Computers and Electronics in Agriculture 164 (2019) 104898

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

A computer vision system to monitor the infestation level of *Varroa destructor* in a honeybee colony

Kim Bjerge[a,*], Carsten Eie Frigaard[a], Peter Høgh Mikkelsen[a], Thomas Holm Nielsen[a], Michael Misbih[a], Per Kryger[b]

[a] School of Engineering, Aarhus University, Finlandsgade 22, 8200 Aarhus N, Denmark
[b] Department of Agroecology – Entomology and Plant Pathology, Aarhus University, Flakkebjerg, Forsogsvej 1, 4200 Slagelse, Denmark

ARTICLE INFO

Keywords:
Computer vision
Deep learning
Varroa destructor
Apis melifera
Multi-spectral illumination

2018 MSC:
00-01
99-00

ABSTRACT

This paper presents a portable computer vision system, that is able to monitor the infestation level of the *Varroa destructor* mite in a beehive by recording a video sequence of live honeybees *Apis mellifera* for 5–20 min. A video monitoring unit with multispectral illumination and camera was designed to be placed in front of the beehive, where bees from a selected frame were shaken off. Subsequently, a computer vision algorithm (denoted as the *Infestation Level Estimator*) based on deep learning analysis of the video stream counted the number of honeybees and found the position of identified varroa mites. In this paper, the design and the algorithm that were used to determine the number of bees and mites are presented. Based on a video sequence with 1775 bees and 98 visual mites, the algorithm measured the infestation level to 5.80% compared to a ground truth of 5.52%. The algorithm had a high F_1-score accuracy for counting bees (0.97), while the F_1-score for detecting varroa mites was lower (0.91). The latter was due to mispredictions, which can be resolved by improving both the trained varroa classifier and the mechanical setup. Overall, the proposed computer vision system and algorithm showed a promising results in nondestructive and automatic monitoring of infestation levels in honeybee colonies and should be considered as an alternative to traditional methods, which require the killing of bees.

1. Introduction

Honeybees, *Apis mellifera*, are important in the pollination of plants and crops (Potts et al., 2010; Klein et al.). The ectoparasitic mite *Varroa destructor* has been identified as a main driver for mortality of honeybees and particularly in honeybee colonies (Kralj anf Fuchs; Le Conte et al.), due to virus transmission (Francis et al.; Di Prisco et al.). Beehives must be inspected for the infestation level of mites to provide the needed treatment before the damage threshold level is reached (Rosenkranz et al.). Due to the exponential growth rate of the mites and the difficulties in detecting low infection levels, optimal timing of varroa treatment is a challenge for many beekeepers (Locke et al.). A variety of methods have been developed including collecting a few hundred bees with mites and washing them with detergent or ethanol to detach the mites, passing the suspension through a sieve to determine the infestation level, or de-lodging the mites with powdered sugar (Dietemann et al.). These methods are the ones most commonly used to monitor the infestation level of honeybee colonies today. Ideally, a sample of 300 bees should be used for comparison between colonies as shown by Lee et al. (2010). The disadvantages of these methods are that the samples of bees and mites are killed and that it is a time consuming operation to wash, separate and count the number of bees and mites. In addition, we observed that in practice, the number of bees is not actually counted, but rather estimated based on volume or weight, which may lead to inaccurate measurements.

In this paper, we describe a new method for automatic counting of bees and mites in order to lower the effort required for the beekeeper to determine the infestation level and minimize the damage done to the colony.

1.1. Related work

Attempts at monitoring *Varroa destructors* and honeybees with computer vision have already been described in various papers.

*Visual pollen monitoring outside the hive:*Rodriguez et al. (2018) present a color vision system for detecting pollen-bearing bees, using a convolutional neural network (CNN) for detecting pollen on bees entering the beehive. Using different CNN layouts, they report accuracies varying from 50% to 96%. They commented on the importance of automatic image preprocessing in the form of alignment of bee image

* Corresponding author.
E-mail address: kbe@ase.au.dk (K. Bjerge).

https://doi.org/10.1016/j.compag.2019.104898
Received 13 July 2018; Received in revised form 4 July 2019; Accepted 14 July 2019

data, which is currently a major source of error for the predictors.

Brood cell monitoring inside the hive: Both Elizondo et al. (2013) and Ramírez et al. (2012) present proposals for developing a system for automatic detection and monitoring of varroa movement in honeybee brood cells, whereas Knauer et al. (2007) use linear discriminant analysis to classify brood cells via image processing. Bauer et al. actually report that a thermal gradient is detectable between affected and unaffected brood cells, and they suggest the use of thermal image sensing for brood image control inspections. Several others used computer vision and sensors in various forms inside the hive to monitor and improve bee colony productivity (Giuffre et al., 2017; Zacepins et al., 2016; Cejrowski et al., 2018).

*Visual mite monitoring outside the hive:*Chazette et al. (2016) describe working towards identifying *Varroa destructors* on bees and ultimately killing them with laser beams once the infested bees enter the beehive. They evaluate different algorithms for detection and localization of mites, including the use of convolutional neural networks. They report a "detection rate of 93%" on live image data. Schurischuster et al. experiment with different camera setups for visual detection infested bees, but their work from 2016 (Schurischuster et al., 2016) only focus on capturing good recordings and do not propose any algorithm for detecting mites. They conclude that mite detection will be a challenging task, and suggest the use of IR or UV illumination. In 2018 Schurischuster et al. (2018) present an image analysis and machine learning technique, where individual bee images are extracted and finally classified into mite- and no mite-infested bees. Their work include an image processing pipeline, consisting of image preprocessing/segmentation, feature extraction and final bee and mite classification based on machine learning. Three machine learning algorithms are analyzed: Naive Bayes, Support Vector Machines and Random Forest classifiers, with the latter yielding the best F_1-score (0.83). Image data is only analyzed with a per-frame algorithm and no final bee monitoring system is presented.

The aim of our work is to synthesize existing methods into a portable, robust and automatic vision system, enabling in-field detection, tracking and counting of infected bees. Thus, our work aims to integrate the use of convolutional neural networks as in Rodriguez et al. (2018) and Chazette et al. (2016), illumination wavelength analysis as suggested in Elizondo et al. (2013), Bauer et al. and Schurischuster et al. (2016), and the construction of a fully functional and automatic mite detection system, extending all methods proposed and preliminary results found in Schurischuster et al. (2018).

In relation to image processing, this paper presents a novelty in the form of an extended image processing pipeline, that also takes the temporal image dimension into account. In comparison with the preliminary results in Schurischuster et al. (2018), we present a fully functional image processing pipeline that utilizes a convolutional neural network for mite detection and localization in the intermediate steps of the pipeline. In the high-level part of the pipeline we incorporate a novel image processing algorithm, named Infestation Level Estimator (ILE), which is able to track bees in the temporal dimension.

Our work also presents the first analysis of multispectral illumination on bee and mites using 19 distinct wavelengths in the range from UV to IR. The multispectral bee and mite data is used for a linear discriminant analysis, which enables us to select three optimal wavelengths for an in-field computer vision system.

2. Materials and methods

In this section, the portable computer vision system, which was used to monitor honeybees infected with *Varroa destructors*, is described. A video monitoring unit (VMU) was developed in the form of an extension system box to the beehive with illumination and camera.

The VMU was connected to a laptop for capturing videos and performing automatic inspections. This automatic computer vision system was based on the design described in Tu et al. (2016), but the computer, camera, bee passageway and illumination were changed as described in the following sections.

