Dealing with large amount of data from automated pest monitoring system

Matej Štefančič, Dejan Jerič, Mateja Štefančič, Peter Čebokli

EFOS d.o.o.
Razdrto 47b
6225 Hruševje, Slovenia
info@efos.si

Abstract: Pest monitoring is time consuming and labor intensive task, since pheromone traps, where insects are being caught, need to be visited on regular basis in order to monitor development of pest population. An innovative product which provides automated and real-time monitoring of pests reduces the need for field-scouting to minimum. In order to develop such product and to keep the whole system running some non-trivial IT issues had to be solved.

1 Introduction

While information technology is more and more used in the field of agriculture some tasks still remain an important challenge. One of such things is monitoring of pests. Due to the fact that there are lots of external factors which can affect occurrence of pests, monitoring still heavily rely on manual inspection of (pheromone) traps. This process is time consuming and labor intensive. Skilled staff need to manually inspect traps, count insects which were caught in the trap and maintain record keeping/statistics of caught insects. As a result, most pest monitoring related field inspections are done on weekly basis, thus results can be misleading (there is not correct estimate when insects were caught) or incomplete - predators can eat insects from the trap. Decisions can also be taken too late due to time delay, especially with ever more popular mating disruption systems.

To deal with challenges listed above, automated and near-real time monitoring of pest occurrence was developed. New solution is a combination of device in the field and software running on the server which handles data sent from devices in the field and provides data to end users. Device is automatically taking image snapshots of sticky surface in modified pheromone trap, sends data to the server, where it is processed and served to users via web or mobile application. Such approach generates on average 8-10 MB of data per trap daily.

Article will focus on brief description of the device which is placed in the field and key factors/problems which needed to be addressed (and which greatly influence the overall solution). Then procedure of processing of data and its organization on the server will be explained.

2 Automated trap for monitoring of insect pests

The device in the field consists of simple trap housing, which is slightly modified large delta from corrugated polypropylene. All needed electronics (battery, CPU, modem, cameras, GPS, etc) is packed into special housing with transparent cover and attached into housing with a strong clip. In the field the device is energy-independent, charged via solar panel. Pests are lured into trap with pheromone or some other lure and retained on sticky surface.

For data communication GPRS is used, since it enables to use infrastructure built by mobile providers. In order to cope with the problem a set of innovative techniques was used/developed in order to reduce amount of data transferred and to make complete transfer as robust as possible. These techniques include segmentation and compression of data, dynamic adjustments of transfer speed, splitting of data in smaller logical chunks and a complete higher level protocol to deal with errors.

3 Data processing

Data processing on device side is limited to device capability so most data/image processing is done on server side (Fig. 1). Additional tasks are performed on device in order to send reliably as much little data possible and still keep good quality of images. These operations are controlled by parameters which can be set on the server or via SMS. There are multiple cameras on device. Raw images for each camera are converted, split by colors to separated files, compressed lossless or lossy with different compression levels (depending on parameters received). Some basic image algorithms are executed in order to avoid sending non-appropriate images (too bright or too dark images for example).

On the server side a process is monitoring location where traps are uploading data in order to start with server-side processing of data as soon as it is uploaded. Files are moved to location created according to defined rules in order to quickly find and manipulate the images later. Color separated images are integrated into one image (per camera). In next step images for each camera are stitched to a single image which is what customer actually can see with his client (web browser, mobile phone, tablet etc.). Multiple image sizes and qualities are also prepared and saved as image thumbnails to optimize response times when serving data to clients. One of the most important modules (currently in beta test phase) is image recognition module which detects targeted insects in image with around 75% probability. Based on that user gets list of traps which need special attention/require further action in the field/orchard. User has also possibility to visually mark targeted insects in the image via modified SVG editor. Both, image recognition module and manual marking are also source of attribute data regarding number of caught targeted insects which is key information for field actions as well as statistical modeling of flight of certain insects.

