Cognitive Data Association for Visual Person Tracking

S. Frintrop[1] and M. Kessel[1]

Abstract— **In this paper, we introduce a cognitive approach for person tracking on a mobile platform. The approach is based on a biologically motivated attention system which is able to detect regions of interest in images based on concepts of the human visual system. A top-down guided visual search module of the system enables to especially highlight features which fit to a previously learned target object. Here, the appearance of a person is learned ad-hoc within the first image in which the person is detected. In subsequent images, the attention system searches for the target features and builds a top-down, target-related saliency map. This enables to focus on the most relevant features of especially this person in especially this scene without knowing anything about a particular person or scene in advance. The system is able to operate in real-time and to cope with the requirements of real-world tasks such as illumination variations.**

I. Introduction

The ability to accurately detect and keep track of people is of large interest in machine vision as well as in mobile robotics. In machine vision, example applications for people tracking include surveillance systems which monitor the behaviour of people in subway stations, in supermarkets or in traffic scenes. In the field of mobile robotics, applications include intelligent driver warning systems, following people with a service robot or guiding people in a museum.

The requirements on the systems as well as the methods which are applicable in a setting vary largely from task to task. In systems with a static camera, it is possible to make use of the fact that the background does not change and to apply methods like background subtraction. If interest is for example in counting people or other statistical investigations which do not require immediate response, it is possible to process the data offline which extends the range of applicable algorithms considerably. On the other hand, systems which shall operate on a mobile platform usually have to operate in real-time and have to deal with more difficult settings. The background changes, illumination conditions vary and platforms are often equipped with low-resolution cameras. All this restricts the methods which are applicable. On the other hand, a mobile platform might be equipped with additional sensors like laser range finders which can be combined with the visual data to improve detection and remove outliers.

Many systems do not have to keep track of individual persons, for example if the task is to monitor people in traffic to avoid accidents or to adapt the robot's velocity to the walking speed of the people in its surrounding. For other tasks like guiding or following individuals, it is essential that the system is able to distinguish between persons and does not interchange its client with someone else. Here, we present a system for tracking people in indoor environments which is able to distinguish individuals and which works on a mobile platform.

The required steps in a visual tracking system are first, the detection of the target of interest, second, the data association, i.e. redetecting the same target in subsequent frames whereas it should be able to deal with environment changes and occlusions, and third, the tracking itself which means the inference about the motion of the target given a sequence of previous measurements. Here, we will focus on the second aspect, the data association. The first aspect is an important and difficult research topic on its own which is not tackled here. We initialize manually by drawing a rectangle around the person in the first frame in which it occurs. When applied to a mobile platform, initialization can be done for example by detecting legs of people in laser data as in [1], [2] or by a visual detection front-end. The third problem, i.e., inferring the motion of the tracked person, can be solved efficiently with Kalman or Particle filters [3]. Note however that in contrast to a tracker based on data from radar or laser scanners, a visual tracker with perfect data association might even be able to work completely without motion inference. But usually, a Kalman or Particle filter helps to reduce false detections and to deal with temporary occlusions of the target. Here, we use a simple motion model which assumes that the tracked person will appear in the next frame in a close neighborhood of its current position. In the future, we plan to integrate tracking with Particle filters.

As mentioned above, the focus of this paper is the data association aspect of person tracking. We follow a feature-based approach based on the biologically motivated attention system VOCUS [4], [5]. The system first determines the most salient region of the person to track. This might be the pullover, the legs, or the head, depending on clothing and background. From this region, VOCUS computes a feature vector which describes the features of the region with respect to the current background. This vector is utilized to search for the person in a top-down manner within the subsequent frames.

