Visual Attention for Robotic Cognition: A Survey

Momotaz Begum and Fakhri Karray

***Abstract*—The goal of the cognitive robotics research is to design robots with human-like cognition (albeit reduced complexity) in perception, reasoning, action planning, and decision making. Such a venture of cognitive robotics has developed robots with redundant number of sensors and actuators in order to perceive the world and act up on it in a human-like fashion. A major challenge to deal with these robots is managing the enormous amount of information continuously arriving through multiple sensors. The primates master this information management skill through their custom-built attention mechanism. Mimicking the attention behavior of the primates, therefore, has gained tremendous popularity in robotic research in the recent years (Bar-Cohen *et al.*, *Biologically Inspired Intelligent Robots*, 2003, and B. Webb *et al.*, *Biorobotics*, 2003). The difficulties of redundant information management, however, is the most severe in case of visual perception of the robots. Even a moderate size image of the natural scene generally contains enough visual information to easily overload the on-line decision making process of an autonomous robot. Modeling primates-like visual attention mechanism for the robot, therefore, is becoming more popular among the robotic researchers. A visual attention model enables the robot to selectively (and autonomously) choose a "behaviorally relevant" segment of visual information for further processing while relative exclusion of the others. This paper sheds light on the ongoing journey of robotics research to achieve a visual attention model which will serve as a component of cognition of the modern-day robots.**

***Index Terms*—Human–robot interaction, joint attention, overt attention, robotic cognition, visual attention.**

I. Introduction

WHEN it comes to information management, the underlying idea of the primates attention mechanism is simple, yet extremely robust: focus on the piece of information which is the most relevant to a given context. A custom-built attentional circuit in the primates helps them to execute this attention behavior [3]. The endeavor of robotics research to design a bioinspired visual attention model for the cognitive robot has strong connectivity with the research in cognitive psychology, computational neuroscience, and computer vision as these are the three disciplines which cultivated the basic research on the artificial modeling of human visual attention. The visual attention models developed for robotic cognition heavily rely on the computational models of visual attention proposed in computer vision and computational neuroscience while the inspiration behind all of these models is rooted in the theories of human visual attention proposed in cognitive psychology and neuroscience. The major motivation for developing computational models of visual attention was two-fold: 1) creating a computational tool to test the validity of the theories/hypothesis of visual attention proposed in psychology and neuroscience; and 2) the potential applications of the principle of focused attention in computer vision, video surveillance, and robotics. Accordingly, we observe the rise of two distinct trends in the research on computational modeling of visual attention. The first one is mostly concerned about simulating the neuronal response of the primates during various attentional activities [4]–[10]. The researchers in computational neuroscience and cognitive psychology are dedicated to develop this type of models. The second kind of computational models are concerned about developing a technical system of visual attention while utilizing the unique properties of the biological attention system [11]–[18]. The researchers in computer vision and robotics are the major developers of the technical models of visual attention. Cognitive robotics, probably, is the most recent user of these models.

The goal of the survey presented in this article is to shed light on the research on visual attention modeling as a component of robotic cognition. A brief discussion on the visual attention models proposed in computer vision and computational neuroscience, their limitations in the context of robotic applications, and how these limitations trigger the evolution of the robotic model of visual attention will also be discussed. The rest of the paper is organized as follows. Section II describes a brief history of development of the computational model of visual attention. Section III discusses the characteristics of the computational model of attention that caused the rise of a different group of visual attention models for the robot. Section IV discusses the existing research on the modeling of robotic visual attention. Section V provides an overall discussion on the success, possibilities, and open challenges related to the design of a visual attention model as a component of robotic cognition. Finally, Section VI draws the conclusion on this survey.

II. Evolution of Research on Computational Modeling of Visual Attention

The visual attention of the primates is not yet a fully understood mechanism and therefore, associated theories and hypothesis are being updated continuously. There exist a number of hypothesis in the literature of cognitive nueroscience and developmental psychology regarding the developmental process of visual attention in the primates [19]–[23]. The researchers, however, are not unified yet about their opinion on the exact mechanism that governs the development of attention in the primates. In spite of this lack of complete understanding, the research on

Manuscript received May 28, 2010; revised October 19, 2010; accepted October 30, 2010. Date of publication December 10, 2010; date of current version March 16, 2011.

M. Begum is with the School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30602 USA (e-mail: mbegum@cc.gatech.edu).

F. Karray is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: karray@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAMD.2010.2096505

visual attention in the past few decades has reached the status where we can comfortably derive a functional framework of the attention related activities. This development inspired the researchers in biology, psychology, computational neuroscience, computer vision, and robotics to develop synthetic models of visual attention which have potential applications in their respective fields. This section discusses about some of the most influential visual attention models in psychology, computational neuroscience, and computer vision.

The pioneering work on visual attention, the *feature integration theory* [24], was proposed in psychology. The generic purpose of the attention models proposed in psychology is to explain human perception and cognition [22], [24]–[26] based on behavioral data (of the primates). The *feature integration theory* advocates the idea that perception of features comes prior to the perception of objects. The features in the visual field register themselves to our visual system and then are processed with the help of visual attention to form the concept of an object. Different visual features, such as color, orientation, spatial frequency, brightness, and direction of movement register their saliency in separate feature-maps. The feature maps also retain the physical location of different features in order to ensure the perfect synthesis of features for each object. In case of computational implementation the feature-maps are summed up to create a master-map of saliency of different features. The master-map is used to direct attention toward different locations in an image in the order of decreasing saliency. The *feature integration theory* went through several facets of development to accommodate the new findings on attention obtained from psychophysical experiments. A comprehensive survey on this popular theory is available in [27]. One major drawback of the early *feature integration theory* is that it considers only the effect of visual strength of different features (commonly known as bottom–up bias) in attentional selection. The *guided search* model [25] of visual attention overcomes this limitation by invoking the effect of top–down selection. The *guided search* is mostly focused on modeling the attention behavior related to visual search. Similar to the *feature integration theory*, gradual upgrading is observed in the *guided search* model [28]–[30] to accommodate the new findings of the primates visual search behavior. Another influential model of visual attention in psychology is the CODE theory of attention [26]. CODE theory is basically an integration of the theory of visual attention [22] with the theory of perceptual grouping by proximity [31]. A major difference of the CODE theory from the *feature integration theory* and the *guided search* model is it considers both space and object as the elemental unit for attentional selection, whereas the latter two only deal with space-based attention. There are many other models of visual attention available in psychology. Please see [32] for a comprehensive survey on the psychophysical models of visual attention.

The synthetic models of visual attention developed in neurobiology and computational neuroscience are based on the findings of the primates visual attention obtained through lesion study and brain imaging techniques [e.g., functional magnetic resonance imaging (fMRI), positron emission tomography (PET)] [4]–[10]. The goal of these models is to faithfully reproduce the results obtained from the study of different attentional networks in the primates brain. Given the fact that study on the brain of live subjects is a very delicate matter and is subjected to different ethical bindings, accurate models in computational neuroscience play a critical role in understanding the operation of different brain networks. A survey on the neurobiological models of visual attention is available in [33]. One of the most popular theories of attention in neuroscience is the *biased competition hypothesis* (BC) (also known as *integrated competition hypothesis* [19], [34]–[37]). The BC hypothesis advocates the idea of a mutually suppressive interaction among visual neurons when excited by different visual stimuli. The attention mechanism of the primates biases this competition in favor of a certain specific stimulus through feedback bias mechanism (also known as top–down bias). The postulates of BC gained widespread popularity due to strong experimental evidences. A number of computational models has been proposed in computational neuroscience based on the postulates of BC [4]–[7], [38], [39]. The BC-based models of visual attention invoke many new findings of attention, e.g., combination of object- and space-based analysis for attentional selection, integration of top–down and bottom–up bias, and integration of covert and overt shift of attention.

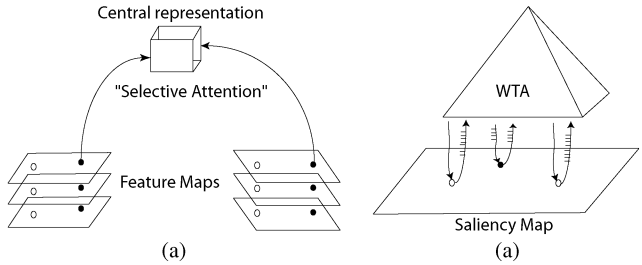


Fig. 1. Schematic drawings illustrating the concepts of saliency map and winner-take-all (WTA) network proposed in the very first computer vision model of visual attention in [12]. (a) The visual strength of different features are registered in individual feature maps. The feature maps are then combined to form a central saliency map. (b) A WTA is used to determine the most salient location as well as to implement the inhibition of return (IOR).

