example, in heavy rain/snowfall.

One shortcoming of research performed on people detection is that the case of a fainted person (person lying flat down on the ground) is typically not considered, many authors only consider persons standing upright [1].

To detect people using only visual information are commonly done by utilizing visual features in combinations with with machine learning techniques [4]. In principle, the problem of people detection and object detection using cameras is very similar. However, to list a complete survey of one of the most active topics in computer vision is beyond the scope of this paper. A survey about pedestrian detection using a single camera is presented by Enzweilier et al. [5].

Worth mentioning and also related to our work is the use of reflectivity values from other active sensors, which emit signals such as light. For example, the state-of-the-shelf navigation system for AGVs utilize reflectors as artificial landmarks in the environment. Most laser scanners provide reflectivity values in addition to the range and bearing values which could be used to detect reflective areas. Essentially the same holds for time-of-flight (TOF) cameras which use active illumination to determine range data and also return intensity values. The key difference here is to combine the camera images with and without the emitted light. The most similar case to the previous mentioned sensors above would be to only utilize data from the image taken with the flash (and not data without the flash).

One notable difference to systems which are used to assist drivers in cars is the field of view the sensor has to cover. A loading / unloading scenario frequently comprises sharp turns, reversing etc. compared to driver assistance systems for cars, which need to observe a relatively narrow cone in front of the car, a crucial requirement for our application scenario is a very large field of view (FOV). This imposes that the system should be able to detect people with a low resolution.

III. Detection System

The system works by comparing two images, one taken with the flash I_f and one taken without I_{nf}. Since the system is mounted on a moving vehicle straightforward background subtraction methods are not applicable here. Instead small interesting sub regions in the images, called keypoints, are used in order to relate the two images I_f and I_{nf} to each other. The method of extracting keypoints is selected based on its ability to detect reflective areas, i.e., local peaks in the intensity values. Each keypoint, please note that keypoints are only extracted from the I_f image, is then tracked in the image without flash I_{nf}. Since I_n and I_{nf} are taken in short succession, the displacement of the tracked keypoints between the two frames should be relatively small, even if during a quick turn or a bumpy ride. Please note that the processing sees the two images as a pair. Based on how the keypoints can be tracked, the system utilizes a window around the keypoint to determine if the intensity change is above a predetermined value. If so the system reports this area to contain reflective material. An overview