Currently, the system works on camera data from a hand-held camera. Thus, it provides all conditions which are necessary to use it on a mobile robot: it is real-time capable and it is able to deal with background changes, viewpoint changes and varying illuminations. Furthermore, no complicated training phase is necessary, the appearance of

[1]The authors are with the Institute of Computer Science III, Rheinische Friedrich-Wilhems-Universität, 53111 Bonn, Germany frintrop@iai.uni-bonn.de

The authors want to thank Prof. A.B. Cremers for supporting this work.

a person is learned ad-hoc from a single frame. This makes the system flexible and also applicable to other objects than people. If the robot shall follow not a person but another robot or if an unknown vehicle in traffic shall be monitored, it is essential to be able to initialize the system quickly without training an object model. In the future, the system shall be implemented on a mobile robot, enabling to use a laser-based person detection for initializing the person tracking.

II. Related work

In mobile robotics, person tracking can be performed with different sensors. Several groups have investigated person tracking with laser range finders [6], [7], [8]. These approaches usually only keep track of the motion of people and do not try to distinguish individuals. One approach which distinguishes different motion states in laser data is presented in [9]. Combinations of laser and vision data are presented in [1] and [2]. Both detect the position of people in the laser scan and distinguish between persons based on vision data. Bennewitz et al. [1] base the vision part on color histograms whereas Schulz [2] learns silhouettes of individuals from training data. This however requires a time-consuming learning phase for each person which shall be distinguished.

In machine vision, people tracking is a well-studied problem. Two main approaches can be distinguished: *model-based* and *feature-based* methods. In model-based tracking approaches, a model of the object is learned in advance, usually from a large set of training images which show the object from different viewpoints and in different poses [10]. Learning a model of a human is made difficult by the dimensionality of the human body and the variability in human motion. Current approaches include simplified human body models, e.g. stick, ellipsoidal, cylindric or skeleton models [11], [12], [13], or shape-from-silhouettes models [14]. When dealing with silhouettes, different closing imposes additional challenges: a woman with a skirt does not have the same silhouette as a person with trousers. An approach which deals with different clothing of people is presented in [15]. While these approaches have reached good performance in laboratory settings with static cameras, they are usually not applicable in real-world environments on a mobile system. They usually do not operate in real-time and rely often on a static, uniform background.

Feature-based tracking approaches on the other hand do not learn a model but track an object based on simple features such as color cues or edges. One approach for feature-based tracking is the Mean Shift algorithm [16], [17] which classifies objects according to a color distribution. Variations of this method are presented in [18], [19]. An interesting extension is presented in [20], in which the authors suggest to on-line select the currently most discriminative features. While most approaches are not especially designed for person tracking, they might be applied in this area as well. One limitation with the above methods is that they operate only on color and are therefore dependent on colored objects.

Visual attention system are especially suited to automatically determine the features which are relevant for a certain object. These systems are motivated by mechanisms of the human visual system and based on psychological theories on visual attention [21], [22]. During the last decade, many computational attention systems have been built [23], [24], [25], [4]. Recently, some systems have been presented which are able to operate in real-time [26], [27], [28]. Most of these systems operate in a purely bottom-up, image-based manner, that means they do not consider pre-knowledge about the scene or a target. Some attention systems which are able to perform visual search for a target in a top-down manner are presented in [29], [30], [4].

Applications of visual attention systems range from object recognition [31] over video compression [32], to robot localization [33], [34]. However, they have rarely been applied to visual tracking. Some approaches track static regions, such as visual landmarks, from a mobile platform for robot localization [33], [34]. This task is easier than tracking a moving object since the environment of the target remains stable. Another approach aims to track moving objects such as fish in an aquarium [35]. In this case however, the camera is static. Furthermore, all these approaches base on bottom-up visual attention and do not include top-down cues to explicitely search for a target. Concerning person detection, we are not aware of any approach which uses a visual attention system to solve the task.

III. Cognitive Data Association

The detection of the features which are used for tracking is performed with the attention system VOCUS (Visual Object detection with a CompUtational attention System) [4], [5]. It is based on concepts of the human visual system [21], [36] and detects the most salient regions in images. VOCUS consists of a bottom-up part which computes saliency purely based on the content of the current image and a top-down part which considers pre-knowledge and target information to perform visual search. The bottom-up part is similar to the well-known approach of Itti et al. [23]. Here, mainly the top-down part of VOCUS is used, but the bottom-up computations are integrated in the learning mode.

