Low cost UAV and image classification for monitoring the deterioration on building façades

Andrea Masiero[1], Francesca Fissore[1], Antonio Vettore[1]

[1] Interdepartmental Research Center of Geomatics (CIRGEO), University of Padua, Viale dell'Università 16, Legnaro (PD) 35020, Italy, (masiero@dei.unipd.it, francesca.fissore@unipd.it, antonio.vettore@unipd.it)

Key words: *monitoring*; *UAV; deep learning; classification*; *deterioration detection; image processing*

ABSTRACT

Due to aging, weathering, infiltration, solar radiation and other factors, building and human infrastructures are subject to deterioration and, consequently, they periodically need to be monitored in order to determine whether they need any maintenance or restoration action. Traditionally, deterioration monitoring is done by means of on-site inspection performed by a human specialized operator. Despite being a reliable monitoring approach for small buildings, it becomes a not so affordable monitoring method for large buildings and infrastructures, where a human operator needs the use of external facilities for a careful visual inspection. The high costs related to such kind of inspection motivate the search for alternative monitoring methods. In particular, the availability of low cost drones, embedded with high resolution cameras, represent a viable way for a visual inspection of areas otherwise difficult to reach by human operators. Actually, this paper aims at investigating the use of low cost drones combined with artificial intelligence recognition methods, which recently proved to reach state of art classification performance in many applications. Such approach is applied for the detection of damaged bricks on the façade of a university building, reaching a good recognition performance.

I. Introduction

Deterioration on buildings, due to several factors such as aging, infiltration and humidity, is typically monitored by means of on-site inspection done by human operators (Bauer, Kraus, Silva, & Zanoni, 2014; Giacomucci et al., 2011; Hällström et al., 2009). As long as such kind of inspection is carefully done by expert operators it usually ensures good results on easily reachable/visible areas. However, for large and high buildings this approach has clear flaws on the reliability of the results, on the difficulty of obtaining them and on the time efficiency of the approach itself. For instance, areas on the upper part of high buildings are often difficult to be directly and carefully observed by a human operator: a reliable quantification of the deterioration can require evaluating the building surface from a very close distance, which usually leads to long procedures in order to enable the operator to be in this kind of working conditions.

This work aims at proposing a different approach which exploits recent development on drones, geomatics and computer vision.

On the one hand, during the last decade the usage of Unmanned Aerial Vehicles increased at an incredible rate, involving a wide range of civil applications. The increasing number of applications implies a significant increase of UAV market size: the global commercial UAV market size is expected to keep on growing at an annual rate of approximately 17% in the next years, making it one of the most rapidly increasing market sectors. The flexibility of usage of UAVs and their worldwide spread are important factors easing the development of low cost solutions based on their use.

On the other hand, the recent developments on machine learning techniques, and in particular on deep learning and convolutional neural networks (Lecun, Bengio, & Hinton, 2015), allowed to improve the level of performance of previously proposed classifiers. Their stunning performance on several recognition applications and their relatively ease of usage make deep learning techniques the state of the art of the current recognition and classification methods.

Taking into account of the above observations, this work proposes the usage of UAVs for monitoring the deterioration level of bricks on a University building. First, images are acquired by a low cost UAV ($ 350) flying close to the building façade. Then, the acquired images are rectified in order to emulate a camera view perpendicular to the building façade. Finally, deep learning-based recognition is used in order to detect deteriorated bricks on the building façade.

The rest of the paper is organized as follows: first, the specific case study considered in this work is presented in Section II. Then, the proposed data processing approach and the obtained results are described in Section III. A discussion on the system performance and a more general point of view taking into account of other remote sensing possibilities is reported in Section IV. Finally, some conclusions are drawn in Section V.

Figure 1. North-East façade of the considered building.

II. CASE STUDY

The case study considered in this paper is the detection of damaged bricks on a façade of the building of the University of Padua (Italy) shown in Fig. 1.

Such work has been motivated by the periodical need of maintenance for the façades of such building, which are subject to significant deterioration, in particular due to infiltration and weathering.

Fig. 2 shows a portion of the building façade, where several damaged bricks can be seen.

Figure 2. Portion of a façade of the building shown in Figure 1.

In this work, the inspection has been carried out by means of a Parrot Bebop 2 quadcopter. Such drone is a very low cost ($ 350, approximately) unmanned aerial vehicle (UAV), which enables low cost acquisition of aerial images at quite high resolution (14 Mpix).

