Improving safety for automated forestry work

Whitepaper by the PENTA MANTIS VISION project (April 2024)

Contributions from: TNO, IDEAS, Adimec, Grass Valley
For questions or further information: contact us through MANTIS website[1]

Introduction

The PENTA MANTIS (iMAgiNg Technologies of Industry and Security) Vision project[1] takes on the challenge to bring image capturing, image quality and processing to the next level for broadcasting & high-end security/safety systems. In the past decade image capture and video-based applications have made major leaps with the transition from HDTV to 4k ultra HD (and beyond), the introduction of high dynamic range, and the broadening of the spectral range. The overall goal of the MANTIS Vision project is to further improve image capture systems for high-end security & safety and broadcast. By combining the latest advances in image capturing, quality & processing technology with focus on system cost reducing implementations, we strive to open paths for new value adding functionality while maintaining ease of use.

An example of potential future use is the MANTIS use case Automated Forestry Work. In this use case the knowledge developed within the MANTIS project from 4 of the partners (IDEAS, TNO, Adimec and Grass Valley) is exploited to a concept of automated person detection for the operators in forestry work.

Background

Forestry work has undergone multiple technological revolutions. Up until the beginning of the 20th century, forestry work was done manually with hand tools. The first revolution started with mechanization and the introduction of the first chainsaw and the beginning of the 20th century. Other steps of tree felling were later mechanized, leading to the wide adoption of "Cut-to-length" (CTL) machine. Later other downstream operations such as loading or extracting of beams were also mechanized, bringing large scale production. Forestry work is now entering the phase of automation. The operator that controls multiple semi-automated machines and takes decisions in a difficult and stressful environment has become the main production bottleneck.

Figure 1 A CTL and a loader at work[2].

The full automation of wood harvesting would allow to further increase production rates. In that context, different technical approaches can be selected to implement the tree felling, loading, and off-loading of trees using unmanned vehicles. One of the major limitations for a broader use of unmanned vehicles is the safety of humans operating in the same area as the automated vehicle. Safety is a major concern for forestry work. The

[1] https://project-mantis.eu/

[2] https://www.stuff.co.nz/science/102182695/robots-are-coming-to-nz-forests

operations are inherently associated with the manipulation of dangerous equipment, such as a chainsaw, in an unsafe environment. This combination of parameters results in a high injury rate, which according to the British Health and Safety Executive is higher than the construction industry[3]. Every year, around 140 people are injured in relation to forestry work. To increase the acceptance of autonomous vehicles for forestry work, those need to ensure that their operations are safe for any person present in the field.

Such as for other autonomous systems that are operating safely together with humans, the situational awareness (SA) around the vehicle is of primary importance. To operate safely, the vehicle must be aware of the different people present near it. With that information the vehicle can then act accordingly, for instance by driving around and avoiding people or by switching off any dangerous tools on board. Depending on the types of objects of interest, multiple sensors and processing strategies can be used and combined to bring a full overview of the relevant area. Cameras are widely integrated in SA systems. Aside from a high angular resolution (when compared to active sensors such as RADAR and LiDAR), they also provide rich contextual information. The use of cameras as a sensor of reference of automated detection of persons carries clear synergies with the PENTA-Mantis project which purpose is to improve image capture systems for high-end security and broadcast, while lowering cost, opening the path for new functionality, and maintaining ease of use.

The IDEAS LWIR sensor that is being developed in the MANTIS project can be of specific additional value for the Forestry case as this sensor is able to spot humans in situations of low light like shade of trees (Figure 2) and degraded atmospheric conditions like fog which are typical conditions for especial Norther climate Forestry.

Figure 2 Loss of visibility of humans in the shades of the trees with VIS (left) but clear visibility with LWIR (right).

Requirements

IDEAS consulted various industrial actors to collect their input and establish a first set of requirements for the system. The goal is to design a system capable of performing forestry tasks in a Nordic climate (mild and cold). The system will obviously operate in a forest area. The background mostly consists of vegetation. The objects of interest, the people present in the area are expected to perform different tasks with different body poses (standing, kneeling, bending, etc..) as shown in Figure 3.