To perform visual search, VOCUS first computes a target-specific feature vector (learning mode, see sec. III-A) and, second, uses this vector to adjust the saliency computations according to the target (search mode, see sec. III-B). Here, we introduce the main concepts of these mechanisms, more details can be found in [4], [5].

A. **Learning mode**

Before the actual tracking can start, the target person has to be detected. This can be done by an arbitrary people detection front-end. Here, we initialize manually by providing a rectangle which includes the person of interest (cf. Fig. 1). Now, VOCUS first computes a bottom-up saliency map (sec. III-A.1), second, determines the most salient region within the rectangle (sec. III-A.2), and third computes a feature vector for this region (sec. III-A.3). This way, the

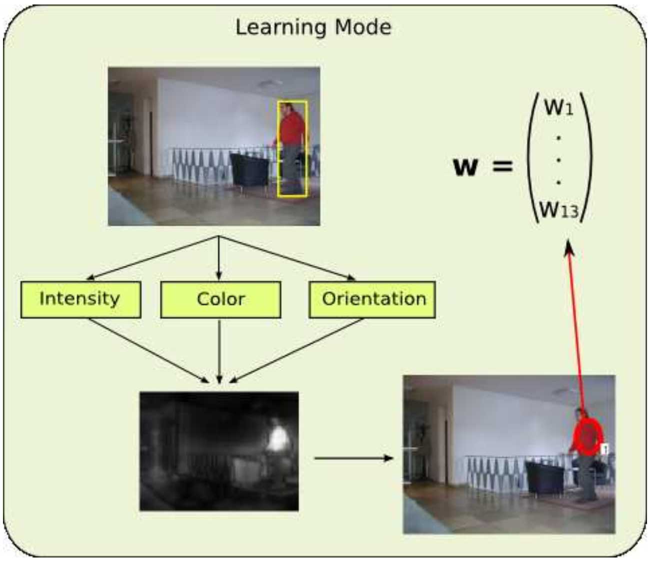

Fig. 1. Learning mode of the attention system VOCUS. Based on strong contrasts and uniqueness in the feature channels intensity, orientation, and color, a bottom-up saliency map is computed. In this map, the most salient region in the manually provided region of interest (yellow rectangle) is determined and a feature vector w is computed which describes the region.

system focuses on the features which are most relevant for particularly this person in particularly this surrounding. This is a big advantage over other approaches which determine the features of interest in advance and are therefore not as flexible as the current approach. In the following, we describe these three steps in more detail.

1) Bottom-up saliency map: First, the bottom-up saliency map is determined by computing image contrasts and uniqueness of a feature. The feature computations for the features intensity, orientation, and color are performed on 3 different scales with image pyramids. The feature intensity is computed by *center-surround mechanisms*; in contrast to most other attention systems [23], [33], on-off and off-on contrasts are computed separately. After summing up the scales, this yields 2 intensity maps. Similarly, 4 orientation maps $(0°, 45°, 90°, 135°)$ are computed by Gabor filters and 4 color maps (green, blue, red, yellow) which highlight salient regions of a certain color. Details about the feature computations can be found in [4].

Before the features are fused, they are weighted according to their *uniqueness*: a feature which occurs seldomly in a scene is assigned a higher saliency than a frequently occurring feature. This is a mechanism which enables humans to instantly detect outliers like a black sheep in a white herd. The uniqueness $\mathcal{W}$ of map X is computed as

$$\mathcal{W}(X) = X/\sqrt{m}, \tag{1}$$

where m is the number of local maxima that exceed a threshold and '$/$' is here the point-wise division of an image with a scalar. The maps are summed up to 3 conspicuity maps I (intensity), O (orientation) and C (color) and combined to form the *bottom-up saliency map*:

$$S_{bu} = \mathcal{W}(I) + \mathcal{W}(O) + \mathcal{W}(C) \tag{2}$$

2) Extracting the most salient region: Since we are only interested in the saliency of the target object, here a person, the most salient region within the provided rectangle is determined. First, the most salient (brightest) point in S_{bu} is determined, and then with *seeded region growing* the surrounding region. This method finds recursively all neighbors with sufficient saliency. We call the resulting region FOA (focus of attention). Although the region is irregularly shaped, we display it for simplicity as an ellipse.

