常常令人無所適從。我們希望能夠解決使用者在拍照時因為要如何構圖而產生的煩惱，因此設計出一款手機相機 APP，藉由使用者最初的拍攝畫面，在螢幕中即時且明確的進行位置及角度的調整指引，在不大幅改變原先畫面內容的情況下，完成最佳的構圖。

這款相機 APP 主要利用物件偵測(object detection)及姿態辨識(pose estimation)的技術，前者會先找出畫面中的人物位置，後者則將人物的 17 個關鍵點定位出來，並加上卡爾曼濾波器(Kalman filter)進行穩定性的優化，讓指引的提示不會因為短暫的偵測誤差而造成大幅度的跳動。我們主要調整的方向為符合最廣為使用的三分之一法，在使用者進行拍攝時及時在螢幕中顯示箭頭來提供位置指引，在螢幕上半部有則有陀螺儀來提供角度指引，幫助使用者藉由改變這兩個部分就能達到最佳的構圖畫面，並在調整到合適位置時出現提示，告知使用者當下為最佳的畫面，建議進行拍攝，同時我們也完善了許多相機的基本功能，像是焦距調整、點擊對焦及自動開啟閃光燈等，讓使用者能更加順利的使用我們的 APP。

我們目前做到的是幫助使用者調整到符合三分之一法則，未來若能夠讓使用者自行微調構圖的推薦參數，自己設定喜歡的拍攝風格，或是更進一步利用分析攝影師的攝影集達到拍攝風格轉移的效果，可將這項專題的實用性大幅提高，並讓大家在進行構圖時能夠有更多的選擇性。

Introduction

Composition is a crucial step when taking a picture. A good composition makes the image harmonic and thus obtains more range for post-processing. On the other hand, a bad composition makes the image unbalanced and makes it hard to use the post-processing process to achieve the expected effect.

We intend to develop a camera APP to help the user compose the picture. Using the techniques such as object detection, pose estimation, etc. Analyzing the photographing environment and the character posture in real time. With the help of the algorithm, we designed to render the recommended composition in the form of an arrow or color on the user interface. Assisting the user to improve their photo composition's quality.

Background

Photo Composition

Photo composition is how a photographer arranges visual elements within their frame. Putting subjects or scenes inside that space may sound easy, yet it's anything but. Composition in your shots can often be difficult, and it's always important.

Composing a good photo is more than just focusing on your main subject. Many perspectives need to be considered, such as object ratio, color contrast, the directionality of visual extension, etc. We can summarize the technique of composition into some common rules.

The rule of thirds

The rule of thirds evenly divided the frame between two equally spaced horizontal and vertical gridlines, creating a three-by-three grid. Then, place the compositional element or interesting point on these lines of the grid intersect or segment. This will make the image more interesting than just centering a subject.

Balance image

Balance is related to, but distinct from, symmetry. The balanced image does not necessarily need to be one-to-one corresponding with respect to the symmetric point. It talks about how when the viewer scans the picture, they can find the space on the other side of their point of interest. For example, things on the left correspond with things on the right.

Passive assisting vs. Active assisting

We distinguish our intended active assisting from the method often used, passive assisting. For

the former, we mean analyzing the environment and indicating how to compose immediately to the user. For the latter, the passive assisting is that the camera simultaneously captures two different focal lengths; one is the view that the screen shows, and the other is a broader image. When we want to adjust the composition after we have already taken the picture. We can crop the image to rearrange the composition.

Problem formulation

We know that the sense of beauty is subjective. Therefore, we need to make some assumptions to quantize this problem. Here, we assume that the rule of thirds is the golden discipline. To further reduce the complexity of this task, we focus the main subject on the single person.

With the above assumption, we can separate the problem into four parts. We are first identifying and locating the person in the frame. Second, analyzing the key points of the person. Third, using the information provided by the above steps and applying the rule of thirds to get the recommended composition. Finally, render the recommended composition indicator to the user interface.

There is no similar product in the market as I know. After our experiment in this work, this idea can be realized.

