

Face Recognition on Drones: Issues and Limitations

Hwai-Jung Hsu

Institute of Information Science, Academia Sinica
hjhsu@iis.sinica.edu.tw

Kuan-Ta Chen

Institute of Information Science, Academia Sinica
swc@iis.sinica.edu.tw

ABSTRACT

Drones, as known as unmanned aerial vehicles (UAV), are aircrafts which can perform autonomous pilot. They can easily reach locations which are too difficult to reach or dangerous for human beings and collect images from bird's-eye view through aerial photography. Enabling drones to identify people on the ground is important for a variety of applications, such as surveillance, people search, and remote monitoring. Since faces are part of inherent identities of people, *how well face recognition technologies can be used by drones* becomes essential for future development of the above applications.

In this paper, we conduct empirical studies to evaluate several factors that may influence the performance of face detection and recognition techniques on drones. Our findings show that the current face recognition technologies are capable of recognizing faces on drones with some limits in distance and angle, especially when drones take pictures in high altitudes and the face image is taken from a long distance and with a large angle of depression. We also find that augmenting face models with 3D information may help to boost recognition performance in the case of large angles of depression.

1. INTRODUCTION

Drones, as known as unmanned aerial vehicles (UAV), are aircrafts without pilots on board [6] that can be piloted remotely or autonomously. They can fly pre-programmed missions without manual controls using autopilot suites [2]. Drones can easily reach locations which are too difficult to reach or dangerous for human beings to take pictures from bird's-eye view. Drones with aerial cameras are widely used in photogrammetry [3], surveillance [23], and remote sensing [9, 11]. In these applications, drones are used to detect or track down specific people on the ground, and to identify individuals from drones is thus a critical feature.

Faces are part of inherent identities of people, and identifying individuals through their faces is human nature [19].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DroNet'15, May 18, 2015, Florence, Italy.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3501-0/15/05 ...\$15.00.

Face recognition is popular in the field of computer vision and can be viewed as a badge of success in image analysis and understanding. Face recognition capability is undoubtedly a key for drones to identify specific individuals within a crowd. For example, to adopt drones in the search of missing elderlies or children in the neighborhood, the drones first need to know who the targets are, and then the search can be launched. Thus, face recognition on drones would be a vital technical component in such applications; consequently, *how well face recognition perform on drones* is a research topic worth to be investigated.

In this paper, we aim to understand the limits of the present face detection and recognition technologies while they are applied on drones, and provide possible guidelines for integrating face recognition into drone-based applications. Since drones may fly in-door or out-door under any kinds of illumination or environment conditions and may take pictures from the air with any possible combination of distance, altitude, and angle of depression. As such, we consider only *unconstrained face recognition* [20,21] technologies in this work. We conduct a series of empirical studies to examine the capability of two popular online face recognition services, Face⁺⁺ [12] and ReKognition [17], in recognizing specific human faces on pictures collected by drones. The influences caused by distances and angles of depression from drones to the subjects are investigated so as to systematically investigate the limits of current face recognition technologies when applied on drones.

The remainder of this paper is organized as follows. Section 2 shows the related works about face recognition on drones. We describe the issues applying face recognition on drones in Section 3, and sketch the experiment design for evaluating the issues in Section 4. In Section 5, our results about evaluating the performance of Face⁺⁺ and ReKognition on drones are shown. The limits of applying current face recognition technology on drones are investigated, and possible approaches for pushing the limits are discussed. Finally, our conclusion and future works are addressed in Section 6.

2. RELATED WORK

The most well-known application of face recognition on drones is that the United States Army combines face recognition with drones for detecting and tracking targets with threat [15]. However, the technology adopted by military is usually confidential and can not be applied for commercial or common use. In early development, thermal images are widely applied on UAVs to track down human targets or vehicles [1, 18]. However, thermal images is insufficient for

accurate identification of people and can only be used for tracking or warning.