The rise of the computer vision models of attention [11]–[18], [40] occurred almost in parallel with the psychophysical models. The model proposed in [12] is the first computer vision model of visual attention and follows the basic postulates of the *feature integration theory*. This model [12] coins the term "saliency map" to define a two-dimensional representation of scene saliency. The model [12] also introduces the concept of a winner-take-all (WTA) network to identify the focus of attention in a saliency map, as well as to implement the property of inhibition of return (IOR) [41]. Fig. 1 shows a schematic diagram of the model. The term "saliency map," as well as its underlying concept have been used extensively in the visual attention literature. Different variants of the WTA network proposed in [12] have been used in several models of visual attention. Probably the most influential computer vision model of visual attention is the neuromorphic vision toolkit (NVT) [13] which has been extensively used in many other models of visual attention in the computer vision and robotics. A number of concepts in the NVT are heavily inspired by the attention model in [12], e.g., feature map, saliency map, and WTA network-based implementation of IOR. The NVT, however,

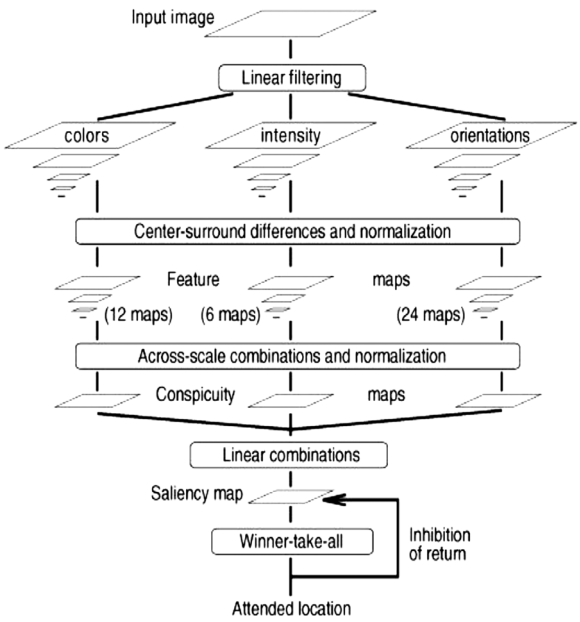


Fig. 2. Architecture of NVT [13]. Multiscale analysis of an input image is performed to evaluate the conspicuities of three image features: color, intensity, and orientation. Individual feature maps are combined to create a saliency map onto which a WTA network operates to identify the next focus of attention.

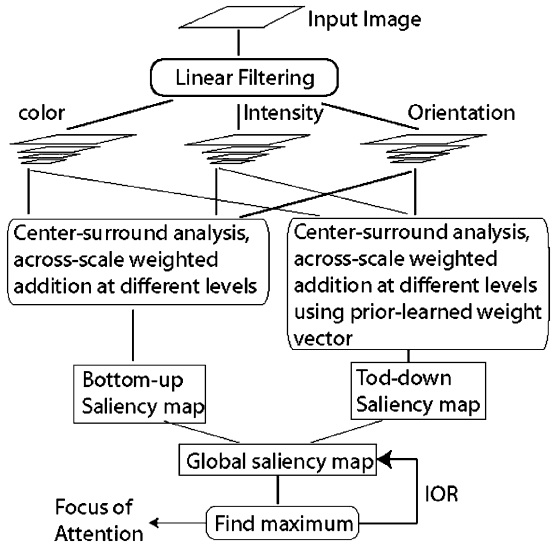


Fig. 3. Architecture of VOCUS [18]. The major differences with NVT [42] lie in a number of sectors in implementation, e.g., the center-surround mechanism, accross-scale addition, total number of image pyramids used for each feature, the learning of the top–down weight matrix, and addition of the top–down and bottom–up saliency map.

proposes a computationally elegant process for calculation of the saliency map. Here, a multiscale analysis of the input images are performed for calculation of individual feature maps which are combined together to create a centralized saliency map. Fig. 2 shows the basic architecture of NVT. The early version of NVT [13] performed only bottom–up analysis of attention but a later modification reported in [14] invokes the effect of top–down selection during visual search.

Some of the shortcomings of NVT have been alleviated in the NVT-based model *visual object detection with a computational attention system* (VOCUS) [18]. Similar to NVT, VOCUS also relies on the attention model in [12] regarding a number of core concepts of visual attention, e.g., saliency map, WTA-based selection of focus, and implementation of IOR. The calculation of saliency map in VOCUS is similar to that of NVT. VOCUS, however, performs a number of improvements over NVT (with respect to implementation and theoretical analysis) which results in better accuracy in identifying the focus of attention in a given image. Fig. 3 shows the basic structure of VOCUS [18].

There is a group of computer vision models which uses connectionist approach for visual attention modeling. Among these connectionist models the most famous is the *selective tuning model* [15]. The *selective tuning model* analyzes four kinds of visual features to identify the focus of attention in an input image: luminance, orientation, color, and motion. The model performs a pyramid style processing of information where the stimuli of interest are located at the top and control an inhibitory beam. This inhibitory beam can inhibit or pass a zone for further processing. The top–down influence is modeled through manipulating the inhibitory beam. A unique characteristics of this model is, in spite of being a technical model of attention, the *selective tuning model* [15] and all of its variants [43], [44] are tightly coupled with biological principles.

There are several other computer vision models of visual attention, but the models discussed in this section are generally considered as the most influential models. This is mostly because the majority of the other existing models of visual attention are, to some extent, derived from them. The following section will shed light on the general characteristics of the computer vision models of visual attention which made them very popular over the last decade and, at the same time, triggered the evolution of a different group of visual attention models for the robot.

III. Computational Models of Visual Attention and the Requirements of Robotic Visual Attention

Among the computational models of visual attention, the models proposed in computer vision gained widespread popularity in robotics. Specially, the computer vision models proposed in [12], [13], [18], and [42] have strong influence on the robotic model of visual attention. This is mostly because the strategies of computer vision models to analyze visual features make them suitable to be applied on the technical system (unlike the model proposed in computational neuroscience). Majority of the existing computer vision model, however, have some characteristics which impose some restrictions on their direct use for robotic visual attention. This section first summarizes the general properties of the computer vision models of visual attention and then sheds light on the issues that restrict their direct use in the robotic systems.

A. General Characteristics of the Computer Vision Model of Visual Attention

The computer vision model of visual attention share some common operating principles. Fig. 4 shows the architecture generally followed by the existing computer vision models of visual attention. The key differences among different models occur in

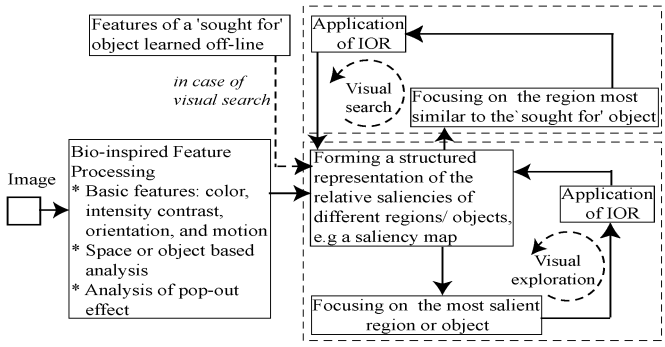


Fig. 4. General architecture of the computational models of visual attention available in the current literature.

Fig. 5. Role of saliency operator in evaluating visual attention. (a) a natural image obtained from the standard image database provided in [42], (b) the saliency map calculated using the method described in [18], and (c) five focus points in the image marked in the order of decreasing saliency.

two sectors: 1) the methodology of implementing the overall architecture, e.g., connectionist approach [15] and [16], filter-based approach [11]–[14], [17], [18], and [40]; and 2) the mechanism of constructing the saliency map. Despite these two sectors of mismatch, the computer vision models of visual attention share a set of common characteristics. They are summarized below.

1) Saliency Operator: A unique, centralized "saliency map" plays a key role to guide attention toward different regions of an input image in almost all of the existing computer vision models of visual attention [13], [14], [16]–[18], [40]. A saliency map, in simple words, is a two-dimensional image (of same size as the input image) in which the intensity value of a pixel represents the relative visual saliency of its corresponding pixel in the original input image. The higher the value is, the more salient the pixel is. A saliency map-based attention model generally reports the most salient pixel in the saliency map as the current focus of attention. In order to prevent the attention from revisiting the same location, the saliency of the current focus of attention is suppressed after being attended and thereby achieving the property of IOR [41]. Fig. 5 shows a typical saliency map corresponding to a natural image (obtained from the freely available standard image database provided in [42]) along with the focuses of attention in the image in the order of decreasing saliency.

2) Covert Shift of Focus: Majority of the computer vision models of visual attention are designed based on the assumption that neither the eye nor the head moves to execute visual attention. The attention mechanism in the primates which obeys this assumption is called covert attention [45]. The absence of eye/head movement during the execution of visual attention has a number of consequences.

- The retinal input remains unchanged throughout the attentional task.
- The frame of reference remains unchanged in the subsequent directions of attention. This simplifies the implementation of the IOR.
- The belief about the scene saliency remains unchanged causing no further requirement to recalculate it after each attentional shift.

Most of the computer vision models of visual attention (e.g., [12]–[18] and [40]) enjoy the simplicity of computation arising from the above three consequences of the covert nature of attentional shift. The covert shift of attention, however, makes it difficult to compare the performance of a model with the ground truth, e.g., with the attention behavior of the human.

3) Bottom–Up and Top–Down Analysis: The early computer vision models of attention (e.g., [12], [13], [15], [40], and [42]) mostly dealt with bottom–up (or stimulus driven) influence in attention selection. To be consistent with the biological findings, the recent models started to invoke the effect of top–down influence [14], [16], [18], [44], [46]. These latter models [14], [16], [18], [44], [46], however, limit the influence of top–down information only to the case of visual search. Thus, the bottom–up cues guide the visual exploration (focusing on the most salient stimuli), while the top–down cues guide the visual search (an active scan of the visual field in search of a prespecified object or stimuli). In almost all of the existing models these two modes of attention (visual search and visual exploration) run in mutual exclusion of each other as shown in Fig. 4. The desired mode of attention (search or exploration) is manually activated by the programmer depending on the task at hand. In some of the models, even the process of generating the saliency map for visual exploration considerably differs from that for visual search.

4) Off-Line Training for Visual Search: Almost all computer vision models of visual attention require an off-line training phase prior to performing visual search. The model learns the target specific visual features during the training phase and the learned information is used to increase the saliency of the target-like features in the test images. This strategy of top–down modulation strongly relies on the efficiency of the off-line training stage: type and quality of the training images, number of training images, etc. [18].

5) Space- and Object-Based Analysis: Inspired by the early psychophysical theories of attention [24], [25], majority of the computer vision models hypothesize "space" as the elemental unit of attention selection [12]–[16], [18], [40], [42]. Accordingly, saliency and task-relevance are investigated at the pixel level without considering the concept of an object. Increasing evidence, in the psychology and cognitive neuroscience, of "object" being one of the elemental units of attention selection [26], [36], [47]–[50] has influenced the recent computer vision models of visual attention. Many of the recent models perform object-based analysis for selective attention [17], [51], [52]. There are, however, only a few efforts which integrate space- and object-based analysis in the same framework [4], [6], [7], [53].

