HAL
open science

Non supervised perceptual model for target recognition in UAVs

Eric Bazan, Petr Dokládal, Eva Dokladalova

To cite this version:

Eric Bazan, Petr Dokládal, Eva Dokladalova. Non supervised perceptual model for target recognition in UAVs. Reconnaissance des Formes, Image, Apprentissage et Perception RFIAP, Jun 2018, Marne la Vallée, France. hal-01790867

HAL Id: hal-01790867

https://enpc.hal.science/hal-01790867

Submitted on 14 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non supervised perceptual model for target recognition in UAVs

Eric Bazan[1] Petr Dokladal[2] Eva Dokladalova[3]

[1,2] *PSL Research University - MINES ParisTech*
[3] *Universite Paris-Est, LIGM, ESIEE Paris*

[1,2]{name.surname}@mines-paristech.fr, [3]{name.surname}@esiee.fr

Abstract

Today, drones play an interesting role in the so-called Revolution 4.0. One of the problems studied by various companies and research groups are the precision landing techniques since this drone feature can be used in applications such as package delivery or object tracking. In this paper, we propose a non-supervised model that allows to detect and recognize a set of landing targets using the Gestalt principles. This proposed method is capable to recognize different coded landing targets in a robust way under outdoor non-controlled light conditions. Comparing to thresholding techniques and other methods, this work deals with image degradations caused by shadows, change of scale, noise and camera target deformation.

1 Introduction

The use of drones to carry out tasks of a civil nature is increasing. Multirotors have attracted the attention of companies and research groups for their cargo transportation ability. However, despite its cost-effectiveness advantage, the package delivery drone task still has a long way to go since it involves many disciplines. In terms of logistics, the drone accurate landing is a well-used strategy since it prevents the drone lands in a wrong place, assuring the integrity of the package and of the drone. Imprecisely landing is usually caused by low accuracy or loss of GPS signal. To avoid the aforementioned, it is common to complement the GPS information by means of an onboard camera. The addition of a camera allows acquiring drone environment information for the precision landing. However, the scenarios conditions where the applications are performed are, in most cases, non-controlled inducing image degradations. Such degradations are generated by changes in lighting, shadows, noise or by the camera manipulation (distance, rotation, abrupt movements, the angle of view, etc.). In that way, an image processing technique that takes into account the mentioned above must be used.

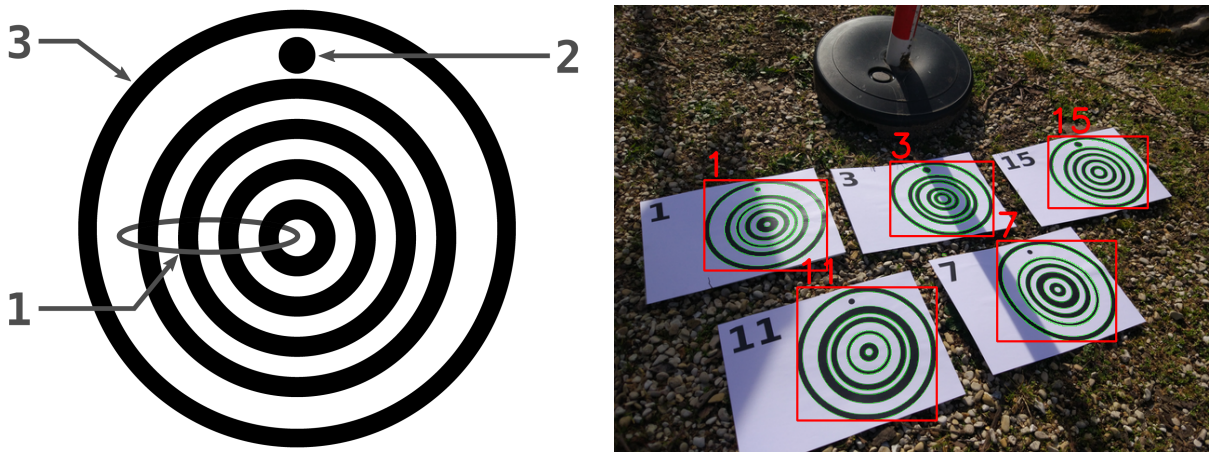
The thresholding methods [**?**, **?**] are one of the first techniques developed for images analysis and they are widely used until today. The objective of these methods is to classify the image pixels as background or foreground according to some property (brightness, color, etc.), however, they are not efficient when light conditions or noise affect the foreground, leading to a bad segmentation. On the other hand, the variational framework [**?**], which deals in a better way the non-controlled light conditions, proposes to minimize a function with a fidelity and a regularization term with which the optimization of the image segmentation is sought, notwithstanding, this method does not solve the image degradations and besides, it cannot be implemented in real time given the number of iterations needed to find the optimal solution. In the case of techniques that use an explicit model, e.g. the stochastic model of Bayes [**?**], it is necessary to have a prior model that characterizes the behavior of real images, and in cases where this model is considered as the null hypothesis [**?**], the number of tests to calculate the number of false alarms per image hinders its use in real time. With the increase of computing power and the amount of data that is generated, machine learning techniques offered the possibility of creating algorithms capable of segmenting, almost exactly, a series of desired objects or classes [**?**]. Some of the problems faced by these methods are that the algorithm must have a learning stage in which a statistical classification model is used to "learn" from a database of images containing the classes. To reach a good segmentation, the collection of images must represent the classes in a wide range of situations. In the precision landing problem with multiple landing points, the training database is quite big since it must contain all possible combinations between the image degradations and the different coded landing targets. For the aforementioned, this work proposes a novel non-supervised model capable to detect and recognize a set of landing targets, in a perceptual manner, that is, taking into account its most representative characteristics.

