論 文 内 容 の 要 旨

Pain is one of the primary feelings that encourage people to seek medical care and attention. Moreover, pain is also defined as both a sensory and emotional experience. Also, reliable and valid pain assessment is a crucial solution for choosing the adequate treatment, monitoring progress, and evaluating the continued treatment. In practice, pain is generally diagnosed by the patient's self-report according to its feeling. Hence, automatic estimation of the sense of pain has lots of applications in medical treatment. However, the feeling of pain is an individual and distinctive experience which differs from each person. According to previous research, facial expression is one of the tools to estimate the pain intensity. Most research on facial expression is based on the Facial Action Coding System(FACS), which falls the whole facial expression apart from basic action units. Each action unit is coded by a five-point scale, which represents the intensity of the facial expression. Although the FACS is suitable for facial expression research, it needs to train the expert to code the patient video, which is time-consuming and inefficient. Until now, there are two main measurements for pain assessment, e.g., self-report scales and pain rating by an observer. In our experiment, we chose the observer's pain rating as the ground truth for training our neural network, e.g., Prkachin and Solomon Pain Intensity(PSPI). Considering the massive application of pain facial expression detection and analysis and medical diagnosis and treatment, we develop the computer-aid technology to help the medical diagnosis and improve the work efficiency in the early treatment and all therapy. We propose to utilize the power of the deep learning technique, which leverages the spatial domain and time domain learning. Attention plays an essential role in human perception. One vital feature of the human visual system is that it does not attempt to process the whole scene at once. In contrast, the vision system of humans tries to analyze a sequence of the partial area and selectively focus on more critical parts to get a better understanding. Recently, there have been several attempts to model the attention ability inside the neural network to improve the performance. Inspired by the previous work in the attention model, we proposed the spatial attention model and inserted it into our convolutional neural network. Our neural network can find the most correlated region on the human face for pain facial expression detection and analysis aiding by our spatial attention model. Experimental results show that our locally spatial attention learning can provide the fine-grained variation on the face region for pain intensity assessment. Our current study extends the prior work in this research area and provides a new method for future studies on painful expression analysis. At this moment, the results demonstrate the performance of our architecture is better than the general structure without locally spatial attention learning and is not outstanding compared with the state-of-the-art methods. In the near future, we will import our architecture by effective network engineering.

Internet of Plant is the application of precision agriculture and yield mapping. In our research, our target agriculture product is green pepper and green citrus. In recent years, the development of image processing, computer vision, and machine vision technology provides a non-contact measurement method. There are many factors that can influence the yield and growth of agricultural products, such as weather, light condition, soil properties, and temperature. Early and precise yield mapping and forecasting can help the farmer and grower to know the fruit growth conditions and aid the marketing strategy. Furthermore, automatic detection of green pepper and green citrus is a worthwhile study for automatic harvesting in agriculture applications. The agriculture, horticulture, and food industry still massively rely on manual labor provided by the farmer and workers. Accompanied by the increasing labor cost and the growth of the aging population in Japan, the automatically harvesting system's development is one of the popular research topics in machine vision, computer vision, and robots. Most sensors can only record the visible part of the electromagnetic spectrum(400nm-700nm). However, the hyperspectral camera can capture not only the visible wavelength but also the near-infrared wavelength. It can provide many more details about living plants and fruits. Compared with the RGB image, the image captured by the hyperspectral camera is a high-dimensional tensor. Motivated by the previous study of the channel attention model and Long short-term memory(LSTM), we proposed our novel neural network-based algorithm for green pepper segmentation. Capture the hyperspectral camera is a time-consuming task, which causes limited image data. Unlike the other research work, our method is a pixel-aware framework for green pepper segmentation. In particular, we select some pixels from the green pepper and some pixel from the foliage. We treat each pixel as a long vector to put into our proposed framework. There are two parts of our deep neural network. The first part is our proposed channel attention module, and the second part is the LSTM. Our channel attention module consists of two fully connected layers with the sigmoid function. Our attention module can generate the attention weight for the input vector and reweight the input vector. LSTM or recurrent neural network(RNN) is one of the important structures of the deep learning family, which is mainly designed to deal with sequential data. Then the reweight vector is fed into the LSTM. By utilizing the memory function of the LSTM, our proposed structure can use the critical wavelength information to distinguish green pepper. The whole network can train in an end-to-end manner. The experimental results on our hyperspectral dataset show competitive performance for our proposed approach. Our proposed channel attention model considers the intrinsic feature of the hyperspectral data. Besides, our proposed methods require a smaller dataset, lower hardware requirements, and faster, compared with the deep convolutional neural network. However, our approach is vector-based machine learning. The hyperspectral camera can provide not only spectral details but also provide the spatial details of the captured scene. In our future plan, we believe an end-to-end convolutional LSTM with channel attention can achieve better performance for green pepper segmentation because it can get much more useful information from the spatial domain.

