On-line counting of pests in a greenhouse using computer vision

Ikhlef Bechar, Sabine Moisan

On-line counting of pests in a greenhouse using computer vision

Ikhlef Bechar and Sabine Moisan
EPI Pulsar, INRIA Sophia Antipolis-Mediterranee
ikhlef.bechar@sophia.inria.fr, sabine.moisan@sophia.inria.fr

Abstract

This article is concerned with the on-line counting of some harmful insects in videos in the framework of in situ video-surveillance that aims at the early detection of pest attacks in greenhouse crops. The challenges mainly concern the tiny resolution and color contrast of the insects of interest in the videos, the outdoor issues and the quasi-real time constraints. Thus, we propose a complete video-processing chain that can cope with such challenges quite satisfactorily while yielding an acceptable performance. The system has been validated off-line against many recorded videos for the whiteflie species (one potential harmful pest species).

1 Introduction

Computer vision has been successfully applied in various real-life surveillance applications [4] because of its various advantages among which one can mention: non-invasiveness, autonomy, accuracy and objectiveness. Recently, motivated by the growing interest for a sustainable Agriculture and a safer alimentation, its area of application has been extended to the automation of the pest monitoring process in greenhouse crops[1, 6]

One commonly used idea consists in endowing a greenhouse with a network of video-cameras that sense during day time some sticky traps (cf. Fig.1(a)) distributed, for instance, uniformly in the greenhouse or according to some priori knowledge about focus regions of harmful insects. Then a video processing makes it possible to recognize the trapped insects belonging to the harmful species of interest, and to describe statistically their spatiotemporal presence within a greenhouse. Such an information is used to predict a pest attack in order to allow the cultivator to use the fighting actions that fall at the early stage.

Despite the used big image size, the objects of interest (harmful insects) look in the videos only like tiny and lowly contrasted objects with unclearly defined borders (see Fig.1(a)-(b)), moreover because of hazardous *in situ* events (illumination variations, reflectances, outliers), this makes it very challenging to recognize on-line the insects of interest. Therefore, we propose to combine some image processing operations with some classical video processing algorithms in order to achieve a final operative system for the on-line counting of pests.

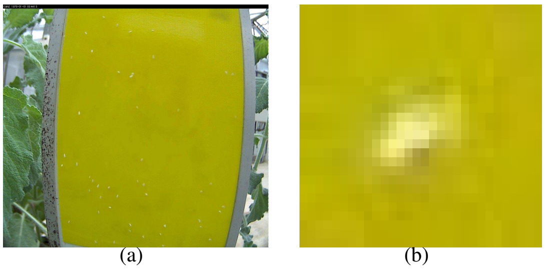
2 Frame-wise detection of harmful insects

Initially, the zone of the sticky trap with respect to each video is extracted automatically once and for all from the first video-frame by using some mere assumptions about similarity of color and compacity. Therefore, all subsequent image and video-processing operations are performed only in this zone of interest in a video (cf. Fig.2(a)).

In the reminder, we firstly describe the two-step algorithm that we developed for the recognition of the harmful insects of interest in individual video-frames, and secondly we show how it can be speeded up and made more robust using some standard video-processing algorithms in order to achieve an on-line operative video-surveillance application.

2.1 RGB-into-gray linear transformation

The first step consists in transforming each RGB video-frame into a gray image where the zones of insects of interest will look as much brighter as possible than their surrounding background. To do so, we consider a linear transformation of the form $I := t_r R + t_g G + t_b B$, and we estimate the linear coefficients t_r, t_g and t_b in such a way to maximize the following (SNR) ratio between the mean contrast over a sample of N_I insect intensities $\mathcal{S}_I = \{(R_i, G_i, B_i); i = 1, \cdots, N_I\}$ and the mean contrast over a sample of N_B background intensities $\mathcal{S}_B = \{(r_j, g_j, b_j); j = 1, \cdots, N_B\}$ as fol-



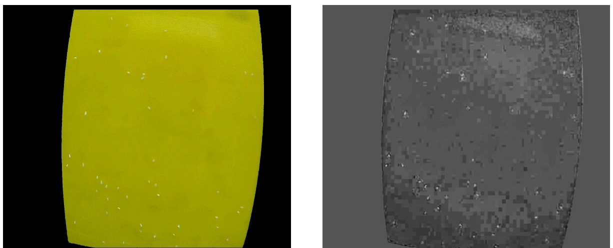
(a) (b)

Figure 1. (a) An example of a typical video-frame (resolution: 1280×960 px). The central yellow zone corresponds to the zone of the sticky trap, and the trapped harmful insects (whiteflies) correspond to the tiny white spots fixed on its surface ; (b) A zoom on the imagette of one insect of interest in the video-frame.

