Set-up of a method for people-counting using images from a UAV.

Daniel R. Parisi[1*], Juan I. Giribet[2,3,4], Claudio Pose[2], and Ignacio. A. Mas[5]

[1] Dto. Ing. Informática, Instituto Tecnológico de Buenos Aires, CONICET, Lavardén 315, C. A. de Buenos Aires, Argentina.
*dparisi@itba.edu.ar

[2] Dto. Ing. Electrónica, Facultad de Ingeniería, Universidad de Buenos Aires, Av. Paseo Colón 850, C. A. de Buenos Aires, Argentina.

[3] Dto. de Matemática, Facultad de Ingeniería, Universidad de Buenos Aires, Av. Paseo Colón 850, C. A. de Buenos Aires, Argentina.

[4] Instituto Argentino de Matemática "Alberto Calderón" – CONICET, Av. Paseo Colón 850, C. A. de Buenos Aires, Argentina.

[5] Dto. de Matemática, Instituto Tecnológico de Buenos Aires, CONICET, Av. Madero 399, C. A. de Buenos Aires, Argentina.

Abstract. We present a new method for obtaining the positions of each person attending an outdoor gathering. From this information it is possible to calculate the density field and, integrating it over an area of interest, the number of people in this area can be quantified. A dual-camera (visible + infrared (IR)) is mounted on an unmanned aerial vehicle (UAV). In this work, we only use the IR channel and present an initial set-up and calibration of the system along with the characterization of a small group of people.

Keywords: people counting, crowd counting, infrared image processing, unmanned aerial vehicles applications

1 Introduction

Counting and monitoring outdoor gatherings, such as street demonstrations, music festivals, parades, religious processions, etc. could be a very important issue for governments, media and, of course, safety managers.

There exist several methods of image processing using visible information from cameras in arbitrary angles [1–3]. In particular, deep neural network approaches are very fulfilling on this task [4–6].

Our current project consists of adding also the IR information to the analysis which could disambiguate visible information and *vice versa*. However, in this first study, we have only considered the IR channel, which is less sensitive to ambient lighting conditions.

In short, our method is based on the following pillars: - Taking images of the crowd using a unmanned aerial vehicle (UAV, also known as a multirotor drone). It is a key issue that the UAV must be equipped with redundant fail-safe mechanisms to avoid it to fall over the people. - The images comprise both visible and infrared (IR) information from the surface. The IR component allows to discriminate people from non-animated objects and highly improves the correct identification of individuals. The camera used is a FLIR DUO PRO R, 640, 19 mm, 9Hz. - Image processing to obtain the (x,y) positions of each pedestrian. - Post-processing. Once the positions of pedestrians are determined, we can calculate the density field. Integrating this field in the area of interest we can count the number of people in this area.

In this work, we present the initial set-up and calibration of the system composed of the UAV, cameras and other sensors. And the first characterization of a small group of people using only the IR information.

In future works, more complex gatherings will be presented. The possible applications of the general method are, first, quantifying with small error the number of people attending a certain event. Second, measure and study the evolution of the density field. And third, consider the dynamics of the crowd "on-line" which can be used for preventing safety issues.

2 UAV description

The aerial vehicle was designed and assembled in our laboratory with the following components.

The frame is the Turnigy H.A.L. hexarotor, a strong, lightweight aluminum frame designed for Heavy Aerial Lift, with a distance between rotors of 775 mm.

The actuators installed on this frame are Multistar 3508-268 KV brushless motors, with 14x5.5" carbon fiber propellers, driven by 40 A, optocoupled Hobbywing electronic speed controllers (ESC).

To power the entire system, a 6S 5000 mAh 20C LiPo battery is used.

The flight controller mounted on the vehicle is the PX4 Pixhawk model, that uses several internal inertial sensors, an N8M GPS plus magnetometer module, and a wireless communication link with a ground station through a 433 MHz, 100 mW telemetry system.

The total weight of the vehicle is 3 Kg, and an additional 500 g for the camera with a mounting case and a 3S 2000 mAh battery to power it, allowing for flights of around 10 minutes.

A picture of the UAV carrying the dual camera is shown in Fig. 1. The final UAV will be provided with redundant fault tolerance mechanisms, one of them was developed by Giribet et al. [7, 8].

3 Infra-Red Image Processing

Pixels vs. cm Calibration A first calibration procedure for the IR image consists in obtaining the function that converts pixels into cm depending on the UAV height. A rectangular 50 cm pattern was recorded from different heights and the ratio between cm and number of pixels in the image could be calculated and it is shown in Fig. 2. The heights were measured with the UAV flight controller barometer.

Fig. 1: The unmanned aerial vehicle with the dual camera (visible and infrared).