Except for using thermal images, Davis et al. develop an LBP-based (local binary patterns) methodology to apply face recognition onto a commercial off-the-shelf UAV for security applications [13]. Davis et al. claim that their system is economic and can be widely applied; nevertheless, they do not evaluate the limits and effectiveness of their system [13]. Korshunov et al. investigate the critical video quality on face recognition that clarifies the limits of face recognition on drones under strict network environment originated from drones flight [10]. Besides, face recognition is also essential for applying drones in rescue missions, and is one of the primary events in robot competition [14].

3. RESEARCH CHALLENGES

To achieve accurate face recognition, the facial images for recognition are recommended to follow the criteria below [16]:

1. 50 pixels between the eye centers is the minimum recommended size for a facial image to perform face template extraction, an necessary pre-process for face recognition.
2. 75 to 90 pixels between the eye centers is the recommended minimal size for a facial image to perform accurate face recognition.
3. The face recognition engine may tolerate the facial image with a certain face posture and still performs good recognition, e.g., $\pm 15^\circ$ in head roll (tilt), $\pm 25^\circ$ in head pitch (nod), and $\pm 30^\circ$ in head yaw (bobble).

The distances from drones and their targets directly affect the size of the facial images in pixels. Since drones take picture from the air, altitudes of drones keep them distant from their targets on the ground. Altitudes also form angles of depression from drones to their targets, and the pitch angles of the facial images collected by drones can thus be large. Besides, speed and flight attitude might also affect the quality of the facial images and degrade the performance of face recognition. Because the influences originated from speed and flight attitude can be compensated with appropriate settings on aerial cameras, we mainly investigate how distances and angles of depression influence the performance of face recognition in this paper.

4. EXPERIMENTS SETUP

Face⁺⁺ and ReKognition are two famous on-line face recognition services. They perform well in corresponding benchmarks [4], and provide open API for development of various kinds of applications. Thus, Face⁺⁺ and ReKognition are adopted for examining how current methods are effective on recognizing faces from the pictures collected by drones.

Figure 1 shows our experiment setup. We took frontal facial images from 11 subjects with GoPro [5], the most popular sports camera for aerial photography in the market. Because controlling drones to take pictures at exact altitudes and positions for various subjects is difficult, we put our GoPro onto a cradle for photographing instead of taking pictures with drones. The GoPro is set up at 3, 4, and 5 meters in heights to simulate the drones flying at the certain altitudes, and the pictures are taken every 0.5 meters between 2 to 17 meters away from the subject on the ground to

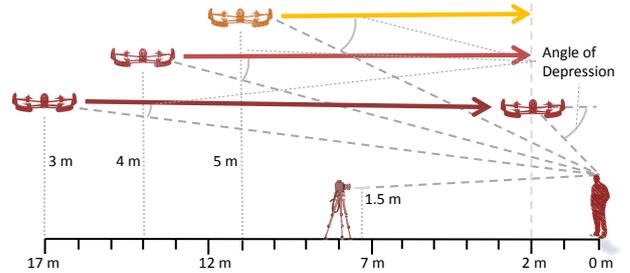


Figure 1: The sketches of the experiment setup.

simulate a 15 meters straight flight toward the subject. We also take the frontal facial pictures at 1.5 meters in heights for comparison. The angles between the horizontal and the line from the GoPro to the subjects' top are considered as the angles of depression between the aerial camera and the subjects. We also ask the subjects to observe the following rules: (1) taking off their glasses to simplify the factors in face recognition, (2) gazing straight ahead to keep a consistent pitch angles of their faces, (3) keeping a deadpan face to suppress influences introduced by facial expressions, and (4) standing still to eliminate the affects caused by movement.

Face⁺⁺ and ReKognition consume facial images of people for model training, and then able to recognize the faces of those people from the other pictures on the basis of the trained model. To collect the source pictures for model training, we use a built-in camera on a smart phone (HTC One M8) to take photos of the subjects. Besides, 7 of the 11 subjects hand in their own glass-free portrait photos for model training. With the photos, the models are trained three-fold:

1. Model \mathcal{J} : with the photos just taken;
2. Model \mathcal{P} : with the portraits provided by the subjects;
3. Model \mathcal{B} : with both of the photos.

