6

Sensor Fusion

In the previous chapters, we have dealt exclusively with the part of attention that is concerned with visual processing. This part is the best investigated one in human behavior, probably because vision is the sense using the most capacity in the human brain: the 32 representations of the retina occupy more than half of the whole cortex [Kandel et al., 1996] and the primary visual cortex V1 has the richest architecture of all cortical areas [Zeki, 1993]. Usually, computational attention systems simulate also only visual attention. One exception is the model of [van der Willigen and von Campenhausen, 2002] which models *audio-evoked orienting* — the orienting behavior in which eyes (and head) are turned to an unexpected sound — with an artificial neural network.

However, human eye movements are not only biased by vision but also by other senses, e.g., the gaze may be directed into the direction of a sound, a smell, or even a touch, [Watanabe and Shimojo, 2005] and the fusion of different cues competing for attention is an essential part of human attention. In robotics, attentional mechanisms might also profit from additional sensor modalities since they yield a richer set of data that enable the detection of more object properties, resulting in more useful foci of attention.

This chapter presents an extension of the attention system VOCUS which enables the fusion of saliencies from different sensor modes: the Bimodal, Laser-based Attention System (BILAS). This allows the detection of different object properties and the detection of a wider variety of saliencies than within a single sensor mode. The modes provided to the attention system are depth and reflection data acquired by a 3D laser scanner in a single scan pass. BILAS takes the data from both laser modes as input and searches both modes for saliencies according to principles described in chapter 4: saliencies of different features, here intensity and orientation, are computed in parallel and fused into one global saliency map on which a single FOA is determined. Most of this chapter was also published in [Frintrop et al., 2005c].

We apply BILAS to laser data of real-world indoor and outdoor scenes and elaborate on the different advantages of range and reflectance values. We

S. Frintrop: VOCUS: A Visual Attention System..., LNAI 3899, pp. 129-147, 2006.

show that these data modes complement each other: contrasts in range and in intensity need not necessarily correspond for one scene element, i.e., an object of similar texture as its background may not be detected in the reflection image, but in the range data. On the other hand, a flat object — e.g. a poster on a wall or a letter on a desk — that could be distinguished in the reflection image, will likely not be detected in the range data. The results indicate that the combination of different modes enables considering a larger variety of object properties. Additionally, we compare the performance of attentional mechanisms on laser data with that on camera data. The comparison reveals the respective advantages of the two kinds of sensors.

Typically, computational models of visual attention use features like intensity, color, and orientation. Depth is rarely considered although it plays a special role in deploying attention. It is not clear from the literature whether depth is simply a feature, like color or motion, or something else (cf. chapter 3.2.2). Definitely, depth is an important feature in human vision; in particular, range discontinuities at the borders of many objects can help to separate objects from each other and from their background and to compute object shapes.

Two approaches that include depth are presented in [Backer and Mertsching, 2000] and [Maki et al., 2000]. They obtain depth data from stereo vision and regard it as another feature. The data obtained from stereo vision is usually not very accurate and contains large regions without depth information. This may justify the integration of the depth values as a feature in the above mentioned models; in our approach the range data come from a special sensor and yield dense and accurate range information, so we regard depth as an additional sensor mode.

The remainder of this chapter is structured as follows: we start in section 6.1 with a description of the data acquisition including a specification of the bimodal 3D laser scanner. In section 6.2, we continue with introducing the extended attention system BILAS. The main part of this chapter are the experimental results in section 6.3 investigating in detail the respective advantages of the two laser modes and of camera data. We finish with a discussion on the presented approach.

6.1 Data Acquisition

The data for the experiments of this chapter were acquired with the AIS 3D Laser Scanner which will be introduced in section 6.1.1. It yields range and reflectance data that are rendered into images (section 6.1.2). In section 6.1.3, we discuss the differences of range data obtained from laser scanners and from stereo vision.

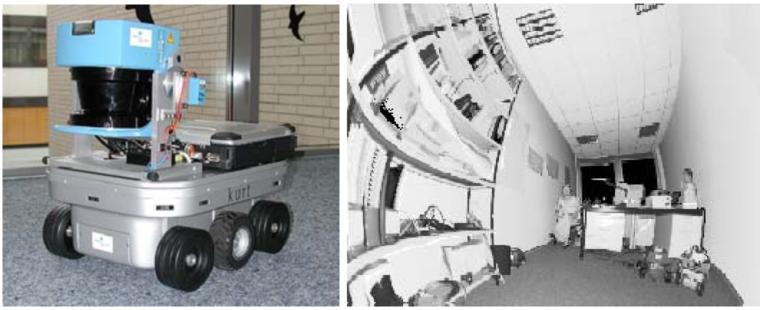

Fig. 6.1. Left: the custom 3D range finder mounted on top of the mobile robot Kurt3D. Right: an office scene imaged with the 3D scanner in reflection value mode, medium resolution (361 × 211 pixels, distortions not corrected)

6.1.1 The 3D Laser Scanner

For the data acquisition in our experiments, we used a custom 3D laser range finder which is mounted on the mobile robot Kurt3D (Fig. 6.1, left). The scanner is based on a commercial SICK 2D laser range finder. In [Surmann et al., 2001], the custom scanner setup is described in detail. The paper also describes reconstruction algorithms and their use for robot applications. Here, we provide only a brief overview of the device.

The scanner works according to the time-of-flight principle: it sends out a laser beam and measures the returning reflected light. This yields two kinds of data: the time the laser beam needs to come back gives the distance of the scanned object (range data) and the intensity of the reflected light provides information about the reflection properties of the object (reflection data). This reflectance measurement is the result of the light measurement by the receiver diode. It measures the amount of infrared light that is returned from the object to the scanner and thus describes the surface properties concerning non-human visible light.

The 2D scanner serially sends out laser beams in one horizontal slice using a rotating mirror (LIDAR: LIght Detection And Ranging). It is very fast and precise: the processing time is about 13 ms for a 180° scan with 181 measurements and the typical range error is about 1 cm. A 3D scan is performed by step-rotating the 2D scanner around a horizontal axis, i.e., the 3D scan is obtained by scanning one horizontal slice after the other. Usually, the area of $180°(h) \times 120°(v)$ is scanned in 1°, 0.5°, or 0.25° steps resulting in the resolutions (181, 361, 721 pts) horizontal and (121, 241, 481 pts) vertical. By restricting the scan area to more narrow angles or by ignoring values at the borders, other resolutions may result. In the experiments in section 6.3, we used resolutions of 152×256 and 361×211.