Although common in almost every computational model, the way these characteristics are achieved differs in different attention models and hence, the variation in performance.

B. Robotic Visual Attention: Issues and Challenges

The computer vision models of attention perform remarkably well in most of the computer vision applications where static images or images from a video stream are manually fed to the model in order to identify the most salient/task-relevant stimuli. In case of some real-time applications where the current visual input of the attention model is to be determined by the decision output of the model (i.e., the focus of attention) at the immediate past, the traditional computer vision models of visual attention faces severe limitations in a number of aspects [53]–[55]. Using a visual attention model as a component of robotic cognition is an example of such applications. In this case, the attention model should be able to locate the behaviorally relevant stimuli in an ongoing stream of visual input and respond to it, perform learning in an on-line fashion and with minimal human supervision, and apply the learned knowledge for guiding the attention behavior in arbitrary environmental settings. The issues and challenges involved with robotic visual attention are summarized below.

Issue 1: Overt Shift of Attention: In majority of robotic applications (e.g., social robots, assistive robots, and entertainment robots) it is desired that selective attention will be accompanied by a saccadic movement of the camera head of the robot. Such movement is necessary to place the object of attention at the center of the camera frame and facilitates the learning of the focused object. A computational model of attention for the robots, therefore, requires an integration of the covert and overt modes in a common framework, much like the same way the primates integrate covert and overt shift of attention. The overt shift of attention leads to the following issues that has to be solved to design a model of attention for the robot.

Issue 1.1: Change of Reference Frame: In the simplest case, the visual attention hardware of a robot consist of a color camera and a two DOF pan-tilt unit (PTU) on to which the camera is mounted. There are at least four coordinate systems involved with the overt attention mechanism: the world coordinate system is fixed while the head coordinate, camera coordinate, and image coordinate systems are changing according to the movements of the PTU. The orientation of the PTU determines which part of the environment the robot will be perceiving through its camera as shown in Fig. 6. A (ϕ, θ) amount of pan-tilt movement of the PTU causes the camera to perceive a different segment of the environment. Thus the content of the robot's visual field changes, although a considerable amount of overlap generally exists between two successive image frames. This makes the "saliency map" calculated prior to the camera movement as partially obsolete and demands either a fresh calculation of saliency or remapping of the previous saliency to the new image coordinate. The remapping supports the experimental evidence that the primates visual attention is not a memoryless process [56].

Issue 1.2: Dynamic IOR: The role of IOR in robotic attention is the same as that in the biological attention system: allowing the shift of attention toward fresh stimuli [41]. Failure to implement the IOR properly might cause a robot to oscillate between two stimuli. In overt attention, camera movement causes the location of a stimulus to shift in the image coordinate. It is, therefore, required to design a dynamic IOR strategy where

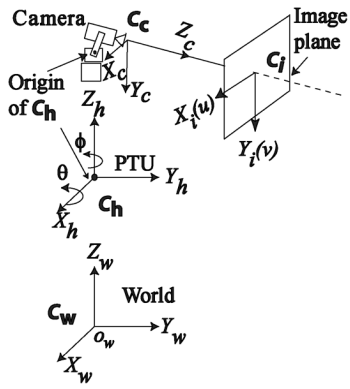
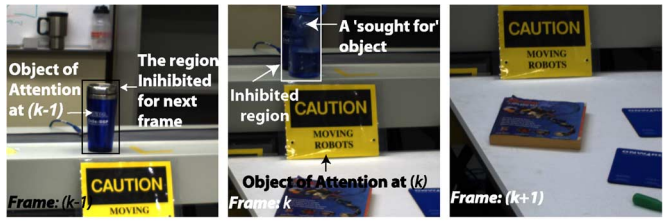


Fig. 6. Coordinate systems involved with robotic overt visual attention. The head coordinate system attached with the PTU is shown separately for clarity purpose.



(a) (b) (c)

Fig. 7. Difficulty in visual search with space-based dynamic IOR. (a) The region of the attended object at frame $(k-1)$ is made inhibited for frame k. (b) The inhibited region is mapped to the new image coordinate system at frame k. A "sough for" object appears within the inhibited region and the robot ignores its presence. (c) A random head movement in search of the "sought for" object causes it to go out of the VF. As a result, the robot requires longer time to find the inquired object.

the location of the recently attended object will me mapped in the new image coordinate in order to inhibit its candidacy as the next focus of attention. The space-based dynamic IOR introduces the complexity that if, between two successive frame capture, a new object appears at the inhibited location of a recently attended object, the robot completely ignores its presence. Fig. 7 demonstrates one instance of this problem. This incurs a longer time to identify a "sought for" object during visual search. It is, therefore, beneficial to integrate space-based IOR with its object-based counterpart. The object-based IOR, however, introduces the problem of object correspondence. In order to inhibit a recently attended object from being attended again, the robot needs to identify it in the shifted image coordinate. This is generally a challenging task due to change in camera perspective, lighting, image blurring due to camera motion, and partial appearance of objects.

Issue 1.3: Partial Appearance: Due to head movement, it is highly likely that a number of objects will partially/completely go out of the camera frame. The probability of this increases when the robot uses a narrow angle optics for the camera or when the objects are located either very close to the camera or near the periphery of the frame. Due to partial appearance, the robot might fail to identify a recently attended object. This, in turn, results in a failure to apply the IOR on it and the robot might reattend the same object. In the worst case, the attention of the robot will start oscillating among a set of objects. The same

trapped situation might also occur if the robot always finds a set of objects "attention worthy" (e.g., because of their novelty) as it can not match the partially perceived features of these objects with its memory database of previously attended and learned features. Application of space-based dynamic IOR on all of the previously attended locations might relax this problem at the expense of worsening the problem stated under *Issue 1.2*.

Issue 2: Integrated Space- and Object-Based Analysis: The space-based analysis, commonly used in majority of the computational models of attention, does not practically fulfill the interest of most robotic applications. Instead of a single pixel reported as the focus of attention, the information about the object underlying that salient pixel is of greater interest in robotic applications. The space-based model of attention, which generally relies on a traditional saliency map, does not preserve the information of the underlying objects. Another serious problem is the common practice of using coarse scales of the input image for space-based saliency map construction [13], [14], [18]. This causes the fading of many attention worthy small regions which do not get chance to be highlighted in the saliency map [57]. The problem of space-based dynamic IOR as stated under *Issue 1.2* is another consequence of space-based analysis. It is, therefore, beneficial to integrate space- and object-based analysis within the same framework.

Issue 3: Optimal Learning Strategy: This issues is particularly related to visual search. To perform a search for an object the robot needs to know the visual features of the target object. Because of the extended number of sensors and actuators, the modern-day robots are blessed with higher degrees of freedom in their visual perception. Even an static object in the environment can be perceived by the robot from arbitrary viewing angle. For a dynamic object the possibilities are even higher. To the best of our knowledge, there is no such image feature which is invariant to arbitrary affine transformation, change in viewing angle and lighting condition. Consequently, in order to identify an object in an arbitrary setting the robot requires to learn "several" views of the object. The precise number to quantify the term "several," however, is not known. The visual attention model of a robot must have a strategy to learn the visual features of an object from different view angles and with minimum human supervision.

Issue 4: Generality: As shown in Fig. 4, in majority of the computer vision models of visual attention, visual search, and visual exploration run in mutual exclusion of each other. The desired loop of attention (visual search or visual exploration) is manually activated by the programmer. Such a manual selection of visual attention mode significantly reduces the generality of an attention model and makes it unsuitable for robotic systems. A robotic visual attention model must be able to switch back-and-forth autonomously between the two modes of attention depending on the behavioral requirement.

Issue 5: Prior Training: The robotic applications can not afford to have a separate off-line training phase for visual search. A robot has a very little use as a task-assistant of human if it requires a precise training to learn every possible object prior to performing a search for it. Rather, it is generally expected in the cognitive robots that they will learn while working, much like the same way we human learn [58]. But unfortunately, the majority of current models of visual attention for the robot use a prior separate training phase to enhance the recognition performance.

Many of the research issues stated above have strong mutual dependency on each other. For instance, a strategy to deal with the changing reference frame (*Issue 1.1*) will inherently provide a solution to implement the dynamic IOR (*Issue 1.2*). Again, for the sake of generality (*Issue 4*) if we integrate visual search and visual exploration in the same framework such that the model can switch back-and-forth between the two modes, there will be no room for prior training (*Issue 5*). In other words, the learning has to be performed on-line in an integrated framework of visual search and exploration. Again, if the target-learning is performed on-line, an intelligent strategy must be devised for learning to ensure that the robot obtains enough information about the target for identification in arbitrary settings (*Issue 3*).

Addressing the *Issues 1–5* is a crucial requirement to design a sound model of visual attention for the robots. In order to meet this requirement, we observe the rise of a separate group of visual attention models dedicated solely for robotic systems. There is no doubt that this new group of models are heavily inspired by the computer vision models of visual attention, specially when it comes to the detail of visual feature processing, but they attempt to address at least some of the research issues stated above.

IV. Robotic Research on Visual Attention

A properly designed attention system provides a task-executing robot with the capacity to blend with human in natural human environment. A number of attempts are observed in robotic literature on the modeling of visual attention for robotic cognition. Many of these models propose general solution to tackle the research issues while some address them in task-specific manner. This survey classifies the existing works on robotic visual attention into two groups based on their inspiration, specific goal, and type of implementation.