To sum up, we take 620 pictures in 3, 680x2, 760 (10 mega pixels) using our GoPro Hero 3⁺ Silver Edition with ultra-wide field of view (170°) among various settings of heights (1.5, 3, 4, and 5 meters) and distances (2 to 17 meters with intervals of 0.5 meters in-between). Along with the settings, total 1,364 facial images from the 11 subjects are collected. Besides, on Face⁺⁺ and ReKognition, Model \mathcal{J} is trained for all the 11 subjects, and Model \mathcal{P} and \mathcal{B} are trained for the 7 subjects who hand in their own portraits.

5. PERFORMANCE EVALUATION

In this section, we evaluate how the challenges mentioned in Section 3 influence the performance of Face⁺⁺ and ReKognition.

5.1 Face Detection

A face needs to be detected before it can be recognized. In this section, the performance of face detection among Face⁺⁺, ReKognition, and the methods in OpenCV [8] (including four Haar and one LBP-based methods) are compared under various settings in altitudes and ground distances between drones and their targets. Table 5.1 shows

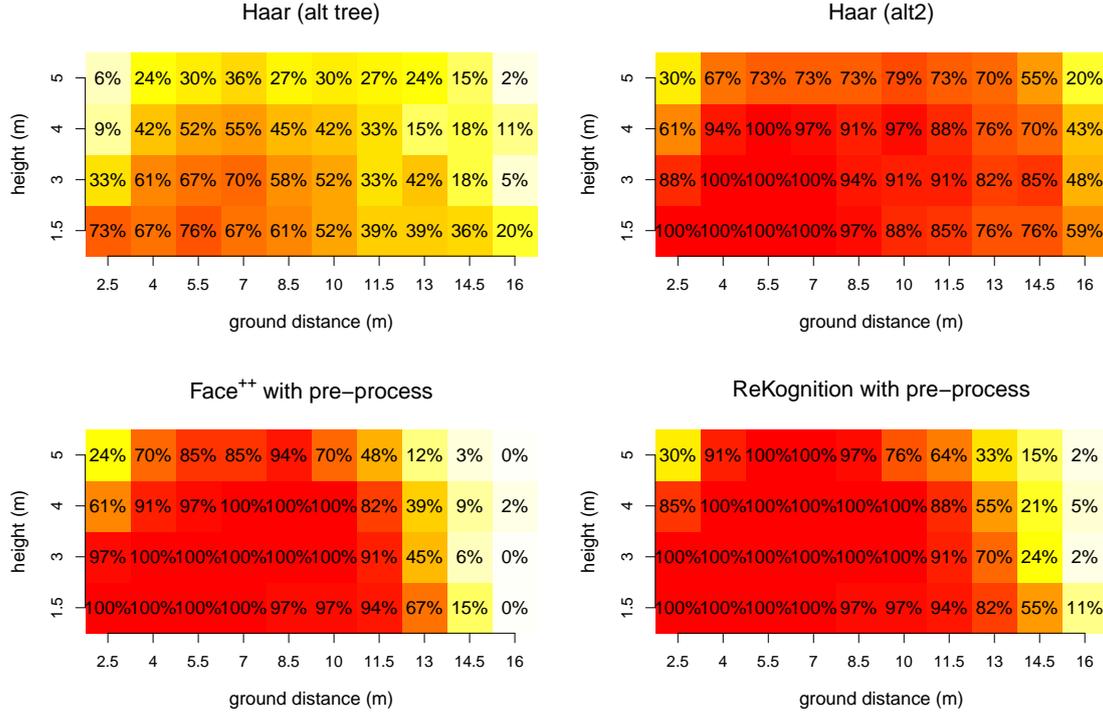


Figure 2: The face detection rate for Haar (alt tree), Haar (alt2), Face⁺⁺, and ReKognition in correspondence to heights and distances.