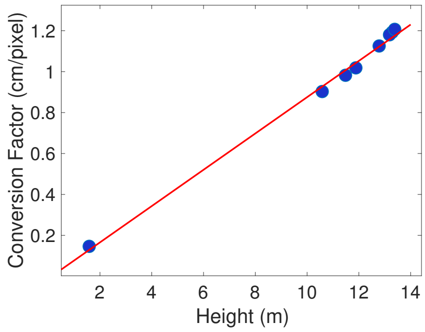

Fig. 2: Linear relation between the factor converting pixels to cm and the height at which the image was obtained.

Filtering Steps We present here a six-step filtering process that allows obtaining the pedestrian positions in the (x,y) plane from a zenith view. This process only considers the IR channel information. The visible information could have greater variations due to external conditions and it is only considered as reference and illustration of the analyzed scenarios. The six steps are the followings:

Step 1: Raw Infrared Image The IR camera provides images having temperature information that can be calculated from the intensity of each pixel. The first processing is to correct the radial distortion of the lens and generate a transformed plane image with temperature information.

Step 2: Entropy Filter Considering a square neighborhood around each pixel j, the entropy $S_j = -\Sigma p_i log(p_i)$ was calculated. Where p_i is the probability of a given intensity interval in each pixel i inside the neighborhood.

Step 3:Entropy and Temperature threshold Only the pixels with values of high entropy (S_j) and human-like temperature (T_j) are kept, i.e., the ones satisfying the condition $S_j > 1 \ \cap \ T_j \in [27°\mathbf{C}, 38°\mathbf{C}]$. The rest of them are set to zero.

Step 4: Smoothing The surviving entropy pixels (S_j) are smoothed with a Gaussian kernel with a standard deviation of $\sigma = 8$, which maximizes the correct identification of people in the current method.

Step 5: Find peaks Then, the maximum values in a square neighborhood of 40 pixels are selected. This number of pixels corresponds approximately to an area occupied by one person in the image.

Step 6: Final Positions The coordinates of peaks found are approximations of the positions of people in the plane.

4 Results under different lighting conditions

As a case study for the proposed method, we have analyzed a small crowd at an outdoor location (Buenos Aires, summer 2019) at different times. Images were taken at different lighting conditions changing from natural (19 hs.) to artificial illumination (20 hs.).

In Fig. 3 the above procedure of six steps (columns) are shown for each of the three lighting conditions (rows). The first column is the visible image displayed only as an illustration, but it is not considered for the six steps analysis.

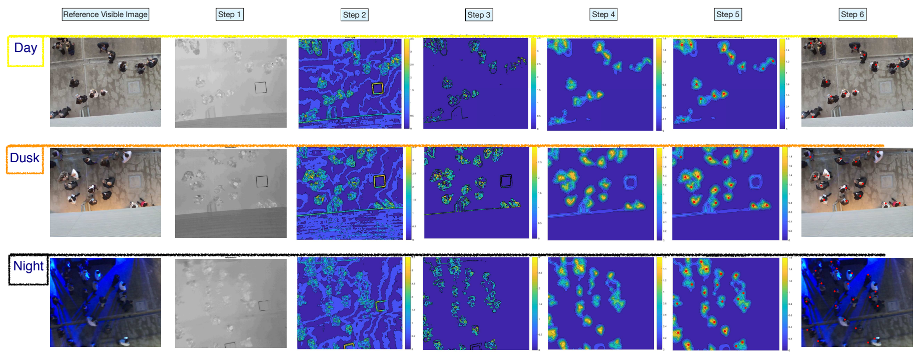

Fig. 3: Six steps procedure described in Sec. 3 for three different lighting conditions from natural to artificial. The first column is the corresponding visible information shown only as reference. The red dots on image from step 5 and 6 are pedestrian positions.

As can be seen, the "day" light is the natural light; "dusk" indicates weaker sunlight after the sunset mixed with few artificial life; "night" scene is completely illuminated by artificial and colored light. Precisely, this abrupt color changing of artificial lights that can be present in some gathering as a music festival, inspire us to include the IR information. The temperature information also suffers from ambient contamination, but it changes smoothly over time. In outdoor locations, if the ambient temperature is high, there is a moment of the day in which the floor could have a similar temperature that the human body and at that time our analysis is not valid. But for all the rest of the day, in particular, during the night, the proposed method should work properly.

It is important to note that the same method, with the same parameters, performs well for the three different lighting conditions studied.