1) Overt attention models: The research works in this group focus on the camera maneuvering mechanism based on the principle of overt visual attention. A considerable number of overt models are inspired by the covert attention models proposed in computer vision.
2) Application-specific visual attention models: The research works in this group develop robotic attention models which are tuned to specific task, e.g., localization, navigation, manipulation, HRI, and joint attention. Many of these task considers the property of selectivity of the primates visual attention as a mere technique to solve the desired task while some others consider visual attention as a component to design cognition in the robots. Most of the works related to HRI and joint attention fall under the second category while attention-based robot navigation, localization and manipulation are generally the members of the first category.

A brief discussion on the works under each group and how they address the research issues in Section III-B are presented in the following sections.

A. Overt Attention Models

The attention mechanism in the primates integrates the overt and covert modes of attention in a highly efficient manner. The stimulus of interest is selected covertly and then placed at the foveal region through overt movement of the eyes [59]. Evidences are also available in favor of the independent occurrence of covert and overt attention [60], [61]. In case of robotics applications, however, direction of attention mediated by eye/head movement is the most suitable choice. The major reason behind this is placing the object of interest at the center of visual field facilitates the learning process. Besides, head/eye movement of the robot provides a way for the user to understand the current gaze of the robot which is specially important in many applications (e.g., HRI). Inspired by these requirements, a number of efforts are observed in the robotic literature for modeling of overt visual attention. At the early stage of this research the principle of overt attention (to place an object of interest at the center of visual field) helped the concept of "active vision" [62], "active perception" [63], or "animate vision" [64] to be established in computer vision. For instance, the theme of "active vision" is to actively position a sensor (preferably a camera) for obtaining enriched information to solve the basic computer vision problems (e.g., shape from shading and depth computation, shape from contour, shape from texture, and structure from motion). A number of active vision models propose mechanism of positioning a camera based on the feedback from a visual attention model [65]–[70]. The major focus of most of these models is the control aspects of saccade generation and/or smooth pursuit tracking. A common practice among these works is to use some well-known covert models of attention (e.g., [12] and [13]) to identify the most interesting/salient region in the image. These active vision models, therefore, are less concerned about the research issues stated in Section III-B.

The overt attention models described in [51] and [71]–[84] are designed to implement in the robots/robotic heads as a component of their cognition. Among them, the models in [51], [71], [72], and [74] adopt different variants of the covert model NVT [13] to identify the visually salient/task-relevant stimuli and introduced different measures to deal with the research issues involved with robotic overt attention. For instance, the model in [72] addresses the *Issue 1.1* by adopting the idea of shifting the entire content of the saliency map in the direction of head movement as suggested in [8]. Another approach to address the *Issue 1.1* is to consider that attention is directed to unparsed regions of space and thereby making the perception independent of space [85]. The object-based overt attention system proposed in [51] implements a simple form of integrated object- and space-based IOR to deal with *Issue 1.2* and *Issue 2*. The overt model described in [74] suggests to remap the location of the recently attended object to the transformed image coordinate in order to implement a space-based IOR (*Issue 1.2*). The problem involved with the partial appearance of objects (*Issue 1.3*) is not noticeable in the experiments demonstrated in [74] due to the use of a wide angle camera. The model in [71] demonstrates few simple cases of overt attention and does not provide any effective solution to any of the research issues.

The neural network-based overt model reported in [77] is tightly coupled with biology (with respect to motor aspects of attention) and is focused on implementing visual exploration behavior guided by the novelty preference characteristics of primates attention. The identification of novelty in [77], however, is achieved though the implementation of space-based IOR, i.e., the robot moves to novel locations (through successive application of space-based IOR) and thereby attends to novel objects. The model [77] also relies on NVT [42] for visual saliency calculation. The issue of dynamic IOR (*Issue 1.2*) is addressed by remembering the locations of the previously visited stimuli. To comply with this strategy the model [77] assumes that all of the stimuli stay within the visual field of the robot at all times. This is a strong assumption which is valid in the experiments demonstrated in [77], but generally does not hold good in most real world scenario. The *Feature Gate* model -based [16] overt model in [78] claims to proposes a general purpose model of visual attention for the humanoid robots but mostly focuses on mimicking the feature-processing attributes of the primates attention system (e.g., log-polar retino–cortical mapping, banks of oriented filter).

All of the overt models discussed thus far follow an image-centric approach where the attention model operates absolutely in the image plane. Focus of attention is evaluated based on the content of a given image and necessary motion command is calculated based on the image dimension and the parameters of the camera optics. As opposed to this traditional image-centric approach, the recent models of overt attention adopt a robot-centric solution for attentional selection [79], [80], [84], [86]. In case of robot-centric approach it is assumed that a robot is a human-like autonomous entity which decides "what to look at?" based on its perception of surrounding with respect to an ego-centric frame of reference. For instance, the model in [84] considers an ego-sphere of infinite radius around a robotic head and the robot is able to project the perceptual information collected through different modality on the surface of the ego-sphere. The concept of the head-centric ego-sphere provides an elegant solution of the issues involved with overt shift of attention (*Issues 1.1, 1.2, 1.3*). The multimodal attention model in [84] considers both acoustic and visual information and combines them into a single head-centric saliency map by taking the maximum value between the two modes. This straightforward methodology of fusing multimodal perception into a single saliency map has several shortcomings, e.g., saliency map from different modes have same influence on the aggregated saliency map. A detailed analysis of this problem is available in [18]. The model [84] operates in a purely bottom–up fashion and performs NVT [42]-like space-based analysis for saliency calculation. The concept of an ego-sphere is also present in the attention model reported in [86]. The model [86], however, uses the principle of attention for updating a sensory ego-sphere with overlapping images perceived by the robot. The multimodal attention model in [82] also uses sensory ego-sphere to focus, learn, and then track the salient stimuli (bright colored moving objects or human faces) in the visual field. The model [82] integrates the visual search and visual exploration in the same framework and thereby eliminates the presence of a training phase during visual search (*Issue 4, 5*). The overt model in [80]

TABLE I
OVERT ATTENTION MODELS FOR THE ROBOTS

References	Synopsis	Issue(s) addressed
[72]	• Relies on [12] for saliency calculation • Considers the bottom-up effect only	• *Issue 1.1*:Re-maps the entire saliency map to the new image coordinate
[51], [73]	• Object-based attention model • Integrates top-down and bottom-up effect • Color-based bottom-up saliency	• *Issue 1.2*: Locations of the last visited objects and the objects' features are memorized • *Issue 3:* Several views of the target are learned
[74]	• Attention model for stereo-vision • Provision for top-down and bottom-up bias • No actual demonstration of top-down effect • Saliency calculation is inspired by NVT [13]	• *Issue 1.2*: Locations of the attended objects are mapped to the new image coordinate
[77]	• Tightly coupled with biology • Focuses on motor aspects of attention • Relies on NVT [13] for saliency calculation	• *Issue 1.2*: Remembers the location of the attended objects and assumes that all objects remain within visual field at all times
[84]	• Robot-centric approach • Only Bottom-up information is processed • Relies on NVT [13] for saliency calculation	• *Issues 1.1, 1.3:* Projects sensor data to an ego-centric frame of reference • *Issue 1.2:* Performs space-based IOR
[86]	• Uses visual attention to register images on a sensory ego-sphere • Relies on *Feature Gate* [16] model of attention	• *Issues 1.1, 1.3:* Projects sensor data to an ego-centric frame of reference
[82]	• Performs learning of visual attention • Object-based analysis • Focuses, learn, and tracks bright colored moving objects and the human faces in the consecutive frames	• *Issues 1.1, 1.3:* Projects sensor data to an ego-centric frame of reference • *Issue 1.2:* Performs object-based IOR • *Issue 5:* Performs on-line learning of target
[80], [87]	• Attention model for tracking • Uses color information only to calculate scene saliency	• *Issue 1.1:* Projects sensor data to a'scene space' expressed in head-centric coordinate system • *Issue 1.2:* Performs object-based IOR
[79]	• Multi-modal attention model • Considers top-down and bottom-up effects • Requires a lot of prior learning to create an 'attention map'	• *Issue 1.1:* Projects sensor data to a global 'attention map' of the surrounding • *Issue 4:* Switches back-and-forth between visual search and exploration through speech command
[55] [88]	• Bayesian overt attention model • Integrate top-down and bottom-up effects • Exhibit preference for novel stimuli • Performs auditory modulation of visual attention through occasional human-interaction	• *Issue 1.1:* Robot-centric analysis of visual features • *Issue 1.2:* Space and object-based dynamic IOR • *Issue 4:* Switches back-and-forth between visual search and exploration through speech command • *Issues 3, 5:* On-line learning mediated by a self-directed learning strategy

uses the term "scene space" instead of "ego-sphere" to represent a two dimensional surface which contains the information perceived by the robot with respect to the robot's head-centric coordinate system. The purpose of the model [80], however, is to track a set of predefined object in the surrounding. To achieve this goal it uses only the color information of the target object and performs object-based analysis to implement the IOR (*Issue 1.2, 2*). Although the model might have the potential to be extended for complex attention scenario, the current implementation in [80] is dealing with only few simple cases. The models in [89] and [90] use scaffolding where the human operator heavily guides the robot to teach what to focus on through speech command and hand-gesture. This solves the problem of prior training (*Issue 5*) and optimal learning strategy (*Issue 3*) with the price of having a dedicated human operator throughout the attention process. Unfortunately, having such a dedicated human operator severs the generality problem (*Issue 4*). A reduced amount of human-dependency for learning of attention is observed in the multimodal overt attention model described in [79]. The model [79] proposes the idea of an attention map, similar to "probabilistic occupancy grid" widely used in robotic mapping [91], to encode the saliency of the robot's surrounding. The attention map can be modulated by the task-demand conveyed to the robot through speech command. The model, however, requires significant amount of prior training and manual work to create an useful attention map for any specific robotic application.

For quick reference, Table I shows a comparative analysis of the overt attention models discussed in this section.