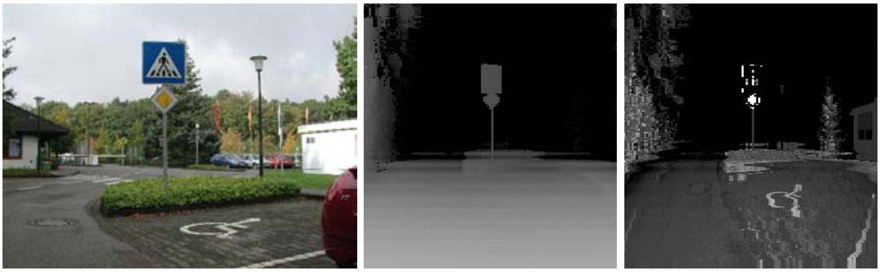

Fig. 6.2. Visualized laser data. Left: scene from camera image, middle: visualized depth data, right: visualized reflection data. Depending on the sensor, the presented images have slightly different extensions, the laser scanner getting a wider angle than the camera in all directions

6.1.2 Rendering Images from Laser Data

The scanner is able to operate in two data modes. In the default mode, it returns only the range data in a predefined resolution. In an alternative mode, it is able to yield the range as well as the reflection data in a single scan pass. The reflection data can directly be converted into a gray scale intensity image as is depicted in Fig. 6.1, right. Here, it shows that the raw data from the scanner is spherically distorted. The distortion was removed in later experiments by rectifying the images as can be seen, e.g., in Fig. 7.12. The visualization of the depth values from the range data requires some transformation. The basic approach is to interpret the depth values as intensity values, representing small depth values as bright intensity values and large depth values as dark ones. Since close objects are considered more important for robot applications, we introduce an additional double proximity bias. Firstly, we consider only objects within a radius $r = 10\,m$ of the robot's location. Secondly, we code the depth values by using their square roots, so pixel p computes from depth value d by:

$$p = \begin{cases} I - (\sqrt{d/max} * I) & : \quad d \leq max \\ 0 & : \quad d > max, \end{cases} \tag{6.1}$$

with the maximal intensity value I and the maximal distance $max = 1000\,cm$. This measure leads to a finer distinction of range discontinuities in the vicinity of the robot and works better than a linear function. If the robot works outdoors and distant objects should be detected, the maximal distance can be increased. Fig. 6.2 shows an example of the visualized laser data.

Since the data from the different sensor modalities result from the same measurement, we know exactly which reflection value belongs to which range value. There is no need to establish correspondences and to perform costly calibration by complex algorithms. The laser data are illumination independent, i.e., the data is the same in sunshine as in complete darkness and no reflection artifacts by external light occur. This yields a robust approach that enables all day operation.

6.1.3 Laser Data Versus Stereo Vision

In current attention systems integrating depth information, the range data is usually extracted from stereo vision. With today's available computing power and advanced stereo algorithms, even real-time stereo vision at frame rate is possible. A 3D scan pass (between 1.2 and 15 seconds, with typically 7.5 s) is slow as compared to the frame rates of CCD cameras. However, for several target applications, for example automatic 3D map building, high frame rates are not needed. In this application, 3D laser range scanning has some considerable advantages over 3D stereo reconstruction.

Firstly, range scanning yields very dense depth information. On the other hand, most 3D stereo vision algorithms rely on matching grey level values for finding pixel correspondences. This is often not possible since, first, correspondences can only be found in textured parts of the stereo images, so large image regions yield no depth data at all; second, ambiguous grey values that cannot be disambiguated result in false matches and, third, shading may prevent finding matches. Hence, the generated depth maps are sparse, often containing large regions without depth information.

Secondly, the precision of the depth measurement of a laser range scanner relies only on the tolerance that its construction foresees. Industry standard scanners like the SICK scanner that we use have an average depth (Z axis) error of 1 cm. The precision error of the Z axis measurement in 3D stereo reconstruction is dependent on a number of parameters, namely the width of the stereo base, the focal lengths of the lenses, the physical width of the CCD pixel, the object distance and the precision of the matching algorithm. The error increases by increased squared object distance, and decreases with increasing focal length (narrowing the field of view). For small robots like Kurt3D, the width of the stereo base is limited to small values ($\leq$ 20 cm), resulting in a typical Z axis error of about 78 cm for objects at a ranging distance of 8 m ($error = d * (d * w)/(b * f)$ with distance $d = 8000\ mm$, pixel width $w = 0,0098\ mm$, stereo base $b = 200\ mm$, $f = 4\ mm$, precision 1 pixel).

And finally, our 3D laser scanner provides a very large field of view and the data of the laser scanner are illumination independent. This enables all-day operation and yields robust data. The named strengths make the 3D laser scanner the sensor of choice in this application. An alternative may be 3D cameras which are about to enter the market. The bimodal attention system can equally be applied to their data as will be briefly discussed in section 6.4.

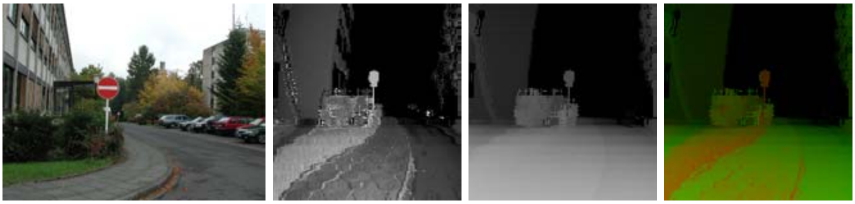

Fig. 6.3. Combining depth (2nd) and reflection (3rd) image into one colorized image (right). Range is coded as intensity, reflection as red-green transition

6.2 The Bimodal, Laser-Based Attention System BILAS

The first plan to build a system of visual attention able to process several sensor modes came from the idea to apply attentional mechanisms on data from a 3D laser scanner. This was a promising idea since the sensor yields dense and precise data and the availability of range and reflection data let us expect the possibility to detect new kinds of saliency. In section 6.3.1 we show that these expectations were fulfilled.