Table 1: Performance of face detection

Method	# of faces	TPR	FPR
Face ⁺⁺	20	0.14	0.05
ReKognition	37	0.27	0.13
Haar (default)	14,777	0.71	0.93
Haar (alt)	1,700	0.77	0.37
Haar (alt2)	2,545	0.78	0.57
Haar (alt tree)	510	0.37	0.002
LBP	2,964	0.63	0.70

*The methods are asked to detect 1,364 target faces from 620 pictures.

the corresponding results composed of (1) the total number of faces detected (# of faces), (2) the true positive rate (TPR), and (3) the false positive rate (FPR).

As a result, the alternative Haar-based method (Haar alt) performs the best with relatively high TPR and low FPR. On the contrary, both Face⁺⁺ and ReKognition perform poorly in detecting faces directly from the pictures we gathered. One possible reason is Face⁺⁺ and ReKognition resize the input pictures into smaller ones for efficiency. Take ReKognition for example, ReKognition allows only up to 800 pixels in widths or heights for the uploaded pictures. The larger-size pictures sent to ReKognition through links are resized internally [17]. On the other hand, as we uploaded the facial images manually cut from the collected pictures onto ReKognition for detection, ReKognition does detect the face which is ignored in the original pictures. Thus, before putting the pictures gathered by drones onto Face⁺⁺ or ReKognition, the pictures need to be pre-processed for ex-

traction of faces. With the assistance of methods in OpenCV and some manual works, we extract all 1,364 target faces from the collected pictures. As a result, Face⁺⁺ and ReKognition detects 885 and 984 faces among them correspondingly. Figure 2 shows the heat map of the face detection rate of Face⁺⁺ and ReKognition among various settings in heights and distances with the assistance of external face detection. Besides, we also attach the results made by Haar (alt2) and Haar (alt tree), the OpenCV methods with the highest TPR and the lowest FPR, in Figure 2 for comparison. The influences introduced by distances and angles of depression are obvious. Haar (alt2) performs better in distances beyond 12 meters, while Face⁺⁺ and ReKognition give a better detection in heights of 3 and 4 meters. All the methods suffer poor performance in combination of short distances (less than 4 meters) and the highest altitude (5 meters), i.e., with large angles of depression.

5.2 Face Recognition

In this section, we evaluate how distances and angles of depression influence the performance of face recognition.

5.2.1 Impact of Distances

First, we investigate how distances between drones and their targets impact the performance of face recognition in this section. Among all the facial images we extract, the faces obtained while the camera is set up at 1.5 meters in heights and 2 to 12 meters in distances are used for the evaluation because within the settings, (1) both Face⁺⁺ and ReKognition show relatively high and stable TPR in face detection, and (2) taking the heights of the subjects into

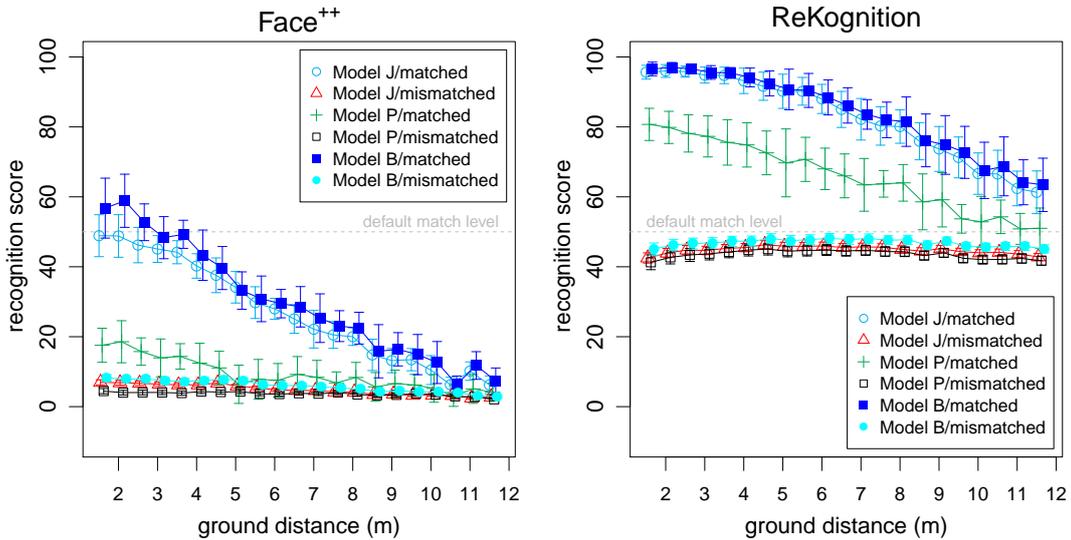


Figure 3: The scores given by Face⁺⁺ and ReKognition under various ground distances between the drones and the subjects.