3) Determining a feature vector: From the FOA, a feature vector $\vec{w}$ (descriptor) is determined from the values of the 10 feature and 3 conspicuity maps of VOCUS. It describes how much each feature contributes to the FOA.

The value w_i for map X_i is the ratio of the mean saliency in the target region $m_{(ROI)}$ and in the background $m_{(image-ROI)}$:

$$w_i = m_{(ROI)}/m_{(image-ROI)}. \tag{3}$$

This computation does not only consider which features are the strongest in the target region, it also regards which features separate the region best from the rest of the image.

After the feature vector for the target person has been computed, this vector can be used to search for the target in other frames.

B. Search mode

In search mode, we determine a top-down, target-related saliency map from which the most salient region is extracted (cf. Fig. 2). This map is composed of an excitation and an inhibition map. VOCUS uses the previously learned feature vector $\vec{w}$ to weight the feature and conspicuity maps according to the target. Depending on the values of the weights w_i, the maps are used to compute the excitation or the inhibition map. The excitation map E is the weighted sum of all feature and conspicuity maps X_i that are important for the learned region, i.e., $w_i > 1$:

$$E = \sum_i (w_i * X_i) \quad \forall\, i \in \{1..13\} : w_i > 1 \tag{4}$$

The inhibition map I shows the features more present in the background than in the target region, i.e., $w_i < 1$:

$$I = \sum_i ((1/w_i) * X_i) \quad \forall i \in \{1..13\} : w_i < 1 \tag{5}$$

Maps with $w_i = 1$ are completely unimportant for the target and are ignored. The top-down saliency map S_{td} results from the difference of E and I and a clipping of negative values:

$$S_{td} = E - I. \tag{6}$$

After the computation of the top-down saliency map, the FOA in S_{td} is determined by *region growing* starting with the

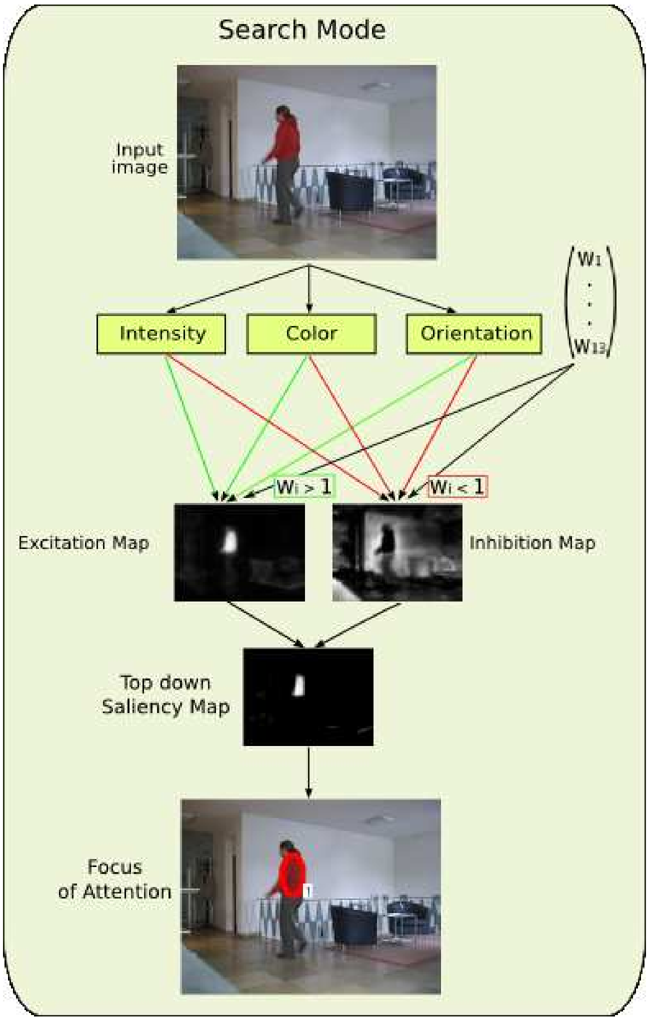

Fig. 2. Search mode of the attention system VOCUS. The feature maps are weighted according to values of the feature vector w which describes the target person. A top-down saliency map is computed which highlights the target-specific regions of interest and the most salient region determines the position of the person.