In first experiments, we applied the bottom-up system of visual attention — at that time the NVT [Itti et al., 1998] since our system didn't yet exist — to each sensor mode image separately. This enabled the investigation of saliencies in laser data and the comparison of the complementary effect of the modes. Nevertheless, it yielded two foci of attention for a single scene instead of one. It was suggesting to combine the results from both sensor modes to yield a single focus of attention especially since the data points directly correspond. Unfortunately, this was not possible with the NVT since this system is only able to process one input image at a time.

To overcome this problem we used a workaround in a first approach (see also [Frintrop et al., 2003b]): the laser data is gray-scale so the color feature channel in the NVT was not used. Utilizing this fact, we fused range and reflection image into one colorized image. To accomplish this, the range data were treated as intensity values of the new input image and the reflection values were coded as color (hue) information. High reflection values were coded in red hues, low ones in greens. This resulted in suitable color images because the color feature computations in the NVT take into account blue-yellow contrasts as well as red-green contrasts. An example scene with range and reflection image as well as the combined colorized image is depicted in Fig. 6.3. This colorized image was fed into the attention system, which computed a single focus of attention based on range and reflection data. In Fig. 6.4 we present this approach.

Although working quite well in our experiments, there were some problems with this approach. First, the approach is restricted to the processing of gray-scale images; the fusion of color images is not possible. Also the extension to more input images is difficult. A third gray-scale image might be coded as blue-yellow transition, but it is questionable whether the processing of

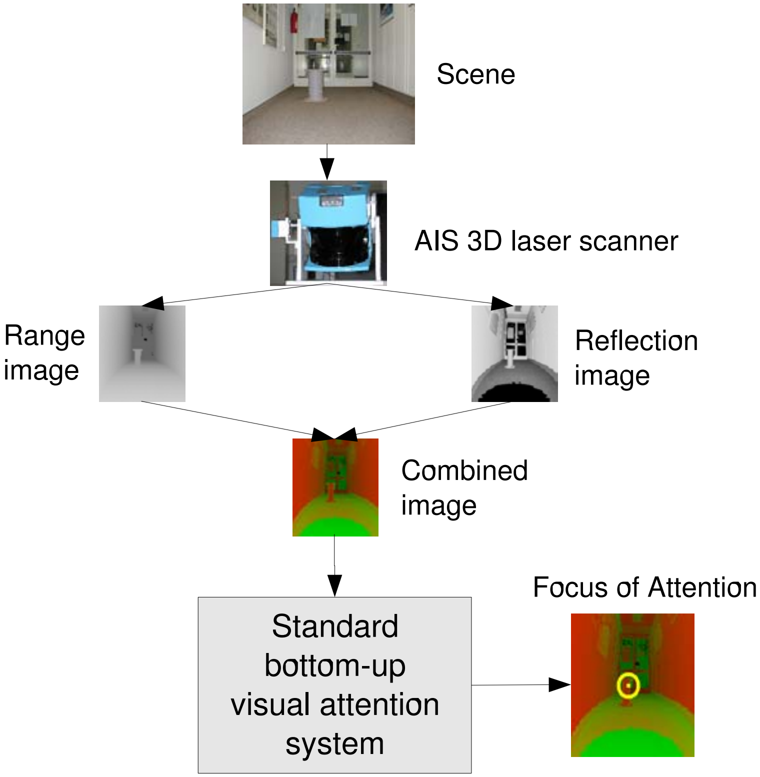

Fig. 6.4. First approach to compute a single focus of attention from range and reflection data: data from both modes is combined into a colorized image by coding range as intensity and reflection as color. On this image, a single focus of attention is computed by a bottom-up attention system (here the NVT [Itti et al., 1998]). A better solution is the new system BILAS which is shown in the following figures

blue-yellow and red-green is independent in the NVT. More than three input images could definitely not be processed with this approach. Second, since in the NVT the computation of the orientation maps works only on the gray-scale data, no orientations are computed for the reflection values. And finally, a new system that computes the saliencies for each mode separately is not only more intuitive but enables also the direct inspection of depth or reflectance saliencies as well as their tuning by top-down mechanisms. These thoughts were the first cause to build an own attention system that is able to process several modes. The single-mode version of the system was introduced in chapter 4, here we show the extension of the system to two modes: the Bimodal Laser-Based Attention System (BILAS) (see also [Frintrop et al., 2005c]).

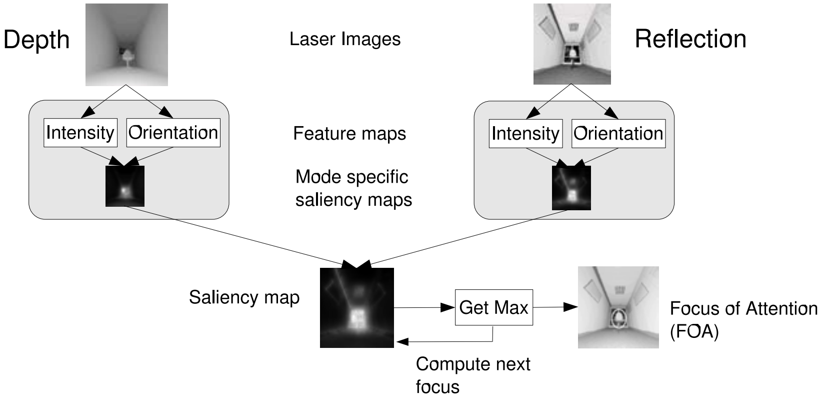

Fig. 6.5. Overview of the Bimodal Laser-Based Attention System (BILAS). The images from the two laser modes "depth" and "reflection" are computed independently. Saliencies according to intensity and orientations are determined and fused into a mode-specific saliency map. After combining both of those maps, the focus of attention is directed to the most salient region. A more detailed figure is shown in Fig. 6.6

BILAS computes regions of interest in the depth and reflection data independently and finally fuses their saliencies yielding a single focus of attention. In Fig. 6.5, we show an overview of this system, in Fig. 6.6 the system is shown in more detail. Since the laser scanner provides only gray-scale data, no color feature is computed and the processing is restricted to intensity and orientation. Notice that depth is not a feature in our approach but a separate sensor mode. Generally, also other sensor modalities may be regarded: all sensor data that are representable in a 2D map might be used as input to the system.