B. Application-Specific Visual Attention Models

The application-specific visual attention models are tuned to the applications they are developed for. Visual attention mechanism has at least two properties which can be tuned in an application specific manner.

- Selectivity: The basic idea of attention is to focus on a relevant visual stimulus for further processing. The "relevancy" of a stimulus can be defined in terms of its similarity with a set of predefined task-relevant features. The irrelevant information in the visual scene are not considered for further processing and thereby reducing the computational load of an artificial system.
- Visual Search: Visual search is an important property of the primates visual attention mechanism which helps to focus on the target-related information in relative exclusion of the others. Thus, the visual search is an special case of demonstration of selectivity. The success of a visual search and the time requirement depends on the number of distractor stimuli present in the visual field and the number of features they share with the target stimulus.

Exploitation of these two properties often causes visual attention to reduce to a tracking problem in many application-specific models of visual attention. In case of attention-based tracking, many of the research issues stated in Section III-B do not arise.

For instance, the object that is to be tracked is learned once and is tracked in the subsequent camera frames. Each incoming camera frame is searched for this specific object. Thus, there is no need to implement the IOR and the change of coordinates does not have any significance effect on the tracking decision [hence, no need to address (Issues 1.1, 1.2)]. An example of such attention-based tracking is demonstrated in [92]. Here, a covert model of visual attention VOCUS [18] is used to perform simultaneous localization and mapping (SLAM) by a mobile robot [92]. The role of the attention model is to identify the most salient stimuli in the scene and then keep on tracking that specific stimuli in the successive frames by adjusting the camera head. Similar strategy of attention-based tracking is also adopted in [93] for vision-based SLAM by mobile robots. To deal with the partial appearance of object (*Issue 1.3*) the model in [92] adopts the strategy that the landmarks that reside at the center of the visual field are given higher priority as it is likely that they can be tracked for an extended period of time.

The attention model in [81] exploits the principle of visual search for robot navigation and mapping. The robot learns the visual features of a set of objects during an off-line training phase. During the autonomous navigation the robot searches for the learned objects, which appear as landmarks, in natural indoor environment. A number of important parameters of the navigation model is chosen based on the off-line training phase. The objects location are projected in an ego-centric frame of reference in order to update a 3-D occupancy grid which contains the information about the landmarks/obstacles in the robot's workspace (*Issues 1.1, 1.2, 1.3*). The model described in [83] is dedicated to design a Bayesian approach of fast visual search for human faces in a video stream. To achieve faster response the attention model sacrifices all other visual information except the intensity feature. Similar to [92], this model [83] also considers each incoming frame as an isolated static image and does not implement IOR. Similar kind of attention model (focusing on the visual search) is also proposed in [94]. Here, the robot is provided with a predefined set of features to search for, e.g., a talking person, human face, human legs located at the closest distance, etc. The robot then uses its multimodal perception to search for this features and attend to them. The task-specific attention model proposed in [105] performs visual search for prespecified object patterns (dominos) and executes manipulative actions based on their 3-D locations. The attention model in [95] is designed for social interaction with human. The model uses omni-directional camera and the nature of images obtained from such cameras enables the visual features to be registered directly in an ego-centric frame of reference. This inherently offers a solution to the problem of coordinate change (*Issue 1.1*), dynamic IOR (*Issue 1.2*), and partial appearance (*Issue 1.3*). The neural-network-based attention models in [106] and [107] performs navigation and object manipulation while combining the top–down and bottom up influence. The learning mechanism employed in these models [106], [107] as a part of the enactive vision system enable them to address the issues of prior training and coordinate change.

The visual attention models developed for HRI mostly consider attention as a step toward making the robots cognitive. Visual attention plays a significant role in HRI in order to establish joint attention [108] between the robot and the human. Establishing joint attention between a human and robot requires that a robot should be able to detect and manipulate the attention of the human, socially interact with the human, and finally see itself as well as the human as intentional agents. Joint attention, therefore, is an excellent tool to build a meaningful HRI system. A basic requirement of joint attention is that the robot should posses a human like attention model with the capacity to manipulate attention of other agent, as well as of being manipulated by other agents. The visual attention models proposed in HRI literature, therefore, have strong emphasis on top–down modulation of attention. A number of approaches, inspired by the cognitive development of human child, are available to model the top–down influence in attention selection, e.g., imitation learning, scaffolding. These methodologies have their own unique way of addressing the issues sated in Section III-B. In some cases, however, their way of addressing one issue severs the consequence of the others.

In case of imitation-based learning of attention, the robot imitates the movements (head/eye/hand) of a person (the user or the operator) to exhibit overt attention behavior [109]. Thus the top–down bias appears as the commands from the human operator conveyed through natural speech, hand gesture, gaze direction, etc. For instance, the models in [96] and [97] evaluate the gaze direction of the user to identify the object of interest to attend. Thus, the model guides a robot to look at the objects to which its user is also looking and thereby establishing simultaneous looking behavior which is a major requirement of joint attention [108]. The work in [99] uses the head pose and eye-gaze direction of the user to identify the object to attend. To further enhance the quality of joint attention it uses pointing behavior by the robot once it attends to an object. The shared attention model in [102] and [103] uses the gaze direction as a cue to decide which object to attend. An integration of imitation learning and visual search is observed in the connectionist model of joint attention reported in [100] and [101] where the robot learns a set of motion patterns in an off-line training phase and reproduces them when it finds similar kind of motion pattern performed by the user. The model introduced in [110] performs overt attention based on gaze direction of the user as well as spoken command. A major complexity of imitation learning is in order to be accurate it requires the robot to have an efficient learning strategy to conceptualize the underlying goal of the imitative actions and form knowledge from that [111]. In other words, the robot has to decide on its own about "what to imitate?" which by itself, is a type of skill that requires cognition.

A bit more relaxed approach (with respect to the amount of cognitive load on the robot) as compared to imitation learning is attention mediated by scaffolding [112]. Here the idea is to explicitly attract the attention of the robot to certain specific stimuli through different kind of actions, e.g., verbal command, hand-gesture, and motionese. For instance, the attention model in [89] uses hand-gesture and verbal command to guide the attention of the robot toward novel objects. Similar approach of attention guiding has been used in [90] in order to perform grasping task by a robot manipulator. The attention model in [104] uses motionese in order to make certain stimuli to appear as extremely salient in the robot's perception. The model

TABLE II
APPLICATION-SPECIFIC MODELS OF ROBOTIC VISUAL ATTENTION

Application	Synopsis (of the attention model)	Issue(s) addressed
Vision-based SLAM [92]	• Extension of VOCUS [18] • Identify the most salient region in a frame and tracks it	• *Issue 1.3:* Higher priority is given to the stimuli located at the center of image
Navigation and mapping [81]	• Attention model for visual search • Learns objects in an off-line training phase and detects them later	• *Issue 1.1:* Update a 3D occupancy grid with locations of the object in
Search for human face [83]	• A Bayesian model of visual search • Searches for face features in each incoming camera frame • Only intensity feature is used	Visual attention reduces to target tracking problem which does not require to address the research issues
People tracking [94]	• Model of visual search to track multiple person • Abstract level information about target are given to the robot and multi-modal data are analyzed to identify people	Visual attention reduces to target tracking problem which does not require to address the research issues
Social interaction with human [95]	• Multi modal model integrating vision and audition to focus on human • Option for top-down and bottom-up influences is available but no actual demonstration of top-down effect	• *Issues 1.1, 1.2, 1.3:* Omni-directional vision is used to maintain a 180^o wide attentional span
Joint attention in HRI [96], [97] [98]	• Considers the caregiver's face as the most salient region • Calculates the gaze direction of the caregiver from his/her face • Attends overtly to the object(s) the caregiver is looking at	• *Issues 1.1, 1.2, 1.3, 4, 5:* Use of imitation learning approach let the human operator to take care of these issues
HRI [99]	• Learns sensorimotor mapping through active interaction • Uses head-pose and eye-gaze direction to identify the next focus	• *Issues 1.1, 1.2, 1.3, 4, 5:* Use of imitation learning approach let the human operator to take care of these issues
HRI [100], [101]	• Learns motion patterns off line and reproduces them when similar motion patterns are observed in the environment	• *Issues 1.1, 1.2, 1.3, 4, 5:* Use of imitation learning approach let the human operator to take care of these issues
Joint attention [102], [103]	• Uses the gaze direction of the human to identify the next focus of attention	• *Issues 1.1, 1.2, 1.3, 4, 5:* Use of imitation learning approach let the human operator to take care of these issues
HRI [89], [90]	• Hand-gesture and verbal command are used to guide the robot's attention to a novel object • Object-based analysis	• *Issues 1.1, 1.2, 1.3, 3:* Uses scaffolding to draw the attention of the robot to the target object • *I4:* Switches back-and-forth between visual search and exploration through speech command
Joint attention [104]	• Relies on NVT for saliency calculation • Use motionese to make some regions highly • salient to the robot and thus drawing • the attention toward that regions	• *Issues 1.1, 1.2, 1.3, 3:* Uses scaffolding to draw the attention of the robot to the target objects
HRI [75]	• Implements Guided Search [25] • Attends to task-specific stimuli (e.g. face, colored toy) based on task demand • Motionese is used to make the task-specific stimuli as the most salient	• *Issues 1.1, 1.2, 1.3, 3:* Uses scaffolding to draw the attention of the robot to the target objects

[104], however, relies on NVT [42] for calculation of saliency, does not implement any form of IOR, and operates in a pure off-line fashion. The attention model developed for HRI in [75] is based on the psychophysical model of visual search proposed in [25]. The model is sensitive to task-specific stimuli (e.g., human face, toys with specific color) and attends to them based on the task-context. This model also uses motionese to guide the robot's attention toward certain specific stimuli. The model performs the IOR and the habituation effect with moving camera but does not mention explicitly how the issues involved with camera movement have been addressed.