maximum of S. The coordinates of the FOA determine the position estimate of the person. Additionally, we use a simple motion model to reduce outliers and search for an FOA only within the local neighborhood (here 32×32 pixels) of the FOA from the previous frame. In future work, this shall be replaced by a Particle filter.

IV. Experiments and Results

To test the cognitive person tracker, we have performed two kinds of experiments. In the first experiment, we have tested the system on a publically available image sequence which was also used for other person tracking experiments. In the second experiment, we have recorded several image sequences ourselves in an office environment. In both experiments, images had a resolution of 320×240. In the first frame of each image sequence, the target person was marked manually and a feature vector was computed for the target. With this feature vector, VOCUS searched for the person in the subsequent frames.

Fig. 3. Person tracking with cognitive data association on a standard test sequence[1] [37]. Top: the top-down saliency map. Bottom: the corresponding input image with the resulting FOA (focus of attention) displayed as red ellipse.

First, the system was applied to a standard test sequence[1] which is usually used to test model-based person trackers [37], [38]. The sequence was recorded with a static camera and consists of 124 images which show a person walking from the right to the left through the image. Our system determines the white shirt as most salient part in this setting and easily tracks the person through the whole sequence without any learned model of a human. Three example images are displayed in Fig. 3, on top the top-down saliency map, below the input image with the FOA displayed as red ellipse.

In a second experiment, we tested the person tracker on several image sequences obtained with a hand-held camera in the corridor of our lab environment. The camera was moved to follow the person and each sequence consists of about 300 images. The results of two of the sequences are displayed in Fig. 4 and 5. On the top, the top-down saliency map is displayed, below the image in which the FOA is shown as a red ellipse. In the first image sequence (Fig. 4), the tracking is easy since the red pullover of the person differs considerably from the surrounding. The top-down saliency map shows a single bright peak which is tracked successfully in all frames. In the second sequence (Fig. 5), the task is more difficult. The clothing of the person is black, a color which is not unique in the setting. The top-down saliency map shows several other black regions as salient. During most of the sequence, the person is tracked successfully anyway. Only during the last part of the sequence, when the person goes through the door on the right, the system confuses the person with the black part of the door. This example shows that it would be useful to verify the current tracking hypothesis from time to time with a person detection module to make sure that the person has not been lost and, if it is lost, to re-initialize the system.

V. Conclusion

In this paper, we have presented a cognitive approach for data association for a visual person tracking system. The appearance of a person of interest is learned from an initially provided target region and the resulting target feature

[1] The sequence is available at www.nada.kth.se/∼hedvig/data.html.

Fig. 4. Person tracking with cognitive data association. Top: the top-down saliency map. Bottom: the corresponding input image with the resulting FOA (focus of attention) displayed as red ellipse. The person is tracked successfully in all frames.

Fig. 5. Person tracking with cognitive data association. Top: the top-down saliency map. Bottom: the corresponding input image with the resulting FOA (focus of attention) displayed as red ellipse. The person is tracked successfully, until it goes through the door on the right. Here, the system confuses the person with the black stripe of the door.

vector is used to search for the target in subsequent frames. The main advantage of the system is that it determines autonomously the most relevant target features in a current setting. If this is a red pullover as in Fig. 1, it mainly focuses on red, if it is the intensity contrast of the head to a white wall, the intensity contrast is highlighted.

Other advantages of the approach are that the system is quickly adaptable to a new target without a time-consuming learning phase, that it works on a mobile platform and does not rely on a static background, and that it works in real-time and under varying illumination conditions. Since the system does not rely on a model of a human, it is able to cope with differences in appearance like a woman with a skirt or a person which is carrying a large object. The system might however currently have difficulties to deal with people with the same appearance (same clothing etc.), if they cross their way. Integrating a Particle filter to estimate the motion trajectory of the people will help to resolve such ambiguities.

