Vision-Based Collision Avoidance for Unmanned Aerial Vehicles by Recurrent Neural Networks

Yao-Hong Tsai

Abstract—Due to the sensor technology, video surveillance has become the main way for security control in every big city in the world. Surveillance is usually used by governments for intelligence gathering, the prevention of crime, the protection of a process, person, group or object, or the investigation of crime. Many surveillance systems based on computer vision technology have been developed in recent years. Moving target tracking is the most common task for Unmanned Aerial Vehicle (UAV) to find and track objects of interest in mobile aerial surveillance for civilian applications. The paper is focused on vision-based collision avoidance for UAVs by recurrent neural networks. First, images from cameras on UAV were fused based on deep convolutional neural network. Then, a recurrent neural network was constructed to obtain high-level image features for object tracking and extracting low-level image features for noise reducing. The system distributed the calculation of the whole system to local and cloud platform to efficiently perform object detection, tracking and collision avoidance based on multiple UAVs. The experiments on several challenging datasets showed that the proposed algorithm outperforms the state-of-the-art methods.

Keywords—Unmanned aerial vehicle, object tracking, deep learning, collision avoidance.

I. Introduction

WITH the rapid development of technology, the technology on the security monitoring system is changing with each passing day. Because of the anti-terrorism and security needs, major cities around the world have set up closed-circuit television or video surveillance systems. At present, Taiwan has security monitoring systems almost in all streets and lanes of the city, ranging from the monitoring system at the door of the home to the monitoring system for road traffic safety and criminal search. They are shown in Fig. 1. Due to the rapid development of the security monitoring system, whenever there is a crime or traffic incident, we can use the storage system of the monitoring system in time to find out the clues needed or to reconstruct the occurrence of things. For example, the Police Department of the Taipei City Government in Taiwan has also established and enabled the closed-circuit television video surveillance system since March 2013. At 11,500 security points, 13,699 sets of camera video surveillance systems were deployed to serve as an electronic wall to protect the safety and security of Taipei residents [1]. It is very effective in detecting cases of violence and theft. The traffic control center of the Traffic Control Engineering Office of Taipei [1] not only monitors video at any time, but also provides instant traffic conditions on the center's webpage, which is shown in Fig. 2.

However, most of the monitoring system equipment is located in the metropolitan area. In the suburbs, there are very few monitoring systems. Therefore, it is difficult to use the surveillance image restoration event in time when events occur in these areas.

Fig. 1 Video surveillance systems on the street

Fig. 2 The website of the Taipei Traffic Control Center [1]

UAVs, as the name suggests, are aircraft that do not require a driver [2]. They usually include flight control modules, network communication modules, and ground control personnel. Additional sensors or cameras can be added for various purposes. In the early days, due to factors such as complex machinery, poor flight performance, difficulty in handling, and high prices, the development of UAVs was almost always carried out by the military and used for military purposes [3]. In recent years, multi-axis UAVs [4] have been actively developed in the field of commercial drones. Two drones used in the study are shown in Fig. 3. The technological innovation of drones has overcome the shortcomings of the past, making performance more stable, flexible, compact, easy to handle, and can be equipped with different components to complete the assigned tasks, and the camera is almost a must-have component.

Yao-Hong Tsai is with the Department of Information Management, Hsuan Chuang University, Taiwan (e-mail: tyh@ hcu.edu.tw).

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:13, No:4, 2019

Fig. 3 Drones used in the study

When a disaster occurs, the drone can arrive at the location where there is no fixed surveillance camera or a dangerous area that has been damaged. Rescue team can understand the area before arrival by the aerial photos [5]. It can improve the safety of rescue workers and make the rescue speed more efficient. UAVs can also be used for discrimination in crop growth monitoring or landform aerial photography. For example, Polo et al. got information of the growth for crops [6], or Chen et al. performed the inspection of the increase and decrease of buildings [7]. The relevant units can use the drone to collect and identify images, which can speed up the verification process and reduce the waste of manpower. In terms of road systems, because of the wide coverage, drones are ideal for such applications [8]-[10].