consideration, the angles of depression from the camera to the targets among the settings are less than 10° , and thus, the influences introduced by angles of depression are alleviated.

Face⁺⁺ gives scores between 50 to 100 for evaluation of whether a face belonging to or not to a designated person. For example, we train a face recognition model for Alan with his own portrait photos. Then, another picture of Alan and a picture of Ben, a person who looks nothing like Alan, are input to Face⁺⁺ for recognition. With the model, Face⁺⁺ rates 75 positive to the face in Alan’s picture and 90 negative to the face in Ben’s to elaborate how they are similar/dissimilar to Alan’s face according to the model trained with Alan’s portrait. In other words, the facial image in Alan’s picture scores 75 and the face in Ben’s scores 10 in similarity to Alan’s face in Face⁺⁺. On the other hand, ReKognition rates between 0 to 1 as the probability of whether a face belonging to a designated person. As the example described above, ReKognition shows that the face in Alan’s picture is Alan with the probability of 0.75 and the face in Ben’s is Alan with the probability of 0.1. ReKognition decides a face belonging to a person if the probability is beyond 0.5. Thus, we can map the scores rated by Face⁺⁺ and ReKognition into a 0 to 100 scale, and use 50 as the *default match level* that Face⁺⁺ and ReKognition decides whether a target face belonging to a designated subject. We train models as Section 4 describes, and ask Face⁺⁺ and ReKognition to rate the 1,364 target faces gained from the experiment for each of the 11 subjects.

To evaluate the distinguishability of Face⁺⁺ and ReKognition, we define *matched* and *mismatched* cases as following. A matched case represents the face being rated belonging to the owner of the model used for recognition. On the contrary, a mismatched case is recognitions between a face and the models belonging to the subjects other than the face owner. The score of a mismatched case is the mean value of the scores for all such recognitions.

Figure 3 shows the results about how Face⁺⁺ and ReKognition perform in recognizing faces from various distances.

Since both Face⁺⁺ and ReKognition rate only the detected faces, we merely consider detected faces in the results. The x-axis of the figure indicates the ground distances, and the y-axis represents the mapped scores given by each method. The lines and dots represent the average scores based on Model \mathcal{J} , \mathcal{P} , and \mathcal{B} for matched and mismatched cases correspondingly. The band markers by the sides of each dot show the 95% confidence intervals.

As a result, Face⁺⁺ rates almost all the matched cases below the default match level, and judges only the matched cases captured within 3 meters as positive while Model \mathcal{B} is applied. However, comparing the average scores and the corresponding confidence intervals between the matched and the mismatched cases, Face⁺⁺ does distinguish the matched cases from the mismatched ones within 9 meters while Model \mathcal{J} or \mathcal{B} is adopted. In other words, alteration of the match level is necessary to apply Face⁺⁺ for face recognition on drones. Face⁺⁺ perform poorly with Model \mathcal{P} . Even with altered match level, Face⁺⁺ can only distinguish the matched cases from the mismatched ones within 4 meters. On the other hand, ReKognition rates almost all the matched cases beyond the default match level and vice versa. The scores between matched and mismatched cases are significantly distinguishable within 12 meters for all the models. As the distances get short, ReKognition shows an even better performance that the scores for matched cases rise and the ones for mismatched cases drop.

5.2.2 Impact of Angles of Depression

Since drones take pictures from the air, their altitudes generate angles of depression between drones and their targets, and thus influence the poses of the faces in the collected pictures. In this section, we study how angles of depression impact the performance of face recognition, and the possible methodology for pushing the limits on recognizing the faces with large angles of depression.