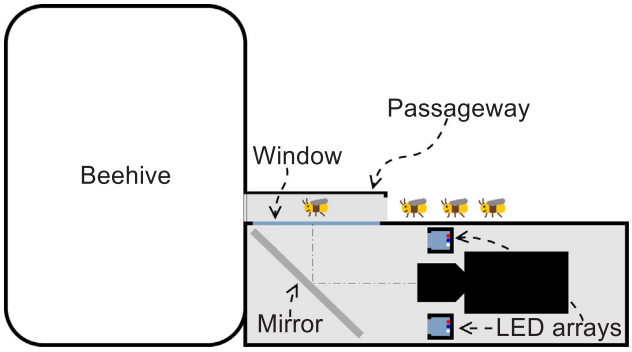

Fig. 1. The video monitoring unit (VMU) at the entrance of the beehive with passageway containing window (acrylic material), mirror (glass), camera and LED illumination arrays. For approximately incident perpendicular light rays the glass mirror had a measured set of reflectivity coefficient given by [0.93 ± 0.020.85 ± 0.030.95 ± 0.02] for the corresponding set of wavelengths [630 nm780 nm470 nm] (R-NIR-B). The acrylic window had a measured reflectance set of [0.048 ± 0.010.039 ± 0.010.040 ± 0.003] for the same set of wavelengths.

2.1. Hardware equipment

The portable VMU could be placed at the entrance of a beehive as illustrated in Fig. 1. The design of the passageway was inspired by Chen et al. (2012) using a black acrylic material with tracks to ensure that the honeybees moved through the passageway in a single file. The passageway and tracks constrained the movement direction and speed of the bees entering back into the hive. The design ensured separation in a perpendicular direction to the tracks, thereby making the image processing and recognition less difficult. However, a queue of bees could still build up along a single track.

While the bees were passing through the narrow tracks, a video stream was captured from below while they were walking over an illuminated window. Distribution of the varroa mites on the bees was analyzed in Bowen-Walker et al. (1997) and Delfinado-Baker et al. (1992), where observations showed that most of the mites are found on the abdomen of the honeybee: "The mites showed a clear preference for the left inter-segmental spaces between the 3rd and 4th abdominal sclerites" (Bowen-Walker et al., 1997). This explains why our passageway was designed so most of the video frames will be taken from the ventral side of the bee. The field of view was adjusted allowing three tracks with three bees in each track to be visible in video frames on a dark background.

Fig. 2 shows a photograph of the actual constructed VMU-unit seen from the outside.

2.2. Bee selection method

Two major ways of inspecting bees were considered: (i) passive monitoring of the natural in- and outflow of bees and (ii) actively taking out individual waxcombs from the hive and shaking off *all* bees just in front of the passageway. The bees would then naturally seek back into the hive and pass through the passageway.

Both methods would lead to some form of bee population selection effect. Method (i) lead to a tendency to lower the frequency of infected bees, since the *Varroa destructors* might tend to stay on the younger bees within the hive. Method (ii) allowed for more bees to pass within a short period of time, but only bees from the selected combs participate in the counting. Adding to this, the passive monitoring via method (1) would place higher requirements on either local video stream buffering or real-time video processing—requirements not yet available with the

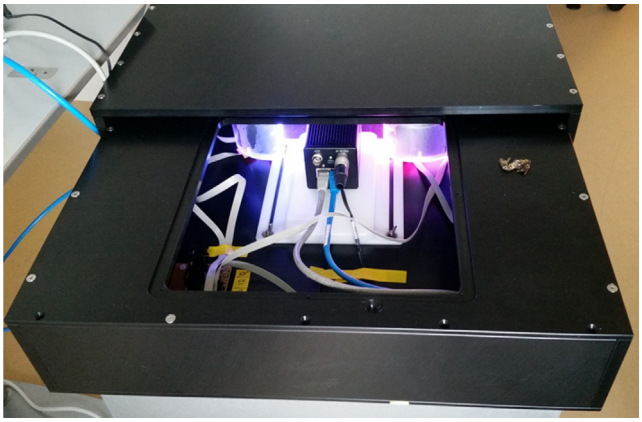

Fig. 2. The video monitoring unit (VMU) seen from the outside with maintenance lid open. Camera and the two multispectral illumination sources can be seen inside the unit.

current VMU system.

2.3. Illumination and camera

It is difficult to distinguish between varroa mites and bees using normal white illumination since they have nearly the same color. Therefore, our approach was to illuminate bees and mites with different spectral wavelengths of light in the spectrum of visible and near-infrared light.

VideometerLab 4 (Videometer, 2016) is a spectral imaging instrument that is used to evaluate the selection of the optimal combination of spectrals. The VideometerLab 4 instrument is able to analyze a sample using 19 distinct wavelengths between the range of 375–970 nm. The sample to be analyzed (bees with mites) was placed inside VideometerLab 4's closed semidome with nearly homogeneous and diffuse lighting conditions, yielding a highly controlled image sampling environment void of specular reflections.

The downside of the VideometerLab 4 was that a full 19-wavelength analysis took around a minute to conduct. This meant that the analysis had to be conducted on live but fixed bees and mites (dead bees yield a different spectrum than live bees). This meant that, due to its physical design, the instrument was incapable of making real-time in-field, video analysis.

The method was thus to select the "best" combination of the 19 distinct wavelengths to be used in the final VMU with a different camera and illumination setup than in the VideometerLab 4 instrument.

Each of the 19 spectral images from the VideometerLab 4 analysis were then manually labeled with a training set of pixel areas containing bees and mites. By combining three different spectral images, a three-dimensional feature vector of randomly labeled pixels was selected for training.

Fig. 3 indicates that the "best" pixel separation of bees and mites was achieved in the near-infrared spectral above 700 nm.

2.3.1. Linear discriminant analysis

To put this into a more formal framework, a linear discriminant analysis (LDA) was used to train and test a classifier using the three-dimensional feature vector of pixels from classes containing bee and mite pixels.

Formally, the combination of the "best" three different wavelengths was found where the classification error on the test set was minimal, with the error being the 0–1 losses on the labeled test data.

When selecting three distinct wavelengths, there was a total of 969 possible wavelength combinations using the binomial coefficient, $\binom{n}{k}\Big|_{n=19,k=3}$. Table 1 shows the 10 best combinations from the LDA analysis. The best separation of bee and mite pixels was obtained by selecting a spectral blue light in the area (435–505 nm), yellow/red light (570–630 nm) and near-infrared light (700–780 nm). The difference of the classification error was less than 0.0045 for these 10 best combinations, and the error for the worst combination was 0.54. A logarithmic relation was observed for the classification error in relation to the sorted ranking of combinations. Among the 50 best combinations, the difference was only 0.009.

2.3.2. Camera and illumination-wavelength selection on basis of the LDA

Since the VideometerLab 4 camera was incapable of real-time in-field analysis, another camera was needed. The multispectral camera AD-130 GE from JAI (JAI) is a specially designed camera equipped with two CCDs, allowing simultaneous capture of color and near-IR spectrals through the same optical path. The camera makes it possible to capture video streams with color and near-infrared (NIR) illumination simultaneously.

The wavelengths 470 nm (blue), 630 nm (red) and 780 nm (infrared) were selected since they belong to the 50 best LDA combinations for pixel classification. Significantly, these wavelengths were also located in the most sensitive area of the color/NIR sensitivity of the JAI camera (see Fig. 4 green component was deliberately omitted since it might affect and overlap with the blue and red components. In Fig. 5, bee and varroa pixel histograms are shown for recordings with the JAI camera with the selected wavelengths.