Figure 3 Variation of body poses encountered during forestry work.

People are also expected to wear different sorts of warm work clothes. The color of the clothes is often green and may not be clearly distinguishable from the vegetation. Finally, the system should be able to detect a person at 90-meter distance with a false positive rate of 0.1%. The main requirements for the detection chain are summarized in Table 1.

[3] https://www.hse.gov.uk/treework/treework-incidents.htm

Table 1 Typical requirements for the detection chain

Maximum range	90 m
Horizontal field of view	360°
Vertical field of view	60°

The challenge is to detect and track a human at large distance. This is non-trivial because the image of the human covers only very few pixels in the overall image, possibly in a cluttered background. The image size is equal to

$$\text{Object Size} \times \left(\frac{1}{1 + \left| \frac{\text{Object Distance}}{\text{Focal Length}} \right|} \right)$$

This means that for the case of focal length << object distance, one finds:

$$\text{Image Size} \approx \text{Object Size} \times \left| \frac{\text{Focal Length}}{\text{Object Distance}} \right|$$

Table 2 Image size computed for 2-meter objects at different distances with 6.2-mm focal length.

Object distance	Image size for 2-m object size	Comments
-100 m	0.124 mm	Our image sensor has 12-µm pixel pitch. The image size of 124µm covers only 10 pixels. Detection is challenging, especially with vibrations.
-10 m	1.240 mm	Ok
-1 m	12.400 mm	Our image sensor has VGA resolution (640 x 480 pixels). The 12.4-mm image size exceeds the image sensor size of 640 * 12µm.

System design

IDEAS and TNO collaboratively worked on defining a vision-based system capable of fulfilling the above-mentioned requirements. Because the system should be able to perform under low light level often encountered during winter in a Nordic climate and that the system should be able to reliably detect persons even if the contrast with the background is low, the choice was made to architecture the system around an infrared sensor. Human bodies emit radiation in the long wave infrared (LWIR) that can be recorded with the appropriate sensor. IDEAS is developing a LWIR camera (Thermal IR) capable of recording the heat signals. The camera not being available yet, IDEAS, together with TNO designed an alternative setup to field test the main parameters of the final system. The prototype will be placed in TNO's tower to image a forested area with people passing. The parameters of the two systems are compared in Table 3.

Table 3 Main parameters of the two systems

Sensor	FLIR A615	IDEAS LWIR core
Array size	640 x 480 pix	640 x 480 pix
Waveband	7 - 13 µm	8 – 14 µm
Pixel pitch	17 µm	12 µm
Frame rate	50 Hz	60 Hz
NETD	0.05 K	0.17 K
IETD	0.02 K	0.07 K

Optics	FLIR A615	TBD6.2mm
Focal length	41.3 mm	6.2 mm
F-number	1	1
Aperture diameter	41.3 mm	6.2 mm
Transmission	-	0.9
IFOV	0.41 mrad	1.94 mrad
FOV (H x V)	15° x 11°	64° x 50°

Using the analytical Triangle Orientation Discrimination) (TOD) model[4], TNO calculated the TOD curve for the foreseen IDEAS LWIR Core camera (red curve in Figure 4). The TOD curve is a relationship between the inverse size of a triangle test pattern (S^{-1} in $mrad^{-1}$) and the minimum temperature difference ΔT (in K) to be able to discriminate the orientation of this triangle. The curve is a system performance property that can be used to calculate Detection, Recognition, and Identification (DRI) ranges for real objects under operational circumstances. For example, for humans the recommended characteristic size and thermal contrast are 0.75m and 3K, and the conversion factors for Detection and Recognition are 1.4 and 5.8 respectively. Detection and Recognition are defined here as discriminating human activities, which will be more difficult tasks than discriminating a human from an animal or an object in the scene (see for example Figure 6). These factors result in a Detection and Recognition range of 166 m and 40 m, respectively. So, for a human observer, the predicted performance is in the right range. Automatic DRI with a state-of-the-art deep learning algorithm may be more challenging. This will be tested using a LWIR camera setup that is installed on the TNO tower.