2 Problem Statement

It is considered a multi-copter type drone that reaches a delivery area thanks to a built-in GPS, then, to perform a precision landing, the drone uses the information given by a perpendicular oriented camera to identify and decode a landing target located at the delivery point. Given the nature of the application, the acquired images are subject to different degradations

e.g. **a)** image noise, due to the sensor quality and the vibrations transmitted by the drone; **b)** presence of occlusion shadows; **c)** target deformation, due to the angle and distance between the camera and the target and **d)** resolution target change, due to the variation of flight height. To achieve a multi-point precision landing, a set of targets is created.

Target description. The proposed landing target (see Fig. 1a) is formed by three main parts: **1.** Number of landing target, coded in a series of B&W rings; **2.** Orientation circle and **3.** Main body, formed by a black ring that encloses the two previous elements. To encode the landing targets, the error-correction Hamming code [**?**] is used to generate a set of binary coded words. Each internal and external ring diameters in part 1 are considered as a message bit. Depending on the bit value (0 or 1), the diameter could increase or decrease creating a unique rings thickness combination for a target. With the current ring configuration in part 1 (4 rings), it is possible to generate a set of up to 16 landing targets.



(a) (b)

Figure 1: (a) The 3 principal landing target parts. (b) Target detection and recognition result of a sample of 5 landing targets of 16 under non-controlled scene conditions. The number of the target is shown on the upper left of the printed sheet; the red rectangle indicates the target's detection and the red number indicates the number of the recognized target.

3 Non-supervised Perceptual Model

Scale-space and outlier edges detector. Despite their real-time capability, the thresholding methods are not convenient to carry out the target detection since they are not robust to the mentioned image degradations. Instead, the proposed strategy obtains the image contours by means of Canny's algorithm [**?**] using as input the positive part of the image Laplacian in three different scales through the filter proposed in [**?**]. The importance of using the scale-space filter lies in the interest of recover all the target contours no matter the size of the target in the image. Using the filter in three different scales increase the number of contours, especially if the image contains a high level of noise or has variations in light intensity. To recover the landing target edges the Gestalt theory [**?**] is used. The visual perception laws state that grouping is the main process in human perception; if some entities share a characteristic in common they get grouped forming a new entity, which correctly represents the scenario studied here. The edges that belong a landing target have grouping characteristics, for example, some of them are concentric, they all are circular and also have a high image gradient magnitude what allows to differentiate them from those edges generated by noise, shadows or other objects in the image scene. Therefore, employing the circularity, the mean gradient intensity and the center of each edge it is possible to create a multi-feature space where the landing target contours behave as outliers. The anomaly detector proposed in [**?**] allows computing the existing correlation between the edges taking the critical value thrown by the *chi-squared* distribution for a 99.9% of confidence as a threshold value.