An extended analysis of the best combination of wavelengths can be found in Ilkiv Misbih and Holm Nielsen (2018). This thesis used a wavelength ranking method and clustering algorithm based on principal component analysis (PCA) to yield a custom *combined measure* (CM) for analyzing the distance between pixel classified as bees and mites respectively. The CM analysis was able to rank all wavelengths combinations, using one, two, three or four district wavelengths to give a ranking list of the "best" combination, including taking the JAI camera spectrum into account.

The CM value of the actual chosen wavelengths combination (470–630-780 nm) gave a rank just below the CM average score. This CM analysis was conducted after selecting the wavelengths actual used, so later versions of the VMU may want to investigate a CM combination with a higher rank.

A specially designed diffuser and a number of narrow spectral LEDs with the selected wavelengths were mounted in the VMU at approximately 45° to the camera focal line to ensure a homogeneous and diffuse illumination to minimize the number of reflections.

Fig. 6 displays the view of the bees traveling along the passageway seen through the JAI camera, with the green image component being replaced with the NIR monochrome component.

2.3.3. Real-time image acquisition and low-level processing

A color and NIR video stream, with a resolution of 1296×966 pixels and 31 fps, was transmitted from the cameras using GigE streaming protocols over two separate Ethernet connections. The necessary sustained bandwidth for transmitting and storing frames real-time equaled 93 MiB/s.

These data were stored persistently for later off-line post-processing. Post-processing consisted of first matching the RGB and NIR stream, temporally coalescing images with equal timestamps and producing a 24 bpp R-NIR-B image stream, see Fig. 7.

Lossless real-time raw stream image compression can be applied if the persistent storage has lower bandwidth than necessary. A primitive compression scheme consists of only looking at the temporal changes between frames and storing the inter-frame difference, which for most frames will be very small. The interframe pixel change could be stored in four bits when the difference is within the range $\pm\,(2^3 - 1)$ (reserving a bit-pattern to indicate the type of the packed data), and in 12 bits when it is greater. Using experimental data this will yield a mean compression ratio of 1.7:1. The current compression implementation is able to compress 24-bpp images with a speed of 100–130 frames/s,

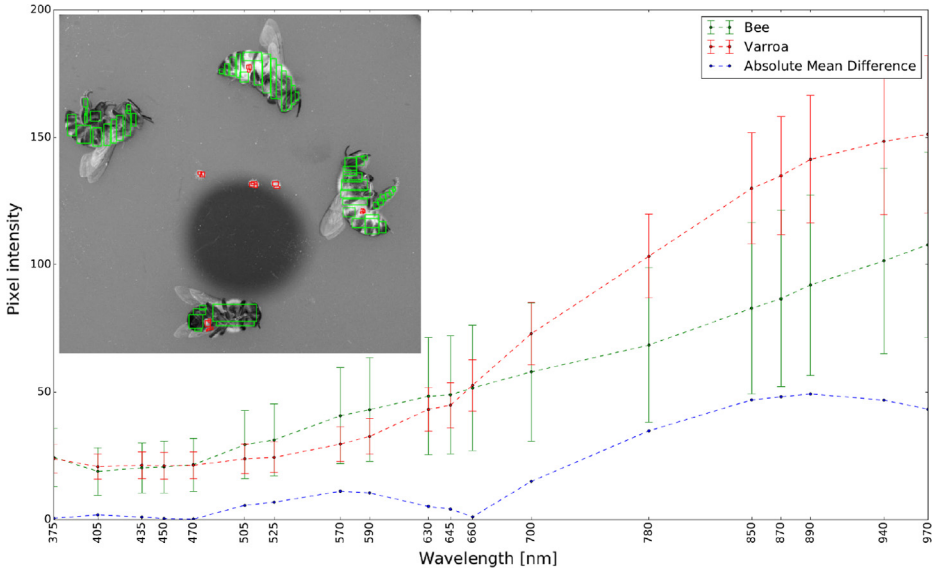

Fig. 3. Mean pixel intensities for pixels in images of bee and varroa mites as a function of wavelength recorded using VideometerLab 4. Bees and mite pixels were manually labeled (see an example of labeling image in upper left corner; actual data used for the plot consist of multiple images) and the μ/σ value pairs for all bee or mite pixels are plotted against wavelength.

Table 1
The 10 best spectral combinations from the LDA analysis with lowest test error, plus the error for the particular combination chosen later (marked by a star).

λ_1/nm	λ_2/nm	λ_3/nm	Error
450	570	780	0.0123
470	570	780	0.0127
450	590	780	0.0131
470	590	780	0.0135
435	570	780	0.0138
435	570	780	0.0146
470	590	700	0.0157
405	570	780	0.0165
450	590	700	0.0168
505	590	780	0.0168
⋮	⋮	⋮	⋮
470	630	780	0.0209*

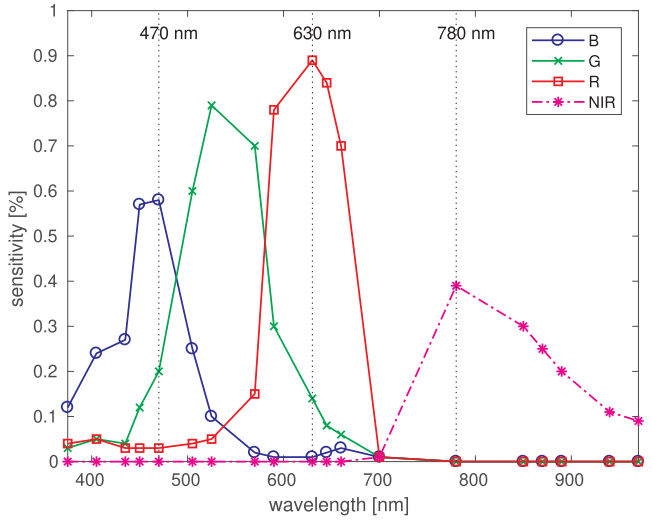

Fig. 4. AD-130 GE Multi-spectral camera pixel sensitivity for the color and near-infrared CCD sensor (JAI, xxxx) at the 19 distinct wavelengths analyzed in the text. The three selected narrow spectral wavelengths for illumination are marked at 470 nm, 630 nm, and 780 nm.

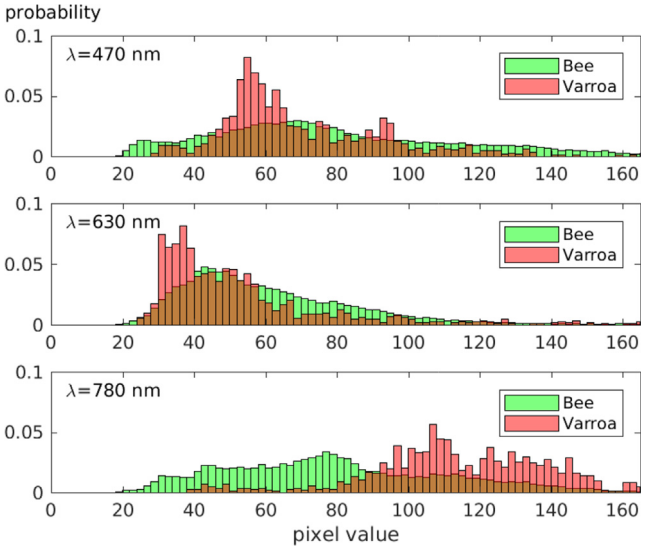

Fig. 5. Histograms of bee and varroa pixel intensity values, for the spectral wavelengths 470 nm, 630 nm, and 780 nm respectively, recorded with the JAI camera. The image path is via the mirror-window-mirror, i.e. data were sampled with the setup given in Fig. 1. The image data for the histogram is the single bee with mite seen in Fig. 6.

which is sufficient for handling both the 12-bpp RGB and 8-bpp NIR raw streams.