System Design

System architecture and flow chart

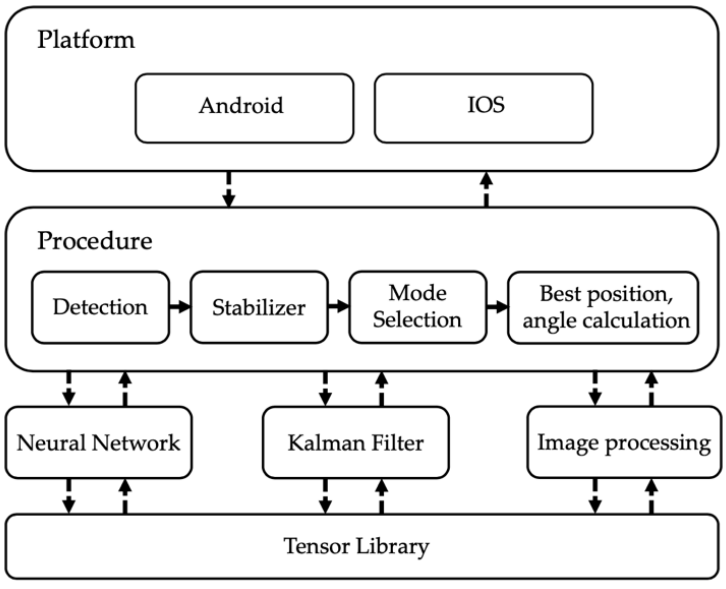

Fig. 1 System architecture

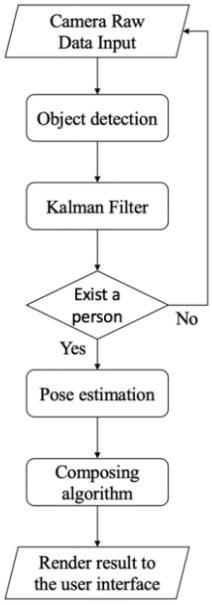

Fig. 2 System flow chart

Object Detection

We need to perform object detection to identify and locate the person in the frame. We choose NanoDet[2] as our detector. NanoDet is an FCOS-style one-stage anchor-free object detection model that uses Generalized Focal Loss as classification and regression loss. Its design is CPU friendly without sacrificing accuracy.

Pose Estimation

After identifying and locating the person, we need to analyze the key points of the person. By knowing the position of the key point, we can estimate the pose. Here, we choose SimplePose[3] as our pose estimator. It runs not only fast but also lightweight.

Kalman Filter

The algorithm we mentioned above will sometimes fail if the input stream is not a still image due to shutter speed, shaking, and other factors, making the image blurred. Hence, we use the Kalman filter to make the algorithms stable. Kalman filter uses a system's dynamic model with multiple sequential measurements to estimate the system's state. There exists a problem. We can estimate a system's state by a unique Kalman filter. However, if there are multiple objects, we need to know how to assign the newer detection result to the previous state to create the sequential measurements that make the Kalman filter work.

Simple Online and Realtime Tracking[1] provides a solution. It is a pragmatic approach to multiple objects tracking where the main focus is associating objects efficiently for online and real-time applications.

Composing Algorithm

We emphasized applying the rule of thirds on different conditions with a single person. The decision tree is the idea we used here to implement the composing algorithm. With the help of object detection and pose estimation, we successfully quantize the problem from "sense of human" to a mathematic problem.

We set the upper-left corner of the image as the origin and express it in a Cartesian coordinate system, the x-axis extends to the right, and the y-axis extends downward. The object detector provides the center coordinate, width, and height of the object, and the pose estimator provides the 17 key points' coordinates of the human.

Therefore, we can express a person with a set of coordinates and normalize them between 0 and 1. We can now build a decision tree to implement the composing algorithm.

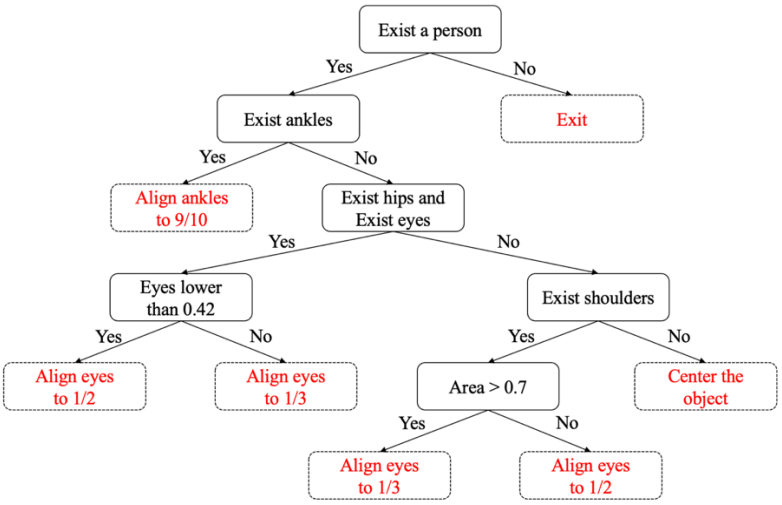

Fig. 3 Y-axis decision tree

Fig. 3 shows the y-axis decision step, and the number in the figure is a normalized y-coordinate. As for the x-coordinate, according to the rule of thirds, placing the subject on two even divided vertical lines will make the picture more interesting. Considering our assumption is a single person, except for two vertical reference lines of the rule thirds, we will add a center line as our reference line. Therefore, the x-coordinate is the x-coordinate of the closest reference line.

