Integrated Computer-Aided Engineering 14 (2007) 73–91
IOS Press

Visual perception in design and robotics

Michael S. Bittermann, I. Sevil Sariyildiz and Özer Ciftcioglu*
Department of Building Technology, Delft University of Technology, Berlageweg 1, 2628 CR Delft, The Netherlands
Tel.: +31 150 278 5035; Fax: +31 150 278 4178; E-mail: o.ciftcioglu@tudelft.nl

Abstract. Studies on human visual perception are described. The visual perception is mathematically modelled as a probabilistic process obtaining and interpreting visual information from an environment. By means of this model some other vision related concepts, such as visual attention and visual openness, are also mathematically defined. The theoretical considerations are implemented in three applications, namely scene description by perception, visual openness measurement for design, and perceptual robotics.

1. Introduction

In design, perception is an important topic, as vision is the major source of information in our comprehension of the environment. Design processes involve virtual reality and virtual agents. Therefore the problem arises how to give human vision to a virtual agent, which plays an important role in the design process we are engaged in. This task directs us to investigate the explicit understanding of perception. Hence, two main goals are pursued in this research. The first one is explicit definition of human perception in mathematical terms. The second one is to equip a virtual agent with such perception abilities. Undoubtedly perceptual virtual agents have a correspondence in robotics, which is known as perceptual robotics [2,3]. Modeling the human vision process in terms of perception is a step in order to effectively and efficiently integrate environmental information into machine-based systems.

Due to the diversity of existing approaches related to visual perception, which emerged in different scientific domains, we provide a comprehensive introduction to be explicit as to both, the objectives, and the contribution of the present research.

Perception is a deeply involved concept. This is indicated by its relation with a fundamental philosophical question, namely whether the reality is independent of human mind [21]. If one takes the position of *mind-dependence*, perception becomes ill-defined. This is the main reason perception is not clearly defined and there is no firm consensus about what it is, as this is elaborated below.

When we take a mind-independent position modelling the human perception is challenging, mainly because it involves not only the eye but also the brain. When we say that we perceived something, the meaning is that we can recall relevant properties of it. What we cannot remember, we cannot claim we perceived, although we may suspect that corresponding image information was on our retina. In this sense the perception definition used in the present research differs from other works, where perception is considered similar to *distance estimation* [36,53], or the *image processing* definition [7,8,31,37] of reconstructing a 3-dimensional scene from 2-dimensional image information.

With this basic understanding it is important to note that the act of perceiving has a characteristic, that is *uncertainty*: it is a common phenomenon that we overlook items in our environment, although they are visible to us, i.e., they are within our visual scope, and there is a possibility for their perception.

This everyday experience has never been exactly explained due to the complexity of the brain processes, which are essentially unknown.

Brain researchers trace visual signals as they are processed in the brain. Although a number of achievements are reported in the literature [16,23,25,26,44,51,52], due to complexity there is no consensus about the exact role of brain regions, sub-regions and individual

*Corresponding author.

ISSN 1069-2509/07/$17.00

nerve-cells in vision, and how they should be modeled. This is because there is enough modality of the brain to accommodate all the research without unification. The difficulty of the task is increased by the involvement of attentional mechanisms in perception, which are driven by diverse conditional interests of the human being. Identification and pin-pointing of these influences on the visual processing in the brain is formidable due to the intimate relation between attention and vision. Therefore, the relation between environmental stimulus and the mental event of *perception* remains unknown.

In the computer vision community attempts were made to model the human perception process starting with a 2-dimensional image, which is definitively stored in a memory [8,31]. In these approaches perception is considered to be the reconstruction of a 3-dimensional representation of a scene from a 2-dimensional one, which is achieved by a number of *image processing* algorithms. However, such reconstruction does not model the characteristic uncertainty of human perception. It is not obvious how some of the retinal image data does not yield the perception of the corresponding objects in our environment. Instead of addressing this fundamental problem of perception, by taking a 2-dimensional image as a starting point for modeling perception, the problem is shifted to a virtual perception situation, where the environment is 2-dimensional as opposed to reality, where it is 3-dimensional.

The justification for the image as a starting point presumably lies in the strategy to model the perception process as a sequence of events, with the retinal image acquisition as initial event. However, modeling the sequence of brain processes, that treat the retinal image by means of algorithmic counterparts, is a formidable endeavor. This holds true even when advanced computational methods are applied for modeling of the individual brain-components' behavior, as in [4,40] for example. This is due to the complexity of the brain processes mentioned above. The attempt is similar to modeling the result of a coin-toss experiment by modeling all of the interaction events the coin undergoes during the procedure in sequence. The ample complexity of the physical process defies reliable outcome of such a deterministic approach.

Since the processing of the retinal stimulus in the brain is highly complex, despite existing work on the functionality of the retina [19], it remains unknown, how a retinal stimulus maps to perception in the human mind. Although there is work on human visual perception related to vision augmentation and man-machine vision [18], where a neuroscience model of vision in the perceptual domain of the brain is proposed, no explicit mechanism of perception is given. This is due to complexity concerning the threshold between physical and mental domain.

In modeling the human vision the involved brain process components as well as their interactions should be known with certainty if a deterministic approach, like the image processing approach, is to be successful. This is currently not the case. Well-known observations of visual effects, such as *depth from stereo disparity* [28,39], *Gelb effect* [12], *Mach bands* [22], *gestalt principles* [17,50], etc., reveal components of the vision process, that may be algorithmically mimicked. However, it is unclear how they interact in human vision to yield the mental act of perception. In particular it remains mysterious, how the combination of the modeled components results in the characteristic uncertainty of perception, which is commonly observed. It may be noteworthy to mention that optical illusions are conditional events, which almost never occur in our everyday experience; whereas the uncertainty of perception is familiar and occurs practically always in our daily life, yet it may remain unnoticed.

In the domain of virtual reality (VR) perception models can be found [24]. In this approach the *visual acuity* of the human eye as well as the *minimum eye resolution* are simulated to model visual perception. Positions in the environment, which are located within a limited central region of the visual scope, are considered to be "perceived", while objects located outside this region are "not perceived". It may be noteworthy to mention that the human ability to attend different regions within the visual scope, focus, rotate the eyes, or even to rotate the head, make it difficult to justify singling out one particular gaze-line in the view as the basis for a human perception model. Next to that, the approach ignores the common phenomenon that we overlook items in our visual scope, even when they are located within a limited central aperture in our visual scope, although such overlooking may be unlikely to occur.

In the area of scene synthesis of the computer graphics a number of methods are proposed to model photographic image formation [15,29,48,54]. Such methods include probabilistic ray tracing algorithms. With respect to visual perception, however, such methods are unable to serve as models, since the human perception not only involves image acquisition on the retina, but notably brain processing of retinal data. The essential question remains unanswered, namely how is it possible that light, which is received on the retina does not yield perception of the corresponding objects, from which the light is emitted or reflected.

The psychology community established the probable "overlooking" of visible information experimentally [32,41], where it has been shown that people regularly miss information present in images. For the explanation of the phenomenon the concept of *visual attention* is used, which is a well-known concept in cognitive sciences [5,11,20,27,38,41,46,47,55]. Treisman analyzed the time people need to fulfill a number of different visual tasks, when looking at images [46]. Based of the results of the experiments concepts termed *feature maps* and *object file* are introduced [45], which are used to explain three observed properties of visual attention; namely, the attention is used to integrate various features of environmental items, it can be directed to differently sized regions in the visual scope, and location of related items is a relevant property. Such qualifications are helpful to understand the concept of attention. However, it remains unclear what *attention* exactly is, and how it can be modeled quantitatively.

The works on attention mentioned above start their investigation at a level, where basic visual comprehension of a scene must have already occurred. For example, in [55] eye saccades are examined showing that people focus more frequently on certain items in a picture than on other items, like the head of a person, and the frequency and focus position are dependent on a given task. However, it is not obvious that such observations indicate anything about the early human perception, where an observer gathers information about where relevant targets located in the view. An observer can execute his/her bias or preference for certain information within the visual scope only when he/she has already a perception about the scene, as to where potentially relevant items exist in the visible environment. This early phase, where we build an overview/initial comprehension of the environment is referred to as *early vision* in the literature, which is omitted in the works on attention mentioned above.

Early vision was specifically addressed in another work [1]. In that work a 7-dimensional variable, termed *plenoptic function*, is proposed as the fundamental element in human vision, which represents properties of the light entering the retina of an observer. However, establishing this variable does not help to understand perception, since the fundamental question, how environmental stimuli yield mental realization of corresponding items remains unaddressed.