With the current off-line post-processing scheme, compression could be paramount for long duration image acquisitions, due to size limitation on fast SSD disk, but also because of the bandwidth limitations of large ordinary disk. Alternatives could be to reduce the image data (bounding box acquisitions, lower frame rate, etc.) or to modify the complete processing pipeline to be real-time capable.

2.4. Intermediate- and high-level processing

The recorded video streams were analyzed with the new developed computer vision algorithm denoted *Infestation Level Estimator* (ILE) to count the number of bees and identify attached varroa mites. The result was the ratio of infested bees in relation to non-infested bees. In the

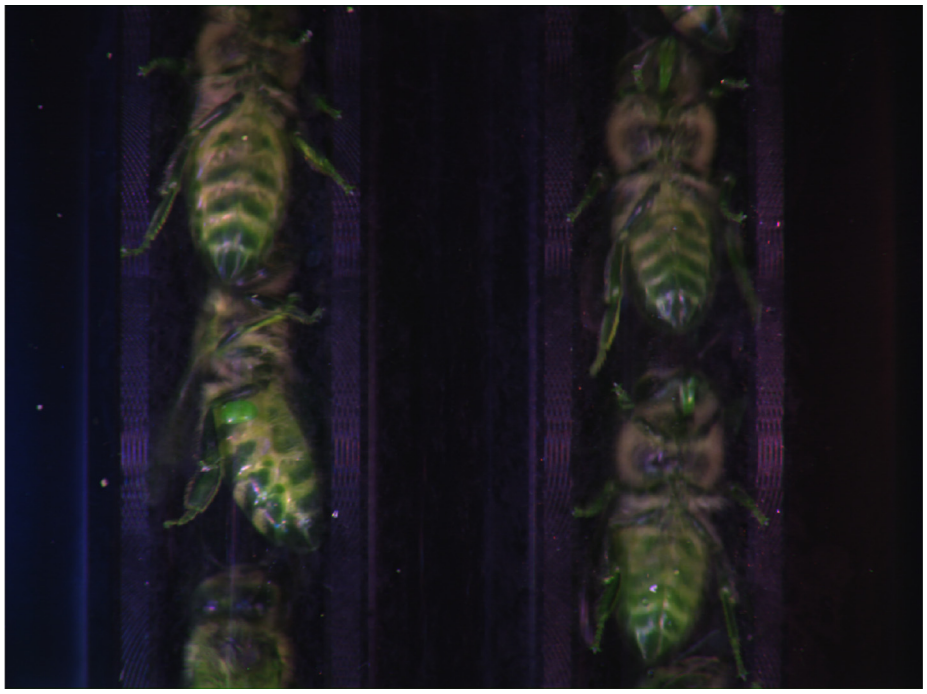

Fig. 6. The actual unprocessed camera view of the bees traveling through the passageway; one bee, to the left in the middle, with a mite. The tracks used to constrain the movements of bees are visible. The image RGB color mapping was R(630 nm)-NIR (780 nm)-B(470 nm), i.e. the normal green component of the image was replaced by the NIR component. With a fixed focal length (8 mm) and constant distance to the objects, the manual setting of the camera *f*-number and shutter speed must strike an optimal value between high illumination (large aperture size/low shutter speed), little motion-blur (high shutter speed) and high depth-of-field (small aperture size).

following subsections, important parts of the algorithm are explained in more detail but first a brief overview of the ILE algorithm is presented.

The ILE algorithm was composed by a number of sequential steps where each image in the video stream was analyzed as illustrated in Fig. 8.

The first two steps were to estimate the location and heading of bees in each image. This was achieved via the image segmentation described in Section 2.4.1, which feature extraction using an elliptic model of the bee. Image features like the *Scale-Invariant Feature Transform* (Lowe) (SIFT) were matched with bee models of different orientations to improve bee separation for regions of bees located very close together. The position and orientation of each bee region in the image were estimated. A detailed description of the bee detection process can be found in Section 2.4.2.

A trained customized *Convolutional Neural Network* (CNN), which used fixed size images of varroa mites, was used to search for mites in the bee regions. This process is described in Section 2.4.3. The actual position of the varroa mite was found by sliding the trained CNN model over a bounding box region of the bee.

The final step in the algorithm was to track bees in the video sequence to ensure that a bee was only counted once as described in Section 2.4.4. For every bee track, it was evaluated whether the bee had a mite attached based on a percentage of mite identifications in all images of the track.

The results were annotated to the video recording with highlighted marks of bee tracks, found varroa mites and statistic counters for bees and varroa mites. The location of detected varroa mites was divided into regions on the bee: thorax and abdomen on the ventral, dorsal side or left/right side.

2.4.1. Background subtraction and segmentation

A grey-scaled foreground image was made by subtracting a fixed image for all frames of the dark background. The grey-scaled foreground image was then segmented by a trained threshold, followed by a morphological open and close operation to create masks of bee regions. Estimating position and heading of bees based on bee masks only worked properly when the bees were not too close to each other as seen in the mask segmentation of Fig. 8. Therefore, a part-based model (Grauman and Leibe) were included to improve the segmentation of bees.

2.4.2. Bee detection

The method for detecting bees was inspired by the implicit shape model (ISM) presented by Leibe et al. (2006). ISM is a well-established method for recognition and segmentation, and it only requires a small number of training samples. As the ISM name suggests, the aim is not to define an explicit model for all possible shapes an object may take, but instead define shapes in terms of which local features are consistent with each other. This allowed a bee to be detected in an image, even though only a subset of its parts was present.

Our method was modified to only focus on detection, as bees were the only objects desired to be detected. An overview of the different phases in detecting objects with this method is seen in Fig. 9. The *training phase* consisted of learning a *codebook* that contained information of which local features may appear on bees. In our case, four images with a size of 240 × 494 pixels were selected, seen from the ventral, dorsal, left and right side of the bee. The codebook was created with feature entries found by a detection algorithm where *Scale-Invariant Feature Transform* (SIFT), *Speeded up robust features* (Bay et al.)

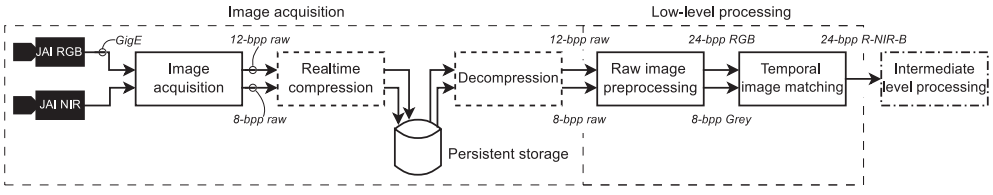

Fig. 7. The low-level image processing pipeline. Raw camera images are stored on disk for later retrieval and post-processing. 12- and 8 bits-per-pixel are used as the raw JAI/Bayer packed pixel format for the RGB and NIR images respectively. Lossless real-time compression can be introduced if persistent storage bandwidth is less than the raw-stream image rate of 93 MiB/s. The 12- and 8-bpp raw images from the network arrives out-of-order with respect to each other, hence the need for the temporal image matching.