In addition to the recommendation of subject coordinates, we also introduce the judgment of shooting angle. To achieve a specific shooting effect, the shooting angle must be considered. Furthermore, we use pose estimation to achieve this correction and cooperate with the built-in electronic gyroscope of the mobile phone to complete the angle recommendation to the user.

Experimental Results

Our system can operate in different user environments and supports multiple portrait modes, such as the full-body mode in Fig. 4 and the half-body mode in Fig. 5. The system will display the current recommended mode on the upper-left corner of the screen to assist users in adjusting for shooting.

Fig. 4(a) and Fig. 5(a) show that the user's composition could be better. We can see the yellow arrow on the screen, indicating the direction we should correct. Besides, the gyroscope on the top of the screen shows a red line which means the angle needs to be adjusted.

After a slight adjustment in Fig. 4(b) and Fig. 5(b), the yellow arrow disappears, meaning the composition is correct. However, the red line in the gyroscope is still displaying, meaning the angle can still be improved.

In the end, we can see the yellow arrow disappears in Fig. 4(c) and Fig. 5(c), and the gyroscope line turns green. Moreover, a green border on the screen indicates that the composition is perfect. The composition obeys the rule of thirds, and the picture seems balanced.

(a) (b) (c)

Fig. 4 Example1 of use (全身、室外)

(a) (b) (c)

Fig. 5 Example2 of use (半身、室內)

Link for demo video:
Introduction: https://youtu.be/UOjVrintHsA
Demo: https://youtu.be/Dtqlv-GSj1I

Conclusion and Future Work

In this work, we formulated the composing problem in terms of object detection, and pose estimation. We proposed a composing algorithm based on the decision tree to achieve the rule of thirds. We design a system within the modular approach, separating the composing system from the user interface, so we can easily port the whole system to other operating systems. Experimental results show our system works well as expected that obeying the rule of thirds.

For future work, we would like to let users fine-tune composition parameters by themselves, making the entire APP more flexible. Moreover, we can build the decision tree in our algorithm from the famous photographer's work to achieve the picturing style transfer effect.

Throughout the process of implementing this work, we have a deeper understanding of object detection, pose estimation, the Kalman Filter, etc., and actually implement these technologies on our APP. In the process of developing this APP, we continuously strengthened the calculation speed and optimized the stability of the guidance prompts, making usage easier and smoother.

Reference

[1] Alex Bewley et al. "Simple online and realtime tracking". In: 2016 IEEE International Conference on Image Processing (ICIP). IEEE, Sept. 2016. doi: 10.1109/icip.2016.7533003. url: https://doi.org/10.1109%2Ficip.2016. 7533003.
[2] RangiLyu. NanoDet-Plus: Super fast and high accuracy lightweight anchor-free object detection model. https://github.com/RangiLyu/nanodet. 2021.
[3] Bin Xiao, Haiping Wu, and Yichen Wei. Simple Baselines for Human Pose Estimation and Tracking. 2018. doi: 10.48550/ARXIV.1804.06208. url: https://arxiv.org/abs/1804.06208.
[4] https://www.adobe.com/creativecloud/photography/discover/photo-composition.html

心得感想

實作專題這堂課不同於一般的課程，教授不會幫我們安排好進度，或是幫我們選好要讀的課本，而是一堂著重於自主學習能力的課程，大部分的事情都必須由我們自行決定，包括尋找資料、確立目標、時間規劃等，在找尋資料的時候發現，對於相同的問題會有許多人提出各種不同的解法，需要我們自行去嘗試，找出最適合我們的，更甚至是去加以改良，在確立目標的時候理解到，教授常會要求我們要把所有要做出來的東西定義清楚，讓我們後續在實作的時候更加順利，在時間規劃的部份，我們需要對自己提出的規劃負責，才不會因為前面任務的延遲而導致後續任務持續的往後延。

從一開始只能處理單張照片的物件偵測及姿態辨識，到後來能夠連續追蹤透過手機鏡頭得到的畫面，再持續去改良運算速度一直到可以完成即時的偵測，之後再想辦法加強偵測的穩定性，最後完成了可以被順暢使用的 APP，在這個過程中會遇到很多的問題，多數時候無法順遂的解決所有錯誤，但在找出錯誤並更正直到得到我們要的部分成果後，會感受到極大的成就感。

感謝這一年下來常常被我們請教拍照時該注意甚麼的同學們，不同使用者所建議的調整機制給了我們許多不同的方向可以去研究。感謝這一年下來教授的指導，在我們卡關或是感受到迷茫的時候，適時的給我們一些指引，讓我們構圖出最佳的屬於我們的實作專題課。