Base of the system is the bottom-up part of VOCUS (chapter 4). First, the images from each mode of the laser scanner are processed independently, i.e., intensities and orientations are computed for the depth as well as for the reflection image. These computations take place as described in chapter 4: the feature maps are computed with center-surround mechanisms and Gabor filters, the maps are weighted according to the uniqueness of the features, they are summed up to conspicuity maps and normalized. The conspicuity maps are weighted again and summed up to a mode-specific saliency map which contains the saliencies according to the specific sensor mode. Finally, the saliencies of each mode are weighted again and fused into a global saliency map.

The fusion of two different kinds of data allows to exploit the respective advantages of both modes: saliencies in one mode correspond not necessarily to saliencies in the other mode. Therefore, a larger variety of object properties is considered and it is possible to detect a pop-out — e.g., in depth — that

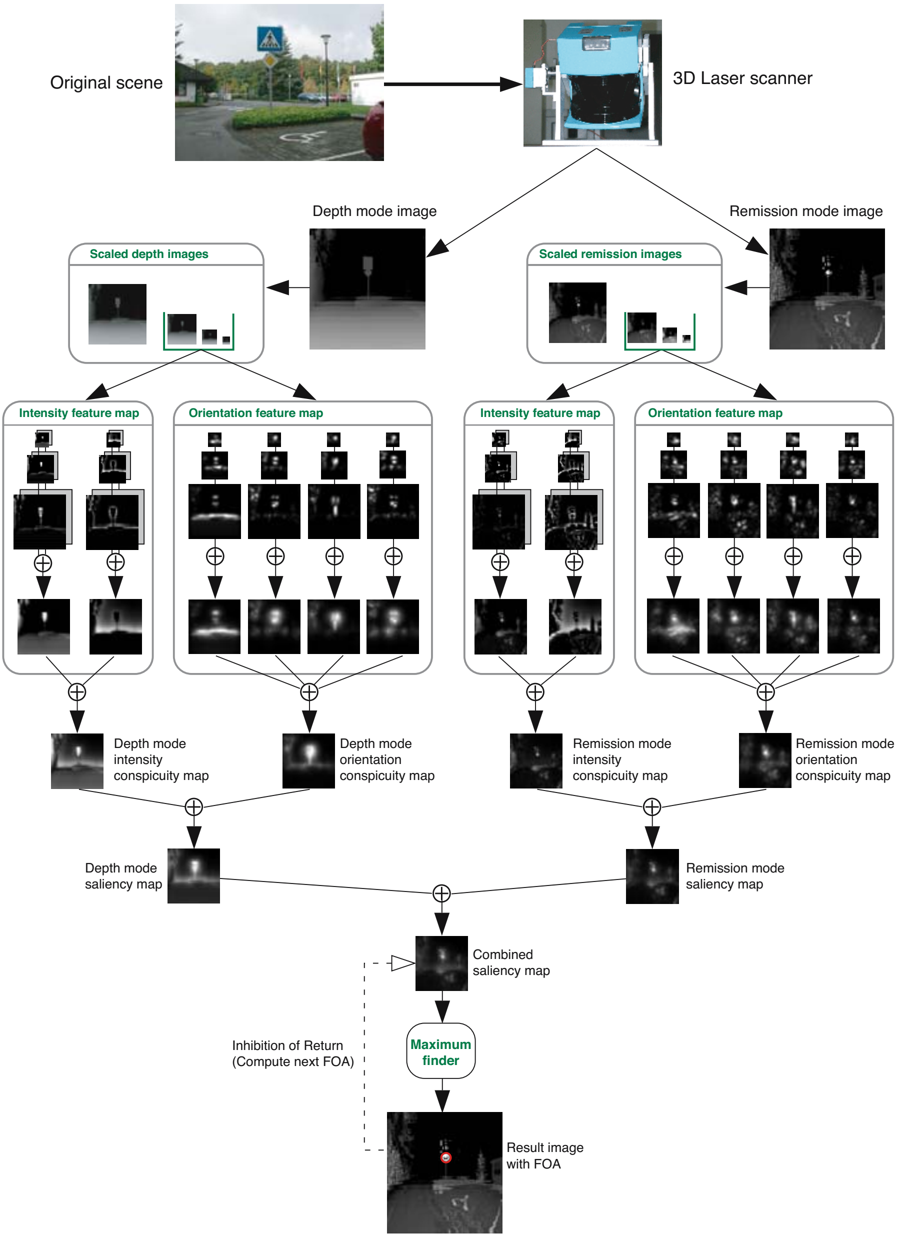

Fig. 6.6. The Bimodal Laser-Based Attention System (BILAS) in detail. The images from the two laser modes "depth" and "reflectance" are computed independently. Saliencies according to intensity and orientations are determined and fused into a mode-specific saliency map. After combining both of those maps, the focus of attention (FOA) is directed to the most salient region (shown as red ellipse)

would be missed otherwise. The saliencies of both modes compete with each other and the focus of attention is directed to the strongest cue.

Note that we do not claim that one sensor mode is better than the other or that laser is better than camera data. Each mode has its advantages and only the combination allows to use all of them.

6.3 Experiments and Results

We have tested our approach on scans of both indoor and outdoor scenes. The laser scans were taken at two different resolutions: 152 × 256 and 361 × 211 data points. From these points, images of sizes 244 × 256 and 288 × 211 were generated. The pixel dimensions do not match exactly the number of data points, since some of the border pixels in horizontal direction are ignored due to distortion effects and in the lower resolution mode the pixels in the horizontal direction were duplicated to yield adequately dimensioned images. The lower resolution proved to be sufficient for the application of attentional mechanisms. The computations of the first focus on both laser images took 230 ms on a Pentium IV with 2400 MHz. The computation of further foci was determined nearly at once (less than 10 ms).