While most of the data belongs to images, attribute data is communicated at each data transfer as well. This data (like current GPS location, battery level, signal strength, ...) is

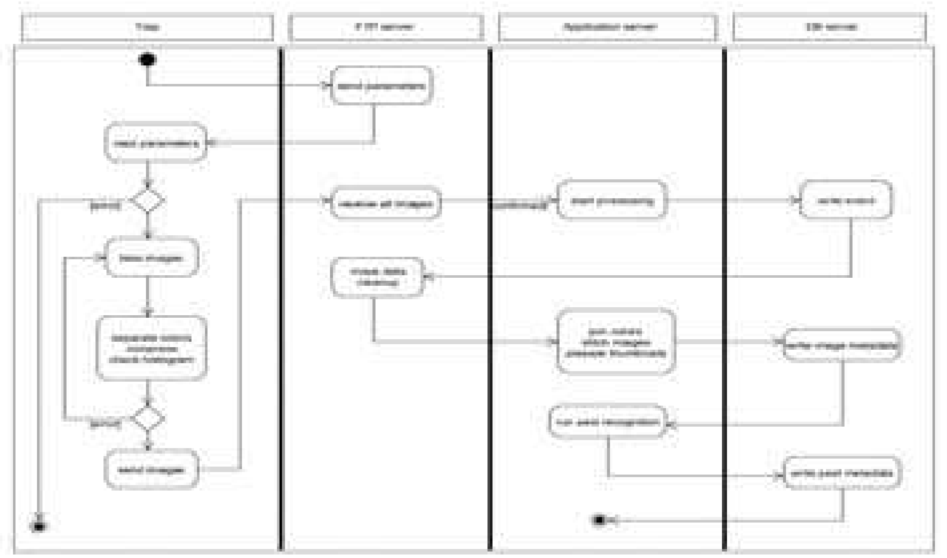

Figure 1: Activity diagram of data processing

also processed and written to central database in order to be used for statistical reporting and controlling of behavior of traps.

Based on experiences with traps being put in the field, it takes on average approximately 30-50 minutes for each complete data transfer over GPRS. This data is usually transferred during the night. When user requests data, response times under 3 seconds are expected. That is why all major image processing tasks are done when trap sends data to the server and not when user requests data, although such approach uses significantly more storage space.

4 Data organization

The following major parts are used on server side: image file storage (images sent by devices); database server (attribute data about system, traps and images); web server (web application and services for clients).

We expect that hardware requirements will grow exponentially with adding devices on fields, but currently this requirements are small, so we need to be agile and adapt accordingly. Our approach is to achieve reliability and performance with in house server clustering of database server and web server. That way we can keep hardware price components relatively low and still have a good performance and scalability of whole system. Good reliability is achieved by having redundant components in the cluster. The fact that we have a central database with attribute data which is relatively small and vast majority of data are images. They are relatively non-related (images taken in one trap do not influence images taken in another trap). This enabled us to organize data in a way that we can use lower cost storage devices (for images stored on file system) as well as lower capacity/cost main servers without compromising data security or response times.

The other tempting possibility is putting whole system in the cloud. In this case we would not have to deal with hardware and would be able to get necessary resources on demand (cloud elasticity). The system cost would be based on actual system usage and load. There are some issues with cloud offers at the moment – relevant to our solution – which should be considered: (a) actual costs are hard to estimate; there are lots of low level parameters which must be examined (like database interactions). While some of them can be estimated reliably, others are not, yet they could have significant impact on overall cost of the system; (b) cloud reliability is a problem [B11]; meaningful SLAs (Service-level agreements) in the cloud market space are rare and toothless [D10]; (c) remote backup is impossible due to large amount of data so cloud provider should do that. When both options were evaluated, non-cloud option proved to be more favorable.

5 Conclusions

Monitoring of insects is an important issue in food production process. Remote and real time monitoring of insects provides an important improvement especially in regard of health (less pesticide residues due to modern pest management techniques) and environment (reduction of field scouting) aspect [WKC10]. System presented in this article is an innovative product which allows such type of monitoring and running of such system requires solving of significant IT related issues which are hidden from end users. Since it is being constantly enhanced with additional functionality and there is more data every day, there will be new challenges to deal with in the future as well.

References

[B11] Blodget, H.: Amazon's Cloud Crash Disaster Permanently Destroyed Many Customers' Data, 2011, http://www.businessinsider.com/amazon-lost-data-2011-4.

[D10] Durkee, D.: Why Cloud Computing Will Never Be Free. Queue - Emulators Queue Homepage archive, Volume 8 Issue 4, April 2010.

[WKC10] Witzgall,P.; Kirsch, P.; Cork, A.: Sex pheromones and their impact on pest management, J. Chem. Ecol., 2010, 36:80-100.