The imitation learning approach and scaffolding relieve a visual attention model from worrying about the issues such as change of image coordinates (*Issue 1.1*), implementation of IOR (*Issue 1.2*), partial appearance of the objects (*Issue 1.3*), and generality (*Issue 3*). The human operator takes care of these issues and the robot's attention model just mimics the operator. Such a huge benefit, however, comes with the heavy price that a human operator must be dedicated for a robot, which is often an unrealistic demand for autonomous robotic applications.

For quick reference, the Table II shows a comparative analysis of the application-specific attention models discussed in this section.

V. GENERAL DISCUSSION

Robotic models of visual attention are generally inspired by the computer vision models. The rich literature of computer vision on the computational modeling of visual attention mostly focuses on developing covert model of visual attention. Although these models are remarkably successful in identifying the most salient stimuli or performing a visual search in a static image, their use for robotic visual attention is restricted by a number of real-world design issues. A number of these issues are involved with the fact that robotic visual attention are generally overt in nature as opposed to the covert notion of attention commonly followed by the computer vision models.

The others are involved with the architecture of the computer vision models of visual attention. Because of these research issues the robotic models of visual attention are steadily drifting apart from the computer vision models with respect to architecture and implementation methodologies.

The survey presented in this article shows that a very popular choice to tackle the problem arising from camera movements in overt attention (*Issue 1*) is the spatio–temporal transformation of the visual features. In case of such spatio–temporal transformation, the saliency map is projected to the new camera coordinates [72]. To solve the IOR issue either the locations of the previously attended objects are mapped to the new frame [74], [77], or an object-based IOR is implemented [51], [73]. A major reason of popularity of the spatio–temporal transformation (to tackle the issues related to camera movements) is it keeps the process of constructing the saliency map same as the way it is generally constructed in the famous covert models, e.g., Koch's model [12], NVT [13], FeatureGate [16], and VOCUS [18]. A parallel approach, however, started to gain even more popularity than the spatio–temporal transformation-based approach. This new approach tackles the whole problem of visual attention from a robot-centric perspective and advocates the idea of an ego-centric saliency map [79], [80], [82], [84], [86]. Some efforts are also observed to develop such map using panoramic camera [95]. The reason of increasing popularity of the robot-centric approach is its natural ability to tackle the issues related to head–eye movements of the robot.

The most popular choice to address *Issue 3–Issue 5* discussed in Section III-B is the learning of visual attention. The learning is generally mediated by human interaction and through using multiple modalities. But the role of human in the learning process differs in different models, e.g., full human guidance as in the imitation learning [96]–[98], [100], [101] and scaffolding [75], [89], [90], [104], [112], and occasional guidance as in [88].

The survey presented in this article clearly shows that the research on robotic models of visual attention is far from perfection. There are only a handful of models which attempts to address the research issues involved with robotic attention in a generic manner. Aside from these research issues, a general constraint of the existing computational models of visual attention is they limit the top–down influence in attentional selection only to the case of visual search. In reality, the top–down influence plays a major role in the primates visual attention mechanism. In case of the primates the top–down influence in attentional selection, however, is the result of a complex interaction among knowledge, emotion, personality, and reasoning. Designing a visual attention model for robotic cognition with such kind of top-down influence will require a dynamic interaction with other components of artificial cognition, e.g., reasoning, planning, emotion, and knowledge-representation. The perfection in the modeling of robotic visual attention, therefore, is closely related to the perfection in the modeling of other cognitive functions. Few efforts are observed in the current literature on the modeling of value system, human-like reasoning and knowledge representation for robotic systems [113]–[116]. These efforts, however, are mostly discrete and do not investigate the effect of other cognitive abilities on the visual attention behavior. A major reason for this lack of investigation is that the underlying neural mechanism of many of the cognitive functions and their mutual effect on the cognitive development are still unknown to the researchers. We can, therefore, hope that further development on the modeling of robotic visual attention will go hand-to-hand with the improvement of our understanding about human cognition. A major consequence of having a full-fledged primates-like visual attention model is that it will advance the current efforts on employing robots for better understanding of the human attention behavior [117]. This will be a magnificent way to use technological advancement for the understanding of human abilities.

VI. Conclusion

This paper has presented a survey of the literature on computational modeling of visual attention with a special focus on the models of visual attention designed for robotic cognition. The paper has identified a set of research issues that are crucial for designing visual attention models which will serve as a component of robotic cognition. The paper then presents an analysis of the existing works on robotic visual attention with respect to these research issues.

In the primates, visual attention is submerged in their perception, action, and in many of the other cognitive functions. In addition to its trivial manifestation in visual exploration and visual search, visual attention works underneath the action execution, planning, reasoning, and decision making process of the primates [3]. Mimicking the visual attention of the primates in the robotic system will not be complete until we explore this hidden influence of attention in the overall cognition of the primates. Besides, the use of visual attention as an stand alone ability of the robot is far less appealing than the case where visual attention works in conjunction with reasoning, decision making and action planning of the robot. This makes the robots a bit more cognitive than the way they are now. Such robots have increasingly growing demand in service industries, assistive and health-care sectors, and entertainment robotics.

References

[1] Y. Bar-Cohen and C. Breazeal, *Biologically Inspired Intelligent Robots*. New York: SPIE Press, 2003.
[2] B. Webb and T. R. Consi, *Biorobotics*. Cambridge, MA: The MIT press, 2001.
[3] D. Drubach, *The Brain Explained*. Englewood Cliffs, NJ: Prentice Hall Health, 2000.
[4] G. Deco and T. S. Lee, "A unified model of spatial and object attention based on inter-cortical biased competition," *Neurocomputing*, vol. 44, pp. 775–781, 2002.
[5] G. Deco and E. T. Rolls, "A neurodynamical cortical model of visual attention and invariant object recognition," *Vis. Res.*, vol. 44, pp. 621–642, 2004.
[6] L. J. Lanyon and S. L. Denham, "A model of active visual search with object-based attention guiding scan paths," *Neural Netw.*, vol. 17, pp. 873–897, 2004.
[7] F. H. Hamker, *Distributed Competition in Directed Attention*. Berlin, Germany: Akademische Verlagsgesellschaft, 2000, pp. 39–44.
[8] P. F. Dominey and M. A. Arbib, "A cortico-subcortical model for generation of spatially accurate sequential saccades," *Cerebral Cortex*, vol. 2, pp. 153–175, 1992.
[9] A. Pouget and T. J. Sejnowski, "Spatial transformation in the parietal cortex using basis functions," *J. Cogn. Neurosci.*, vol. 9, pp. 222–237, 1997.