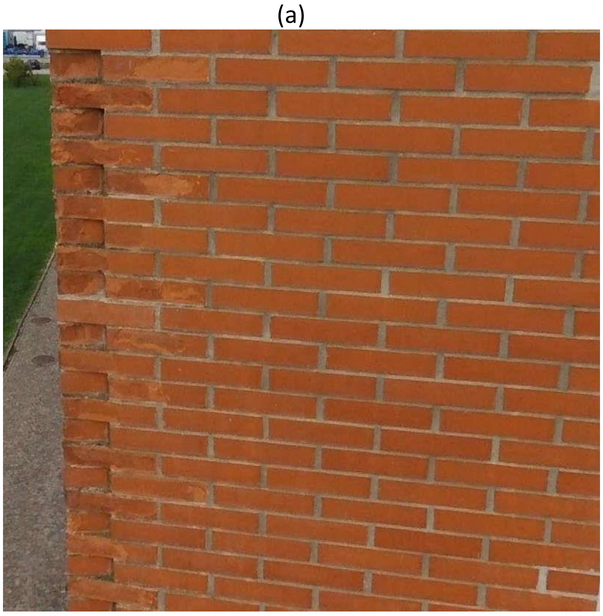
Despite such drone can be used as an unmanned vehicle, actually, in this work the quadcopter was expected to fly very close to the building. Given the risk of crash, the possibility of human presence close to the considered area, and the quite unusual planned flight, in this work the drone has been remotely piloted by a human operator.

The drone flew at 1-2 m from the building façade during all the data acquisition. Camera and drone orientations were set in order to have an approximate frontal view of the façade.

Motivated by the need of properly surveying all the area of interest and given the difficulty of ensuring a reasonable overlapping between successive single images, visual data acquisition was performed in video mode, by exploiting the 1080p@30 Hz full HD video acquisition mode of the Bebop 2 camera.

Despite the data acquisition has been done in video mode, the quality of images acquired by the drone is relatively good, in particular in slowly moving (quasi-still) conditions, as shown in Fig. 3.

(a)



(b)

Figure 3. (a) Example of Parrot Bebop 2 view at approximately 3-4 m far from the façade. (b) Portion of the façade extracted by a Parrot Bebop 2 camera view at 1-2 m distance from the building.

Despite the Bebop 2 camera is relatively good for a low cost drone, it is quite clear the quality difference between the image detail shown in Fig. 2 (acquired at ground level by a professional digital camera) and the portion of the Bebop 2 image shown in Fig. 3. Nevertheless, the results shown in the following Section experimentally prove that the considered approach allows to obtain reasonable results even with such low cost drone and camera.

III. Image processing and damage detection

Frames acquired by the Parrot Bebop 2 in video mode: such frames are 2 Mpix images that correspond to the central part of the wide-angle camera view.

Due to the presence of some wind the drone slightly changed its orientation during the data acquisition. Since in the proposed procedure it is assumed to feed a neural network with frontal views of the bricks to be analyzed, the video frames are corrected in order to simulate the view from a perfectly orthogonal direction to the planar wall surface. To this aim, the interior camera parameters have been pre-calibrated (actually, such calibration has been done by processing the video frames).

A. *Image Processing*

Each video frame is processed in order to:

1) Correct the perspective point of view
2) Extract the area corresponding to each brick.

First step for obtaining an approximate perspective correction is the estimation of the camera reference frame orientation with respect to the wall plane. To such aim, lines corresponding to the brick borders are extracted.

Edge detection is implemented by using a Sobel filter and then extracting the most probable lines by means of the Hough transform (Illingworth & Kittler, 1988; Kanopoulos, Vasanthavada, & Baker, 1988).

Figure 4 shows an example of Sobel filter result on a portion of a drone video frame.

Hough transform requires a parametric representation of the lines to be detected. A standard line representation has been considered:

$$\rho = u \cos\theta + v \sin\theta \quad (1)$$

where u, v = horizontal and vertical image coordinates (in pixels)
ρ = distance between the origin of the reference system and the line
θ = angle of the shortest line connecting the origin with the considered line (which hence has an angle of $\theta + 90°$

A discrete parametrization for θ and ρ in (1) has to be considered in order to apply the Hough transform. Angles have been sampled at a 1 degree resolution, whereas ρ has been sampled at a 1 pixel resolution.