This paper focuses on the data association aspect of tracking. To complete a visual person tracker on a mobile robot, several important aspects still have to be done: first and most important, the manual initialization of the system which provides the target region to the learning mode has to be automatized. We plan to apply a laser-based person detection as in [2]. Second, the current system sometimes confuses the target person with a similar region of the surrounding as in Fig. 5. Several extensions could help to resolve these ambiguities, e.g. integrating a motion channel into VOCUS, considering several salient regions for a person in parallel (e.g. the pullover as well as the shoes), or re-initializing with a person detector from time to time. Finally, as mentioned before, we plan to integrate a Particle filter to distinguish persons with similar appearance and to deal with temporary occlusions of people. Testing the system for longer sequences and with more people will also be a topic for future work.

References

[1] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, "Learning motion patterns of people for compliant robot motion," *The International Journal of Robotics Research*, vol. 24, no. 1, pp. 31–48, 2005.

[2] D. Schulz, "A probabilistic exemplar approach to combine laser and vision for person tracking," In Proc. of the International Conference on Robotics Science and Systems (RSS 2006), 2006.

[3] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT Press, 2005.

[4] S. Frintrop, "VOCUS: a visual attention system for object detection and goal-directed search," Ph.D. dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn, Germany, July 2005, published 2006 in Lecture Notes in Artificial Intelligence (LNAI), Vol. 3899, Springer Verlag Berlin/Heidelberg.

[5] S. Frintrop, G. Backer, and E. Rome, “Goal-directed search with a top-down modulated computational attention system,” in *Proc. of the Annual meeting of the German Association for Pattern Recognition (DAGM 2005)*, ser. Lecture Notes in Computer Science (LNCS). Conference: Wien, Austria: Springer, Sept. 2005, pp. 117–124.

[6] M. Montemerlo, S. Thrun, and W. Whittaker, “Conditional particle filters for simultaneous mobile robot localization and people-tracking,” in *Int’l Conference on Robotics and Automation (ICRA)*, 2002.

[7] A. Fod, A. Howard, and M. J. Matari, “Laser-based people tracking,” In Proc. of the IEEE International Conference on Robotics and Automation (ICRA), 2002.

[8] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, “People tracking with mobile robots using sample-based joint probabilistic data association filters,” International Journal of Robotics Research, 22(2), 2003.

[9] G. Taylor and L. Kleeman, “A multiple hypothesis walking person tracker with switched dynamic model,” in *Australasian Conference on Robotics and Automation (ACRA)*, 2004.

[10] K. Rohr, “Towards model-based recognition of human movements in image sequences,” *CVGIP – Image Understanding*, vol. 59, no. 1, pp. 94–115, 1994.

[11] C. Breglera, J. Malik, and K. Pullen, “Twist based acquisition and tracking of animal and human kinematics,” *International Journal of Computer Vision (IJCV)*, 2004.

[12] R. Urtasun, D. J. Fleet, and P. Fua, “Temporal motion models for monocular and multiview 3d human body tracking,” *Computer Vision and Image Understanding (CVIU), special issue Modeling People*, 2006.

[13] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman, “Human body model acquisition and tracking using voxel data,” *International Journal of Computer Vision*, vol. 53, no. 3, pp. 199–223, 2003.

[14] K. Cheung, S. Baker, and T. Kanade, “Shape-from-silhouette across time part II: Applications to human modeling and markerless motion,” *International Journal of Computer Vision (IJCV)*, 2005.

[15] B. Rosenhahn, U. Kersting, K. Powell, T. Brox, and H. P. Seidel, “Tracking clothed people,” in *Human Motion - Understanding, Modeling, Capture, and Animation*. Springer, 2007.

[16] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, 2002.