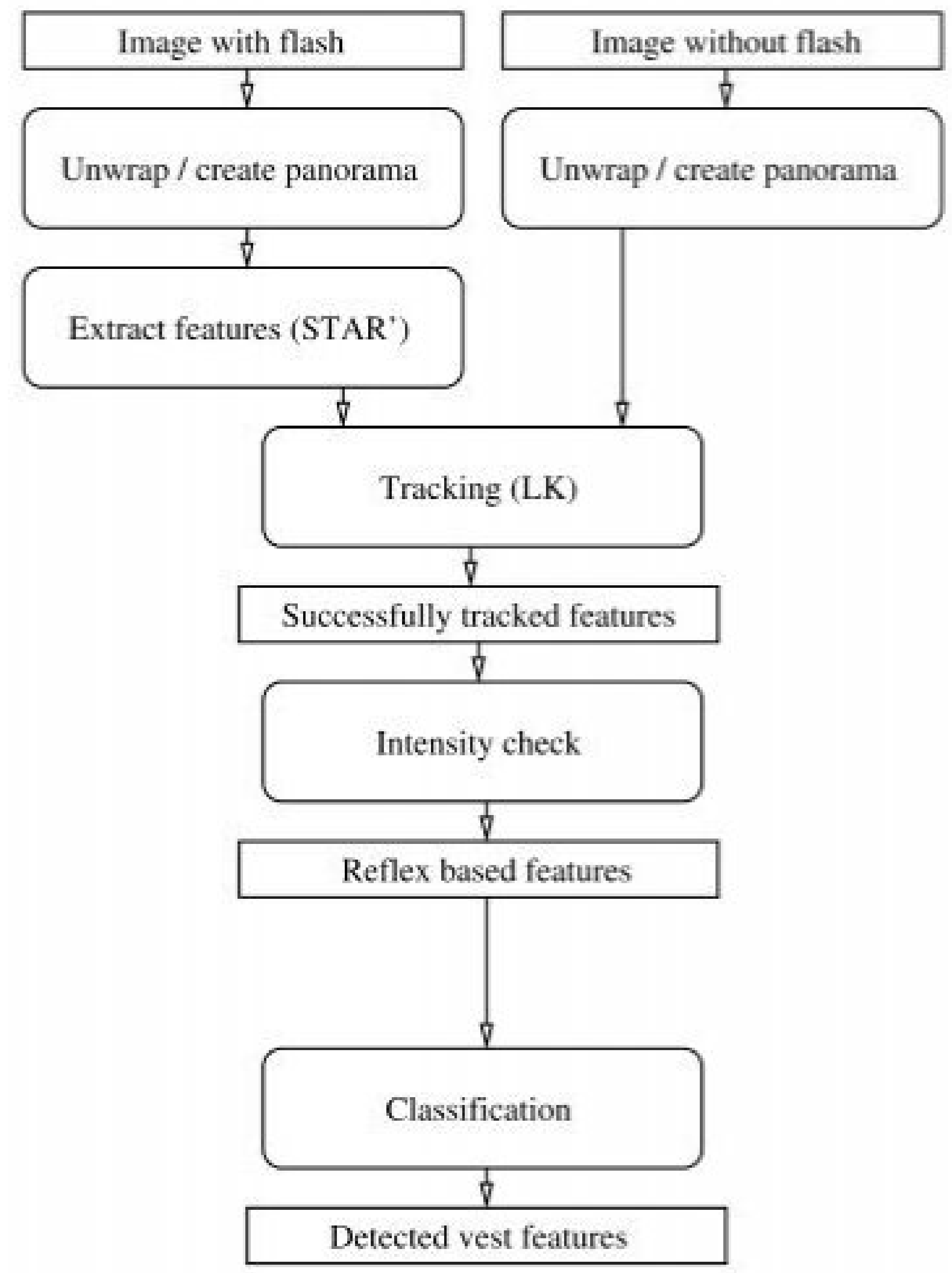

Fig. 2. An overview of the data flow within the system. A squared boxes indicates the data type and a box with rounded corners a method.

diagram can be seen in Fig. 2 and an example detection result can be seen in Fig. 3. One important criterion is to perform the detection at "real-time", which for computer vision applications sometimes is referred to be at least 10 Hz [6].

A. Hardware

The system consists of a standard of-the-shelf monochrome CMOS sensor, IDS imaging USB UI-1228LE, with a resolution of 752 $\times$ 480 pixels. A band pass filter with a center wavelength of 852 nm with a full width-half max of 10 nm is mounted between the lens and the sensor. The lens is a fish-eye type with an approximate FOV of 180 degrees. Next to the camera 8 IR LEDs with a wavelength of 850 nm are placed around the camera, see Fig. 4. The emission characteristic of the LEDs is such that the emission reaches its maximum in the direction normal to the LED and falls down to 50% at an angle of 60 degrees to normal direction. This assures a wide and relatively uniform illumination coverage of the camera's FOV. The LEDs are mounted in a way that directs weaker illumination to the top and bottom of the image. These directions are less important for people detection since they typically correspond to partial views of the vehicle itself (when looking downward) and the sky (when looking upward). Further on, the camera images are rectified to only contain a panorama where the top and bottom parts of the image are removed.

One difficulty of having a large FOV is that object size decreases very rapidly in the image coordinates with larger distance, i.e. only a small amount of pixels might represent the object to be detected. For cameras the width x given