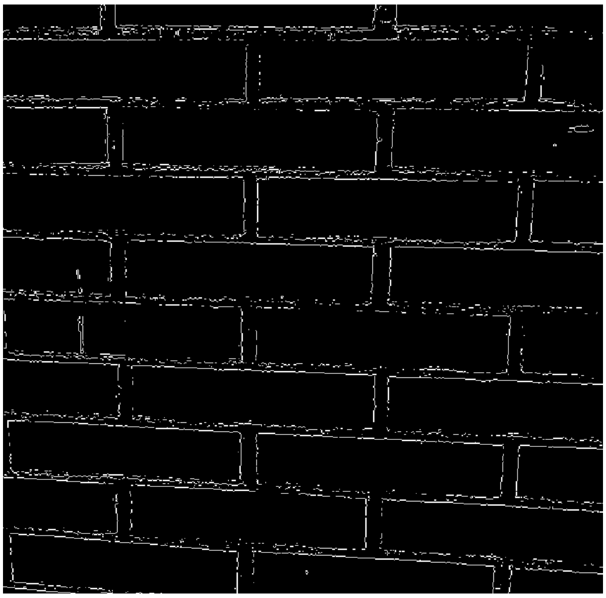


Figure 4. Result of the Sobel filter on a portion of video frame acquired by the drone.

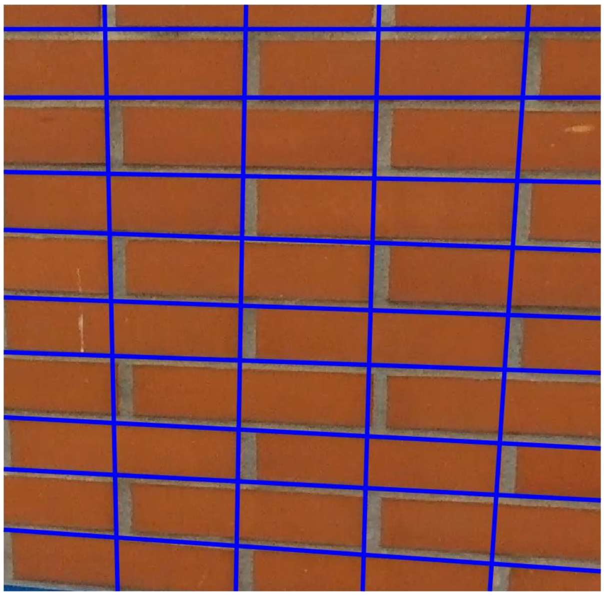


Figure 5. Lines detected with the Hough transform on the area shown in Figure 4.

The lines detected by means of the Hough transform in the image portion of Figure 4 are shown in Figure 5.

Once lines have been detected camera-wall relative orientation is estimated and corrected in order to have an approximately orthogonal camera view of the wall with the horizontal axis of the camera aligned with the horizontal brick borders: in such conditions the brick borders are approximately aligned with both the vertical and horizontal image axes (check an example of the obtained results in Figure 6). Such transformation is similar to the commonly used image rectification, i.e.

undistorted image coordinates are remapped as follows:

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} \propto KR^T K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{2}$$

where K = intrinsic camera matrix
R = rotation matrix describing the camera-wall relative orientation

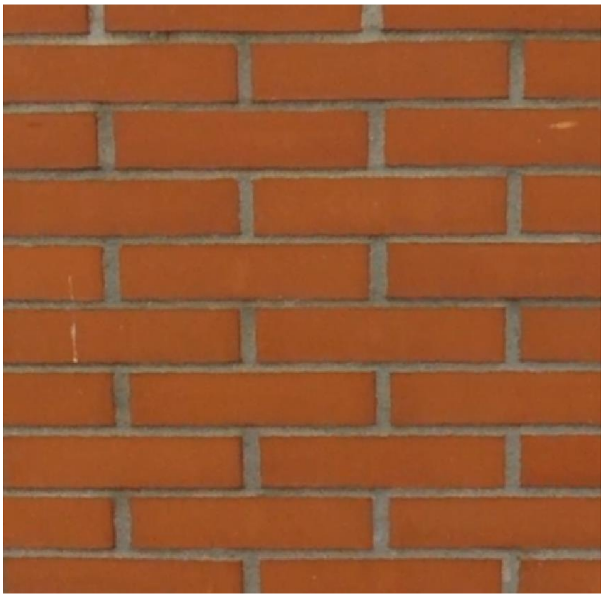


Figure 6. Remapped image.

Then, brick segmentation is performed by the applying the following steps:

- Thresholding the image (e.g. by using Otsu's method)
- Finding the list of the connected regions in the thresholded image.
- For each connected region with a sufficient area (i.e. in order to consider only bricks which are entirely, or almost entirely, visible in such image) determining the smallest polygon (e.g. rectangle) containing such connected region.

The result of such procedure applied to Figure 6 is shown in Figure 7.