[17] D. Comaniciu, V. Ramesh, and P. Meer, “Real-time tracking of non-rigid objects using mean shift,” Proc. Conf. Computer Vision and Pattern Recognition (CVPR), vol. 2, Hilton Head Island, South Carolina, 2000.

[18] G. R. Bradski, “Computer vision face tracking for use in a perceptual user interface,” Intel Technology Journal, 1998.

[19] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, “Color-based probabilistic tracking,” Proceedings of the 7th European Conference on Computer Vision (ECCV) London, UK, Springer-Verlag, 2002.

[20] R. Collins and Y. Liu, “On-line selection of discriminative tracking features,” in *Proceedings of the 2003 International Conference of Computer Vision (ICCV ’03)*, Oct. 2003.

[21] A. M. Treisman and G. Gelade, “A feature integration theory of attention,” *Cognitive Psychology*, vol. 12, pp. 97–136, 1980.

[22] J. M. Wolfe, “Guided search 2.0: A revised model of visual search,” *Psychonomic Bulletin and Review*, vol. 1, no. 2, pp. 202–238, 1994.

[23] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[24] G. Backer, B. Mertsching, and M. Bollmann, “Data- and model-driven gaze control for an active-vision system,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 23(12), pp. 1415–1429, 2001.

[25] Y. Sun and R. Fisher, “Object-based visual attention for computer vision,” *Artificial Intelligence*, vol. 146, no. 1, pp. 77–123, 2003.

[26] S. Frintrop, M. Klodt, and E. Rome, “A real-time visual attention system using integral images,” in *Proc. of the 5th International Conference on Computer Vision Systems (ICVS)*, Bielefeld, Germany, March 2007.

[27] S. May, M. Klodt, E. Rome, and R. Breithaupt, “GPU-accelerated affordance cueing based on visual attention,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2007.

[28] M. Björkman and J.-O. Eklundh, “Vision in the real world: Finding, attending and recognizing objects,” *Int’l Journal of Imaging Systems and Technology*, vol. 16, no. 2, pp. 189–208, 2007.

[29] F. H. Hamker, “The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision,” *Journal of Computer Vision and Image Understanding (CVIU), Special Issue on Attention and Performance*, vol. 100, no. 1-2, pp. 64–106, 2005.

[30] V. Navalpakkam, J. Rebesco, and L. Itti, “Modeling the influence of task on attention,” *Vision Research*, vol. 45, no. 2, pp. 205–231, 2005.

[31] D. Walther, U. Rutishauser, C. Koch, and P. Perona, “Selective visual attention enables learning and recognition of multiple objects in cluttered scenes,” *Computer Vision and Image Understanding (CVIU)*, vol. 100, no. 1-2, pp. 41–63, 2005.

[32] L. Itti, “Automatic foveation for video compression using a neurobiological model of visual attention,” *IEEE Transactions on Image Processing*, vol. 13, no. 10, Oct 2004.

[33] N. Ouerhani, A. Bur, and H. Hügli, “Visual attention-based robot self-localization,” in *Proc. of European Conference on Mobile Robotics (ECMR 2005)*, Ancona, Italy, Sept. 2005, pp. 8–13.

[34] S. Frintrop and P. Jensfelt, “Active gaze control for attentional visual SLAM,” in *accepted for Proc. of the IEEE International Conference on Robotics and Automation (ICRA’08)*, 2008.

[35] M. Veyret and E. Maisel, “Attention-based target tracking for an augmented reality application,” International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, 2006.

[36] M. Corbetta and G. L. Shulman, “Control of goal-directed and stimulus-driven attention in the brain,” *Nature Reviews*, vol. 3, no. 3, pp. 201–215, 2002.

[37] H. Sidenbladh, M. J. Black, and L. Sigal, “Implicit probabilistic models of human motion for synthesis and tracking,” in *Proc. of European Conference of Computer Vision (ECCV)*, 2002.

[38] T. Brox, B. Rosenhahn, U. Kersting, and D. Cremers, “Nonparametric density estimation for human pose tracking,” Pattern Recognition (Proc. DAGM) Berlin, Germany, Springer, LNCS, Vol. 4174, 546-555, 2006.