Object detection technology is one of the most advanced technologies in the field of computer vision technology. The use of UAVs for related surveillance and its application is a very popular and important research in recent years. The reason is that the high mobility capabilities of UAVs can help enhance the effectiveness of surveillance and can be developed in a wide variety of related applications. The most basic core technology lies in the object tracking method in the image. Based on the deep learning method, this paper proposes the vision-based collision avoidance for UAVs by recurrent neural networks.

II. OBJECT DETECTION OR TRACKING BY UAVS

Reilly et al. [11] proposed solutions for important issues such as large camera movement, strong parallax, a large number of moving objects, a small number of target pixels, a single image source channel, and a low frame rate. They also completed the test on the CLIF (Columbus Large Image Format) database. Wang et al. [12] overcame the problem of background image changing with the aerial photography platform under the application of drone image acquisition. The authors adjusted the traditional object detection method and established the background model with Gaussian mixture model. The model of the background motion vector is established by the amount of pixel movement vector variation between two adjacent image images.
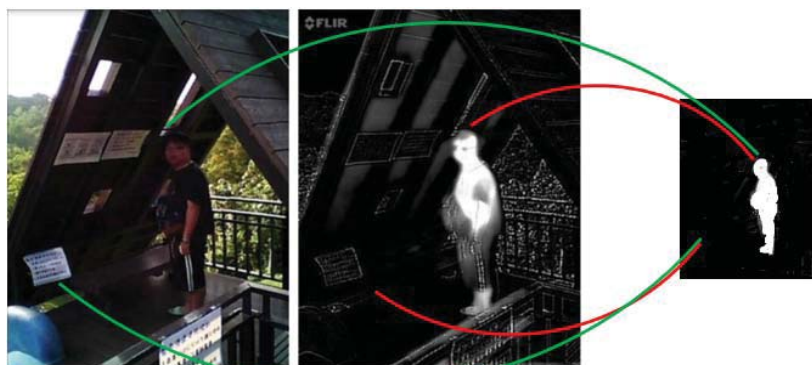
Different sensors can be mounted on the drone for a variety of data collection. It is usually combined with satellite positioning to obtain information about the location for subsequent path planning and task assignment. Lin et al. [13] used the Geographic Information System (GIS) to construct a panoramic image of the scene, and then obtained relevant information corresponding to the background image of the target, so that the spatial relationship between space and time in the environment of the target can be effectively and accurately tracked. Zhou et al. [14] further proposed the method of real-time positioning of roads in the aerial video of UAVs at low and medium altitudes. Their concept is to combine homography-alignment-based tracking and online graph cut detection to locate the region of roads.
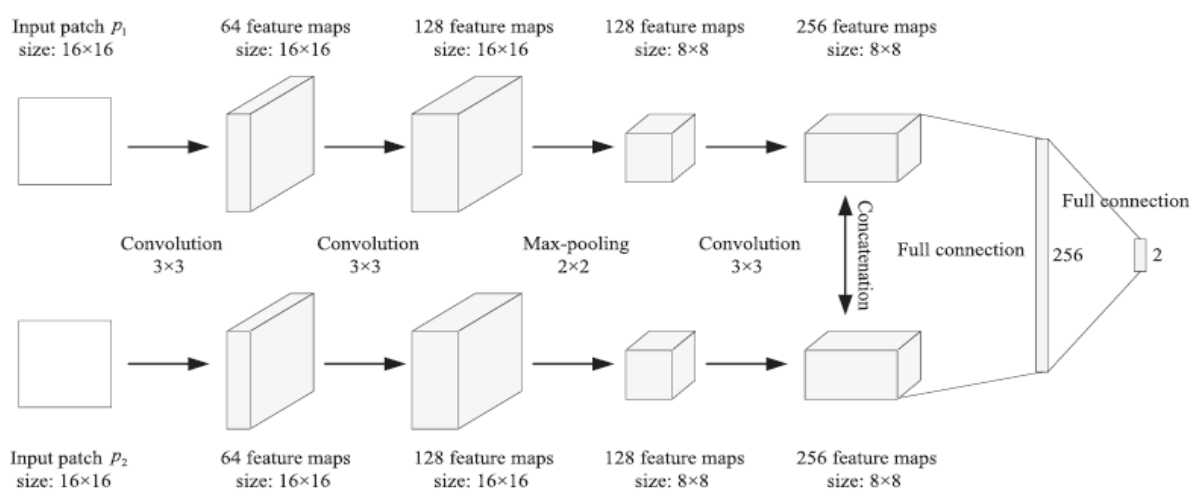
Moranduzzo and Melgani [15] proposed the method to detect ground vehicles and calculate the number of them by drones. First it searched for the asphalt pavement area and narrowed the detection range to this area. This can effectively reduce the proportion of error detection. Then, the image feature points of the scale invariant feature transform (SIFT) are extracted from the input image, and the difference of Gaussian (DoG) is used to obtain a more complete high-frequency signal. Parameswaran et al. [16] proposed technical improvements in object detection for maritime security surveillance issues. In the absence of preset foreground objects and background characteristics, the technology can overcome different problems such as the size, type, and color of the ship. In addition, for the background part, there will be strong sunlight and sea water reflection, waves and water ripples, etc., which will affect the detection results. The authors used the adaptive hypothesis test algorithm to search for small targets in the film and to effectively counter the effects of noise in the background. Then, the feature is captured by the discrete cosine transform (DCT) with the mask, and the final result is determined by the quadratic discriminant analysis (QDA) classifier. In fact, the drone is also flying in the same way while taking the image and is easily disturbed by the instantaneous change of the external wind speed and the airflow, showing an unstable flight state. Therefore, when the object tracked is performed, it must operate under the changing background conditions, and thus greatly increasing the difficulty.