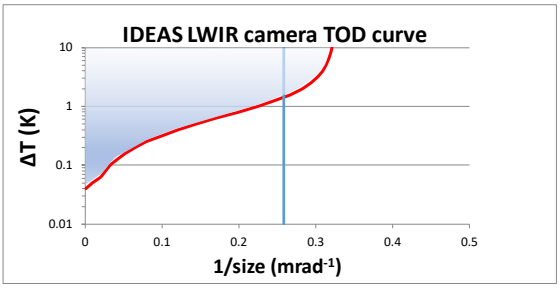

Figure 4 TOD curve (red curve) for the IDEAS LWIR core from Table 3 with the human-in-the-loop. The blue vertical line indicates the Nyquist frequency of the camera.

Once the IDEAS LWIR configuration was determined, TNO derived the working distance at which the TNO tower LWIR camera configuration would perform similarly as the IDEAS system. The 90-meter requirement would convert to a working distance of 420 meter for the TNO tower LWIR camera configuration. The working distance is defined as the distance between the camera and the object. This working distance is just within the measurement range of the tower as shown in Figure 5.

Figure 5 The TNO tower test location.

The system was installed on the tower and oriented towards the forest such that the zone between 200 and 400 meter is present in the field of view (Figure 6).

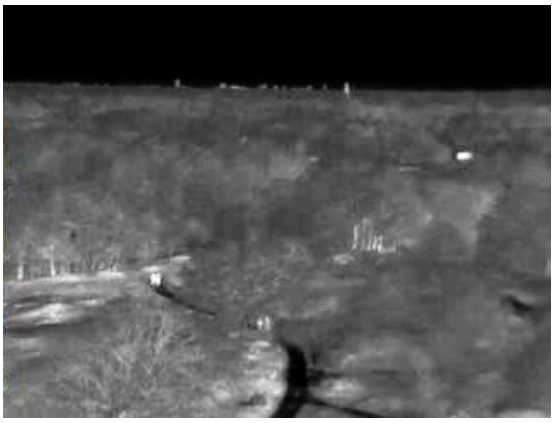

Figure 6 LWIR view of the scene. One can recognize people (blobs) walking on the cold path (black track).

Object detection and range estimation

With the system design in place, one of the first question that researchers want to answer is “Does the system perform as it should?” Or can a human be automatically detected and tracked at a 90-meter distance from the camera under all atmospheric and seasonal conditions? This is non-trivial because the human appears in only very few pixels in the image, and this is challenging for current state

[4] Bijl, P., Hogervorst, M.A. & Toet, A. (2017). Progress in sensor performance testing, modeling and range prediction using the TOD method: an overview. Proc. SPIE 10178, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXVIII, 101780U (May 3, 2017); doi:10.1117/12.2266788.

of the art Deep Learning algorithms. This is of especial importance as the thermal contrast of the human will vary with environmental conditions like temperature (daily/ seasonal related) and atmospheric conditions. The preliminary computations made during the design phase are intended to ensure that the prototype performance reflects the final system performance. However, it does not guarantee that the result produced by the detection pipeline is of sufficient quality to fulfil the requirements. To answer that question, TNO adapted the setup to record a dataset capturing the seasonal variations of the scene. The goal is to collect enough data to be able to validate the performance of the automated person detection pipeline. By regularly capturing short footages, one has the chance to capture different weather conditions (wind, snow, rain, haze, etc.), lighting conditions (low illumination, diffuse, reflections, low elevation, lens flare, etc.), different seasons and vegetation conditions (over a year) and different person with different clothing and doing various activities (walking, running, sitting, et cetera). Given the fixed position of the camera, the background cannot be varied. The collection of this dataset was primarily intended for the validation of the deep learning algorithm and not to train the algorithm.

