梁家愷博士演講

Mobile Computational Photography at Google

文／劉柏聲 多媒體工程研究所碩士生、洪翊筑 資訊科學與工程研究所碩士生

梁家愷博士在 Google Android Camera Algorithm Team 擔任 Principal Software Engineer 及 Director。主要的研究項目為在智慧手機上的相機應用，有鑑於手機有限的記憶體及運算處理能力，針對任何演算法及軟硬體交互功能皆需要在資源與效能之間取捨，相較於傳統相機的能力可能可以差到 40 倍，但手機同時也受益於強大的處理器及方便的使用、編輯及分享等功能。這次的演講主題是關於在 google pixel 4 中對相片進行後續處理的運算難題，內容分為三個不同主題，分別是 Distortion-Free Graphie, Super Resolution Zoom 和 Fused Video Stabilization。

第一個主題中提到在過去的手機照片中，位於邊緣的人物會因為與鏡頭角度過大，不同的投影方式產生不同的拉伸變形。這部分能透過首先將照片中的主角物體篩選出來，並透過網格優化，產生整張圖的變形曲線，再對應回影像中。產生的結果會導致人臉附近的線彎曲，再對彎曲的線做額外的處理，得出一張優化後的照片。

第二個主題是利用手持拍照所產生的些微移動產生一張高解析度的影像。舉 4×4 的影像為例，有三張對同樣的位置拍的照片，彼此間有些微的移動，將這些經過位移的影像對齊，並經過不同影像之間訊息的交換融合，有些影像包含一些清晰部分，就能幫助剩下影像在同位置產生高解析度影像。

第三個主題是改進影片的穩定度，在電影場景中，人物移動時相機也會需要跟著移動，而在片場中，會需要大型機械手臂跟許多專業人士的輔助，以呈現影片的穩定度。在手機攝影中，也希望能達到這種穩定性，利用動作估計來達成連續動作的順暢度，利用模糊估計及動作解析來使畫面平滑化，並更正畫面擷取的位置，對歪斜的畫面也擷取相同方向的歪斜方框作為影像最後的結果以得到抵銷手震及過度晃動的效果。

以上這些主題都伴隨著實際應用上的一些問題，像是手機硬體與軟體的結合，需要開發出適合目前系統在效能與花費資源間平衡的最佳方式，特別是在現今使用者對自己手機拍出來的照片或影片都用高標準來要求 (4k@60fps)。以及針對不同的使用者產生不同的回饋，並利用可控制的變數來讓使用者自由的調整想要的照片感覺。縱使目前手機的相機功能在經過這幾年的發展已經日趨成熟，未來還是有很多值得探討與改善的議題，在圖片品質上，若有一張在低光源環境下的照片，如何將圖片細節與雜訊完美的分辨出來，抑或是在不同場景下如何決定最適合的色調，又或者移動中的物體與手持移動的區分…等。而在應用程式層面則是希望能設計出一套能適用在任何地區任何環境任何使用者都能使用具備通用性的軟體，並能在有限的計算資源下決定最佳的策略。將解決這些議題的技術轉化成實際的產品於手機中，讓手機相機在未來能夠擁有更多實用的功能，勝制能夠取代攝影機的地位。

在經過今天演講的內容之後，對於世界上最頂尖軟體公司之一的 Google 在影像處理及電腦視覺領域的發展有更深入的了解，當我們還在試圖利用各種方法提升我們研究的準確度時，Google 身為一家公司，更在意演算法是否能完整的複製到各種環境當中，尤其是需要考量手機硬體與大型機器的運算能力之間的差距，並研究如何生產出 user-friendly 的產品，讓使用者能在不改變原本使用習慣，甚至是讓人在使用中毫無察覺之下，盡可能的提高相片或影片的品質，也希望能達到功能即時反應避免等待時間過長的使用不方便。另一方面，梁博士也跟我們分享申請 Google 需要具備的能力以及如何對症下藥去準備，有鑑於 Google 對每位員工的期待是希望每個人都能具備獨當一面解決問題的能力，在面試的準備上，除了要應徵職位所需的專業知識外，軟體工程師所需的程式能力以及資料結構是基本外，電腦視覺工程師也必須對除了自己專業領域之外的電腦視覺任務有一定的認識，畢竟在 Google 工作是工程師取向，也就是相較於主管指派工作，Google 更希望是由工程師自發性的對工作內容產生興趣並在這基礎上不斷精進自己，不斷累積研發能量。很感謝有這次的機會能與美國 Google 內的資深工程師有這樣的交流互動，學習到業界將學術成果在實際產品上應用的經驗以及許多無法從課本或網路上得知的秘辛。