The camera images depicted in this section represent the same scenes as the laser scans to facilitate the scene recognition for the reader and to enable comparison between the sensor modalities. It has to be remarked that camera and laser images do not show identical parts of the scene, since the apex angles and their fields of view are different.

In this section, we focus on three aspects. Firstly, we show the general performance of attentional mechanisms on laser data (section 6.3.1). Secondly, the different qualities of the two laser modes are shown (section 6.3.2), and finally, we compare the performance of attentional mechanisms on laser images with those on corresponding camera images (section 6.3.3).

6.3.1 Regions of Interest in Laser Data

Here, we briefly demonstrate the general performance of attentional mechanisms on laser data to indicate that it makes sense to determine salient regions in laser data with an attention system since the regions are of potential interest in robotic applications. Fig. 6.7 shows four scenes, a camera image as reference on the left and the laser image combined from both laser modes on the right.

In the first three laser images, the FOAs point to objects that also a human observer would consider as salient: a traffic sign, two flower pots and a statue with flowers. These objects are focused because they are highly salient in laser images: the traffic sign has strong reflection properties that yield high saliencies in the reflection image. Furthermore, it pops out in depth and shows a vertical orientation (cf. the maps in Fig. 6.5). Similar effects are true for

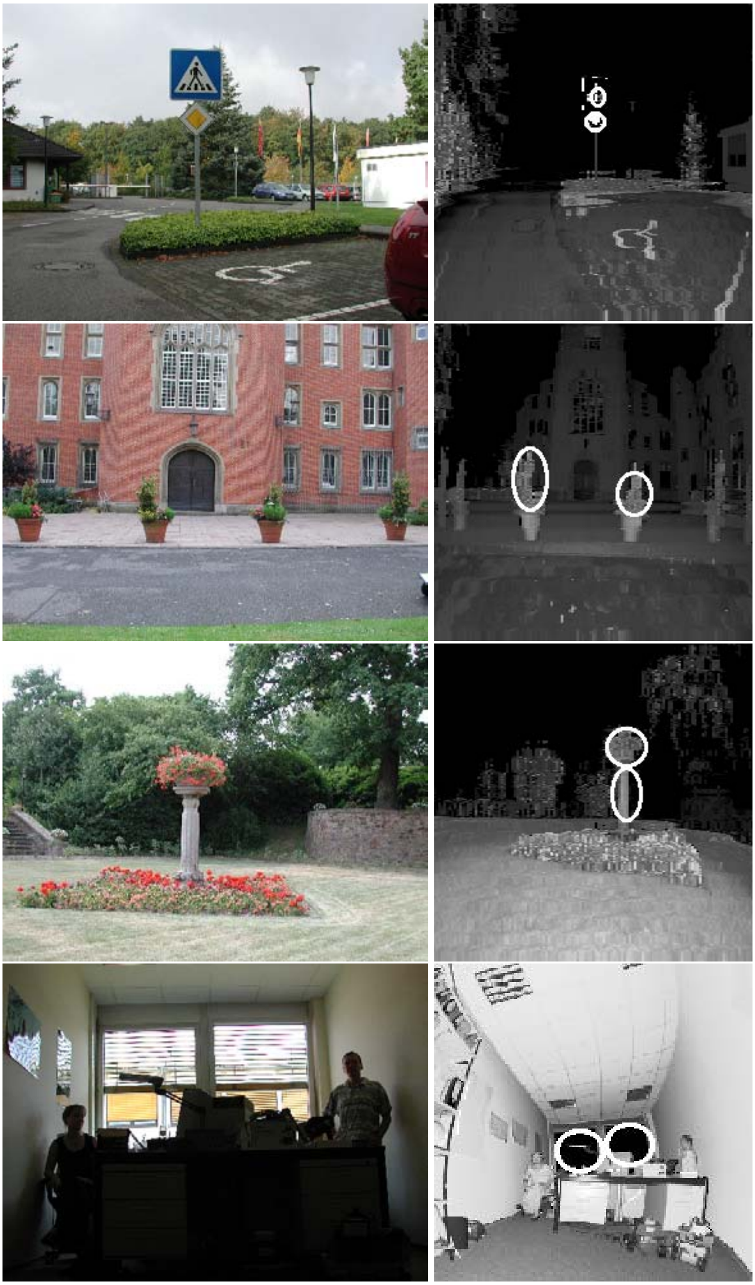

Fig. 6.7. The first two foci of attention computed by BILAS on laser scanner data. Left: the scene in a camera image. Right: foci on the combination of range and reflection data

the objects in the next two images. The last row shows an example of a scene in which the foci point to regions, the windows, that most human observers would not consider as conspicuous, since they are not useful to most tasks. However, in a pure bottom-up approach the window region is highly salient in the laser data, because the glass is transparent for the laser scanner, yielding black regions in both laser modes. Note that similar effects would arise in the processing of the camera image, which shows the window region much brighter than the rest of the image.

6.3.2 Fusing Two Laser Modes: Depth and Reflection

This section shows the different qualities of the two laser modes. For that purpose, we applied our system separately to range and reflection data. Additionally, we applied it to the simultaneous input of both modes, showing how their different properties influence the detection of salient regions. We start with the presentation of some scenes where certain saliencies are only detected in the range data and other saliencies only in the reflection data. The shown examples (Fig. 6.8–6.11) are presented in reading order as follows: depth image, reflection image, combined image, and camera image as a reference of the scene.

The advantages of the depth mode are illustrated in Fig. 6.8 and 6.9. The example in Fig. 6.8 shows a rubbish bin in a corridor. The rubbish bin is highly salient in the depth image, but not in the reflectance image. Here, the vertical line of the door attracts the attention. In the combined image, the influence of the depth focus is stronger, resulting in a focus on the rubbish bin. Remember that the influence of the maps is determined by the weighting function $\mathcal{W}$ that strengthens maps with few salient regions (cf. eq. 4.9). Of course, the focus in the combined image is not always on the desired object since this is a task-dependent evaluation. The region with the highest bottom-up saliency wins and attracts the FOA.