While the early perception process is unknown, identification of attention in perception, that is due to a task specific bias, is limited. This means, without knowledge of the initial stage of perception its influence on later stages is uncertain, so that the later stages are not uniquely or precisely modeled and the attention concept is ill-defined.

Since attention is ill-defined, ensuing perception is naturally also merely ill-defined. Some examples of definitions on perception are *"Perception refers to the way in which we interpret the information gathered and processed by the senses,"* [30] and "*Visual perception is the process of acquiring knowledge about environmental objects and events by extracting information from the light they emit or reflect,*" [34]. Philosophers also offered a number of verbal definitions on perception and related concepts, for example, see [21,42,43]. Such verbal definitions above are helpful to understand what perception is about; however they do not hint how to tackle the perception beyond qualitative inspirations. Although we all know what perception is apparently, there is no unified, commonly accepted definition of it.

As a summary of the previous part we note that visual perception and related concepts have not been exactly defined until now. Therefore, the perception phenomenon is not explained in detail and the perception has never been quantified, so that the introduction of human-like visual perception to machine-based system remains as a soft issue.

In the present paper a newly developed theory of perception is introduced. In this theory visual perception is put on a firm mathematical foundation. This is accomplished by means of the well-established probability theory. The work concentrates on the early stage of the human vision process, where an observer builds up an unbiased understanding of the environment, without involvement of task-specific bias. In this sense it is an underlying fundamental work, which may serve as basis for modeling later stages of perception, which may involve task specific bias. The probabilistic theory can be seen as a unifying theory as it unifies synergistic visual processes of human, including physiological and neurological ones, as well as philosophical aspects of vision. Interestingly this is achieved without recourse to neuroscience and biology. It thereby bridges from the environmental stimulus to its mental realization. At this point we do not elaborate our definition of attention and perception, since this will be established naturally as result of the vision model, which is described in the next section.

Through the novel theory twofold gain is obtained. Firstly, the perception and related phenomena are understood in greater detail, and reflections about them are substantiated. Secondly, the theory can be effectively introduced into advanced implementations since perception can be quantified. It is foreseen that modeling human visual perception can be a significant step as the topic of perception is a place of common interest that is shared among a number of research domains, including cybernetics, brain research, virtual reality computer graphics, design and robotics [13]. Robot navigation is one of the major fields of study in autonomous robotics [6,33,49]. In the present work, the human-like vision process is considered. This is a new approach in this domain, since the result is an autonomously moving robot with human-like navigation to some extent. Next to autonomous robotics, this belongs to an emerging robotics technology, which is known as perceptual robotics [2,3]. From the human-like behaviour viewpoint, perceptual robotics is fellow counterpart of emotional robotics, which is found in a number of applications in practice [10]. Due to its merits, the perceptual robotics can also have various applications in practice.

Other main application areas of perceptual robotics can be categorised into two parts as reality and virtual reality. In the real-life, autonomous robot movement with obstacle avoidance is a big challenge. Perceptual robotics can easily address this issue with major improvements along this line. For instance, by measuring the distance to obstacles directly in real-time instead of using different localization algorithms in a static environment. In the virtual reality, perceptual robotics can be important means of perception-based estimations for accuracy and precision, like distance estimation, for instance. These are demanded features in game industry, flight simulation, as well as consistent spatial design quality assessment in architectural design.

From the introduction above, it should be emphasized that, the research presented here is about to demystify the concepts of perception and attention as to vision from their verbal description to a scientific formulation. Due to the complexity of the issue, so far such formulation is never achieved. This is accomplished by not dealing explicitly with the complexities of brain processes or neuroscience theories, about which more is unknown than known, but incorporating them into perception via probability. We derive a vision model, which is based on common human vision experience explaining the causal relationship between vision and perception at the very beginning of our vision process. Due to this very reason, the presented vision model precedes all above referenced works in the sense that, they can eventually be coupled to the output of the present model. The novel theory introduced here is based on the axiom, which is presented in the following section, and this is enough to explain a number of perception related phenomena that we commonly experience as result of our vision process. In this way the integrity of the theoretical results are verified by means of the practical implications.

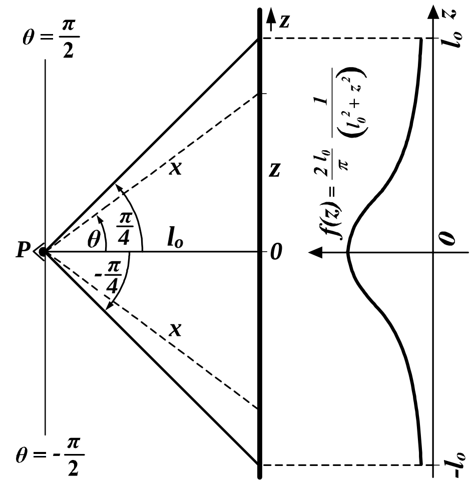

Fig. 1. The geometry for visual perception from top view, where P represents the position of eye, looking at a vertical plane with a distance l_o to the eye; $f_z(z)$ is the probability density function of the visual perception.

It is noteworthy to mention that in the brain models referenced above, all models are different due to the different focus of attention that refer to uncountable number of modalities in the brain and therefore they are inconclusive as to understanding of a particular brain process like perception and attention on a common ground. As a state of the art, they try to form a firm clue for perception and attention beyond their verbal accounts.

The organisation of this work is as follows. Section two describes the theoretical visual perception model development. Section three deals with the implementation of the perception model via a virtual agent. Section four describes the implementations both in architectural design and autonomous robotics, which is followed by conclusions.

2. Perception model development

2.1. Basic probabilistic vision model

Vision is our essential source of information while we interact with environment. Based on vision, many derivative concepts of vision can be defined, such as visual perception, visual attention, visual privacy, and so on. However, since the concepts, which have been derived from the vision process, are only verbally described up till now, the definition of these concepts does not have precise consensus in literature, although they roughly coincide in essence. Having noticed that, this work endeavors to establish a mathematical vision model, so that the ensuing concepts of vision derivatives are mathematically defined. As result of this it is foreseen that several elusive concepts like visual perception and visual attention are no longer elusive but subject to quantification and computation. By doing so several vision-related concepts can be effectively introduced into advanced implementations, like scene description by perception, visual openness measurement for design, or perceptual robotics, having insight into the role of perception in such tasks.

As a complex process, vision involves image acquisition with the eye and interpretations in different regions in the brain. Since this process is formidably complex and details of it are not known, its modeling is difficult when classical, deterministic methods are used. The well-established probability theory is particularly suitable to handle such high complexity and uncertainty. The main principle of modeling by probability theoretic considerations is to absorb complexity and uncertainty into probability. To illustrate this, let us consider a coin-toss experiment as an example. When a coin is tossed it undergoes a large amount of physical interactions with air molecules around, the surface it lands on, etc. Modeling this process is a formidable task if we endeavor to model each of the interaction events. However, if we are interested in modeling the result of the process, and if we can afford to ignore the detailed states of the coin during the process, we can model the situation perfectly by stating, that the probability of the outcome is 0.5 for the event that the coin lands on either side. The meaning is, that when we make a single coin toss, we cannot predict with certainty, which result will appear, however we attach a probability to the possible outcomes, which absorbs the total complexity of the experiment. When we make a large amount of experiments, the model is verified as the ratio of the results approaches to the number assigned as probability.

With respect to human vision, a probabilistic model is most appealing, as it absorbs the complexity of the eye/brain process by the probability theoretic considerations, resulting in a robust match between model outcome and the perception phenomena we commonly experience.

We start with the basics of the perception process using a simple, yet fundamental visual geometry. This is shown in Fig. 1. In the figure an observer is facing and viewing a vertical plane from the point denoted by P. By viewing, the observer has chance to receive visual data from all directions within his/her scope with equal probability in the first instance. That is, the observer visually experiences all locations on the plane with uniform probability density as to the angle θ that defines the viewing direction, without any preference for one direction over another. This axiomatic consideration ensures that there is no visual bias at the beginning of perception as to the direction, from which information in the environment may be obtained. In particular, the unbiasedness above is with respect to the direction, and therefore the model concerns the very starting instance of a vision process, before any directional preference for certain information in the environment can be executed, such as in visual search or object recognition, for example. The axiomatic starting point of the theory is based on basic human vision experience and it is not an ad hoc approach trying to explain a certain vision phenomenon. Instead this is a starting point from scratch to arrive at some results through further derivations, which follow. Therefore the justification will follow afterwards.