lows:

$$\frac{\sum_{i=1}^{N_I}(t_r R_i + t_g G_i + t_b B_i)^2}{\sum_{j=1}^{N_B}(t_r r_j + t_g g_j + t_b b_j)^2} \to \max_{t_r, t_g, t_b} \qquad (1)$$

Both color samples $\mathcal{S}_I$ and $\mathcal{S}_B$ were previously obtained by manual sampling of insect imagettes in sample videos (here, a dataset of 500 insect imagettes is used) in such a way that the linear coefficients t_r, t_g, t_b can be estimated off-line by solving the generalized eigen value problem which is deduced from (1).



(a) (b)

Figure 2. (a) Projection of an RGB video-frame on the recognized zone of the sticky trap ; (b) Its RGB-into-gray transformation.

2.2 Recognition of potential locations of insects of interest in a video-frame

Now, we would like to automatically extract the bright spots in a given gray video-frame (cf. Fig.2(b)) that may correspond to the insects of interest in the original video-frame. Such a problem of automatic extraction of small spots from unspecific backgrounds has been extensively studied in various application contexts such as fluorescence videomicroscopy [3], genetic spot array images [5], and feature extraction in active vision amounting to extracting interest points in video-frames[2]. In this paper, we propose a new parametric approach for the extraction of tiny spots in videos and we apply it to the problem of the recognition of harfmul insects of interest in videos.

To do so, we model a bright spot as a contrasted rectangular pattern $\mathcal{R} := \mathcal{R}(r, w, \theta, f(\cdot,\cdot))$, with r and w standing for its half-width and its half-length respectively, θ which stands for its tilt angle, and $f(x, y)$ which stands for a $2D$ function describing the gray intensity level at any point (x, y) of the plane. For simplicity's sake, we shall assume that $f(x, y)$ is a piecewise constant function which is equal to a constant $h+a$ inside $\mathcal{R}$, and to a constant a outside $\mathcal{R}$ as follows:

$$f(x,y) = \begin{cases} h+a, & \text{if } \quad |x\cos(\theta) + y\sin(\theta)| \le w \text{ and} \\ & \quad |-x\sin(\theta) + y\cos(\theta)| \le r; \\ a, & \quad \text{otherwise.} \end{cases}$$

where h stands for the gray contrast of $\mathcal{R}$ and a stands for the gray level of its surrounding background. Now, in order to yield a continuously differentiable $2D$ image which can show "singularities", namely local maxima at the rectangular zones of interest in the image and which can then be extracted efficiently by using a geometric differential technique, for instance by using the Karush-Kuhn-Tucker (KKT) local maximality criterion, we propose to convolve $f(x, y)$ with a gaussian kernel $K_\sigma(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x^2+y^2)}{2\sigma^2}}$ to obtain the following $2D$ scale space intensity profile

$$f_\sigma(x,y) = h \times \\ \Big(\Phi_\sigma(x\cos(\theta) + y\sin(\theta) + w) - \Phi_\sigma(x\cos(\theta) + y\sin(\theta) - w)\Big) \\ \Big(\Phi_\sigma(-x\sin(\theta) + y\cos(\theta) + r) - \Phi_\sigma(-x\sin(\theta) + y\cos(\theta) - r)\Big)$$

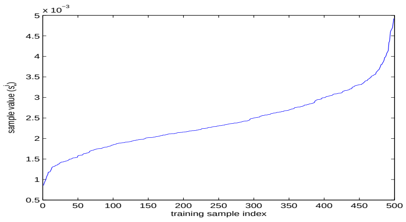
with $\Phi_\sigma(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{t} e^{\frac{-u^2}{2\sigma^2}} \mathrm{d}u$.

We show then that a condition on σ which is of the form: $\sigma \ge \frac{\max(w,r)}{\sqrt{3}}$ is sufficient in order for $f_\sigma(x, y)$ to show a clearly defined local maximum at the centroid of $\mathcal{R}$. Therefore σ is chosen as follows $\sigma := \frac{\ell_1}{\sqrt{3}}$, with ℓ_1 standing for the prior about maximum half-width or maximum half-length of an insect of interest in a video, and the latter can be obtained from a sample of manually segmented insect imagettes.