The Speech of Dr. Chia-Kai Liang

Mobile Computational Photography at Google

Dr. Chia-Kai Liang is a director/principal software engineer leading the Android Camera Algorithm Team at Google, focusing on mobile computational photography. Since mobile phones are limited in computing power and memory, any algorithm design or hardware-software interaction function has to take into consideration a tradeoff between system resources and performance. Although the capabilities of mobile phones may be 40 times less than traditional cameras, mobile phones would benefit from powerful processors and easy-to-use editing and sharing functions. Dr. Liang's talk covered the computational photography problems of subsequent photo processing in Google pixel 4. Three topics were highlighted in this talk: Distortion-Free Graphic, Super Resolution Zoom, and Fused Video Stabilization.

The first topic mentioned that subjects near the edge of phone photos, in the past, were stretched outward to cause different degrees of distortion due to the properties of the wide-angle lens. The solution was to detect the salient objects in the photo, generate deformation curves of the whole image by the grid optimization, and calculate the corresponding bended lines around the face. We then conduct additional processes based on these lines to get an optimized image.

The second topic was that a very slight movement of a hand-held camera is advantageously utilized to produce a high-resolution image. Take the example of some 4x4 images. Three photos are taken in the exact same place with slight movement between each other. By aligning these shifted images and merging information between them, the combination of multiple overlapping low-resolution images can produce a high-resolution image.

The final topic of this talk is bringing video stabilization to all those shaky-hand videos. Tracking shots normally move the camera side to side following the movement of characters through a scene. In the studio, the stability of the filming would rely on the cooperation of a large robotic arm and many technicians. Likewise, mobile photography would bring stabilization to shaky videos. It comprises several techniques, such as motion estimation to achieve the fluency of continuous motion, fuzzy estimation and motion analysis to smooth the images and correct the position of the captured frame, and the compensation frames for the skewed images, in order to offset hand-shaking effect and excessive shaking effect. However, the implementation of the solutions to the above topics on mobile phones still encounters some challenges, such as the high integration between hardware components and software modules. We have to develop an effective way to strike a balance between overall efficiency and resource utilization for specific mobile systems, especially while users are more demanding about capturing high-quality images and videos (4k@60fps) nowadays. To provide adaptive user experience for different users, the mobile phone system lets users adjust parameters to produce the image with the desired size and quality. Even though the functions of camera phones have become more mature in recent years, many unsolved issues of phone cameras would be worth exploring and improving. For example, how to reserve the details of an image while removing the random noise from the image taken in low light conditions, and how to suggest the best hue in different scenes, as well as distinguish the differences between a moving object and a hand-held movement, etc. At the application level, we hope to design a software suitable for all users with the effective usage of limited computing resources according to universal design principles. The solution to these issues would be implemented as applications running on mobile phones so that the mobile phone camera would become powerful enough to replace the standard camera in the future.

After today's speech, we have a deeper understanding of the development in image processing and computer vision in Google, one of the top software companies in the world. While we consider using different methods to improve the accuracy of our research, Google, from a real product perspective, is more concerned about the feasibility of algorithms that can be applied to different platforms, especially facing the performance gap between mobile hardware and high-performance computers. In the meantime, the design team spends a lot of effort on user-friendly products so that users can get much higher video or picture quality when they have no need to change their usage habits, or even when the functions are running behind the scenes. Furthermore, users can also receive real-time responses to avoid the inconvenience of long-waiting. Beside technical discussion, Dr. Liang shared with us the ability needed for technical jobs at Google and the tips for the preparation of the application. Since Google expects employees to solve problems independently, an applicant must equip not only with the professional knowledge for the position, such as programming skills and data structure, but also common sense regarding computer vision beyond his expertise, while preparing for the interview. After all, Google is famous for its free and self-motivation atmosphere. Compared to fulfilling the duties assigned by supervisors, Google engineers are encouraged to proactively seek interest in the daily jobs and continuously improve themselves to level up their technical skills. We are very grateful to have such a great opportunity to interact with a senior engineer in Google from the United States in person to gain industrial experience in commercializing academic research findings as well as the firsthand engineering tricks that cannot be learned from textbooks or the Internet.