The example in Fig. 6.9 shows a hallway scene. The depth image shows a FOA on an open door — visible as dark region — which could be interesting for a robot as a passage. In the reflection image the foci point to other regions. Here again, the influence of the depth image is stronger, resulting in FOAs on the open door in the combined image, too.

Please note that the foci in the combined image are not a union of the foci of both modes. In the combined image, the first focus might point to a region that is the most salient region neither in the depth nor in the reflection image. This might happen for a simple reason: if the depth image has its most salient point at location a and the reflection image at location b, whereas both images have a point with lower saliency at location c, then the saliency of location c sums up to the highest saliency in the combined image, yielding the primary focus of attention.

The advantages of the reflection mode are shown in Fig. 6.10 and 6.11. Although the traffic sign in Fig. 6.10 attracts the first FOA in both laser

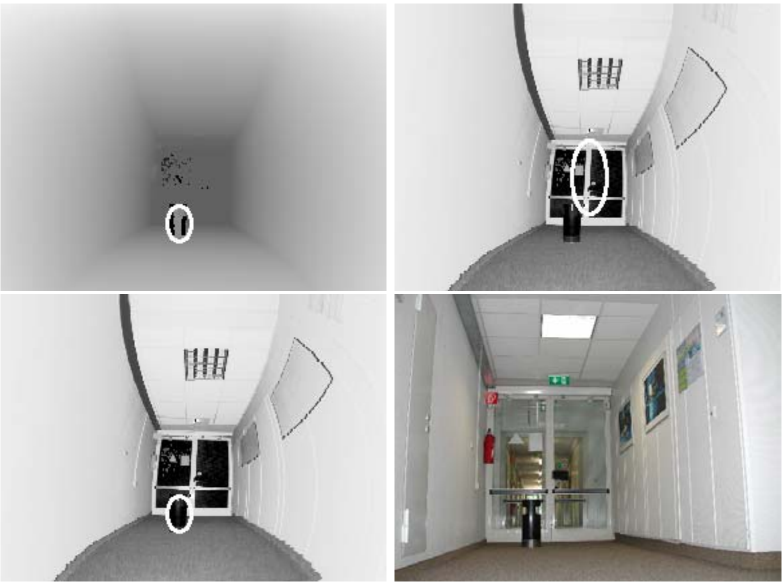

Fig. 6.8. The foci in laser data show some advantages of the depth mode. In reading order: depth image, reflection image, combined image, camera image. The rubbish bin is salient only in the range data. Here, the stronger influence of the depth image causes the first focus to point to the rubbish bin in the combined image, too

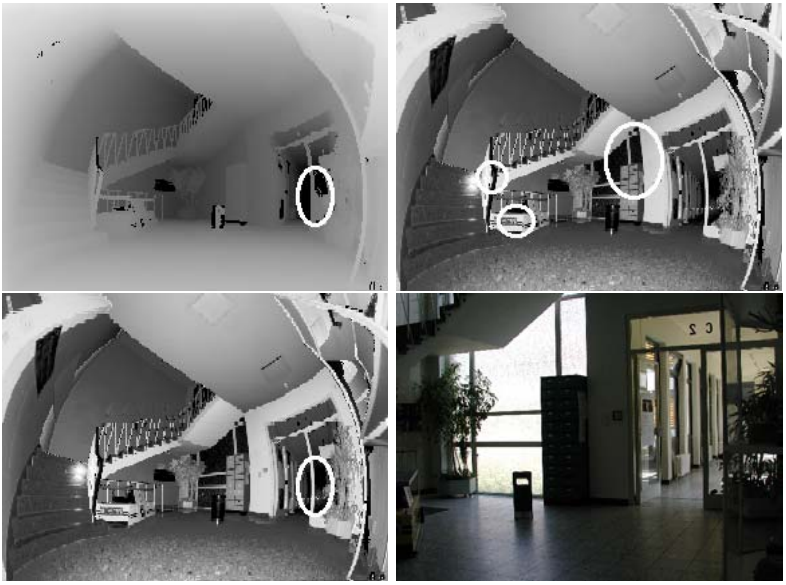

Fig. 6.9. The foci in laser data show some advantages of the depth mode. In reading order: depth image, reflection image, combined image, camera image. The open door is salient only in the range data

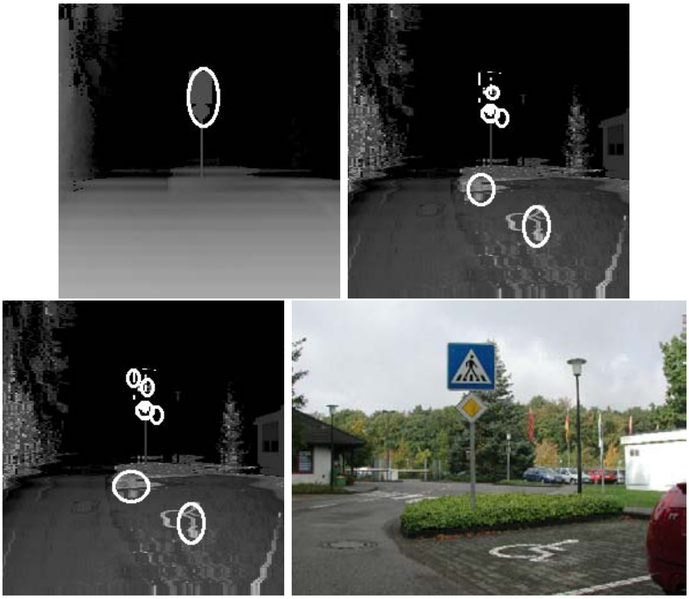

Fig. 6.10. The foci in laser data show some advantages of the reflection mode. In reading order: depth image, reflection image, combined image, camera image. The handicapped person sign is salient only in the reflection data