Now, one expresses the KKT sufficient conditions of local maximimality of $f_\sigma(u, v)$ at the centroid point $(0, 0)$ of $\mathcal{R}$ as follows:

$$\nabla f_\sigma(0,0) = 0 \qquad (2)$$

$$\nabla^2 f_\sigma(0,0) < 0 \qquad (3)$$



(a)

Figure 3. Sample values of $\min(s_1(\sigma,w,r,h), s_2(\sigma,w,r,h))$ computed for 500 insects imagettes and arranged in an ascending order.

with ∇ and ∇^2 standing respectively for the gradient and the Hessian operators with respect to x and y. For detection purposes, we shall focus more on the second KKT condition of local maximality (3) of $f_\sigma(u,v)$ at the centroid of some rectangular pattern which can also be seen as a measure of its saliency. Such a criterion amounts then to saying that both eigen values of $\nabla^2 f_\sigma(0,0)$ are negative. However, because of the presence of noise in the image, such a criterion should be replaced with a robust one consisting in saying that both eigen values of $\nabla^2 f_\sigma(0,0)$ should lie below a certain negative threshold '$-s_*$', which in plain language means that only salient enough rectangular patterns which may correspond to insects of interest in some video-frame should be considered. Note, in particular, the invariance of such criterion w.r.t. to θ. As a consequence, knowledge of the tilt angle θ of a rectangular pattern is not necessary. In practise, s_* may be estimated off-line from manually segmented sample imagettes of the insect of interest (cf. Fig.1(b)) as we shall explain it hereafter. First of all, one shows easily that the respective eigen values of $\nabla^2 f_\sigma(0,0)$ are given by:

$$s_1(\sigma,w,r,h) \quad := \quad \frac{-2hw}{\sqrt{2\pi\sigma^3}} e^{-\frac{w^2}{2\sigma^2}} \left(2\Phi_\sigma(r)-1\right)$$
$$s_2(\sigma,w,r,h) \quad := \quad \frac{-2hr}{\sqrt{2\pi\sigma^3}} e^{-\frac{r^2}{2\sigma^2}} \left(2\Phi_\sigma(w)-1\right)$$

Therefore, given a sample of N manually segmented imagettes (insect pixels *versus* background pixels), one can easily estimate the parameters r, w and h (we estimate h as the difference between the mean gray intensity over insect pixels and the mean gray intensity over background pixels in an imagette) to obtain, using the latter formulae, a set of candidate values for s_* as follows $\mathcal{S}_* = \{s_*^i, i = 1,\cdots,N\}$, with $s_*^i = \min(s_1^i(\sigma,w,r,h), s_2^i(\sigma,w,r,h))$. Hence, one might choose s_* as the minimum value of $\mathcal{S}_*$ (cf. Fig.3) .

Thus, the algorithm for the extraction of potential insect locations in a given video frame that we propose consists in two main steps. Firstly, extract the centroid locations by using the fact that for some pixel (x,y), the eigen values of matrix $\nabla^2 f_\sigma(x,y)$ lie below $-s_*$. Secondly, such locations are grouped by means of the connected components algorithm and a quick local conquer-and-merge segmentation strategy in the RGB frame allows to complete the detections (cf. Fig.4(c)).

(a) (b)

Figure 4. (a) A video-frame ; (b) Extraction of the insect locations in it.

3 An on-line video-processing solution

3.1 The basic online video-processing idea

In order to achieve an on-line video-surveillance solution, we exploit the fact that the objects of interest in a video (harmful insects) may appear in it only in a sparse way during day time. Thus, initially a video-frame is divided into a number of $k \times k$ slightly overlapping virtual image patches (e.g. $k = 10$) that can be processed very quickly by the insect detector described above. Then, a quick background subtraction algorithm in the spirit of the Mixture-of-Gaussians algorithm (MoG) [7] which runs permanently with respect to each video allows to detect any significant intensity changes in it. When a change is detected at some pixel, the pixel undergoes a second test referred as the insect presence detection test which will try to classify it as 'likely' or 'unlikely' to be an insect of interest pixel. This is achieved by learning off-line the space of color intensities of insects of interest by means of a Principal Components Analysis (PCA). The frame patch with maximum number of pixels that passed the insect presence detection test is then submitted to the insect detector described in section 2 in order to realize independently a precise detection of any recently trapped insect of interest, provided that this number exceeds some user-defined threshold. A detection is validated if and only if it intersects with a minimal number of pixels that passed the presence detection test. Since a recently trapped insect may manage to displace slowly from its initial location in the sticky trap (due to its continuous efforts to escape from it), so in order to keep track of it without having to detect it again in subsequent video-frames, hence a quick TBD (Track-Before-Detect) type tracking algorithm is used. Its principle is explained in the following subsection.