Based on the results in Section 5.2.1, the scores rated by Face⁺⁺ and ReKognition are relatively stable while the ground distances are within 4 meters. Thus, the target faces

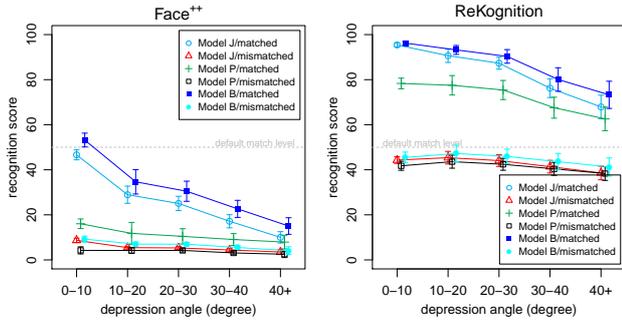


Figure 4: The scores given by Face⁺⁺ and ReKognition under various angles of depression with different training models.

collected with 1.5, 3, 4, and 5 meters in heights and 2 to 4 meters in ground distances are used for evaluating the performance of Face⁺⁺ and ReKognition in face recognition among various angles of depression. As the evaluation done in Section 5.2.1, the average scores for matched and mismatched cases are separately calculated. Considering only the detected faces, Figure 4 shows the results generated by Face⁺⁺ (left) and ReKognition (right). For both of the methods, scores drop as angles of depression get large. Even so, ReKognition gives scores of all the matched cases beyond the default match level, and significantly distinguish the matched cases from the mismatched ones. Face⁺⁺ still performs poorly while adopting Model \mathcal{P} , and rates low to both the matched and mismatched cases with angles of depression more than 40° for all the models. Therefore, although Face⁺⁺ looks able to distinguish faces collected with large angles of depression, some augmentation may still be required.

One of the possible approaches for such augmentation is adopting 3D modelling technique to generate photos posing additional pitch angles for model training. Kemelmacher-Shlizerman et al. reconstruct a 3D face from a single uncontrolled facial image that the 3D face can present positioning angles not presented in the original image [7]. If the face recognition model is trained with extra images of large pitch angles, the distinguishability to faces collected with large angles of depression might be augmented. To examine this idea, we use FaceGen Modeller 3.5 [22] made by Singular Inversions, Inc. to generate 3D facial models from the subjects' portraits. For each subject, 10 frontal face images with pitch angles from 0° to 45° are generated. The additional face images are put into Face⁺⁺ and ReKognition together with the original ones for model training. The augmented models are annotated as Model \mathcal{J}' , \mathcal{P}' and \mathcal{B}' in the following paragraphs for short.

As a result, Figure 5 shows how 3D augmentation influences the distinguishability of Face⁺⁺ and ReKognition in various angles of depression. Although 3D augmentation helps little to ReKognition, Face⁺⁺ is undoubtedly beneficial from it. The scores for the matched cases are obviously ascended among almost all the angles of depression, especially for large angles of depression. However, the scores for mismatched cases rise for both methods after 3D augmentation is introduced. Rises on scores of mismatched cases might weaken the distinguishability of the methods. One of the possible reasons for the phenomenon is the faces gener-

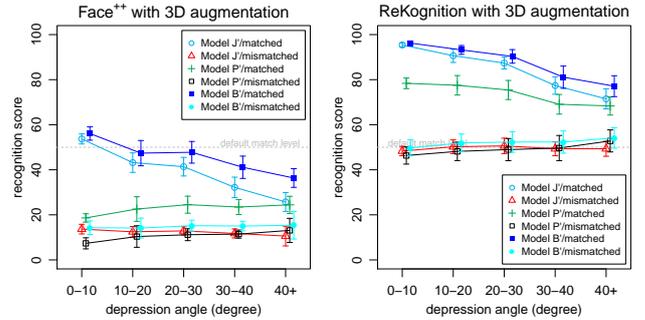


Figure 5: The scores given by Face⁺⁺ and ReKognition under various angles of depression with 3D augmentation.

ated from FaceGen Modeller might not be sufficiently authentic, and confuse the scoring mechanism in Face⁺⁺ and ReKognition for the mismatched cases.