[10] R. P. N. Rao, "Bayesian inference and attentional modulation in the visual cortex," *Cogn. Neurosci. Neuropsychol.*, vol. 16, pp. 1843–1848, 2005.
[11] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Rev.: Neurosci.*, vol. 2, pp. 194–203, 2001.
[12] C. Koch and S. Ullman, "Shifts in selective visual attention: Toward the underlying neural circuitry," *Human Neurobiol.*, vol. 4, pp. 219–227, 1985.
[13] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
[14] V. Navalpakkam and L. Itti, "Top-down attention selection is fine-grained," *J. Vis.*, vol. 6, pp. 1180–1193, 2006.
[15] J. K. Tsotsos, S. Culhane, Y. Winky, L. Yuzhong, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artif. Intell.*, vol. 78, pp. 507–545, 1995.
[16] K. R. Cave, "The featuregate model of visual selection," *Psychol. Res.*, vol. 62, pp. 182–194, 1999.
[17] Y. Sun and R. Fisher, "Object-based visual attention for computer vision," *Artif. Intell.*, vol. 146, pp. 77–123, 2003.
[18] S. Frintrop*, VOCUS: A Visual Attention System for Object Detection and Goal-directed Search*. Heidelberg, Germany: Springer-Verlag, 2006, vol. 3899, Lecture Notes in Artif. Intell. (LNAI), 3-540-32759-2.
[19] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annu. Rev. Neursci.*, vol. 18, pp. 193–222, 1995.
[20] R. Fantz, "Visual experience in infants: Decreased attention to familiar patterns relative to novel ones," *Science*, vol. 146, pp. 364–370, 1964.
[21] L. E. Bahrick, M. H.-R. , and J. N. Pickens, "The effect of retrieval cues on visual preferences and memory in infancy: Evidence for a four-phase attention function," *J. Exp. Child Psychol.*, vol. 67, pp. 1–20, 1997.
[22] C. Bundense, "A theory of visual attention," *Psychol. Rev.*, vol. 97, pp. 523–527, 1990.
[23] R. E. Walley and T. D. Weiden, "Lateral inhibition and cognitive masking: A neuropsychological theory of attention," *Psychol. Rev.*, vol. 80, pp. 1284–302, 1973.
[24] A. M. Treisman and G. Gelade, "A feature integration theory of attention," *Cogn. Psychol.*, vol. 12, pp. 97–136, 1980.
[25] J. M. Wolfe, K. Cave, and S. Franzel, "Guided search: An alternative to the feature integration model for visual search," *J. Exp. Psychol.: Human Percept. Perform.*, vol. 15, pp. 419–433, 1989.
[26] G. D. Logan, "The CODE theory of visual attention: An integration of space-based and object-based attention," *Psychol. Rev.*, vol. 103, pp. 603–649, 1996.
[27] P. T. Quinlan, "Visual feature integration theory: Past, present, and future," *Psychol. Bulletin*, vol. 129, pp. 643–673, 2003.
[28] J. M. Wolfe, "Guided search 2.0: A revised model of visual search," *Psych. Bulletin and Rev.*, vol. 1, pp. 202–238, 1994.
[29] J. M. Wolfe and G. Grancarz*, Guided Search 3.0: Basic and Clinical Applications of Vision Science*. Norwell, MA: Kluwer Academic, 1996, pp. 189–192.
[30] J. M. Wolfe, "Guided search 4.0: A guided search model that does not require memory for rejected distractor," *J. Vis.*, vol. 1, p. 349a, 2001.
[31] M. P. Oeffelen and P. G. Voss, "Configurational effects on the enumeration of dots: Counting by groups," *Memory Cogn.*, vol. 10, pp. 396–404, 1982.
[32] D. Heinke and G. W. Humphreys*, Computational models of visual selective attention: A review*. London, U.K.: Psychology Press, pp. 273–312.
[33] J. K. Tsotsos, L. Itti, and G. Rees*, A Brief and Selective History of Attention*. Amsterdam, The Netherlands: Elsevier, 2005.
[34] L. Chelazzi, J. Duncan, E. K. Miller, and R. Desimone, "Responses of neurons in inferior temporal cortex during memory guided visual search," *J. Neurophysiol.*, vol. 80, no. 6, pp. 2918–2940, 1998.
[35] S. Kastner and L. G. Ungerleider, "The neural basis of biased competition in human visual cortex," *Neuropsychologia*, vol. 39, pp. 1263–1276, 2001.
[36] J. Duncan, "Selective attention and the organization of visual information," *J. Exp. Psychol.*, vol. 113, pp. 501–513, 1984.
[37] S. Luck, L. Chelazzi, S. A. Hillyard, and R. Desimone, "Neural mechanisms for directed visual attention," *J. Neurophysiol.*, vol. 77, pp. 24–42, 1997.
[38] G. Deco*, Biased Competitive Mechanisms for Visual Attention in a Multi-Modular Neurodyanmical System*. Berlin, Germany: Springer-Verlag, 2001, pp. 114–126, LNAI 2036.
[39] L. J. Lanyon and S. L. Denham, "A biased competition computational model of spatial and object-based attention mediating active visual search," *Neurocomputing*, vol. 58–60, pp. 655–662, 2004.
[40] R. Milanese, "Detecting Salient Regions in an Image: From Biological Evidence to Computer Implementation," Ph.D. dissertation, Univ. of Geneva, Geneva, Switzerland, 1993.
[41] M. I. Posner and Y. Cohen*, Components of Visual Orienting*. Hillsdale, NJ: Erlbaum, 1984, pp. 531–556.
[42] L. Itti and C. Koch, "A saliency based search mechanism for overt and covert shift of visual attention," *Vis. Res.*, vol. 40, pp. 1489–1506, 2000.
[43] S. J. Dickinson, H. I. Christensen, J. K. Tsotsos, and G. Olofsson, "Active object recognition integrating attention and viewpoint control," *Comput. Vis. Image Understand.*, vol. 67, pp. 239–260, 1997.
[44] J. K. Tsotsos, Y. Liua, J. C. M. Trujilloc, M. Pomplund, E. Siminea, and K. Zhoua, "Attending to visual motion," *Comput. Vis. Image Understand.*, vol. 100, pp. 2–40, 2005.
[45] W. James*, Principle of Psychology*. New York: Holt, 1890.
[46] R. Milanese, H. Wechsler, S. Gil, J. Bost, and T. Pun, "Integration of top-down and bottom-up cues for visual attention using non-linear relaxation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn.*, Seattle, WA, 1994, pp. 781–785.
[47] R. A. Rensink, "The dynamic representation of scenes," *Vis. Cogn.*, vol. 7, pp. 17–42, 2000.
[48] S. Yantis and J. T. Serences, "Cortical mechanisms of space-based and object-based attentional control," *Current Opinion Neurobiol.*, vol. 13, pp. 187–193, 2003.
[49] M. Goldsmith and M. Yeari, "Modulation of object-based attention by spatial focus under endogenous and exogenous orienting," *J. Exp. Psychol.: Human Percept. Perform.*, vol. 29, pp. 897–918, 2003.
[50] B. J. Schol, "Objects and attention: The state of the art," *Cognition*, vol. 80, pp. 1–46, 2001.
[51] F. Orabona, G. Metta, and G. Sandini, "Object-based visual attention: A model for a behaving robot," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, San Diego, CA, 2005.
[52] T. Wu, J. Gao, and Q. Zhao, "A computational model of object-based selective visual attention mechanism in visual information acquisition," in *Proc. IEEE Conf. Inform. Acquisition*, 2004, pp. 405–409.
[53] M. Begum, G. Mann, R. Gosine, and F. Karray, "Object- and space-based visual attention: An integrated framework for autonomous robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Nice, France, 2008, pp. 301–306.
[54] M. Begum, F. Karray, G. Mann, and R. Gosine, "A biologically inspired probabilistic model of visual attention for cognitive robots," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, 2010, 10.1109/TSMCB.2009.2037511), accepted for publication.
[55] M. Begum, F. Karray, G. Mann, and R. Gosine, "Re-mapping of visual saliency in overt attention: A particle filter approach for robotic systems," in *Proc. IEEE Int. Conf. Robotic. Biomimetics*, Bangkok, Thailand, 2008.
[56] M. S. Peterson, A. F. Kramer, R. F. Wang, D. E. Irwin, and J. S. McCarley, "Visual search has memory," *Psychol. Sci.*, vol. 12, pp. 287–292, 2002.
[57] M. Z. Aziz and B. Mertsching, "Fast and robust generation of feature maps for region-based visual attention," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 633–644, May 2008.
[58] G. Sandini, G. Metta, and D. Vernon, "The icub cognitive humanoid robot: An open-system research platform for enactive cognition," *Lecture Notes Comput. Sci.*, vol. 15, pp. 358–369, 2007.
[59] H. Deubel and W. X. Schneider, "Saccade target selection and object recognition: Evidence for a common attentional mechanism," *Vis. Res.*, vol. 36, pp. 1827–1837, 1996.
[60] R. Johnason, G. Westling, A. Backstrom, and J. Flanagan, "Eye-hand coordination in object manipulation," *J. Neurosci.*, vol. 21, pp. 6917–6932, 2001.
[61] J. M. Findlay and I. D. Gilcrist*, Active Vision Perspective*. Berlin, Germany: Springer-Verlag, 2001, pp. 83–103.
[62] J. Aloimonos, I. Weiss, and A. Badyopadhyay, "Active vision," *Int. J. Comput. Vis.*, vol. 1, pp. 333–356, 1988.
[63] R. Bajscy, "Active perception," *Proc. IEEE*, vol. 76, pp. 996–1005, 1988.
[64] D. Ballard, "Animate vision," *Artif. Intell.*, vol. 48, pp. 57–86, 1991.
[65] J. J. Clark, "Spatial attention and saccadic camera motion," in *Proc. IEEE Int. Conf. Robot. Autom.*, Leuven, Belgium, 1998, pp. 3247–3252.
[66] J. J. Clark and N. J. Ferrier, "Modal control of an attentive vision system," in *Proc. IEEE Int. Conf. Comput. Vis.*, Tarpon Springs, FL, 1988, pp. 514–523.
[67] L. Manfredi, E. S. Maini, P. Dario, C. Laschi, B. Girard, N. Tabareau, and A. Berthoz, "Implementation of neurophysiological model of saccadic eye movements on an anthropomorphic robotic head," in *Proc. IEEE/RSJ Int. Conf. Human. Robot.*, Genova, Italy, 2006, pp. 438–443.

[68] D. Coombs and C. Brown, "Real-time smooth pursuit tracking for a moving binocular robot," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, San Francisco, CA, 1992, pp. 23–28.

[69] C. Brown, "Prediction and cooperation in gaze control," *Biol. Cybern.*, vol. 63, pp. 61–70, 1990.

[70] J. A. Driscoll, R. A. Peters, and K. R. Cave, "A visual attention network for a humanoid robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Victoria, BC, Canada, 1998, pp. 1968–1974.

[71] A. Ude, V. Wyart, L.-H. Lin, and G. Cheng, "Distributed visual attention on a humanoid robot," in *Proc. IEEE-RAS Int. Conf. Human. Robot.*, Tsukuba, Japan, 2005, pp. 381–386.

[72] S. Vijayakumar, J. Conrad, T. Shibata, and S. Schaal, "Overt visual attention for a humanoid robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. System.*, Maui, HI, 2001, pp. 2332–2337.

[73] G. Metta, "An attentional system for humanoid robot exploiting space variant vision," in *Proc. IEEE-RAS Int. Conf. Human. Robot.*, Tokyo, Japan, 2001.

[74] A. Dankers, N. Barnes, and A. Zelinsky, "A reactive vision system: Active-dynamic saliency," in *Proc. Int. Conf. Comput. Vis. Syst.*, Bielfeld, Germany, 2007.

[75] C. Breazeal, A. Edsinger, P. Fitzpatrick, and B. Scassellati, "Active vision for sociable robots," *IEEE Trans. System, Man, and Cybern., A, Syst., Humans*, vol. 31, no. 5, pp. 443–453, Sep. 2001.

[76] R. Fay, U. Kaufmann, and A. Knoblauch, *Combining Visual Attention, Object Recognition and Associative Information Processing in a Neurobotic System*. Berlin, Germany: Springer-Verlag, 2005, pp. 117–142.

[77] J. Vitay, N. P. Rougier, and F. Alexandre, *A Distributed Model of Spatial Visual Attention*. Berlin, Germany: Springer-Verlag, 2005, pp. 54–72.

[78] O. Stasse, Y. Kuniyoshi, and G. Cheng, "Development of a biologically inspired real-time visual attention system," in *Proc. IEEE Int. Workshop Biol. Motivated Comput. Vis.*, Seoul, Korea, 2000, pp. 150–159.

[79] J. L. Crespo, A. Faina, and R. J. Duro, "An adaptive detection/attention mechanism for real time robot operation," *Neurocomputing*, vol. 72, pp. 850–860, 2009.