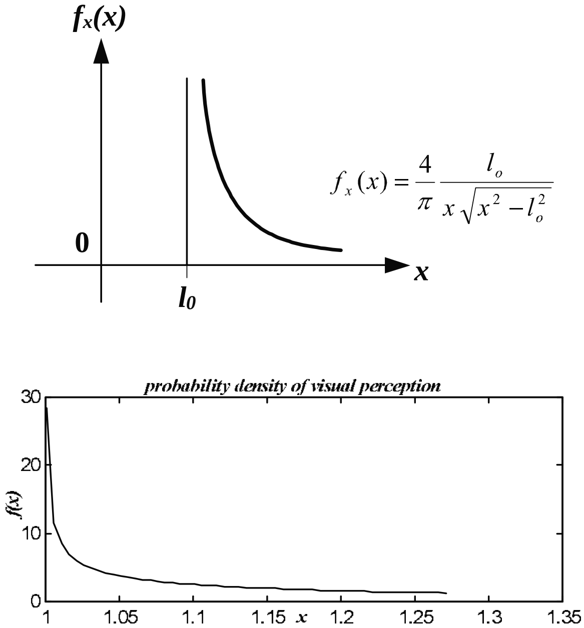

Fig. 2. Variation of the probability density function of the random variable representing the distance x between eye and a location on the plane shown in Fig. 1; upper plot is a sketch; lower is a computed plot with $l_o = 1$.

From above, the axiom entails that the probability is the same for each single differential visual resolution

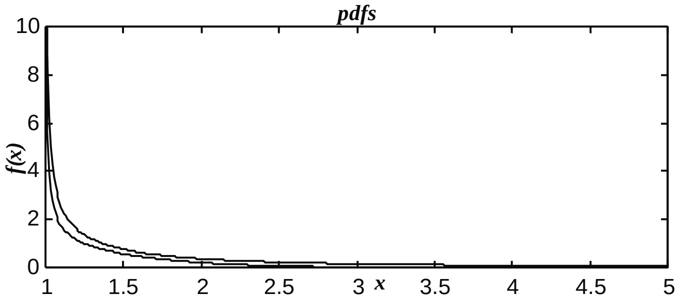

Fig. 3. The visual perception pdf $f_x(x)$ and its modified alternative (lower) due to approximation of $f_z(z)$ as a Gaussian.

angle $d\theta$. This means that the probability density function (pdf), which belongs to the angle θ is uniformly distributed. This is shown in Fig. 1, where the scope of vision is taken as $\theta = \pm\pi/4$.

The angle θ is a *random variable* in the terminology of the probability theory. Since θ is trigonometrically related with each point on the plane, the distances x or z, which are indicated in Fig. 1, are also random variables.

$$x = \frac{l_o}{\cos(\theta)} \tag{1}$$

Assuming the scope of sight is defined by the angle $\theta = \pi/4$, the pdf f_θ is given by

$$f_\theta = \frac{1}{\pi/2} \tag{2}$$

Since θ is a random variable, the distance x in Eq. (1) is also a random variable. The pdf $f_x(x)$ of this random variable is computed as follows.

Theorem on the *function of random variable* [35]: To find $f_x(x)$ for a given x we solve the equation

$$x = g(\theta) \tag{3}$$

for θ in terms of x. If $\theta_1, \theta_2, \ldots, \theta_n, \ldots$ are all its *real* roots,

$$x = g(\theta_1) = g(\theta_2) = \ldots\ldots = g(\theta_n) = \ldots.$$

Then

$$f_x(x) = \frac{f_\theta(\theta_1)}{|g'(\theta_1)|} + \ldots + \frac{f_\theta(\theta_2)}{|g'(\theta_2)|} + \ldots + \frac{f_\theta(\theta_n)}{|g'(\theta_n)|} + \ldots \tag{4}$$

Clearly, the numbers $\theta_1, \theta_2, \ldots, \theta_n, \ldots$ depend on x. If, for a certain x, the equation $x = g(\theta)$ has no real roots, then $f_x(x) = 0$.

According the theorem above,

$$g'(\theta) = \frac{l_o \sin(\theta)}{\cos^2(\theta)} \tag{5}$$

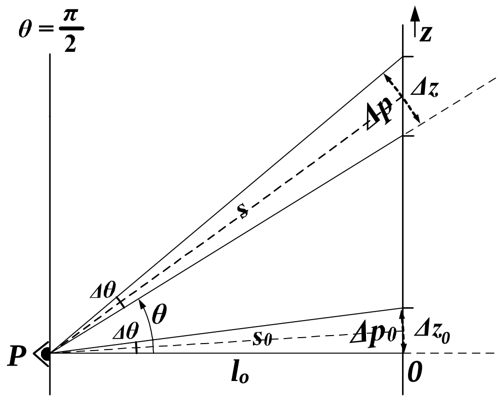

Fig. 4. Sketch explaining the relative importance of the viewing direction for visual perception.

Between $\theta = -\pi/4$ and $\theta = +\pi/4$,

$$g(\theta) = \frac{l_o}{\cos(\theta)} \tag{6}$$

has two roots, which are equal and given by

$$\theta_{1,2} = \arccos\left(\frac{l_o}{x}\right) \tag{7}$$

Using Eq. (7) in Eq. (5), we obtain

$$g'(\theta) = \frac{x\sqrt{x^2 - l_o^2}}{l_o} \tag{8}$$

Substituting Eqs (2), (7) and (8) into Eq. (4), we obtain

$$f_x(x) = \frac{4}{\pi} \frac{l_o}{x\sqrt{x^2 - l_o^2}} \tag{9}$$

for the interval

$$l_o \leqslant \ x \ \leqslant \frac{l_o}{\cos(\pi/4)}.$$

For this interval, the integration below becomes

$$\int_{l_o}^{\sqrt{2}\,l_o} f_x(x)\,dx = \frac{4}{\pi} \int_{l_o}^{\sqrt{2}\,l_o} l_o \frac{dx}{x\sqrt{x^2 - l_o^2}} = 1 \tag{10}$$

as it should be as a pdf. The sketch of $f_x(x)$ vs x is given in Fig. 2 and its variation for $l_o = 1$ is also given in the same figure. The derivations above and the further derivations below are motivated to arrive at the abstract definition of attention and perception in mathematical terms, as this will naturally appear in the text.

In place of a plane geometry, for a circular geometry, the pdf $f_x(x)$ in Eq. (9) takes a uniform distribution, as this is shown in the Appendix. This exemplifies the perception occurring in a different geometry.

It is interesting to note that for the plane geometry in Fig. 1, the probability density for x is concentrated close to $\theta \cong 0$, that is in perpendicular direction to the plane, where $f_x(x)$ approaches to infinity. We extend our calculations to derive the pdf along the z direction in Fig. 1. In this case, proceeding in the same way as before, we write

$$\begin{aligned} tg(\theta) &= \frac{z}{l_o} \\ z &= g(\theta) = l_o\, tg(\theta) \end{aligned} \tag{11}$$

$$\begin{aligned} g'(\theta) &= \frac{dz}{d\theta} = \frac{l_o}{\cos^2(\theta)} = l_o(1 + tg^2(\theta) \\ &= l_o\left(1 + \frac{z^2}{l_o^2}\right) \end{aligned} \tag{12}$$

$$\theta_1 = \mathrm{arctg}(z/l_0) \tag{13}$$

$$f_z(z) = \frac{f_\theta(\theta_1)}{|g'(\theta_1)|} = \frac{l_o}{\pi(l_o^2 + z^2)} \tag{14}$$

for the interval $-\infty \leqslant z \leqslant \infty$. For this interval, the integration below becomes

$$\int_{-\infty}^{+\infty} f_z(z)\, dz = \frac{l_o}{\pi} \int\limits_{-\infty}^{+\infty} \frac{dz}{z^2 + l_o^2} = 1 \tag{15}$$

as it should be. For the interval

$$l_o \leqslant x \leqslant \frac{l_o}{\cos(\pi/4)} \tag{16}$$

$$f_z(z) = \frac{2}{\pi} \frac{l_o}{(l_o^2 + z^2)} \; (-l_0 \leqslant z \leqslant +l_0) \tag{17}$$

so that

$$\int_{-l_0}^{+l_0} f_z(z)\, dz = \frac{2\, l_o}{\pi} \int\limits_{-l_0}^{+l_o} \frac{dz}{z^2 + l_o^2} = 1 \tag{18}$$

as one has to expect. The variation of $f_z(z)$ is shown in Fig. 1. This result clearly explains the relative importance of the front view as compared to side views in unbiased human visual experience of the plane. This coincides with our common experience that we perceive/remember better the details of the frontal region of a plane we are facing, compared to the side regions.

In the applications of the perception model, which we deal with in Section 4, $f_z(z)$ is approximated by a Gaussian function due to computational convenience. Therefore, the implication of this on the approximation to the exact probability density $f_x(x)$ given by Eq. (9) is presented below, beforehand.