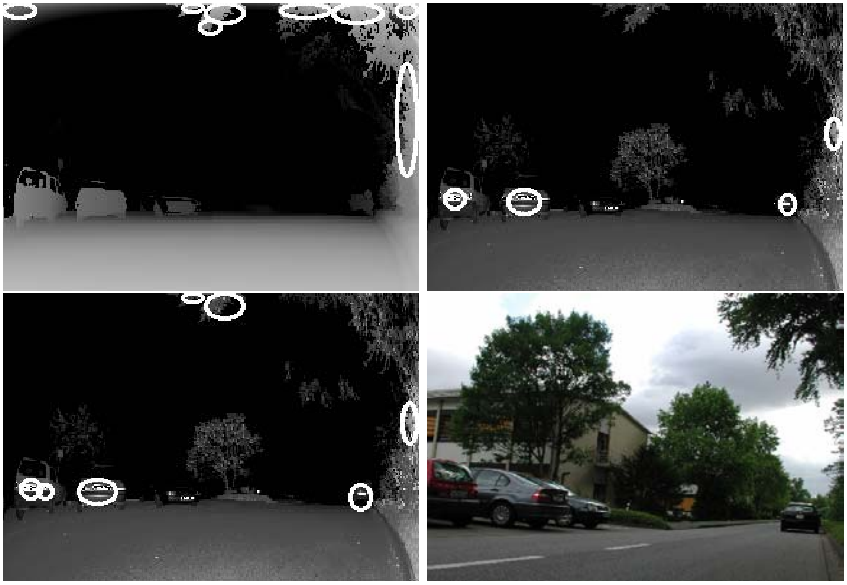

Fig. 6.11. The foci in laser data show some advantages of the reflection mode. In reading order: depth image, reflection image, combined image, camera image. All of the four cars are among the first six focus regions in the reflection data

modes, in the reflection image the 5th FOA is directed to the handicapped person sign on the floor. In the depth data this sign is completely invisible. In the combined data this detection occurs later: the 6th FOA is on the handicapped person sign. Another example is shown in Fig. 6.11. Three of the four cars in the scene are among the first four FOAs in the reflection image and within the first seven FOAs in the combined data. Obviously, the strongly reflecting license plates are the reason for high saliency in these regions. In the depth image, the cars are not focused, because the saliency of the nearer tree is stronger.

These examples show the respective advantages of the two laser modes and their complementary effect, enabling to consider different object properties.

6.3.3 Camera Versus Laser

Usually, computational visual attention systems take camera images as input. In this section, we compare this approach to the here introduced method, considering the respective advantages of the sensors.

We present three different cases: FOAs that are similar in both kinds of sensor data, those that are unique in camera images and those being unique in laser data. Fig. 6.12 shows two examples of scenes where both sensor modalities yield the same results: the traffic signs attract the attention in both scenes. We remark that this is due to different reasons: the camera FOAs are attracted by the color of the traffic sign, the laser FOAs by its depth and reflection properties. Obviously, the design of traffic signs is carefully examined since they attract bottom-up attention of different kinds.

One of the advantages of a camera is its ability to obtain color information. Although laser scanners exist that are able to record color and even temperature information, ours is not. Both scenes in Fig. 6.13 show cases in which color properties alone produced saliencies in image regions (the car in the upper image, the telephone box in the lower one) that would hardly be salient in the laser mode data.

On the other hand, Fig. 6.14 shows objects that are only focused in the laser images. The person (top) and the rubbish bin (bottom) are only focused in the laser image. The bottom image is a good example of a scene showing advantages of both, camera and laser. Whereas the focus in the laser data is on the rubbish bin — an interesting region during obstacle avoidance or cleaning up — it is in the camera image on fire extinguisher and emergency exit signs — important regions in security-relevant tasks.

Since each sensor enables the detection of different object attributes, best results should be achieved by a combination of both sensors, inducing a much richer variety of salient regions; this remains subject for future work.

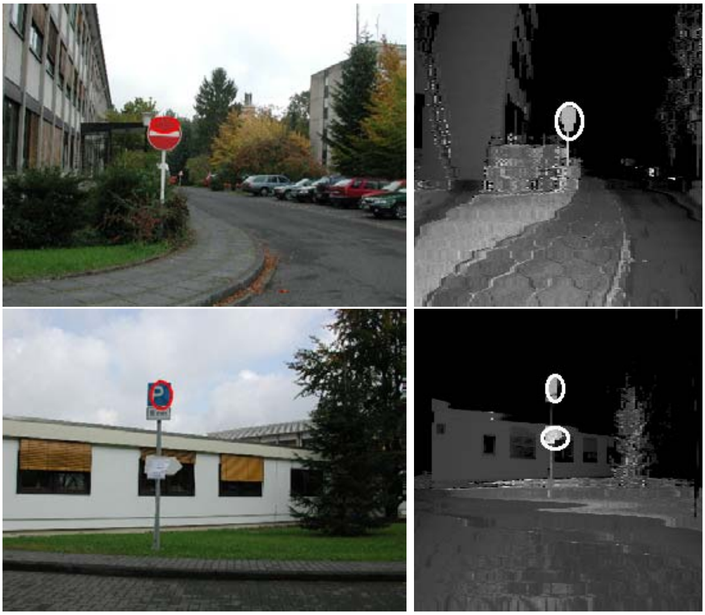

Fig. 6.12. Foci showing the same regions in camera and in laser data. Some FOAs on camera images (left) and laser images, combined from depth and reflection data (right). The FOAs are attracted due to different object properties: by color and intensity in the camera images and by depth contrast and reflection properties in the laser data

6.4 Discussion

In this chapter, we have introduced an extension of VOCUS to several sensor modalities: the Bimodal Laser-based Attention System (BILAS). The bimodal input data for the attention system, depth and reflection, were provided by a 3D laser scanner. Both data modes were processed independently considering different saliencies for the respective modes.