B. *Damage Detection*

Once bricks have been properly segmented in the video frames, they are resized and used as input to feed a neural network classifier.

In this work the very commonly used AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) has been modified and re-trained in order to properly distinguish between damaged bricks and those in good conditions.

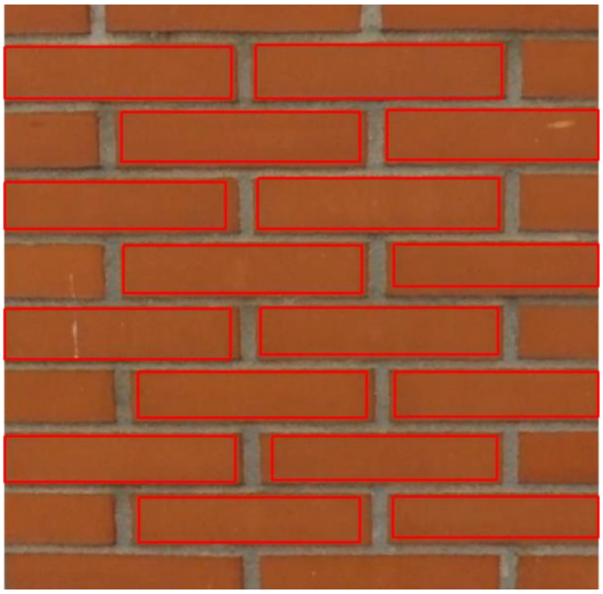


Figure 7. Bricks segmented in Figure 6.

To this aim a set of 92 bricks have been properly manually labeled. In order to ensure a satisfying performance with both such categories, half of the learning samples were damaged bricks, and viceversa. Furthermore, since 92 samples are clearly not enough to properly train the network, data augmentation was obtained as follows: the original images were rotated and symmetrically flipped in order to increase the cardinality of the training set, as often done when dealing with not so large dataset to be used in deep learning applications.

Hence, the network has been trained by using 736 samples (larger than the original dataset by a factor 8).

Then, other 192 bricks have been manually labeled and used for validating the previously trained network, leading to the results shown in Table 1. It is worth to notice that, in order to check the network performance as much as possible the same kind of data augmentation applied to the brick images in the learning phase has been applied to the validation dataset. Consequently, the cardinality of the validation dataset increased to 1536 bricks: 1168 in good conditions (i.e. 146 according to the original dataset) and 368 damaged ones (i.e. 46 in the original dataset).

Table 1. Brick classification results

Real class	Classification result	
Brick conditions	Good	Damaged
Good	1157 (99.06%)	11 (0.94%)
Damaged	8 (2.2%)	360 (97.8%)

C. *Combining Different Frames*

A standard vision based approach is used for the estimation of the relative pose between the camera in two successive video frames, i.e. feature matching, epipolar geometry estimation, and motion parameter extraction from the essential matrix.

It is worth to notice that in this case the aim of such step is rather that establishing the proper correspondences between bricks, in order to avoid to re-analyze the already examined ones, than to accurately estimate the camera motion.

IV. Discussion

A number of works have been published during the last years showing the effectiveness of neural network based classifiers (Deng, 2014; Masiero et al., 2019; Pierdicca et al., 2018; Simonyan & Zisserman, 2014). In particular the spread of pre-trained networks has tremendously increased both the range of applications and the effectiveness of neural network-based approaches.

This paper aimed at showing the potential of even a very widely used network, such as the AlexNet, once properly re-trained in order to work on a very specific damage detection classification problem: to such aim, a very low cost drone has been used to acquire video frames of the façade of a university building. The drone flew at quite close distance from the façade, with the camera optical axis orientated approximately orthogonal to the façade surface (Figure 8).

Figure 7. Parrot Bebop 2 drone acquiring a video close to a façade of the building of Figure 1.

The acquired video frames were processed in order to compensate the variations of drone orientation with respect to such orthogonal direction and to properly segment bricks in the acquired images.

Then, the last three layers of the AlexNet have been re-trained on a relatively small learning dataset (736 data samples, obtained by simple transformations of 92 brick images, half of which representing damaged bricks and half bricks in good conditions). It is worth to mention that, despite the availability of a much larger training dataset is clearly a desirable working condition to ensure more reliable training results (and, for instance, reduce the chance of overfitting issues), actually the possibility of reaching reasonable training results also with a relatively small training dataset is clearly very interesting in terms of practical applicability of the approach, in particular in the applications where data acquisition before the system use is quite hard.