From Fig. 1, we can write

$$x^2 = l_o^2 + z^2 \tag{19}$$

where

$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} z^2} \tag{20}$$

From Eq. (19), we write $z^2 = x^2 - l_o^2$, so that

$$z_{1,2} = \pm\sqrt{x^2 - l_o^2} \tag{21}$$

$$g(z) = \sqrt{l_o^2 + z^2} \tag{22}$$

$$g'(z) = \frac{dg}{dz} = \frac{1}{\sqrt{\left(\frac{l_o}{z}\right)^2 + 1}} \tag{23}$$

Substituting $z_{1,2}$ from Eq. (21) into Eq. (23) yields

$$g'(z_{1,2}) = \frac{1}{\sqrt{\frac{l_o^2}{x^2 - l_o^2} + 1}} \tag{24}$$

From the function of a random variable theorem

$$\begin{aligned} f_x(x) &= \frac{f_z(z_1)}{|g'(z_1)|} + \ldots + \frac{f_z(z_2)}{|g'(z_2)|} + \ldots \\ &\quad + \frac{f_z(z_n)}{|g'(z_n)|} + \ldots \end{aligned} \tag{25}$$

$$f_x(x) = \frac{f_z(z_1)}{|g'(z_1)|} + \frac{f_z(z_2)}{|g'(z_2)|} \tag{26}$$

we obtain

$$f_x(x) = \sqrt{\frac{2}{\pi\sigma^2}} \frac{x}{\sqrt{x^2 - l_o^2}} e^{-\frac{1}{2\sigma^2}(x^2 - l_o^2)} \tag{27}$$

which is the modified form of exact $f_x(x)$ in Eq. (9) due to approximation of $f_z(z)$ in Eq. (17) by a Gaussian Eq. (20). Both pdfs of $f_z(z)$ given by Eq. (9) and given by Eq. (27) are shown together for comparison in Fig. 3, for $\sigma = l_o = 1$.

The result of the relative importance of the front view as compared to side views in human vision perception can be explained easily as sketched in Fig. 4.

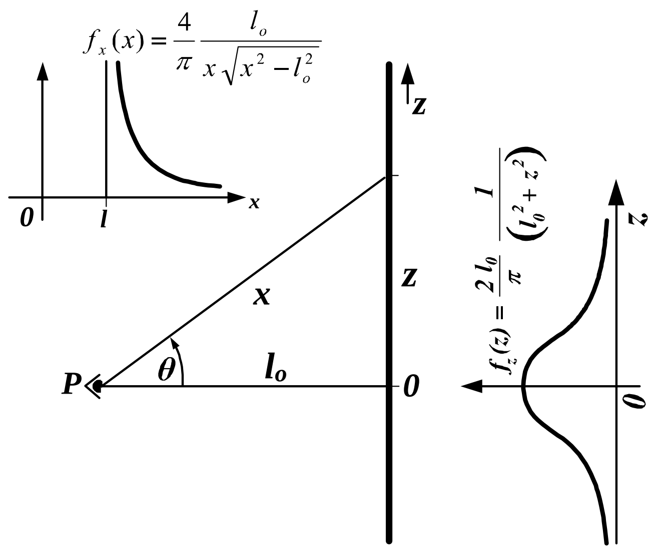

Fig. 5. Sketch showing the dual representations $f_x(x)$ and $f_z(z)$ of the visual perception of the vertical plane shown in Fig. 1.

In Fig. 4,

$$\Delta z_0 \cong \frac{\Delta\theta\ s_0}{\cos\theta} \quad , \ \Delta z \cong \frac{\Delta\theta\ s}{\cos\theta} \tag{28}$$

$$s_0 = \frac{l_o}{\cos\theta_0}, s = \frac{l_o}{\cos\theta}, \text{ so that} \tag{29}$$

$$\Delta z_0 \cong \frac{\Delta\theta\ l_o}{\cos^2\theta_0}, \Delta z \cong \frac{\Delta\theta\ l_o}{\cos^2\theta} \tag{30}$$

Noting that for $\theta_0 = 0$, we obtain

$$\Delta z_0 \cong \Delta\theta\ l_o\ \Delta z \cong \frac{\Delta\theta l_o}{\cos^2\theta} \tag{31}$$

Since $\Delta z_0 \leqslant \Delta z$, this clearly shows that the visual resolution is higher for the case with θ_0 relative to the case with θ. This implies that one gets more visual details at the origin as the visual resolution is higher there, and consequently the general shape of pdf $f_z(z)$, seen in Fig. 1, i.e., $f_z(z)$ is highest. The result in Eq. (30) can be obtained from Eq. (11) by differentiation. Namely,

$$z = l_o\, tg(\theta) \tag{32}$$

and by differentiation

$$dz = \frac{l_o\, d\theta}{\cos^2\theta} \tag{33}$$

is obtained.

One should note that, the density functions $f_x(x)$ given in Eq. (9) and $f_z(z)$ in Eq. (17) are dual representations of the same phenomenon. One can easily pass from one to another as shown in the following demonstration via Fig. 5.

From Fig. 5, we write

$$z = f(x) = \sqrt{x^2 - l_o^2} = g(x) \tag{34}$$

The roots of Eq. (34) are

$$x_{1,2} = \pm\sqrt{z^2 + l_o^2} \text{ and } x_1 = \sqrt{z^2 + l_o^2} \tag{35}$$

where only the positive root is valid, since x is always positive.

$$g'(x) = \frac{2x}{2\sqrt{x^2 - l_o^2}} = \frac{x}{\sqrt{x^2 - l_o^2}} \tag{36}$$

$$\begin{aligned} f_z(z) &= \frac{f_x(x_1)}{|g'(x_1)|} = \frac{\frac{1}{\pi}\frac{l_o}{x_1\sqrt{x_1^2 - l_o^2}}}{\frac{x_1}{\sqrt{x_1^2 - l_o^2}}} \\ &= \frac{l_o}{\pi}\frac{1}{(z^2 + l_o^2)} \end{aligned} \tag{37}$$

or

$$f_z(z) = \frac{l_o}{\pi}\frac{1}{(z^2 + l_o^2)}\ (-\infty \leqslant z \leqslant +\infty) \tag{38}$$

which is given by Eq. (14).

2.2. *From attention to perception*

Based on the results of the basic theoretical considerations above, some vision related concepts can be terminologically defined in mathematical terms. Namely, vision is the ability to see. All vision-related concepts, such as visual attention, perception, openness and others are the sub-areas of vision. This means, vision is the essential necessity for the further considerations in

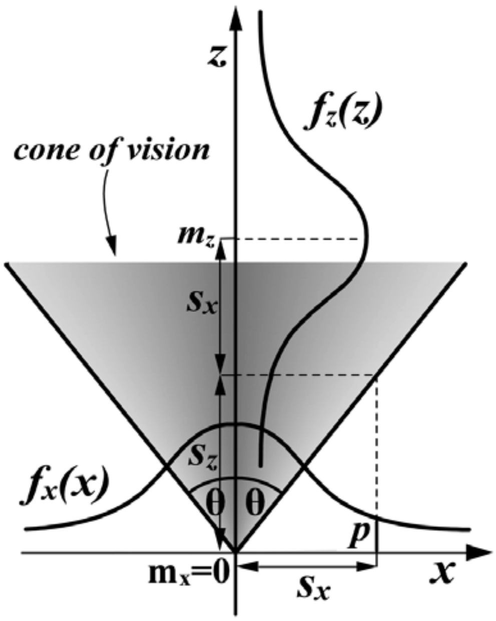

Fig. 6. The Gaussian visual perception in forward direction with a visual scope in the shape of a cone with the angle 2θ, where z is the forward direction.

these sub-areas. The probability density functions derived in the preceding subsection correspond to *visual attention*, which is a well-known concept in cognitive psychology [5,11,20,27,32,38,41,45–47,55], yet never formally defined with consensus. From the derivations above we note that, visual attention refers to a density within an infinitesimally small region, and per unit length or interval. The integration of the probability density functions $f_x(x)$ or $f_z(z)$ within a finite interval we define as *perception*, which becomes a probability; namely the probability that an observer obtains information from a region in the environment by vision. Perception is real and physical. Important conclusion is that, it is given by a probability, which means it is a probabilistic event. The definition of attention and perception above may not exactly coincide with those used in the psychological literature, due to vague and verbal definitions in the latter case. However, the mathematical formulations presented above are decisive and they are subject to adoption into diverse fields. The integration of visual attention in a finite small interval where $f(x)$ is approximately constant, gives perception (P) as

$$P = \int_{\Delta x}^{f_x(x)\,dx} \cong f_x(x)\,\Delta x \tag{39}$$

or alternatively in a finite small interval where $f(z)$ is approximately constant, gives perception (P)

$$P = \int_{\Delta z}^{f_z(z)\,dz} \cong f_z(z)\,\Delta z \tag{40}$$

so that perception becomes proportional to attention in this case. For the forward direction of viewing the infinite plane where $\theta \approx 0$, for the interval given in Eq. (16), the perception becomes

$$P \approx \frac{2}{\pi}\Delta\theta \tag{41}$$

where $\Delta\theta$ is the vision resolution of the human eye. Although in the theoretical development the region over which attention is integrated to form perception can be infinitesimally small, in the practical case the region is restricted by the human eye resolution, which is defined by the differential angle $\Delta\theta$ in Fig. 4.