Affinity clustering. With the previous strategy, the resulting edges that have an important value of mean gradient and are circular are recovered, however, all those contours that are not part of a target but are circular, they also remain. Given the form and the description of the landing target, we know that the rings are approximately concentric (at least part 1 and 3 in Fig. 1a) and also contain a fixed number of circumferences. Moreover, due to the shape of the landing target, it is possible to compute Jaccard index between each contour and an ellipse. Therefore, we use the centroid of each contour and its Jaccard index as grouping characteristics (proximity and similarity respectively) to perform an affinity clustering method [**?**].

Target decoding. The most related contours are grouped in clusters. For the drone to recognize a landing point it is necessary to decode the clusters. This stage uses the diameters (or major axes of the ellipse if the target is deformed) of the circles as binary values applying the reverse procedure to the Hamming coding [**?**]. The use of error-correction Hamming

code allows to identify up to 2 errors in the message and correct one erroneous bit of the message frame. Normally, this feature is well used when the drone is at high altitudes and the size of the landing target is not large enough to recognize the different ring combinations and the diameter size change.

4 Experiments and Results

To evaluate the results of the proposed method, it is compared to 10 different thresholding techniques [**?**] using a collection of synthetic images that simulate the four image disturbances mentioned in section 2. The F1-score, which calculates the weighted average between precision and recall, is used as an evaluation metric. The results obtained show that some thresholding techniques (e.g. clustering-based thresholds) work fine under noise degradations while others (e.g. local adaptive thresholds) work under the presence of shadows but none of them is robust to all the degradations. In a real scenario the disturbances are not isolated, that is, in a single image the landing target may be affected by noise, shadows, deformations and scale changes as is shown in figure 1b. Some experiments that were carried out in real conditions can be seen in `https://youtu.be/igsQc7VEF2c`.

5 Conclusions

Thresholding and energy-based techniques are not suitable in the target detection task because the presence foreign objects and non-controlled light conditions affect the result. On the other hand, machine learning techniques need to be trained taking into account all (if possible) image degradations and the number of the landing targets in the set to be successful. In this work, the landing target detection and recognition is achieved by way of a non-supervised perceptual model. The landing target model is built using the Gestalt's visual perception principles under the idea that the contours characteristics behave as outliers in the multi-feature distribution. The advantages of this method lie in the versatility of characteristics that can be used to carry out the "groupings" without the need to establish regulation parameters or to have a training phase. In addition, the proposed method is robust to changes in lighting, the presence of shadows and noise in the image and has the auto-correction capacity.

References

[1] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[2] A. Desolneux, L. Moisan, and J.-M. Morel, *From Gestalt Theory to Image Analysis: A Probabilistic Approach*, ser. Interdisciplinary Applied Mathematics. New York: Springer-Verlag, 2008.

[3] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science (New York, N.Y.)*, vol. 315, no. 5814, pp. 972–976, Feb. 2017.

[4] R. W. Hamming, "Error detecting and error correcting codes," *The Bell System Technical Journal*, vol. 29, no. 2, pp. 147–160, Apr. 1950.

[5] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, Jul. 1989.

[6] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[7] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.

[8] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proceedings Ninth IEEE International Conference on Computer Vision*, Oct. 2003, pp. 10–17 vol.1.

[9] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation." *J. Electronic Imaging*, vol. 13, no. 1, pp. 146–168, Jul. 2010.

[10] M. Wertheimer, "FormsUntersuchungen zur Lehre von der Gestalt II," *Psycologische Forschung*, vol. 4, pp. 301–350, 1923.

[11] A. Witkin, "Scale-space filtering: A new approach to multi-scale description," in *ICASSP '84. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 9, Mar. 1984, pp. 150–153.

[12] T. Zhao and R. Nevatia, "Bayesian human segmentation in crowded situations," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, Jun. 2003, pp. II–459–66 vol.2.