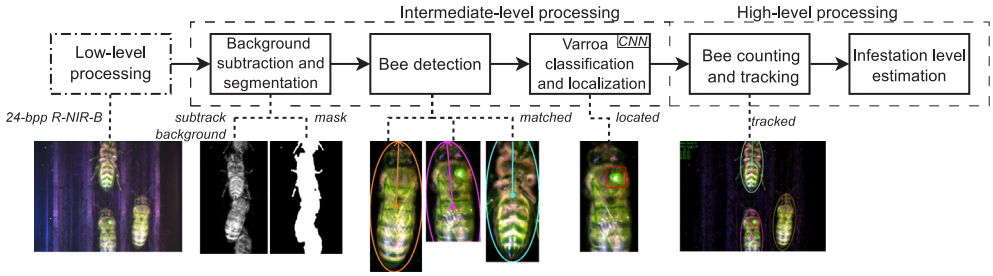

Fig. 8. The processing pipeline of the intermediate- to high-level image processing algorithms to analyze and count the number of bees with *Varroa destructor*. A trained convolutional neural network (CNN) was used for the Varroa classification and localization stage.

(SURF) and *Oriented Fast and Rotated BRIEF* (Rublee et al., 2011) (ORB) were evaluated. The next step was to learn the actual ISM that specified where on the bees the codebook entries may occur, i.e. for every codebook entry a spatial relationship to the bee center and heading was stored.

When the training phase was done, a set of SIFT, SURF or ORB features were extracted from a raw input image. The extracted features were then matched against features contained in the codebook from training images bees seen from different orientations. The matching was performed by calculating the Euclidean distance between features in the image and codebook. In Fig. 9, feature matching shows significantly more matches with the training image of a bee with the same orientation at the ventral side. When a feature was matched with an entry from the codebook, the activated entry cast a vote for a possible object center position and heading. The vote was based on the spatial relationship from the matched entry learned during the training phase. All votes with a center position outside the segmented foreground mask or with an impossible heading due to the mechanical setup were discarded.

Each vote described a potential bee object in the image, but as many of the votes were relatively similar, the votes were clustered to reduce the number of bee hypotheses. Leibe et al. (2006) used *Mean-Shift Mode Estimation* (MSE) to search for hypotheses as maxima in the voting space, and all parts that contributed to it were collected. The votes were therefore clustered using MSE to initially reduce the number of hypotheses. The final step was to reduce the remaining hypotheses to only the ones that best describe the image. To find the optimal number of hypotheses, each hypothesis was assigned a score based on its internal consistency. It was based on the product of the number of supporting votes and the mean distance between feature locations that voted for the hypothesis as features scattered in a larger area had a higher likelihood of being an actual bee. The hypotheses were sorted in descending order, based on its score, to indicate the confidence level in each of the hypotheses. To remain a valid hypothesis, each of them had to comply with different criteria, such as foreground coverage, image overlap and ellipse overlap.

2.4.3. Varroa identification and localization

It was a challenging task to identify varroa mites on the bee. Several traditional computer vision and machine learning methods were investigated, such as separating the bee and *Varroa destructor* at pixel level. The overall idea was to use a classifier such as a *Support Vector Machine* (SVM) or LDA. The classifier was trained with pixel areas of mites and bees. The classification output was a segmented black and white image in which each white region should represent a *Varroa destructor*. This method seemed promising regarding its potential for a high recall classification of varroa mite and bee pixels. However, the major drawback was a low precision of the varroa classification, where a score of 0.71 was achieved measured at the pixel level. This means that about 29% of the varroa pixels would be classified as areas on the

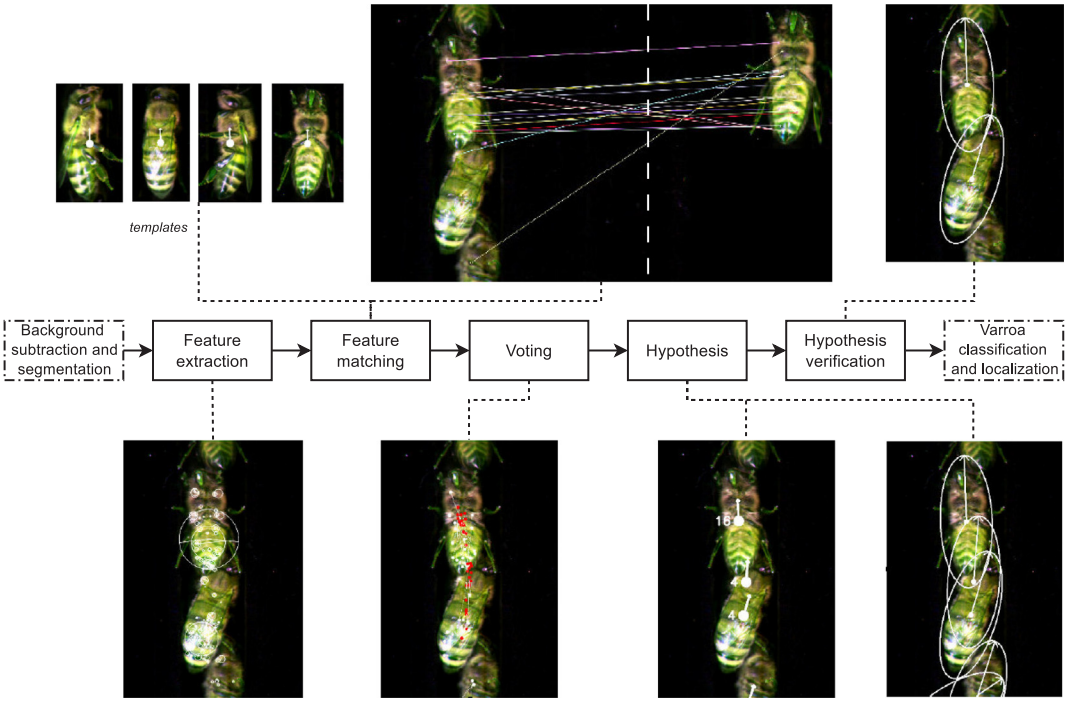

Fig. 9. The bee detection processing stage expanded into sub-pipeline methods using a method inspired by ISM.

bee. Therefore, we proposed the following customized CNN architecture. In the field of deep learning, especially architectures of *Convolutional Neural Network* have provided particularly good results in many areas of Computer Vision (Liu et al.). The CNN method uses both pixel intensity values and spatial information about objects in the image. Several CNN architectures were investigated and hyperparameters of the networks were explored in finding the optimal network architecture to classify varroa mites. Since the camera recorded images of bees with a constant working distance, the image of varroa mites will not change in size. An image window from the bee region with the size $32 \times 40 \times 3$ pixels was chosen to cover the maximum size of a varroa mite with the given camera setup, where the image window had three channels one for each selected wavelength. Several images were used for training the CNN network, with and without varroa mites. The optimal architecture was found by using combinations of the hyperparameters below to train different CNN's for varroa classification:

- Kernel size $n \times n$, $n \in \{3, 4, 5\}$
- Pool size, $n \times n$, $n \in \{2, 3\}$
- Convolutional depth n, $n \in \{8, 16, 32, 64, 128\}$
- Fully connected size n, $n \in \{64, 128, 256, 512\}$

The optimal chosen CNN architecture is shown in Fig. 10. It had four layers for feature detection and a fully connected neural network for final varroa mite classification. The first two layers performed convolution using 256 kernels with a kernel size of 4×4 followed by maximum pooling of size 2×2 and stride 2. The next two layers performed convolution using 32 kernels with a kernel size of 4×4, with the same pooling as mentioned above. All convolutional layers used the ReLu activation function. The last fully connected layer had one hidden layer with 128 neurons and a sigmoid activation function in the output layer. This architecture was chosen because it achieved average precision, recall and an F_1-score above 99% which indicated a good model in detecting image areas with varroa mites. More results regarding the varroa classifier are described in Section 3.2.