To maximize the amount of collected data and ease the selection process, TNO automated the data collection. First, the setup was extended by integrating two visual (VIS) cameras (MicroCam TMX55) provided by Adimec. The VIS camera was setup to record synchronously with the LWIR camera. Then the different sensors were optically aligned regarding to each other. Once aligned, coordinates measured in one of the sensor frames could be converted to coordinates in other sensor frames. Then an automatic annotation was done on the high-resolution video and transferred to the other sensors. The automatic annotation relies on the small object detection developed by TNO within the Mantis project. When the recording is made and annotated, the system also logs weather data (rain intensity, haze, wind, etc.) corresponding to the time at which the recordings are made. Finally, a manual step is performed to control the quality of the annotations. In case needed, the annotations are manually adapted. With that setup in place, a wide dataset is being collected and used for the validation of the person detection. An illustration of the diversity of the recordings is shown in Figure 7.

Figure 7 Seasonal variations of the scene. From top to bottom, March, April, and May.

To decide whether the person enters the prohibited zone, one needs to estimate the distance between the vehicle and the person. The goal is not to precisely measure the distance between the person and the vehicle but to decide whether the machine should stop or not. That requirement means that the system should provide a distance map such that every pixel is labelled with a depth. Laser devices such as laser scanners or range finder will produce a sparse depth map, but their measurements will be accurate. With the development of deep learning, approaches designed to build a depth map based on a single camera position have emerged. They produce a relative depth map that can be used in combination with other sparse measurements to create an absolute depth map. TNO and Grass Valley have been evaluating the benefit of this innovative technology for different applications in both broadcast and security domain. For this specific use case, TNO applied a monocular depth estimation technique to derive a relative depth map from the images recorded from the tower. Figure 8 shows that depth map corresponding to the scene visible from the tower. The color shown on the map goes from yellow (close) to purple / black (far/infinity). One can observe the correct relative position of the different trees. The experiment shows that monocular depth estimation could be used to decide if the person is too close to continue operating.

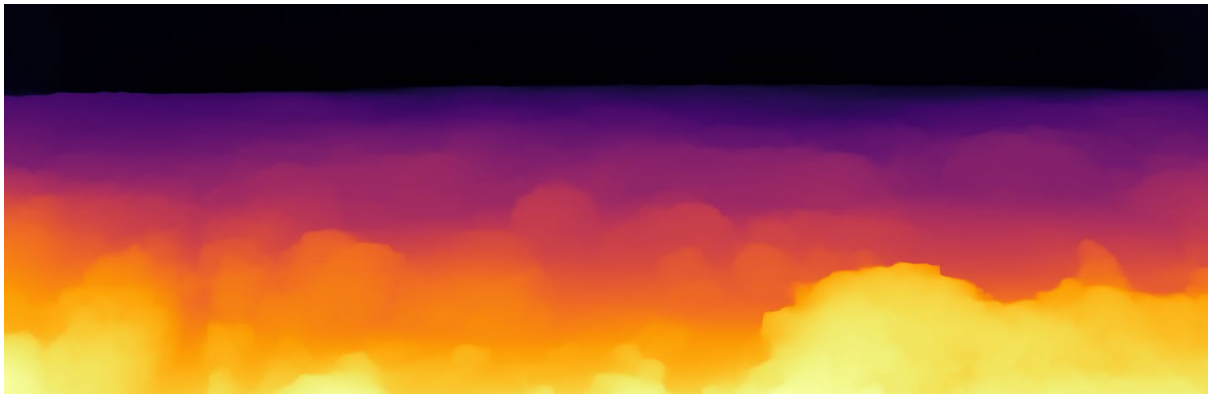

Figure 8 Depth map extracted using monocular depth estimation.

---

iMAgiNg Technologies of Industry and Security Vision (MANTIS Vision) takes on the challenge to bring image capturing, image quality and processing to the next level for broadcasting & high-end security/safety systems.

A project under the PENTA (20005 MANTIS Vision).
https://project-mantis.eu/

Partnering companies, Universities and RTOs:

Netherlands: Adimec, TU Delft, Grass Valley, TNO

Belgium: ams OSRAM

Norway: ideas