III. THE PROPOSED METHOD

We combine a thermal image with a color image to construct a background and foreground image fusion model. The method can effectively eliminate the influence of shadow or light changes in the image, and then develop a filter suitable for object tracking based on the background model. The model will use CNN's deep learning training method to obtain the parameter values of each channel under the double reference image. In the part of the dual reference image fusion, the requirements of real-time computing will also be considered. The model will be generated by multi-level analysis, and the proposed architecture is shown in Fig. 4. Fig. 4 (a) shows the position of the reference model corresponding to the color image and the thermal image. The source image of the double reference is used to construct the schematic position of the hot zone of the target. Fig. 4 (b) shows the network architecture of CNN. Each time, 16×16 image blocks are used as input, 8 × 8 image blocks are used for sub-sampling, and core masks are trained and tested with 3×3 and 2×2. The experimental images are from the Image Fusion Resource website [17], and the actual image collection is performed using a thermal sensing camera, Flir One [18].

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:13, No:4, 2019

(a)

(b)

Fig. 4 The fusion network for thermal and color image

In recent years, there have been tremendous changes in the field of computer vision and machine learning research, of which graphic recognition is the most obvious example. In some cases, the traditional hand-crafted features are significantly inferior to the learning-based features generated by deep learning. Combining nonlinear classifiers such as Softmax [19] can get good results. The CNN model LeNet-5 proposed by Yann Lecun is shown in Fig. 5. Therefore, under the real time consideration, we try to construct high-order image features in the image in a deep learning manner to improve system performance. In 1998, Yann Lecun proposed the convolutional neural network (CNN), which can train deep learning algorithms for multi-layer network structures and apply them to text recognition [20]. Their proposed CNN model is limited by time and convergence, and the application effect has been poor. Until later, Geoffrey Hinton proposed a pretraining method, which led to breakthrough developments in CNN related applications and launched a wave of deep learning [21]-[23].
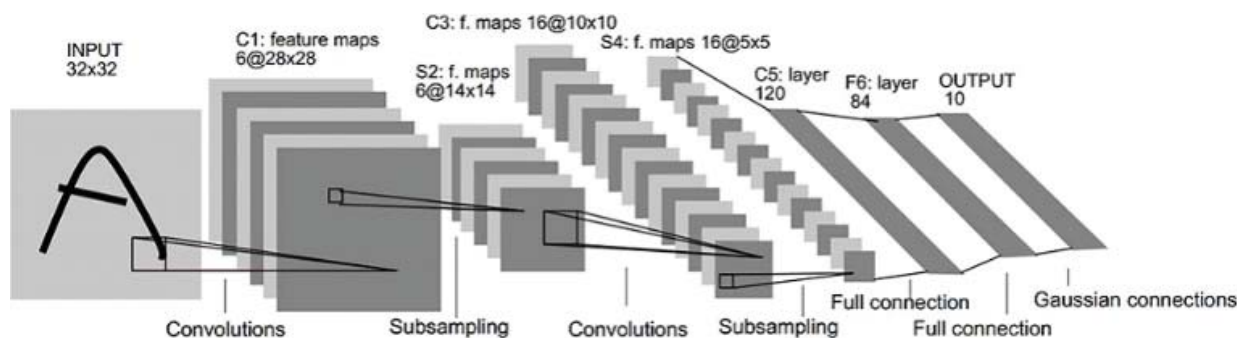