Computational Optics provides a way to co-design optics devices, camera sensors, and spectral illumination. Recently, the proposal of end-to-end optimization of optics, sensor, and camera pipeline is attracting widespread interest. However, the similar color between the green pepper and foliage becomes a challenging problem in the automatic harvesting system. This phenomenon is called metamerism. The reason arose of metamerism is because the surface reflectance and scene illumination have more degrees of freedom than trichromatic vision responses. Multispectral camera and hyperspectral camera are promising tools for monitoring crop plants or vegetables in the agriculture and food industry because it can provide rich information in the invisible wavelength. In precision agriculture, the hyperspectral camera can offer a rich clue for site-specific fertilization and special plant protection. Thus, it can access the architecture, biochemistry of plants, and the reflectance characteristics of the target materials in a non-invasively way. The reflected light of the target plants or materials is recorded by a high spectral and spatial resolution by a three-dimensional tensor. The disadvantage of the multispectral camera and the hyperspectral camera is the high price. So, it is not available for most farmers and workers. The optical filter is one of the critical devices in the camera. The spectral feature of the optical filter can permit the monochrome camera to capture more information on the object surface than color imaging. Besides, the optical filter is a widely used optic device in machine vision, chemical analysis, etc. It can improve the performance of the imaging system or help to inspect the object in some particular task, such as polarizing filters. In our research, we proposed to utilized an optical filter to enhance a red-green-blue(RGB) camera system to detect green pepper or immature citrus. According to the previous study, not every camera spectral response function is optimal for a specific computer vision task. Rather than select the best camera spectral response function, we propose to add an optical filter to enhance the RGB camera system. Our proposed method can generate any available camera model. We use the power of the deep neural network to help us to find the best optical transmission curve of the optical filter. Specifically, we reported the relationship between the depth-wise convolutional kernel and the transmission curve of the optical filter. Motivated by the previous study, we can represent the camera spectral response function by three convolutional kernels. As a result, the weights of the first layer and weights of the second layer in our neural network represent the optical filter and color filter array (CFA) inside the camera. Hence, we can represent the physical device by the depth-wise convolutional layer and convolutional layer. To our best knowledge, it is the first proposal to represent the optical device by the deep neural network layer. Consequently, we can use the deep neural network to find the optimal transmission curve for the optical filter in a data-driven way. The whole deep neural network structure is optimized by the segmentation loss, which is detecting the green pepper and immature citrus from a similar background. Both optical filter and color filter array must satisfy specific physical requirements that constrain both the spectral transmission curve and camera spectral response. First, the weight of the transmission curve must be non-negative. Second, the smooth curve is more feasible than other arbitrary curves in the implementation process. According to the above feature of the transmission curve, we proposed the non-negative and smoothness constraint for the weights of the transmission curve of the optical filter. The experimental results show our proposed framework can achieve better detection results than an RGB camera without an optical filter. Besides, our proposed camera system is cheaper and easier to use than multispectral and hyperspectral cameras. Replacing the expensive hyperspectral camera with a more affordable camera-optical filter system, we can significantly reduce the cost of an automatic harvesting system for most farmers. Thanks to the later optical filter production technology, we can implement our optical filter. We also extend the optical filter design from the visible wavelength to the near infrared wavelength in our follow-up work.

The above paragraph summarizes the main contribution of our research. Our research topic is not only for pain facial expression detection but also for optical filter design. It has proven to be the power of the deep neural network and the wide application in the computer vision research area. The rest of this thesis is organized as follows. Chapter 1 gives details about frame-by-frame pain estimation by using locally spatial attention learning. Chapter 2 presents our proposed attention module and LSTM structure for green pepper segmentation. Chapter 3 specifies optical filter design by deep neural network for green pepper segmentation. Chapter 4 describes An Optical Filter Enhanced RGB Camera System, and its application to green citrus segmentation. Last is the acknowledgment.

Keywords: deep neural network, attention, pain face, computational optics, optical filter design