5.3 Discussion

From the results in Section 5.1, we know that both Face⁺⁺ and ReKognition suffer poor face detection rate in large angles of depression. Since we do not consider the face detection rate in Section 5.2, the influences introduced by low face detection rate at large angles of depression are discussed in this section.

Assuming both the methods give 0 points to the undetected faces for matched and mismatched cases. Figure 6 shows the AUC (area under curve) of ROC (receiver operating characteristic curve) representing the capability of both the methods in distinguishing the matched cases from the mismatched ones among various combination of heights and distances while Model \mathcal{B} is applied. As a result, the influences from angles of depression and distances are significant. Take 0.75 as the standard of acceptable distinguishability, Face⁺⁺ is applicable on drones while the distances are within 12 meters, and so does ReKognition within 14 meters. Both the methods show no distinguishability in large angles of depression (with 5 meters in heights and ground distances less than 3 meters), and need some distances away from the targets to prevent the influences introduced by angles of depression. Face⁺⁺ needs about 3 and 5 meters on the ground for heights in 4 and 5 meters correspondingly, and ReKognition needs 3 meters on the ground for heights in 5 meters.

6. CONCLUSION AND FUTURE WORK

In this paper, we investigate how altitudes, distances, and angles of depression are influential to the performance of face recognition on drones. Through the empirical studies on Face⁺⁺ and ReKognition, we conclude that the present face recognition technologies are able to perform adequately on drones. However, some obstacles need to be conquered before such techniques can unleash their full potentials:

1. The small-sized facial images taken by drones from long distances do cause trouble to both face detection and recognition.
2. The pose variances introduced by large angles of depression dramatically weaken the capability of both face detection and recognition.

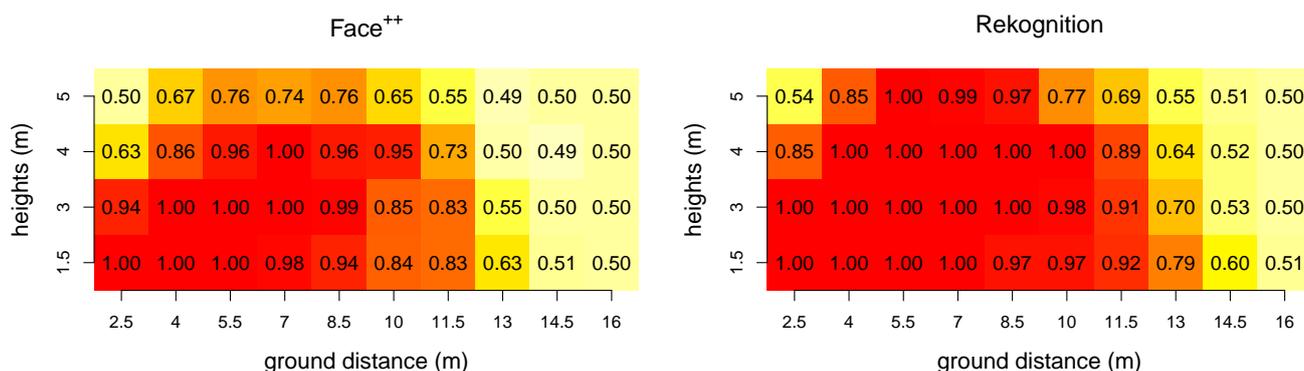


Figure 6: The capability of distinguishing the matched cases from the mismatched ones for Face++ and ReKognition considering both detected and non-detected faces.

3. A recognition model augmented with 3D modelling techniques might increase the performance of face recognition in the case with large angles of depression. However, this augmentation may also decrease the distinguishability of faces in common cases, and thus requires further investigation.