The striking results from the above computations, where we derived the probability densities for x and z in the plane case, are in conformity with the common human vision experience as to visual perception. Namely, for this geometry the visual attention and thereby perception is strongest along the axis of the cone of vision, relative to the side directions. This means that details of the front plane have a greater chance to be seen, and hence remembered, than the side parts.

One can easily verify the results above from the probabilistic vision model intuitively by basic vision experience. For instance, in a scene, when one considers only the geometry and not any other aspects like color, comparatively an object with bigger dimensions is more likely to be seen, and therefore perception of this object is high. If there are more than one objects in the scene, since the attention, that is the discrete probability density, would be distributed among these objects, then the final seeing action for these objects will be merely subject to probability, which means subject to perception. The bigger object is more likely to be seen in this case.

Another example is reading action of a text. In order to recognize a letter, or a word, or pairs of words it has to be perceived first. In this process the scene in question becomes restricted to the very area of each letter or word or pairs of words, which means the whole attention is restricted to this area, with the associated probability density. The integral of this probability density within that area is one. The area defines the range of the probabilistic space, where attention is defined. By this, the perception of each letter, or word, or pairs of words becomes almost certain, and consequently reading and comprehension are likely to occur correctly. However, if the letters, or words, or pairs of words are not distinct enough matching the eye resolution, or alternatively

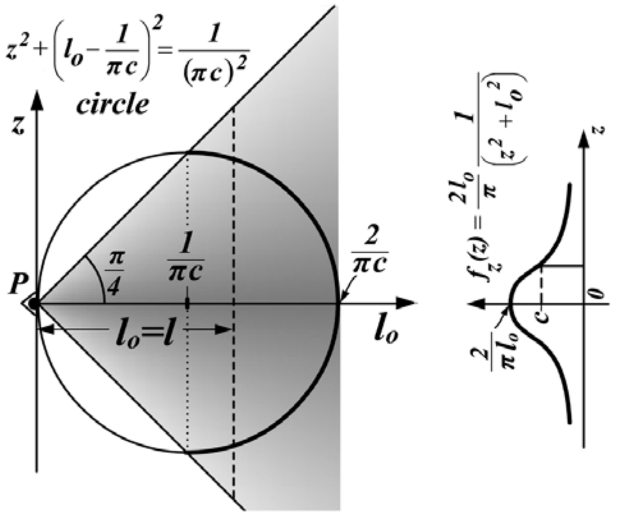

Fig. 7. The vision cone associated with the probability density $f_z(z)$; *circle of attention* and corresponding attention probability density are also indicated.

in a fast reading action, irregularities in reading and comprehension may occur. The reason is that in this case more letters, words, or pairs of words are considered at a time than that involved in the slow reading. This means the probabilistic space may be larger than it should be for recognition, which means there may not be satisfactory attention to each item, thereby lower perception as to each letter or word or pairs of words of the text, we consider. After the perception process symbol recognition process follows [56].

As the perception presented here involves integration of attention in a particular domain, the perception theory is closely related to gestalt principles. Gestalt theory addresses the phenomenon that an observer experiences a number of elemental visual items as belonging together, so that they form another visual entity, which is referred to as *gestalt* [50]. With respect to the role of gestalt theory for perception it should be clearly noted that an observer executes *grouping* based on perception, as introduced here, so that he/she should have mentally realized beforehand the elemental visual items to be grouped. The number of domains and their shapes in a scene to be integrated is subject to probabilistic considerations. In these considerations the elementary probability of a gestalt quality denoted by p can be assessed by means of perception considerations presented here, and further computations can be carried out with underlying probability theory [17]. In this way seamless coupling of the gestalt theory to the perception theory presented here is an interesting relevance.

The vision model presented in this research is by no means a simplistic description of the vision process, although the geometry considered is rather simple. The simplicity is for the sake of the description of the process, and the conclusions from this simple geometry are far-reaching. The basic geometry given in Fig. 1 can be extended to other geometrical cases for more involved perception computations, for example, see [14].

3. Perception-based human vision modelling

When we endeavour to implement the perception model elaborated above to a virtual agent in virtual reality or a perceptual vision robot it is important to realize that there is a difference between human perception and robot perception. A human mentally receives information from the environment in a probabilistic way, which is perception as it is defined here, whereas in the case of our virtual reality implementation a robot or agent probes the environment randomly, and the randomness is shaped to fit to the human perception. To accomplish this we model forward vision in 3D-space. For this purpose we use three Gaussian functions, each of which delivers independently a random number for the three orthogonal coordinate axes. These numbers are used as the components of the direction vectors for the vision rays. The randomness mentioned above has to be shaped according to Eq. (17). This needs some attention to accomplish, and it is noteworthy to give some details of this treatment, which is presented below.

For the sake of simplicity of explanation we restrict the case to two-dimensional space. The problem is to adjust the parameters of the Gaussian functions belonging to each coordinate axis in such a way that the vision is modelled as depicted in Fig. 6; that means the intensity of vision rays are within the cone of vision as shown in Fig. 1, and their distribution in forward direction is Gaussian. Let us select two coordinate axes as x and z, where z is the axis pointing forward direction. Let us designate the angle of the visual scope as 2θ, as shown in Fig. 6.

From the figure we can write

$$s_z = \frac{1}{tg\theta} s_x \tag{42}$$

We assume that, in the Gaussian distribution, the number of rays beyond a certain probability is negligibly small. This probability is indicated as p, which corresponds to $x = s_x$. To restrict the rays within the cone we have to ensure that the z component of the Gaussian distribution has a minimal or larger value all the time within the probability limit designated as s_x in the figure. We call this value as bias along the axis z, which is designated as $m_z = s_x + s_z$ in Fig. 6. For $\theta = \pm\pi/2$, $s_z = 0$ and $m_z = s_x$. Further, to restrict the rays within the cone with the angle θ, the bias should have an additional value, which has to be equal to s_z given in Eq. (47). This implies that, the z component of the Gaussian should have a bias given by

$$m_z = s_z + s_x = [1 + 1/tg(\theta)]s_x \tag{43}$$

for the required θ. The probability p is connected to s_x by

$$\frac{1}{\sqrt{2\pi}\sigma} \bar{e}^{\frac{1}{2}\left(\frac{s_x}{\sigma}\right)^2} = p \tag{44}$$

From here we obtain

$$s_x = \sqrt{-2\ln\left[\sigma\, p\sqrt{2\pi}\right]} \tag{45}$$

Using this value in Eq. (19) we obtain

$$m_z = \left(1 + \frac{1}{tg\theta}\right)\sqrt{-2\ln\left[\sigma\, p\sqrt{2\pi}\right]} \tag{46}$$

One notes that to have a solution in Eq. (51) σ and p must have the condition

$$\sigma^2\, p^2 \ < \ \frac{1}{2\pi} \tag{47}$$

to obtain a real value for m_z. The parameter p essentially represents the probability after which the Gaussian function, which extends to infinity in both directions, is chopped off to model a solid angle limiting the vision. When p is chosen sufficiently small, so that the inequality

$$\sigma^2\, p^2 \ << \ \frac{1}{2\pi} \tag{48}$$

is satisfied, then the vector addition of the stochastic x and z components provides a Gaussian pdf modeling the vision. When p is chosen excessively small the forward direction of vision becomes very much accentuated within the cone. Choosing p so that inequality Eq. (53) is not satisfied then the distribution of rays within the cone is relatively less peaked in the forward direction than the reverse situation mentioned above.

4. Perception applications

4.1. Scene description by perception

The close relationship between attention and perception can be geometrically described to get insight into this intimacy. By means of this, the dependency of perception on distance is demonstrated below. Let us consider the probability density $f_z(z)$ in the case of viewing an infinite plane remains constant, which is denoted by c. In this case, from Eq. (17) we can write

$$\frac{2}{\pi}\frac{l_o}{z^2 + l_o^2} = c \tag{49}$$

which yields

$$z^2 + (l_o - \frac{1}{\pi\, c})^2 = \frac{1}{(\pi\, c)^2} \tag{50}$$

Taking both z and l_o as variables Eq. (50) represents a circle. This is illustrated in Fig. 7, together with the cone of vision of the observer denoted by P. We call this circle as *circle of attention*.