[80] J. M. Canas, M. M. Casa, and T. Gonzalez, "An overt visual attention mechanism based on saliency dynamics," *Int. J. Intell. Comput. Medical Sci. Image Process.*, vol. 2, pp. 93–100, 2008.

[81] F. Saidi, O. Stasse, and K. Yokoi, "A visual attention framework for search behavior by a humanoid robot," in *Proc. IEEE Int. Conf. Human. Robot.*, Genova, Italy, 2006, pp. 346–351.

[82] L. Aryananda, "Attending to learn and learning to attend for a social robot," in *Proc. IEEE Int. Conf. Human. Robot.*, Genova, Italy, 2006, pp. 618–623.

[83] N. J. Butko, L. Zhang, G. W. Cottrell, and J. R. Movellan, "Visual saliency models for robot cameras," in *Proc. IEEE Int. Conf. Robot. Autom.*, Pasadena, CA, 2008, pp. 2398–2403.

[84] J. Ruesch, M. Lopes, A. Bernardino, J. Hornstein, J. S. Victor, and R. Pfeifer, "Multi modal saliency-based bottom-up attention: A framework for the humanoid robot icub," in *Proc. IEEE Int. Conf. Robot. Autom.*, Pasadena, CA, 2008, pp. 962–967.

[85] A. Murata and H. Iwase, "Visual attention models-object-based theory of visual attention," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Tokyo, Japan, 1999, pp. 60–65.

[86] K. A. Fleming and R. E. B. R. A. Peter, II, "Image mapping and visual attention on a sensory ego-sphere," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Beijing, China, 2006, pp. 241–246.

[87] J. M. Canas, M. M. Casa, P. Bustos, and P. Bachiller, "Overt visual attention inside jde control architecture," in *Proc. Port. Conf. Artif. Intell.*, Covilhã, Portugal, 2005, pp. 226–229.

[88] M. Begum, F. Karray, G. Mann, and R. Gosine, "A probabilistic approach for attention-based multi-modal human-robot interaction," in *Proc. IEEE Int. Symp. Robot Human Interact. Commun.*, La Jolla, CA, 2009.

[89] A. Haasch, N. Hofemann, J. Fritsch, and G. Sagerer, "A multi-modal object attention system for a mobile robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. System.*, Barcelona, Spain, 2005, pp. 2712–2717.

[90] P. Mcguire, J. Fritsch, J. J. Steil, F. Roethling, G. A. Fink, S. Wachsmuth, G. Sagerer, and H. Ritter, "Multi-modal human-machine communication for instructing robot grasping tasks," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. System.*, 2002, pp. 1082–1088.

[91] A. Elfes, "Sonar-based real-world mapping and navigation," *IEEE J. Robotic. Autom.*, vol. 3, pp. 249–265, 1987.

[92] S. Frintrop and P. Jensfelt, "Attentional landmarks and active gaze control for visual slam," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 1054–1065, Oct., 2008.

[93] A. Davison and D. Murray, "Simultaneous localization and map-building using active vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 865–880, Jul. 2002.

[94] S. Lang, M. Kleinehagenbrock, S. Hohenner, J. Fritsch, G. A. Fink, and G. Sagerer, "Providing the basis for human-robot-interaction: A multi-modal attention system for a mobile robot," in *Proc. Int. Conf. Multi Modal Interfaces*, Vancouver, BC, Canada, 2003, pp. 28–35.

[95] O. Déniz, M. Castrillón, J. Lorenzo, M. Hernyández, and J. Méndez, "Multimodal attention system for an interactive robot," in *Lectures Notes in Computer Science, vol.2652. First Iberian Conf. Pattern Recognition and Image Analysis*, 2003, pp. 212–220.

[96] Y. Nagai, K. Hosoda, and M. Asada, "Joint attention emerges through bootstrap learning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Las Vegas, NV, 2003, pp. 168–173.

[97] Y. Nagai, M. Asada, and K. Hosoda, "A developmental approach accelerates learning of joint attention," in *Proc. IEEE Int. Conf. Develop. Learn.*, Cambridge, MA, 2002, pp. 277–282.

[98] H. Sumioka, K. Hosoda, Y. Yoshikawa, and M. Asada, "Acquisition of joint attention through natural interaction utilizing motion cues," *Adv. Robot.*, vol. 21, pp. 983–999, 2007.

[99] M. W. Doniec, G. Sun, and B. Scassellati, "Active learning of joint attention," in *Proc. IEEE/RAS Int. Conf. Human. Robot.*, Genova, Italy, 2006, pp. 34–39.

[100] M. Ito and J. Tani, "On-line imitative interaction with a humanoid robot using a dynamic neural network model of a mirror system," *Adapt. Behav.*, vol. 12, pp. 93–115, 2004.

[101] M. Ito and J. Tani, "Joint attention between a humanoid robot and users in imitation game," in *Proc. IEEE Int. Conf. Develop. Learn.*, La Jolla, CA, 2004, pp. 277–282.

[102] A. P. Shon, J. J. Storz, and R. P. N. Rao, "Toward a real-time bayesian imitation system for a humanoid robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, Rome, Italy, 2007, pp. 2847–2852.

[103] M. W. Hoffman, D. B. Grimes, A. P. Shon, and R. P. N. Rao, "A probabilistic model of gaze imitation and shared attention," *Neural Netw.*, vol. 19, pp. 299–309, 2006.

[104] Y. Nagai and K. J. Rohlfing, "Computational analysis of motionese toward scaffolding based robot action learning," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 1, May 2009.

[105] M. Bollmann, R. Hoischen, M. Jesikiewicz, C. Justkowski, and B. Mertsching, "Playing domino: A case study for an active vision system," in *Proc. 1st Int. Conf. Computer Vis. Syst.*, Las Palmas, Spain, 1999, vol. LNCS 1542, pp. 392–411.

[106] M. Suzuki and D. Floreano, "Enactive robot vision," *Adapt. Behav.*, pp. 122–128, 2008.

[107] S. Jeong, M. Lee, H. Arie, and J. Tani, "Developmental learning of integrating visual attention shifts and bi-manual object grasping and manipulation tasks," in *Proc. IEEE Int. Conf. Develop. Learn.*, Ann Arbor, MI, 2010, pp. 165–170.

[108] F. Kaplan and V. V. Hafner, "The challenges of joint attention," *Interact. Stud.*, vol. 7, pp. 135–169, 2006.

[109] M. Oginoa, H. Toichia, Y. Yoshikawaa, and M. Asadaa, "Interaction rule learning with a human partner based on an imitation faculty with a simple visuo-motor mapping," *Robot. Autonom. Syst.*, vol. 54, pp. 414–418, 2006.

[110] Y. Yoshikawa, T. Nakano, M. Asada, and H. Ishiguro, "Multimodal joint attention through cross facilitative learning based on x principle," in *Proc. IEEE Int. Conf. Develop. Learn.*, Monterrey, CA, 2008, pp. 226–231.

[111] M. Lopes and J. S. Victor, "A developmental roadmap for learning by imitation in robots," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 308–321, Apr. 2007.

[112] J. S. Bruner and L. Postman, "On the perception of incongruity: A paradigm," *J. Personality*, vol. 18, pp. 206–223, 1949.

[113] X. Huang and J. Weng, "Novelty and reinforcement learning in the value system of developmental robots," in *Proc. Int. Workshop Epigenetic Robot.*, Edinburgh, Scotland, 2002, pp. 47–55.

[114] Y. Zhang and J. Weng, "Action chaining by a developmental robot with a value system," in *Proc. IEEE Int. Conf. Develop. Learn.*, Cambridge, MA, 2002, pp. 53–60.

[115] W. Dodd and R. Gutierrez, "The role of episodic memory and emotion in a cognitive robot," in *Proc. IEEE Int. Conf. Robot Human Interact. Commun.*, Nashville, TN, 2005, pp. 692–697.

[116] K. Kawamura, W. Dodd, P. Ratanaswasd, and R. Gutierrez, "Development of a robot with a sense of self," in *Proc. IEEE Int. Symp. Comput. Intell. Robot. Autom.*, Edmonton, AB, Canada, 2005, pp. 211–217.

[117] J. L. Krichmar and G. M. Edelman, "Brain-based devices: Intelligent systems based on principles of the nervous system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot Syst.*, Las Vegas, NV, 2003, pp. 940–945.

Momotaz Begum received the B.Sc. degree in electrical engineering from the Bangladesh University of Engineering and Technology (BUET), in 2003, and the M.Eng. degree in mobile robotics from the Memorial University, Newfoundland, Canada, in 2005. She also received the Ph.D. degree in intelligent robotics from the University of Waterloo, Waterloo, ON, Canada, in 2010.

She is currently working as a Postdoctoral Fellow at the School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA. Her research interests include artificial cognition, human–robot interaction, developmental robotics, simultaneous localization and mapping for mobile robots, and image processing.

Fakhri Karray received the Ing. Dipl. degree from the University of Tunisia, Tunisia, in 1984, and the Ph.D. degree from the University of Illinois, Urbana, in 1989, in the area of systems and control.

He is currently a Professor of Electrical and Computer Engineering at the University of Waterloo, Waterloo, Canada, and the Associate Director of the Pattern Analysis and Machine Intelligence Laboratory. He has authored extensively in journals and conferences proceedings, and holds 13 U.S. patents in various areas of intelligent systems. He is the coauthor of a textbook on soft computing, *Soft Computing and Intelligent Systems Design*, (Addison Wesley Publishing, 2004).

Dr. Karray serves as the Associate Editor of a number of journals in his filed, including the IEEE TRANSACTIONS ON MECHATRONICS, the IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS (PART B), the *IEEE Computational Intelligence Magazine*, the *International Journal of Robotics and Automation*, the *Journal of Intelligent Systems and Control*, and the *International Journal on Smart Sensing and Intelligent Systems*. He is the Waterloo Chapter Chair of the IEEE Control Systems Society and the IEEE Computational Intelligence Society.