We have tested our system on both indoor and outdoor real-world scenes. The results show that range and reflection values complement each other: some objects are salient in depth but not in reflection data and vice versa. The comparison between the 3D laser scanner and a camera as input sensors exhibited that their data also contain complementary features. In camera images, regions may be salient due to color contrast, which is not existent in laser data. On the other hand, laser data allow the detection of salient regions that cannot be identified in camera data. Best results will be achieved by a combination of laser and camera data, a topic we consider for future work. Due to the distortions of the laser data and the different fields of view of laser and camera, this fusion is not a trivial task and has to be examined

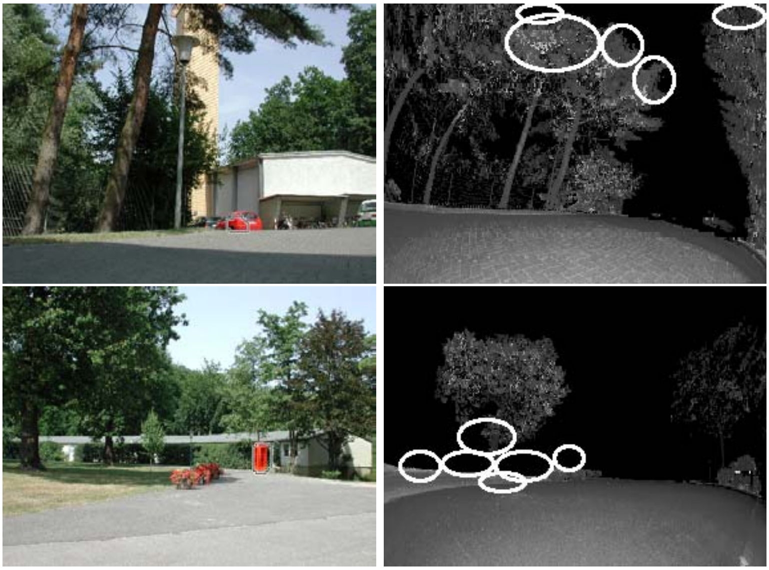

Fig. 6.13. The foci show some advantages of camera images over laser data: the red car (top) and the red telephone box (bottom) are only focused in the camera images (left), but not in laser data (right)

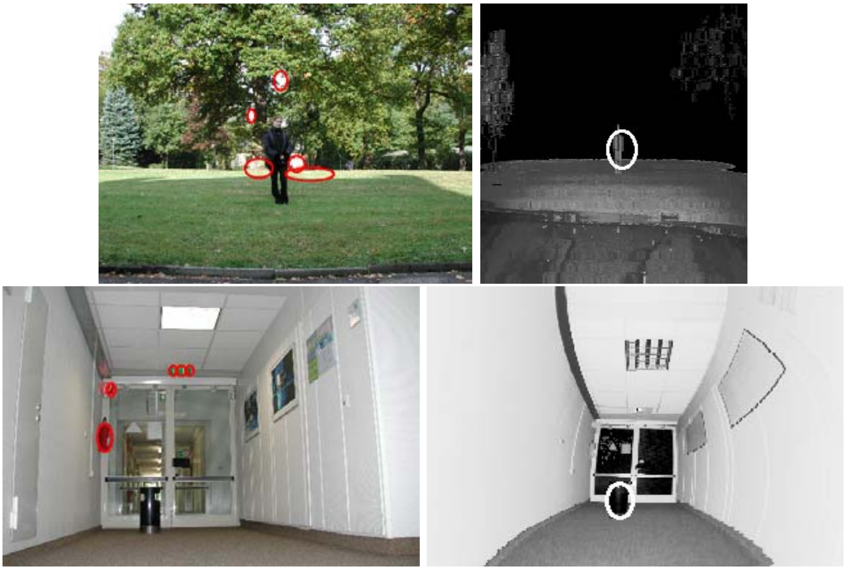

Fig. 6.14. The foci show some advantages of the laser data: the person (top) and the rubbish bin (bottom) are only focused in the laser data (right), but not in camera images (left). The bottom example shows the respective advantages of the sensors: the FOA in the laser data is on the rubbish bin whereas the FOAs in the camera image are on the fire extinguisher and the emergency exit sign

carefully [Sequeira et al., 1999]. First results can be found in [Pervölz et al., 2004].

Considering two sensor modes is a first step for the integration of multiple sensors in an attention system. The same way the two laser modes are fused, the system can be augmented to combine information of arbitrary sensors that provide the possibility to locate the sensor information in the environment. Not only camera and laser data, also auditory information could be depicted in a map and searched for salient regions provided that the direction of the sounds are known. Another possibility is to use infrared cameras to facilitate the detection of humans, a task we consider for future work [Hennig, 2004]. However, the integration of different sensor information requires careful examination.

An advantage of the laser scanner data is that it is independent of illumination variances. Different lighting conditions are a big problem in computer vision applications that rely on camera images. The laser scanner can be applied even in complete darkness, yielding the same results and providing a visual impression of the scene based on the reflection data. This can be an advantage in applications like surveillance in which the robot has to operate at night.

A limiting factor for the application of a scanning device in robot control is the low scan speed. The minimum speed of the scanner is 1.7 seconds for a low resolution 3D scan. Therefore, data from other sensors have to be used for robot navigation in quickly changing environments. On the other hand, the 3D scanner is well-suited for applications in low dynamics environments, like security inspection tasks in facility maintenance, interior survey of buildings and 3D digitalization. A much faster way to acquire range and reflection values are 3D laser "cameras", that use a sensor array to measure these values in parallel.

Several research prototypes of 3D "cameras" are known, e.g., the CSEM range camera [URL, 14], the PMD camera [URL, 15], and the 3D camera at KTH [Carlsson et al., 1999], At the moment, these cameras are still expensive, are mostly restricted to shorter ranges and very low resolutions, and usually yield results that are less precise than those of a laser scanner, but in future such devices might be the sensor of choice for such systems as the one presented here. The application of our system to data from a 3D camera is straightforward: the depth information is extracted and rendered into an image as described in 6.1.2, the color information forms a second image, replacing the reflectance data of our system. This approach has also the advantage of corresponding values and it furthermore provides color information and mainly undistorted data. One approach of applying attentional mechanisms to the data of a 3D camera is presented in [Ouerhani and Hügli, 2000].

In this chapter, we focus on the bottom-up computation of saliencies. Obviously, the next step will be the combination of this approach with the top-down guidance of the previous chapter, a topic we leave for future work. Note that one weight vector has to be computed for each sensor mode. Inevitably, the search will be less successful than the experiments in chapter 5 since it is hard to detect targets only from the two features intensity and orientation. Nevertheless, it might be possible to distinguish obstacles (bright regions in range data) and passages (dark regions in range data). Best results are to be expected from performing goal-directed search on the data from several sensor modes.