3.2 Frame-to-Frame insect tracking

Obviously, one needs firstly to specify which discriminative feature vector Y of an object of interest to use for the tracking task. However, since precise shape information about the insects is lacking in the recorded videos, hence one is left with no many options for choosing Y. Thus, we propose to choose the feature vector Y of any detected insect of interest as follows. Firstly, a square bounding box $\mathcal{B}_t$ is computed around the insect in the video-frame 't' where it has been firstly detected as the smallest square bounding box which contains it. A feature vector Y_0^t is then computed w.r.t. such a bounding box as follows. Firstly, take all RGB intensities lying inside $\mathcal{B}_t$, transform them into gray intensities by using the linear transformation as explained in subsection 2.1, then sort them in an ascending order and finally arrange them in a vector Y_0^t. One notes in particular the invariance of Y_0^t with respect to a rotation or a translation of the insect. In the next video-frames, such a bounding box is updated by sliding its centroid in the neighorhood of its current position (typically, in a small square window F of some predefined size $H \times H$ px.), and for each slid bounding box $\mathcal{B}_{t+1}$, a new feature vector Y^{t+1} is computed in the same way as explained for Y_0^t. $\mathcal{B}_t$ might then be updated by choosing for example which $\mathcal{B}_{t+1}$ maximizes a similarity criterion between Y^t and Y^{t+1} of the form $\mathcal{S}(Y^t, Y^{t+1}) := \min\left\{\frac{\|Y_0^t)\|}{\|Y^{t+1}\|}, \frac{\|Y^{t+1}\|}{\|Y_0^t\|}\right\} \frac{{Y^t}^T Y^{t+1}}{\|Y^t\|\|Y^{t+1}\|}$, where $\frac{{Y^t}^T Y^{t+1}}{\|Y^t\|\|Y^{t+1}\|}$ stands for the correlation ratio between vectors Y^t and Y^{t+1}, and $\min\left\{\frac{\|Y_0^t)\|}{\|Y^{t+1}\|}, \frac{\|Y^{t+1}\|}{\|Y_0^t\|}\right\}$ is a factor which favors vectors with comparable modules so as to fight more robustly against noise. However, such a tracking criterion $\mathcal{S}(Y^t, Y^{t+1})$ has shown in our tests to be somewhat sensitive to image noise, hence we propose to replace the maximum of the criterion $\mathcal{S}(Y^t, Y^{t+1})$ over F with its empirical mean.

4 Method's evaluation

The currently developed version of our vision application has been tested off-line against 8 video sequences representing the whitefly species and recorded under realistic *in situ* conditions during daytime for periods ranging from 20 minutes to 1 hour (the insects trap process was accelerated by placing the cameras near highly infested plants). An assessment of the results against ground truth revealed that the false positive rate is negligible (namely, one false positive has been found against about 250 found others), whereas the false negative rate is of order of 3%, and the latter concerns mainly some insects that were not detected by the algorithm because of their too low signal to noise ratio or because of a highly illuminated neighboring background. Finally, the system has shown to be able to process videos at a satisfactory video frame rate for the application (nearly 2 frames every 3 seconds).

5 Conclusion

We developed a new full on-line computer vision prototype for in-situ pest monitoring and we showed its feasibility for the case of one potential harmful pest species (the whitefly), nevertheless, its extension to other harmful pest species of interest (e.g. the greenfly species) is straightforward. An *in situ* systems validation at a larger scale has been scheduled with one project partner (INRA).

6 Acknowledgements

The authors are grateful to CREAT (La Gaude) and INRA (Avignon) for their participation in this work.

7 REFERENCES

References

[1] C. Bauch and T. Rath. A prototype of a vision based system for measurements of whitefly infestation. *Acta Horticulturae (ISHS)*, 0(691):773–780, Jan. 2005.

[2] E. Z. G. Loy. A fast radial symmetry transform for detecting points of interest. *7 th Euproean Conference on Computer Vision*, Oct.. 2002.

[3] H. B. H.-Y. Chen, N. Brandle and H. Lapp. Robust spot fitting for genetic spot array images. *Image Processing,Proceedings. International Conference on Pattern Recognition and Image Processing*, 3:412–415, Oct.. 2000.

[4] L. V. N. Haering, A. Peter and A. Lipton. The evolution of video surveillance: an overview. *Machine Vision and Applications*, 19:279–290, Jun. 2008.

[5] J. Olivo-Marin. Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9):1989–1996, Sept. 2002.

[6] V. M. P. Boissard and S. Moisan. A cognitive vision approach to early pest detection in greenhouse crops. *Int. Journ.of Comp. Elect. in Agric.*, 2(62):81–93, Jul. 2008.

[7] C. Stauffer and W. Grimson. Background mixture models for real-time tracking. *CVPR*, Aug. 1999.