The inner part of this circle corresponds to the space, where the perception is deemed to be significant and relatively insignificant outside of this circle. Outside of the circle the attention is $f_z(z) < c$. This is illustrated in Fig. 8 via $f_z(z)$, where the heavy shaded area (right) is the perception, which corresponds to the viewed space shown as shaded area within the circle of attention (left). In this figure the distance between the observer and the plane, which is viewed is $l_o = l_a$. The same figure with $l_o = l_b$ is shown in Fig. 9. Clearly, $l_a < l_b$. The viewed spaces in Figs 8 and 9 are approximately equal as this is illustrated in Fig. 10, where the summation of two light shaded areas is assumed to be approximately equal to the heavy shaded area.

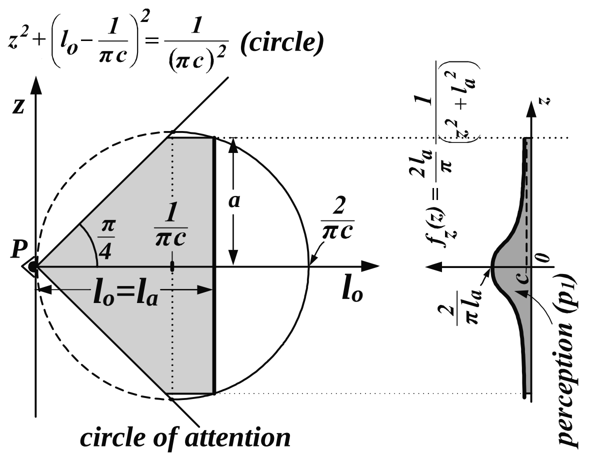

Fig. 8. Illustration of the perception of a scene, whose boundary is located at distance $l_o = l_a$ from the eye; the circle of attention, which belongs to a minimum visual attention threshold c (left), and corresponding computation of visual attention and perception of the scene are also shown (right).

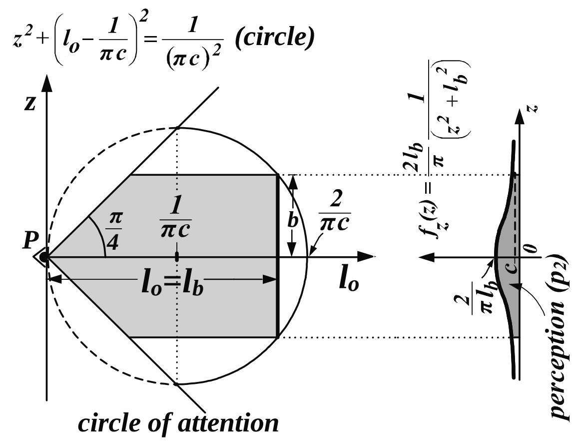

Fig. 9. Illustration of the perception of a scene, whose boundary is located at distance $l_o = l_b$ from the eye; the circle of attention, which is the same as in the situation shown in Fig. 8 (left), and corresponding computation of visual attention and perception of the scene are also shown (right).

The perceptions p_1 and p_2 respectively in the cases of $l_o = l_a$ and $l_o = l_b$ are indicated by the shaded areas in the $f_z(z)$ vs z plots. To compare the scenes as to perception p_1 and p_2 are computed as follows. For $l_o = l_a$, from Fig. 8, the integral of attention, as perception, yields

$$p_1 = \int_{-a}^{+a} f_z(z)\,dz = \frac{2\,l_a}{\pi}\int\limits_{-a}^{+a}\frac{dz}{z^2+l_a^2} \tag{51}$$

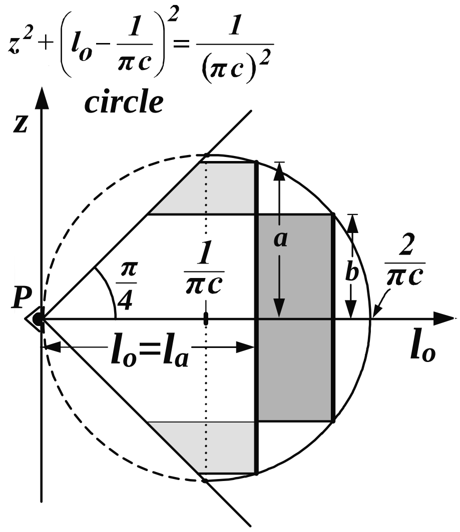

Fig. 10. Illustration of the effect of changing the viewing distance from l_a to l_b as to the viewed space in the preceding figures.

$$= \frac{2}{\pi}\,\mathrm{arctg}\frac{a}{l_a}$$

For $l_o = l_b$, from Fig. 9, the integral of attention yields

$$p_2 = \int_{-b}^{+b} f_z(z)\,dz = \frac{2\,l_b}{\pi}\int\limits_{-b}^{+b}\frac{dz}{z^2+l_b^2}$$
$$= \frac{2}{\pi}\,\mathrm{arctg}\frac{b}{l_b} \tag{52}$$

As an illustrative example, assuming that $a = 0.916$ R where R is the radius of the circle of attention, l_a is computed from Fig. 8 as

$$l_a = R + \sqrt{R^2 - a^2} = 1.4R \tag{53}$$

so that from Eq. (56) the perception p_1 becomes

$$p_1 = \frac{2}{\pi}\,\mathrm{arctg}\frac{0.916R}{1.4R} = \frac{2}{\pi}\,0.65 = 0.42 \tag{54}$$

Assuming that $b = 0.8R$, l_b is computed from Fig. 9 as

$$l_b = \mathrm{R} + \sqrt{\mathrm{R}^2 - \mathrm{b}^2} = 1.6R \tag{55}$$

so that from Eq. (57) the perception p_2 becomes

$$p_2 = \frac{2}{\pi}\,\mathrm{arctg}\,\frac{0.8R}{1.6R} = \frac{2}{\pi}0.5 = 0.32 \tag{56}$$

Summarizing the computations, for the geometry in Fig. 8, the perception is higher relative to the case seen in Fig. 9. In these illustrations, the geometrical proportions are given by

Fig. 11. Illustration of an architectural design, which corresponds to the perception computation shown in Fig. 8.

$$\frac{\mathrm{a}}{l_o} = 0.65 \text{ and } \frac{\mathrm{b}}{l_b} = 0.5 \tag{57}$$

and the respective perceptions are given by

$$\mathrm{p}_1 = 0.42 \text{ and } \mathrm{p}_2 = 0.32 \tag{58}$$

The result in Eq. (58) shows that the perception p_1 is about 30% more than p_2. This perception difference is clearly exemplified in two scenes given in Figs 11 and 12, where each scene corresponds to the same circle of attention shown in Figs 8 and 9. The probability of getting the visual information present in the scene shown in Fig. 11 is higher than for the scene shown in Fig. 12. In Fig. 11, perception given by p_1 is higher compared to $p_{2,}$, which means that the details can be better obtained from the scene as compared to Fig. 12.

In Fig. 11, the counter is far but it is in the frontal view, therefore it has higher perception. On the other hand, the other objects at the side parts are nearer to the observation point. Therefore they have comparable perception with the frontal objects. The overall result is a balanced view about the scene. In Fig. 12, the same scene is at greater distance, and therefore perception of it is less compared to Fig. 11. In Fig. 12, the side walls purposely restrict the vision of the observer to establish the same minimum attention c, defined in Fig. 7, for the front wall as that used in Fig. 11. The side walls normally cannot have any perceptual information including the texture of the walls. However, without a texture it is not possible to represent a wall or to restrict the vision. This is the reason why the side walls are left as plain as possible for minimal perceptual impact.

4.2. *Visual openness measurement for design*

In this application, visual openness perception is considered, which has essential implications on general design process as well as robot movement. The visual openness perception is obtained from visual perception data, which are derived from the distances between observer and environment. That is, the visual openness of a space is perceived in mind with the association of the distances. As a measure of openness we define the *mean-perceived-distance* $\overline{l}_o$, which is in mathematical terms, given by

$$\begin{aligned}\overline{l}_\mathrm{o} &= \int_{l_o}^{\sqrt{2}l_o} x f_x(x) dx = \frac{4}{\pi} \int_{l_o}^{\sqrt{2}l_o} l_o \frac{dx}{\sqrt{x^2 - l_\mathrm{o}^2}} \\ &= \frac{4l_\mathrm{o}}{\pi} \cosh^{-1}\sqrt{2} \approx 1.12 l_o\end{aligned} \tag{59}$$

considering Eq. (10).