Fig. 5 The CNN model LeNet-5 proposed by Yann Lecun [19]

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:13, No:4, 2019

The Recurrent Neural Network (RNN) [24] is a neural network that connects the output signal to the input. The nodes in the middle of the hidden layer will be connected to each other. The purpose of this approach is to pass back the output value from the last point in time and record it in the neuron. The output value of the next time point can be related to the previous time point to achieve a memory-like effect on the neural network. RNN can be used to describe the dynamic behavior over time, and its network architecture is shown on Fig. 6 (b). The simple cyclical neural network is difficult to deal with the exponential growth or disappearance of weights, and it is difficult to extract features in a long time. The model usually combines different long short-term memory (LSTM) [25] and uses different functions to calculate the state of the hidden layer. LSTM can be used to process and predict important events with very long intervals and delays in time series. In the past, it has been successfully applied in the recognition and speech recognition of handwriting [26]. Fig. 6 (b) shows the structure of the cyclic neural network and the extended architecture over time. Fig. 6 (a) introduces functions on the input, output and forget gates to calculate the node state for the neural network combined with LSTM.
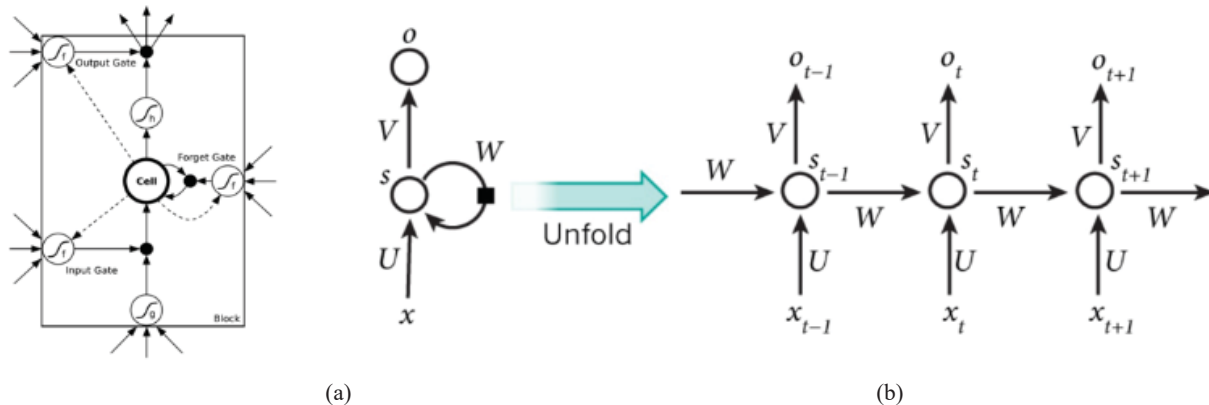

(a)   (b)

Fig. 6 The Network architecture of cyclic neural networks and long-term and short-term memory models

This study established a mathematical model for the operation of two drones based on the Newton's second theorem to design the equation of motion. The parameters of the equation will be measured instantaneously by the flight process, and the mathematical model of the parameters can be dynamically input, which can be used for the flight attitude of the drone to accomplish the vision-based collision avoidance. By designing the state feedback control law and the proportional integral derivative (PID), the parameter adjustment of the control is realized in the closed loop system, and the result of the object tracking is achieved for collision avoidance control. Finally, Matlab's built-in Simulink is applied to draw the overall system control chart. It is convenient to simulate and analyze the single-drone flight and the two-drone formation flight. The results of the simulation verification are used to judge whether the designed controller is feasible and can be applied to the actual hardware. Rewrite the open source APM flight control program and the ground station program Mission Planner for verification testing [27].