Once the bee region was located in the image, a window of 32×40 pixels was slid over the entire bee region using the trained CNN for prediction. The result was a binary vector that was reshaped to create a binary image with blobs of areas representing varroa mites. By combining the varroa placement in relation to the ellipse model of the bee, it was possible to give an estimate of where on the bee the *Varroa destructor* was found. The selected areas were: head, thorax and abdomen on the ventral, dorsal and left/right side of the bee. Varroa mites found on the head region were excluded based on the analysis by Bowen-Walker et al. (1997) which indicated an extremely small probability (0.1%) of mites located on the head region. The problem was that the glossa (tongue) of the bee had similar shape, size, and color as the *Varroa destructor* and will in many cases be classified as false positive detections.

2.4.4. Bee tracking and counting

Even though a *Varroa destructor* was found on multiple images, it should only be counted once if it was on the same bee. It was, therefore, necessary to track the bees in order to correctly count the bees.

The position and heading of each bee were estimated for every single frame, and tracking could, therefore, be solved by finding the optimal assignment of bees in two consecutive images. The *Hungarian Algorithm* (Munkres) was our chosen method for finding the optimal assignment for a given cost matrix. In this application, the cost matrix should represent how likely it was that a bee in the previous image had moved to a given position in the current image. The cost function was defined as the Euclidean distance between bee center position in the previous and current image. The *Hungarian Algorithm* required the cost matrix to be squared, and in our case was defined as an $N \times N$ matrix, where each entry was the cost assigning bee_i in previous image to bee_j in current image. Dummy rows and columns were added to the matrix to ensure that it was always squared. All entries in the dummy row or column had to have a cost significantly larger than all other entries in the cost matrix to ensure that the algorithm did not make a "wrong" assignment to a dummy. The resource or task being assigned to a dummy could be used to determine which bee from the previous image had left, or which bee had entered into the current image. The bee counter was incremented when a bee left the top of the image, and it was evaluated whether the bee was infested.

3. Results

One experiment was conducted at Aarhus University in Flakkebjerg, Denmark in September 2017 under varying weather conditions with both sun and rain during the experiment. The experiment with one infected beehive was performed by capturing 10 video recordings with a duration between 1 and 10 min. The setup of the experiment is shown in Fig. 11. Infected and healthy bees were shaken directly off a frame onto the black plate of the VMU at the entrance of the beehive. Frames were selected randomly from the beehive. After waiting a while, the bees started seeking back into the beehive, passing through the specially designed and illuminated passageway while they were filmed. To help the process along, smoke was used to force the bees into finding their way back to the hive.

Bees from an infected beehive with varroosis were then used in the experiment. A sample of 697 bees was washed with soap water to estimate the level of infestation of the entire colony. Here, 130 mites were counted manually and the number of bees was measured by comparing the weight of 60 bees to the total weight. The number of bees was estimated to be 697, which was the total weight divided by the average weight of one bee. The infestation level by alcohol wash was estimated to be $130/697 = 18.7\%$. A number of video recordings were used to train and develop the algorithm to count the ratio and position of *Varroa destructors*. One selected video recording of 4:46 min (**T1**) was used for testing and evaluation of the computer vision system and accuracy. Varroa mites found in all other videos (**TR**) of 40:27 min were used to train the CNN classifier.

A second final test (**T2**) of the system was conducted by a beekeeper in an infested beehive with queen No. 579 in late September 2018 in Vitved, Denmark. Several video sequences were recorded (32:33 min in total) using the same VMU setup as in Flakkebjerg, with the only modification being an increase in LED lights.

The infestation level was manually measured by the beekeeper to be

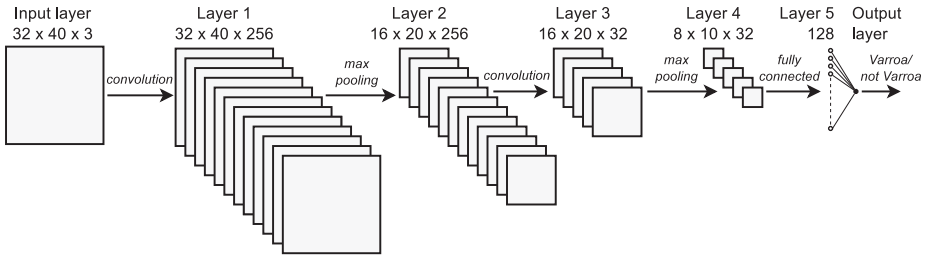

Fig. 10. CNN architecture for *Varroa destructor* classification used an input of a 32×40 R-NIR-B image.

Fig. 11. Photo of the video monitoring unit in front of a beehive with bees seeking back into the hive.

10.3% one week before the video recordings of **T2**. The infestation level was measured using an alcohol wash, where the number of bees was estimated to be 300 based purely on volume. This test was made after development of the system was completed, including VMU and ILE.

In the following sections, the results from the experiments of the change in illumination and the last stages of the algorithm concerning intermediate to high-level processing (as described in Fig. 8) will be presented.

3.1. Illumination

The number of LEDs was doubled in the **T2** experiment, and the actual increase in light intensity was measured to be around 80% in (20% loss due to the diffuser setup), yielding an increase in the input image quality. Compared to Fig. 6, the aperture size was lowered slightly giving both better depth-of-field and a slightly more illuminated image due to the absolute light intensity increase. Other camera settings, like shutter speed, were not altered from **T1** to **T2**. In software only a single image preprocessing parameter (image intensity scaling) was adjusted to accommodate for the new, brighter image stream.

Quality analysis of the raw camera image streams on **T2**, however, still showed signs of per-wavelength anisotropic illumination and missing inter-balancing of the R-NIR-B channels, which should have been accomplished in the low-level image processing stages of Fig. 7. The NIR channel had a mean pixel value approximately 40% lower than the R and B channels. Levelling out this inter-wavelength difference might increase the overall precision.

Analysis of the reflectance coefficients of the mirror and the acrylic window (see Fig. 1) is currently suffering from a high degree of uncertainty. The reflectance of the mirror was found to be in the order of 85–95% but with uncertainty in the physical measurement being at least of the order of 2–4%. The air-acrylic material reflectance was found to be around 4%, with a tendency to yield a lower mirror and higher air-acrylic reflection in the NIR band compared to the R and B wavelength.

The standard deviation in both sets of coefficient comes from a set of six different physical setups for measuring the mirror reflectivity and three different setups for calibrating the air-acrylic reflectance. The real uncertainty in the measurements is judged considerable higher due to the primitive physical constraints of the experiments.

Overall the importance of these reflections coefficients were, however, dwarfed by the absolute increase in illumination intensity (80%) where the high-level image processing stages of VMU displayed a strong tendency towards being relatively independent of illumination variance.

3.2. Varroa classification

A dataset using **TR** videos for training and testing the varroa classifier was based on 3,476 32×40 cropped R-NIR-B images of varroa and 4819 areas of bees. This gave a total of 8295 labeled images found in **TR** video recordings used for training. From this dataset, 80% was used for training and 20% for testing. Data argumentation was applied to all images with a flip vertical, horizontal and rotation of 180 degrees. This operation provided four times more training and test data and improved the CNN classifier. To find the best CNN architecture for varroa classification, different hyperparameters were adjusted, including kernel size, pool size, convolutional depth and fully connected hidden layer size as described in Section 2.4.3. All architectures were trained by performing 5-fold cross-validation, early stopping and dropout probability of 0.5 after each hidden layer. The average F_1-score was used as a measure for a given architecture's performance. Table 2 shows a summary of the architectures with the highest F_1-scores.