For the visual openness perception measurement we consider the laser beam technology, where the length of each visual ray between human eye and an object in the environment is represented by a laser ray spanning the ray source and the object. This is accomplished in virtual reality by a virtual agent. In this case, the laser source provides rays not in scanning mode but as a random source of rays with certain statistical properties, as described above and illustrated in Fig. 6. Explicitly the perception computations are used to simulate the human vision and equip the virtual agent with human-like vision. The averaging to obtain the mean-perceived-

Fig. 12. Illustration of an architectural design, which corresponds to the perception computation shown in Fig. 9.

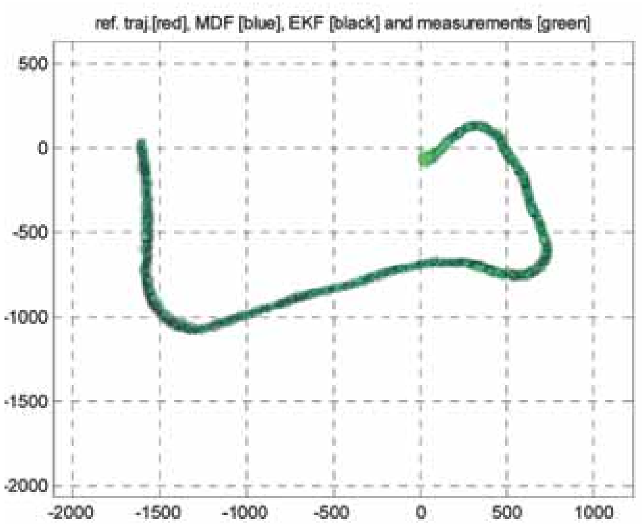

Fig. 13. Total robot trajectory in openness-perception based autonomous navigation; measurement, Kalman filtering, and multiresolutional filtering estimation are in the figure, but due to scale, only the bunch of cross symbols is seen; such details are shown in Figs 14 and 15.

distance is performed by exponential averaging in real-time on a sample by sample basis. If the time constant of the exponential averaging is kept sufficiently small then the measurement outcome can be used for robot navigation due to minimal latency of the measurement. In the case of a moving robot, it experiences human-like interaction with the environment for autonomous movement [13].

For the openness measurement in architectural design a sigmoidal mapping function is used to scale the mean-perceived-distance between zero and one [9]. Such ad hoc approach has certain merits in architectural design due to the soft nature of the design task, and the integration of the openness measurement into design in this way is a valid pragmatic approach. In this context, in place of sigmoidal mapping, some other appropriate function can be considered. Such mappings represent supposedly the brain processes to some extend, and therefore it is a complex issue. To make more robust visual openness measurements, fuzzy logic is appealing, because it absorbs the imprecision involved in the measurement procedure, where the imprecision

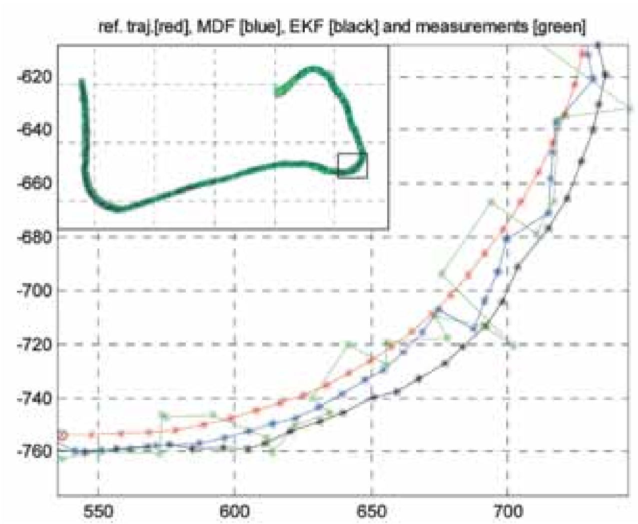

Fig. 14. Robot trajectory, measurement, Kalman filtering, and multiresolutional filtering estimation, in bending mode.

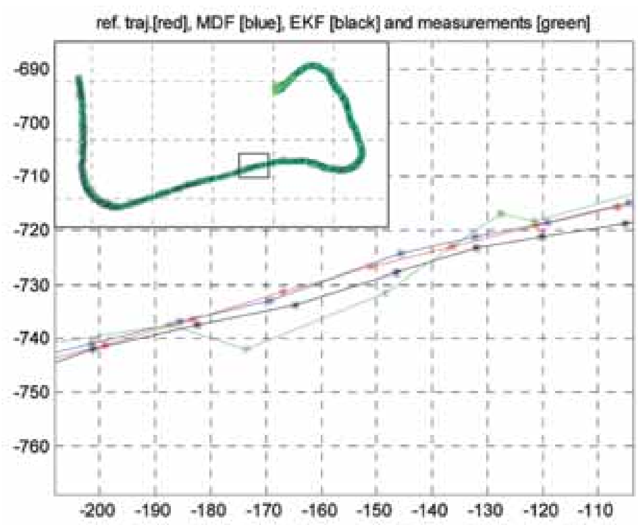

Fig. 15. Robot trajectory, measurement, Kalman filtering, and multiresolutional filtering estimation in straight-ahead mode.

is due to the involvement of the brain processes that are responsible for the attribution of semantic labels to percepts.

4.3. Perceptual robotics

From perceptual robotics viewpoint it may be desirable to equip a robot with human like interpretation of the perceptions for path planning with humanoid behaviour at the same time. The perception computations presented in the previous sections are important to simulate the human vision and equip the virtual agent with this. In this application we consider a vision cone for the robot as seen in Fig. 6, meaning that we consider the perpendicular infinite plane geometry. By doing so the robot has a main focus to forward direction fitting to its task to move forward without hitting any wall delineating the trajectory. The forward movement is based on the mean-perceived-distance, which is defined earlier as measure of *openness*.

The perception outcomes from the model are implemented in an avatar-robot in virtual reality. The perceptual approach for autonomous movement in robotics is important in several respects. On one hand, perception is very appropriate in a dynamic environment, where predefined trajectory or trajectory conditions, like occasional obstacles or hindrances, are duly taken care of. On the other hand, the approach can better deal with the complexity of environments by processing environmental information selectively. For the verification of the theoretical study done in this work, a trajectory is defined and the same trajectory is estimated from the simulated perception measurements by means of sensor fusion, which involves Kalman filtering and wavelet transform for the multiresolutional representation of the sensory data. The environment delineating the trajectory is defined as a 3-D corridor of 100 units wide in Fig. 13 and the delineation lines are not shown in the figure. The robot is expected to move through that space, and the trajectory is defined as the central line of this space in 2-D. The measurement data, the estimated robot trajectory, and the trajectory itself are in Fig. 13, however they are not seen due to the scaling made for the sake of showing the total space. Kalman filtering is especially effective due to modelling the measurement noise and the process noise in the estimation process.

Perception data are subjected to decomposition and information fusion and the implementation is carried out by means a virtual agent in a virtual reality environment [13]. The rays stemming from the agent's eye interact with the environment and their lengths are obtained. In reality this length is obtained via laser sensor, for instance. The interaction points with the environment are recorded and the position is identified as the exponentially averaged value of the coordinates. This means there is some delay in the measurements, as delay of perception, depending on the time constant involved in the exponential averaging.

Presently, the experiments have been done with simulated measurement data since the multiresolutional filtering runs in a computationally efficient software platform, which is different than the computer graphics platform of virtual reality.

For the simulated measurement data, first the trajectory of the virtual agent is established by changing the system dynamics from the straight ahead mode to bending mode for a while three sharp bending modes are seen in Fig. 13, whose start and stop are indicated vaguely by small circles. with the complete trajectory of the perceptual agent. The state variables vector is given by

$$X = [x, \dot{x}, y, \dot{y}, \omega] \tag{60}$$

where ω is the angular rate and it is estimated during the move. When the robot moves in a straight line, the angular rate becomes zero. For explicit illustration of the experimental outcomes the same figure with a different zooming range and the zooming power is given in Fig. 14 for bending mode and 15 for a straight-ahead case. In details, there are three lines plotted in the figures. The black line is the extended Kalman filtering estimation at the highest resolution of the perception measurement data. The cross symbols connected by the green lines represent the measurement data set. The outcome of the multiresolutional fusion process is given with the blue line.

From the experiments it is seen that, the Kalman filtering is effective for estimation of the trajectory from perception measurement. Estimation is improved by the multiresolutional filtering. Estimations are relatively more accurate in the straight-ahead mode.