From the (x,y) positions obtained, the density field can be calculated. There are many ways of doing this, we choose the coarse-graining method in which the pedestrian mass is considered to be spread over the space, we take a Gaussian with its mean value matching the position of the pedestrian and spanning over the whole space, in our case, with a standard deviation $\sigma_g = 0.46$ m. For details on the method see Ref. [9–13]. Figure 4 displays the corresponding density fields. One property of the density fields is that integrating it over the whole area provides the number of pedestrians.

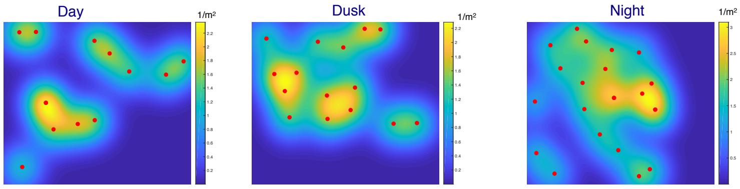

Fig. 4: Density fields for the three different lighting conditions indicated.

5 Conclusions and next steps

We are developing a solution for people counting and crowd characterizations, based on an fault-tolerant UAV equipped with a visible and an infrared camera.

As a first step, we analyzed a small crowd at an outdoor location at different illumination conditions from natural to artificial light. This first analysis only considers de IR channel, which is successively filtered until obtaining the people's positions in the plane. This approach is robust under different lighting conditions, such as shadows and artificial colored light, representing an improvement over processing visible information.

Future works will consider combined visible + IR images processing for increased accuracy in the detection of still pedestrians. Also, higher densities and bigger crowds will be studied.

From the people's positions, the density field can be calculated. The convolutional neural network methods mentioned in the introduction, provide as output the density field, taking as input visible images. As a goal for future work, it could be an enhancement for this deep neural network methodology to incorporate infrared information as a four-channel of an extended image (RGB+IR).

Acknowledgements

The authors acknowledge financial support via project PID2015-003 (Agencia Nacional de Promoción Científica y Tecnológica, Argentina; Instituto Tecnológico de Buenos Aires; Urbix Technologies S.A.) and from ITBACyT-2018-42 (Instituto Tecnológico de Buenos Aires).

References

1. Subburaman, V.B., Descamps, A., Carincotte, C.: Counting people in the crowd using a generic head detector. In: 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, pp. 470–475. IEEE (2012)
2. Bansal, A., Venkatesh, K.: People counting in high density crowds from still images. arXiv preprint arXiv:1507.08445 (2015)
3. Rodriguez, M., Laptev, I., Sivic, J., Audibert, J.Y.: Density-aware person detection and tracking in crowds. In: 2011 International Conference on Computer Vision, pp. 2423–2430. IEEE (2011)
4. Boominathan, L., Kruthiventi, S.S., Babu, R.V.: Crowdnet: A deep convolutional network for dense crowd counting. In: Proceedings of the 24th ACM international conference on Multimedia, pp. 640–644. ACM (2016)
5. Wang, C., Zhang, H., Yang, L., Liu, S., Cao, X.: Deep people counting in extremely dense crowds. In: Proceedings of the 23rd ACM international conference on Multimedia, pp. 1299–1302. ACM (2015)
6. Zhang, Y., Zhou, D., Chen, S., Gao, S., Ma, Y.: Single-image crowd counting via multi-column convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 589–597 (2016)
7. Giribet, J.I., Sanchez-Pena, R.S., Ghersin, A.S.: Analysis and design of a tilted rotor hexacopter for fault tolerance. IEEE Transactions on aerospace and electronic systems **52**(4), 1555–1567 (2016)
8. Giribet, J.I., Pose, C.D., Ghersin, A.S., Mas, I.: Experimental validation of a fault tolerant hexacopter with tilted rotors. International Journal of Electrical and Electronic Engineering & Telecommunications **7**(2), 58–65 (2018). DOI 10.18178/ijeetc.7.2.58-65
9. Helbing, D., Johansson, A., Al-Abideen, H.Z.: Dynamics of crowd disasters: An empirical study. Physical review E **75**(4), 046,109 (2007)
10. Goldhirsch, I.: Stress, stress asymmetry and couple stress: from discrete particles to continuous fields. Granular Matter **12**(3), 239–252 (2010)
11. Zhang, J., Behringer, R.P., Goldhirsch, I.: Coarse-graining of a physical granular system. Progress of Theoretical Physics Supplement **184**, 16–30 (2010)
12. Weinhart, T., Labra, C., Luding, S., Ooi, J.Y.: Influence of coarse-graining parameters on the analysis of dem simulations of silo flow. Powder technology **293**, 138–148 (2016)
13. Garcimartín, A., Pastor, J.M., Martín-Gómez, C., Parisi, D., Zuriguel, I.: Pedestrian collective motion in competitive room evacuation. Scientific reports **7**(1), 1–9 (2017)