The best three architectures had very high F_1-scores, which only varied by 0.0005, but had a varying number of learnable parameters. The architecture with the smallest amount of learnable parameters (570,000) was chosen in the end, because it was faster to train and minimized the risk of overfitting. The chosen model shown in Fig. 10 had an F_1-score of 99.5%, which indicated that the trained CNN was very accurate in its predictions. The model had a recall of 99% indicating that 1% of the varroa mites in the test set were missed.

3.3. Bee counting and tracking

The test (**T1**) video recording was used to evaluate the tracking and counting of bees. The result concerned both bee counting and tracking as a combined overall evaluation. An F_1-score of 0.97 was achieved by detecting 301 out of 303 observed bees. The **T1** video recording was divided into two parts, where the second part had many bees that were located close together. This part of the recording with 180 bees had a recall of 96%, meaning that 4% of the bees were missed. SIFT was used for feature extraction and bee models in the final test and evaluation.

3.4. Infestation level estimator

The evaluation of the *Infestation Level Estimator* was based on the test

Table 2
The three best CNN architectures with highest F_1 score of the classifications. The Hyperparams column shows dictionary values of: {kernel size, pool size, convolutional depth layer 1, convolutional depth layer 3, fully connected size}.

Rating	Hyperparams	Learnable parameters	F_1-score
1.	4, 2, 64, 128, 51	1,400,000	0.9974
2.	4, 2, 256, 64, 256	1,200,000	0.9972
3.	4, 2, 256, 32, 128	570,000	0.9969

Table 3
Scores of the ILE's *Varroa destructor* counts. All training (**TR**) videos (9 videos of 40:27 min in total) with 961 bees recorded at Aarhus University. Test video **T1** (4:46 min) with 302 bees, recorded at same beehive as the videos used for training. Test videos **T2** (32:33 min) with 1775 bees, recorded at infested beehive with queen No. 579 in Vitved.

Video	Mites	ILE	Precision	Recall	F_1-score
TR	78	8.12%	0.90	0.86	0.88
T1	26	8.61%	0.85	0.88	0.86
T2	103	5.80%	0.88	0.93	0.91

(**T1**, **T2**) video recordings. It is important to emphasize that the recording for these tests was not used to train varroa classification. The *ground truth counts* were created through manual inspection of the video by an expert specialized in honeybees.

For **T1** the ILE counted 26 infested bees as opposed to the ground truth's count of 25. The infestation level was measured to be 8.61% (26/301) by the ILE, whereas the ground truth was 8.25% (25/303). Of the ground truth's of 25 varroa mites, only 22 were counted correctly, meaning that four mites were false positive predictions. The estimate of the varroa position was very accurate, as 20 out of 22 mites were located correctly on the bee.

The ILE measurements, precision, recall and F_1-score for the training (**TR**) and tests (**T1**, **T2**) are shown in Table 3. For the final test **T2**, the infestation level was measured to be 5.80% (103/1775) by the ILE. By comparison, the ground truth was 5.52% based on 98 mites and 1,775 bees found in 32:33 min of video. The precision, recall and F_1-score were very similar for the training and test video recordings. Videos used to train (**TR**) the varroa CNN classifier had a better precision than test video, which was to expected. The recall and F_1-scores for test **T2** was a little higher than **TR** and **T1** due to improved uniform illumination inside the VMU.

4. Discussion

The *Infestation Level Estimator* was good at tracking and counting bees even when the bees were close together. The results showed that the most challenging task was to detect varroa mites correctly as a number of mispredictions were observed. Missing varroa detections occurred three times where two situations were due to the fact that varroa mites were barely visible in the **T1** video recording.

Training of the varroa CNN classifier was a challenging task. In the first attempt, 3476 random parts of the bee were selected, and a low F_1-score of only 0.76 was measured based on the **T1** video. To improve the classifier, additional images containing parts of the bee, causing the false predictions, were added from the training videos. In total, 1340 additional images of the bee including mouth, wings and abdomen were manually labeled and added to the training data. The most significant remaining error derived from the dorsal tip of the abdomen of the back side being classified as a varroa mite, which explains two out of the four false positive predictions. An improvement of the ILE will be to train it with even more data from the bottom of the abdomen, which was the area causing most of the false predictions. It will also be desirable to somehow create more uniform illumination and inter-balaced R-IR-B wavelength levels as attempted in the final test **T2**, which resulted in an improved recall.

Results from **T2** recordings measured 12 false mites which was equal to a false infestation level of only 0.68% (12/1775). This means that the ILE will measure the infestation level in a beehive without varroa mites to be ~1%.

The ILE was accurate in detecting varroa mites on infested bees, even though there were some challenges with the current setup. The lack of visibility of the varroa mites was a general problem since the bees were only filmed from a single point of view. Once the bees entered the passageway, it was very unlikely that they would twist or turn, meaning only one side of the bee would be visible. The ground truth infestation level of 5.52% was far from the washed sample of bees, where a level of 10.3% was measured. This difference could be due to both the visibility of the bees and the fact that the washed sample was not the same bees as the ones recorded in the video. The washed sample was made a week before the recorded video, which could also explain some of the difference.

In order to conduct better statistical analysis and draw empirical conclusions about the proposed system, there is a need for more video recordings from different beehives. Further the mechanical setup must be improved in order to ensure that the bees do not rush and try to pass each other, causing errors in bee counting.

5. Conclusion

In this paper, a computer vision system to monitor the infestation level of *Varroa destructors* in a bee colony was presented. A *Video Monitoring Unit* recorded the infested bees which were manually shaken off a selected frame from the hive and dropped on the unit at the entrance of the beehive. A multispectral camera and illumination with blue, red and infrared LED light were used in video recordings to ease the separation of varroa and bee pixels. A novel computer vision algorithm based on deep learning named *Infestation Level Estimator* (ILE) analyzed the video by tracking and counting the number of bees and estimating positions of found varroa mites.

The infestation level was measured to be 5.80% by the ILE, while to the ground truth was measured to be 5.52% based on a 32:33-min video recording with 1775 bees, out of which 103 were infected with mites. The percentage of false varroa detections was ~1%. The ILE had an F_1-score accuracy of 0.97 for counting bees and only 0.91 for detecting infected bees due to misprediction. The latter can likely be resolved by improving the ILE algorithm and the mechanical setup.

Acknowledgments

We would like to thank Peter Ahrendt who made this project possible with his initial investigations and the idea of using multispectral illumination. The work was supported by the beekeeping program 2016–2019 funded by the Danish Government and EU [Grant No. 16-80611-000003].

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compag.2019.104898.

References

Bauer, D., Wegener, J., Bienefeld, K., Recognition of mite-infested brood by honeybee (apis mellifera) workers may involve thermal sensing. J. Therm. Biol. 74. doi:https://doi.org/10.1016/j.jtherbio.2018.04.012.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., Speeded-up robust features (SURF). Comput. Vis. Image Understand. doi:https://doi.org/10.1016/j.cviu.2007.09.014. arXiv:arXiv:1011.1669v3.

Bowen-Walker, P.L., Martin, S.J., Gunn, A., 1997. Preferential distribution of the parasitic mite, Varroa jacobsoni Oud. on overwintering honeybee (Apis mellifera L.) workers and changes in the level of parasitism. Parasitology 114 (2), 151–157. https://doi.org/10.1017/S0031182096008323.