The re-trained network showed very good performance on the validation dataset (192 bricks), leading to a correct classification in the 98.8% of the considered validation cases. False positive (bricks in good conditions classified as damaged) chance was 0.94%, whereas false negative (damaged bricks classified as in good conditions) probability was 2.2%.

Despite the proposed approach gave quite good results, it is clear that the integration with other sensor information might help in providing more information to the human operators in charge of evaluating the need for maintenance.

For instance, the introduction of 3D information, as for instance terrestrial laser scanning such as in (Błaszczak-Bąk, 2016), can be useful in order to determine the geometric level of the damages, or to check the potential position changes of bricks (possibly due to several factors, such as building structural instabilities, infiltrations, and so on). Furthermore, the use of multi-spectral/hyper-spectral/thermal cameras can be useful for instance in order to determine more information about infiltration level (Bauer, Pavón, Oliveira, & Pereira, 2016).

Consequently, an integrated approach, which considers information acquired by different sensors, shall be considered in our future research works. Such approach, which can use a neural network-based sensor integration method, can lead to the development of a more informative system, enabling the simultaneous acquisition/estimation of multiple information on the elements which form the building façades.

Furthermore, it is also worth to mention that the considered network is not necessarily the best one for the considered problem. Actually, this work shows the quite impressive adaptability of neural networks, however a more in depth investigation on the network optimization should be done in the future in order to improve further the obtained results.

V. Conclusions

A number of work recently showed the effectiveness of deep learning based approaches: in particular, the introduction of pre-trained neural network allowed to dramatically improve the performance of classifiers.

This paper presented the use of a deep learning-based method in order to detect damaged bricks on a building façade (the building of the University of Padua used as case study in this work is shown in Figure 1). The proposed approach, which exploited a very commonly used neural network properly re-trained with a relatively small training dataset, confirmed the high potential of such kind of approach in this application as well.

Actually, our future work will be dedicated on the integration of the standard camera information with that acquired by different sensors in order generalize and extend such approach in order, for instance, to improve the comprehension of the causes leading to the detected damages.

References

Bauer, E., Kraus, E., Silva, M. N. B., & Zanoni, V. A. G. (2014). Evaluation of damage of building facades in Brasília. In *XIII DBMC - International Conference on Durability of Materials and Components*.

Bauer, E., Pavón, E., Oliveira, E., & Pereira, C. H. F. (2016). Facades inspection with infrared thermography: cracks evaluation. *Journal of Building Pathology and Rehabilitation*. https://doi.org/10.1007/s41024-016-0002-9

Błaszczak-Bąk, W. (2016). New optimum dataset method in LiDAR processing. *Acta Geodynamica et Geomaterialia*, *13*(4), 381–388. https://doi.org/10.13168/AGG.2016.0020

Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*. https://doi.org/10.1017/atsip.2013.9

Giacomucci, L., Bertoncello, R., Salvadori, O., Martini, I., Favaro, M., Villa, F., … Cappitelli, F. (2011). Microbial Deterioration of Artistic Tiles from the Façade of the Grande Albergo Ausonia & Hungaria (Venice, Italy). *Microbial Ecology*. https://doi.org/10.1007/s00248-011-9812-0

Hällström, J., Barup, K., Grönlund, R., Johansson, A., Svanberg, S., Palombi, L., … Conti, C. (2009). Documentation of soiled and biodeteriorated facades: A case study on the Coliseum, Rome, using hyperspectral imaging fluorescence lidars. *Journal of Cultural Heritage*. https://doi.org/10.1016/j.culher.2008.04.008

Illingworth, J., & Kittler, J. (1988). A survey of the hough transform. *Computer Vision, Graphics and Image Processing*. https://doi.org/10.1016/S0734-189X(88)80033-1

Kanopoulos, N., Vasanthavada, N., & Baker, R. L. (1988). Design of an Image Edge Detection Filter Using the Sobel Operator. *IEEE Journal of Solid-State Circuits*. https://doi.org/10.1109/4.996

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). AlexNet. *NIPS*. https://doi.org/10.1145/3065386

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*. https://doi.org/10.1038/nature14539

Masiero, A., Chiabrando, F., Lingua, A. M., Vettore, A., Marino, B. G., Fissore, F., & Guarnieri, A. (2019). 3D MODELING OF GIRIFALCO FORTRESS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. https://doi.org/10.5194/isprs-archives-xlii-2-w9-473-2019

Pierdicca, R., Malinverni, E. S., Piccinini, F., Paolanti, M., Felicetti, A., & Zingaretti, P. (2018). Deep convolutional neural network for automatic detection of damaged photovoltaic cells. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. https://doi.org/10.5194/isprs-archives-XLII-2-893-2018

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.