IV. EXPERIMENTATIONS

In order to measure the performance of object detection and tracking in security monitoring applications, several commonly used image databases are tested in this paper. At present, the research team collects the well-known object tracking video database including OTB50 [28], OTB100 [29], VOT2014, VOT2015, TC128 [30], ALOV300++ [31] and the test database UAV123 with UAVs [32]. However, most common image or video databases are currently designed for specific methods and are usually incomplete in type and quantity. We also collect the latest standard video database by drones with cameras for thermal and color information and construct a relatively complete and feasible test database.

In terms of performance measurement, the two indicators of false rejection and false acceptance are mainly used, because both indicators will be affected by the threshold and will affect each other. One decline will lead to another increase. In order to obtain the overall system performance, this paper adopts the HTER (Half Total Error Rate) index proposed by Anjos and Marcel [33], [34], which is based on the above two types of errors, false rejection rate (FRR) and false acceptance rate (FAR). The system was able to recognize moving objects in the data set with 97.5% of accuracy (526 out of 540 objects), while only 16 instances were incorrectly categorized (14 airplanes were not identified, and 2 objects were wrongly identified). The positive prediction value for our CNN was calculated to be 99.6%, false discovery rate was 0.4%, true positive rate was 97.4%, and false negative rate was 2.6%.

V. CONCLUSION

This paper successfully developed real-time moving object tracking and vision-based collision avoidance technology that can be applied to drones for surveillance applications and the research results fit industrial considerations. Combining deep learning to solve high-level and low-level features extraction and noise problems and integrating spatial and frequency domain identification scores can increase the overall system identification performance. The integrity and comparison of the experimentations will continue to be strengthened in the future.