Cejrowski, T., Szymański, J., Mora, H., Gil, D., 2018. Detection of the bee queen presence using sound analysis. In: Nguyen, N.T., Hoang, D.H., Hong, T.-P., Pham, H., Trawiński, B. (Eds.), Intelligent Information and Database Systems. Springer International Publishing, Cham, pp. 297–306.

Chazette, L., Becker, M., Szczerbicka, H., 2017. Basic algorithms for bee hive monitoring and laser-based mite control. In: 2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016, doi:10.1109/SSCI.2016.7850001.

Chen, C., Yang, E.C., Jiang, J.A., Lin, T.T., 2012. An imaging system for monitoring the in-and-out activity of honey bees. Comput. Electron. Agric. 89, 100–109. https://doi.org/10.1016/j.compag.2012.08.006.

Delfinado-Baker, M., Rath, W., Boecking, O., 1992. Phoretic bee mites and honeybee grooming behavior. Int. J. Acarol. 18 (4), 315–322. https://doi.org/10.1080/01647959208683966.

Dietemann, V., Nazzi, F., Martin, S.J., Anderson, D.L., Locke, B., Delaplane, K.S., Wauquiez, Q., Tannahill, C., Frey, E., Ziegelmann, B., Rosenkranz, P., Ellis, J.D., Standard methods for varroa research. J. Apicult. Res. doi:https://doi.org/10.3896/IBRA.1.52.1.09.

Di Prisco, G., Annoscia, D., Margiotta, M., Ferrara, R., Varricchio, P., Zanni, V., Caprio, E., Nazzi, F., Pennacchio, F., A mutualistic symbiosis between a parasitic mite and a pathogenic virus undermines honey bee immunity and health. In: Proceedings of the National Academy of Sciences. doi:https://doi.org/10.1073/pnas.1523515113.

Elizondo, V., Brice no, J., Travieso, C., Alonso, J., 2013. Video monitoring of a mite in honeybee cells. doi:https://doi.org/10.4028/www.scientific.net/AMR.664.1107.

Francis, R.M., Nielsen, S.L., Kryger, P., Varroa-virus interaction in collapsing honey bee colonies. PLoS ONE. doi:https://doi.org/10.1371/journal.pone.0057540.

Giuffre, C., Lubkin, S.R., Tarpy, D.R., 2017. Automated assay and differential model of western honey bee (Apis mellifera) autogrooming using digital image processing. Comput. Electron. Agric. 135, 338–344. https://doi.org/10.1016/j.compag.2017.02.003.

Grauman, K., Leibe, B., Visual Object Recognition, Synthesis Lectures on Artificial Intelligence and Machine Learning. doi:https://doi.org/10.2200/S00332ED1V01Y201103AIM011.

Ilkiv Misbih, M., Holm Nielsen, T., 2018. Estimating Varroa Destructor Infestation Level in a Bee Population Using Computer Vision, Master's thesis. Aarhus University, Department of Engineering.

JAI. Ad-130ge is a 2-ccd multi-spectral prism camera (Accessed on 05.06.2018). < https://www.jai.com/products/ad-130-ge > .

Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., Importance of pollinators in changing landscapes for world crops. In: Proceedings of the Royal Society B: Biological Sciences. doi:https://doi.org/10.1098/rspb.2006.3721.

Knauer, U., Zautke, F., Bienefeld, K., Meffert, B., 2007. A comparison of classifiers for prescreening of honeybee brood cells.

Kralj, J., Fuchs, S., Parasitic Varroa destructor mites influence flight duration and homing ability of infested Apis mellifera foragers. Apidologie. doi:https://doi.org/10.1051/apido:2006040.

Le Conte, Y., Ellis, M., Ritter, W., Varroa mites and honey bee health: can Varroa explain part of the colony losses? Apidologie. doi:https://doi.org/10.1051/apido/2010017.

Lee, K.V., Moon, R.D., Burkness, E.C., Hutchison, W.D., Spivak, M., 2010. Practical sampling plans for varroa destructor (acari: Varroidae) in apis mellifera (hymenoptera: Apidae) colonies and apiaries. J. Econ. Entomol. 103 (4), 1039–1050. https://doi.org/10.1603/EC10037.

Leibe, B., Leonardis, A., Schiele, B., 2006. An Implicit Shape Model for Combined Object Categorization and Segmentation (May). pp. 508–524. doi:https://doi.org/10.1007/11957959_26.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., A survey of deep neural network architectures and their applications. Neurocomputing. doi:https://doi.org/10.1016/j.neucom.2016.12.038.

Locke, B., Semberg, E., Forsgren, E., De Miranda, J.R., Persistence of subclinical deformed wing virus infections in honeybees following Varroa mite removal and a bee population turnover. PLoS ONE. doi:https://doi.org/10.1371/journal.pone.0180910.

Lowe, D.G., Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. doi:https://doi.org/10.1023/B:VISI.0000029664.99615.94.

Munkres, J., Algorithms for the assignment and transportation problems. J. Soc. Ind. Appl. Math. doi:https://doi.org/10.1137/0105003.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E., 2010. Global pollinator declines: trends, impacts and drivers. doi:https://doi.org/10.1016/j.tree.2010.01.007.

Ramírez, M., Prendas, J.P., Travieso, C.M., Calderón, R., Salas, O., 2012. Detection of the mite varroa destructor in honey bee cells by video sequence processing. In: 2012 IEEE 16th International Conference on Intelligent Engineering Systems (INES), pp. 103–108. https://doi.org/10.1109/INES.2012.6249811.

Rodriguez, I.F., Megret, R., Acuna, E., Agosto-Rivera, J.L., Giray, T., 2018. Recognition of pollen-bearing bees from video using convolutional neural network. IEEE Winter Conference on Applications of Computer Vision (WACV) 2018, 314–322. https://doi.org/10.1109/WACV.2018.00041.

Rosenkranz, P., Aumeier, P., Ziegelmann, B., Biology and control of Varroa destructor. J. Invertebr. Pathol. doi:https://doi.org/10.1016/j.jip.2009.07.016.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: an efficient alternative to SIFT or SURF. In: Proceedings of the IEEE International Conference on Computer Vision. doi:https://doi.org/10.1109/ICCV.2011.6126544. arXiv:0706.3355.

Schurischuster, S., Zambanini, S., Kampel, M., Lamp, B., 2016. Sensor Study for Monitoring Varroa Mites on Honey Bees (Apis mellifera).

Schurischuster, S., Remeseiro, B., Radeva, P., Kampel, M., 2018. A preliminary study of image analysis for parasite detection on honey bees. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). doi:https://doi.org/10.1007/978-3-319-93000-8_52.

Tu, G.J., Hansen, M.K., Kryger, P., Ahrendt, P., 2016. Automatic behaviour analysis system for honeybees using computer vision. Comput. Electron. Agric. 122, 10–18. https://doi.org/10.1016/j.compag.2016.01.011.

Videometer A/S, 2016. Lyngsøe Allé 3, DK-2970 Hørsholm, Denmark, VideometerLab 4, pp. 1–2 (Accessed on 4.07.18). < https://videometer.com/Portals/0/pdfs/VideometerLab4.pdf?ver=2016-05-30-152146-040 > .

Zacepins, A., Kviesis, A., Ahrendt, P., Richter, U., Tekin, S., Durgun, M., 2016 (2016). Beekeeping in the future – smart apiary management. In: Proceedings of the 2016 17th International Carpathian Control Conference, ICCC, pp. 808–812. https://doi.org/10.1109/CarpathianCC.2016.7501207.