In the future, since the sizes of facial images greatly influence the performance of face recognition, how the parameters of aerial cameras (e.g., resolutions and compression rate) may impact the performance of face recognition on drones should be further studied. Besides, cameras with large FOV (field of view) not only capture wide scenes into pictures, but also generate morphs at the margin of the pictures. To compensate the negative influences caused by such morphs is also worth of investigation. Last but not least, although we conclude that the current face recognition techniques are capable on drones, applying online services such as Face++ and ReKognition directly on drones may be practically infeasible. Constraints from network bandwidth, batteries, and computation power of the embedded system carried by drones limit how face recognition can be applied in this scenario. Developing a face recognition enabled drone-based system which is balanced in accuracy, computation, network transmission, and power consumption will be part of our future plans.

7. REFERENCES

- [1] A. GÅEszczak, T. P. Breckona, and J. Hana. Real-time People and Vehicle Detection from UAV Imagery. In *IS&T/SPIE Electronic Imaging*, pages 78780B–1–13, 2011.
- [2] APM. APM Autopilot Suite, <http://ardupilot.com/>.
- [3] F. Nex and F. Remondino. UAV for 3D mapping applications: a review. *Applied Geomatics*, 6(1):1–15, March 2014.
- [4] G. B. Huang, M. Ramesh and T. Berg and E. Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
- [5] GoPro, Inc. GoPro, <http://gopro.com>.
- [6] International Civil Aviation Organization. CIR328 AN/190 Unmanned Aircraft Systems (UAS), 2011.
- [7] Ira Kemelmacher-Shlizerman, Ronen Basri. 3D Face Reconstruction from a Single Image Using a Single Reference Face Shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):394–405, 2011.
- [8] Itseez. OpenCV, <http://opencv.org/>.
- [9] J. Paneque-GÅalvez, M. K. McCall, B. M. Napoletano S. A. Wich, and L. P. Koh. Small Drones for Community-Based Forest Monitoring: An Assessment of Their Feasibility and Potential in Tropical Areas. *Forests*, 5(6):1481–1507, 2014.
- [10] P. Korshunov and W. T. Ooi. Video quality for face detection, recognition, and tracking. *ACM Trans. Multimedia Comput. Commun. Appl.*, 7(3):14:1–14:21, Sept. 2011.
- [11] M. Bartholmai, E. Koeppe, and P. P. Neumann. Monitoring of Hazardous Scenarios using Multi-Sensor Devices. In *Proceedings of SENSORDEVICES 2013*, pages 9–13, 2013.
- [12] Megvii, Inc. Face++, <http://www.faceplusplus.com/>.
- [13] N. Davis, F. Pattaluga, and K. Panetta. Facial recognition using human visual system algorithms for robotic and UAV platforms. In *Proceedings of TePRA 2013*, pages 1–5, 2013.
- [14] N. Dijkshoorn et al. Amsterdam Oxford Joint Rescue Forces - Team Description Paper - Virtual Robot Competition - Rescue Simulation League - RoboCup 2011. In *Proceedings of the 15th RoboCup Symposium*, pages 1–8, 2011.
- [15] N. Shachtman. Army Tracking Plan: Drones That Never Forget a Face, <http://www.wired.com/2011/09/drones-never-forget-a-face/>.
- [16] Neurotechnology. Basic Recommendation for Facial Recognition, <http://www.neurotechnology.com/face-image-recommendations-constraints.html>.
- [17] Orbeus, Inc. ReKognition API, <https://rekognition.com/>.
- [18] P. Rudol and P. Doherty. Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery. In *2008 IEEE Aerospace Conference*, pages 1–8, March 2008.
- [19] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.
- [20] S. K. Zhou, R. Chellappa, and W. Zhao. *Unconstrained Face Recognition*, volume 5. Springer Science & Business Media, 2006.
- [21] S. Zafeiriou, I. Kotsia, and M. Panti. Unconstrained face recognition. *Face Recognition in Adverse Conditions*, 2, 2014.
- [22] Singular Inversions, Inc. FaceGen, <http://www.facegen.com/>.
- [23] T. Wall and T. Monahan. Surveillance and violence from afar: The politics of drones and liminal security-scapes. *Theoretical Criminology*, 15(3):239–254, 2011.