REFERENCES

[1] The website of the Taipei Traffic Control Center http://tms.bote.taipei.gov.tw/main.jsp?lang=zh_TW.
[2] "Unmanned Aircraft Systems". ICAO. Accessed 2nd August, 2016 http://www.icao.int/Meetings/UAS/Documents/Circular%20328_en.pdf.
[3] S. A. Cambone, K. J. Krieg, P. Pace and L. Wells II, "Unmanned aircraft systems (UAS) roadmap 2005–2030," *USA: Office of the Secretary of Defense*, 2005.
[4] M. Corcoran, "Drone wars: The definition dogfight". Accessed 2nd August 2016. http://www.abc.net.au/news/2013-03-01/dronewars-the-definition-dogfight/4546598.
[5] A. Ahmed, M. Nagai, C. Tianen, and R. Shibasaki, "Uav based monitoring systemand object detection technique development for a disaster area," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, pp. 373–377, 2008.
[6] J. Polo, G. Hornero, C. Duijneveld, A. García and O. Casas, "Design of a low-cost Wireless Sensor Network with UAV mobile node for agricultural applications," *Computers and Electronics in Agriculture*, vol. 119, pp. 19–32, 2015.
[7] B. Chen, Z. Chen, L. Deng, Y. Duan and J. Zhou, "Building change detection with RGB-D map generated from UAV images," *Neurocomputing*, vol. 208, pp. 350–364, 2016.
[8] B. Coifman, M. McCord, R. Mishalani, M. Iswalt and Y. Ji, "Roadway trafficmonitoring froman unmanned aerial vehicle," *IEE Proceedings-Intelligent Transport Systems*, vol. 153, no. 1, pp. 11–20, 2006.
[9] K. Kanistras, G. Martins, M. J. Rutherford and K. P. Valavanis, "Survey of unmanned aerial vehicles (uavs) for traffic monitoring," *in Handbook of Unmanned Aerial Vehicles*, pp. 2643–2666, 2015.
[10] P. J. Hiltner, "Drones Are Coming: Use of Unmanned Aerial Vehicles for Police Surveillance and Its Fourth Amendment Implications," *The. Wake Forest JL & Pol'y*, vol. 3, pp. 397, 2013.
[11] V. Reilly, H. Idrees and M. Shah, "Detection and tracking of large number of targets in wide area surveillance," *Computer Vision ECCV*, pp. 186-199, 2010.
[12] Y. Wang, Z. Zhang and Y. Wang, "Moving Object Detection in Aerial Video", *11th Inter-national Conference on Machine Learning and Applications*, pp. 446-450, 2012.
[13] C. Lin, S. Pankanti, G. Ashour, D. Porat and J. R. Smith, "Moving camera analytics: Emerging scenarios, challenges, and applications", *IBM Journal of Research and Development*, vol. 59, pp: 5:1-5:10, 2015.
[14] H. Zhou, H. Kong, L. Wei and D. Creighton, "Efficient Road Detection and Tracking for Unmanned Aerial Vehicle", *Transactions on Intelligent Transportation Systems*, vol. 16, pp. 297-309, 2015.
[15] T. Moranduzzo and F. Melgani, "Automatic Car Counting Method for Unmanned Aerial Vehicle Images", *Geoscience and Remote Sensing*, vol. 52, pp. 1635 – 1647, 2014.
[16] S. Parameswaran, C. Lane, B, Bagnall and H. Buck, "Marine Object Detection in UAV full-motion video", *Proc. SPIE 9076 Airborne Intelligence, surveillance, Reconnaissance Systems and Applications*, XI, 907608, 2014.
[17] The website of ImageFusion.Org, The Online Resource for Research in Image Fusion, http://www.imagefusion.org/.
[18] The website of Flir camera, http://www.flir.tw/flirone/.
[19] The website of Softmax, http://ufldl.stanford.edu/tutorial/supervised/SoftmaxRegression/.
[20] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
[21] A. Mnih, and G. E. Hinton, "Learning nonlinear constraints with contrastive backpropagation," *In: Neural Networks, IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on. IEEE*, p. 1302-1307, 2005.
[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 38, no. 1, pp. 1–1, 2015.
[23] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke and J. Schmidhuber, "A Novel Connectionist System for Improved Unconstrained Handwriting Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 855–868, 2009.
[24] R. Girshick, "Fast r-cnn," *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.
[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory". *Neural Computation*. vol. 9, no. 8, pp. 1735–1780, 1997.
[26] G. E. Hinton, et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
[27] The website of Mission Planner, http://ardupilot.org/planner/docs/mission-planner-overview.html.
[28] Y. Wu, J. Lim and M. H. Yang, "Online Object Tracking: A Benchmark," *In: 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2411-2418, 2013.
[29] Y. Wu, J. Lim and M. H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834-1848, 2015.
[30] P. Liang, E. Blasch and H. Ling, "Encoding color information for visual tracking: Algorithms and benchmark," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5630-5644, 2015.
[31] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan and M. Shah, "Visual tracking: An experimental survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442-1468, 2014.
[32] M. Mueller, N. Smith and B. Ghanem, "A Benchmark and Simulator for UAV Tracking," *ECCV 2016: European Conference on Computer Vision*, pp. 445-461, 2016.
[33] A. Anjos, and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," *in Proc. IJCB*, pp. 1–7, 2011.
[34] Wu, H.Y., M. Rubinstein, E. Shih, J. Guttag, F. Durand and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, Art. ID 65, 2012.

Yao-Hong Tsai received the M.S. and Ph.D. degrees in information management from the National Taiwan University of Science and Technology (NTUST), Taipei, Taiwan, R.O.C., in 1994 and 1998, respectively. He was a Researcher with the Advanced Technology Center, Information and Communications Research Laboratories, Industrial Technology Research Institute (ITRI), Hsinchu. He is currently an Associate Professor with the Department of Information Management, Hsuan Chuang University, Hsinchu. His current research interests include image processing, pattern recognition, personal communication systems, and cloud computing.