5. Conclusions

The visual perception is investigated by means of a probabilistic mathematical model and applied to design and robotics. In design, the visual perception of scenes is quantified, so that they can be consistently compared, and visual openness is measured. The probabilistic model is most appropriate since the human vision system deals with natural images using their statistical properties rather than dealing with each piece of image information in order to be able to cope with the complexity of information. By means of the model, the characteristic aspects of visual perception are substantiated providing insight into the complex visual process; thereby the gap between environmental stimulus and its mental realization is spanned without recourse to the realm of neuroscience.

There are a number of application areas for probabilistic perception modelling. As to this research, the perception is a commonplace of two important diverse research and application areas which are of concern as the motivation in this work. The first application considered in this research is spatial design in architecture. Since the substantial part of information that human receives from the environment is visual, the modelling of vision and implementation on a virtual agent/robot can provide important desirable features to be able to quantify the visual aspects in an architectural design, as this is exemplified in this work. The second application considered in this research concerns the branch of autonomous robotics, where the robot navigation is based on visual perception, so that robot navigation becomes similar to that of human. Among other surmised applications, the research has important implications for robotics in general, including industrial, perceptual and/or intelligent robotics. From the industrial robotics viewpoint, robot-driven transportation is a potential application, for instance. Another important application of common interest is autonomous robotics where the robot moves in an environment without collision by having real-time visual openness perception information during the move without any predefined trajectory. This approach is rather unique as to the novelty of visual openness perception concept presented here for robotics with a realized prototype in virtual reality. Additionally, in the virtual reality, autonomous robotics presumably is an alternative for animation-based indoors or outdoors move that are prescribed in advance, while in the autonomous robotics case it becomes more explorative, as it is not prescribed beforehand.

As the state of the art, the autonomous perceptual robotics is actively engaged to meet the computational challenge of the soft issues of vision/perception [2,3, 13]. Therefore, the novel probabilistic theory of vision is a major demanded contribution to robotics technology.

As the state of the art, the soft issues of vision in architectural design are dealt with the subjective decisions, without any amenable quantification subject to computations. Therefore, the novel probabilistic theory of vision is a major contribution to architectural design.

The novel probabilistic vision/perception theory is a common place for architecture and robotics technology, where the latter includes engineering technologies at large. The establishment of a firm commonplace for such seemingly distant disciplines is an exemplary interdisciplinary research.

Acknowledgements

The contribution of Paula T.M. van den Bergh and Sjade A. van Arum to this research is gratefully acknowledged.

Appendix

In the case of viewing a circular geometry from the central point of the circle, the pdf of the visual perception becomes uniform as one intuitively concludes. Referring to Fig. A1, this is shown mathematically as follows. In circular geometry, the random variable connected to θ is ω where

$$f_\theta(\theta) = 1/(\pi/2)$$

$$\omega = g(\theta) = l_o\theta \tag{A1}$$

$$g'(\theta) = \frac{dg(\theta)}{d\theta} = l_o \tag{A2}$$

Using the theorem on the function of random variable, given by Eq. (4) in the text, we write

$$f_\omega(\omega) = \frac{f_\theta(\theta_1)}{|g'(\theta_1)|} \tag{A3}$$

The root of Eq. (A1) is given by $\theta_1 = \frac{\omega}{l_o}$, which gives

$$f_\omega(\omega) = \frac{2}{\pi\ l_o} \tag{A4}$$

as uniform pdf of visual perception, which satisfies $\int\limits_{-l_o\pi/4}^{l_o\pi/4} f(\omega)d\omega = \int\limits_{-l_o\pi/4}^{l_o\pi/4} \frac{2}{\pi l_o} d\omega = 1$, as it should be.

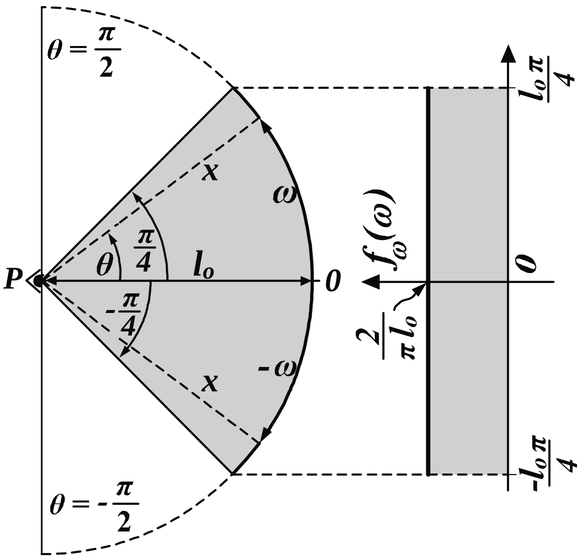

Fig. A1. Computation of visual attention in the case of a circular geometry.

References

[1] E.H. Adelson and J.R. Bergen, The Plenoptic function and the elements of early vision, in: *Computational Models of Visual Processing*, M. Landy and J.A. Movshon, eds, MIT Press, Cambridge, MA, 1991, pp. 3–20.

[2] E. Ahle and D. Söffker, *A cognitive-oriented architecture to realize autonomous behavior – part I: theoretical background*, Proc. 2006 IEEE Conf. on Systems, Man, and Cybernetics, October 8–11, Taipei, Taiwan, 2215–2220.

[3] E. Ahle and D. Söffker, *A cognitive-oriented architecture to realize autonomous behavior – part II: application to mobile robots*, Proc. 2006 IEEE Conf. on Systems, Man, and Cybernetics, October 8–11, Taipei, Taiwan, 2221–2227.
[4] M.A. Arbib, ed., *The Handbook of Brain Theory and Neural Networks*, Cambridge: MIT Press, second edition, 2003.
[5] M.H. Ashcraft, *Cognition, Pearson Prentice Hall*, 2006.
[6] M. Beetz, T. Arbuckle, T. Belker, A.B. Cremers, D. Schulz, M. Bennewitz, W. Burgard, D. Hähnel, D. Fox and H. Grosskreutz, Integrated, plan-based control of autonomous robots in human environments, *IEEE Intelligent Systems* **16**(5) (2001), 56–65.
[7] M. Bertero, T.A. Poggio and V. Torre, Ill-posed problems in early vision, *Proceedings of the IEEE* **76**(8) (1988), 869–889.
[8] J. Bigun, *Vision with direction*, Springer Verlag, 2006.
[9] M.S. Bittermann, I.S. Sariyildiz and Ö. Ciftcioglu, *Visual space perception model identification by evolutionary search*, Proc. 15th Int. Conf. Design 2006, May 15–18, 2006, Dubrovnik, Croatia, 185–192.
[10] B. Adams, C. Breazeal, R.A. Brooks and B. Scassellati, Humanoid robots: a new kind of tool, Intelligent Systems and Their Applications, *IEEE* **15**(4) (2000), 25–31.
[11] C. Bundsen, A theory of visual attention, *Psychological Review* **97**(4) (1990), 523–547.
[12] J. Cataliotti and A. Gilchrist, Local and global processes in surface lightness perception, *Perception & Psychophysics* **57**(2) (1995), 125–135.
[13] Ö. Ciftcioglu, M.S. Bittermann and I.S. Sariyildiz, *Autonomous robotics by perception*, Proc. ISCIS & ISIS 2006, Joint 3rd Int. Conf. on Soft Computing and Intelligent Systems and 7th Int. Symp. on Advanced Intelligent Systems, September 20–24, 2006, Tokyo, Japan, 1963–1970.
[14] Ö. Ciftcioglu, M.S. Bittermann and I.S. Sariyildiz, *Towards computer-based perception by modeling visual perception: a probabilistic theory*, Proc. 2006 IEEE Int. Conf. on Systems, Man, and Cybernetics, October 8–11, 2006, Taipei, Taiwan, 5152–5159.
[15] R.L. Cook, Distributed ray tracing, *Computer Graphics* **18**(3) (1984), 137–145.
[16] A. Cowey and E.T. Rolls, Human cortical magnification factor and its relation to visual acuity, *Exp Brain Res* **21** (1974), 447–454.
[17] A. Desolneux, L. Moisan and J.-M. More, A grouping principle and four applications, *IEEE Trans on Pattern Analysis and Machine Intelligence* **25**(4) (2003), 508–513.
[18] R. Eckmiller, O. Baruth and D. Neumann, *On human factors for interactive man-machine vision: requirements of the neural visual system to transform objects into percepts*, Proc. IEEE World Congress on Computational Intelligence WCCI 2006 – Int. Joint Conf. on Neural Networks, July 16–21, 2006, Vancouver, Canada, 699–703.
[19] R. Eckmiller, D. Neumann and O. Baruth, Tunable retina encoders for retina implants: why and how, *Journal of Neural Engineering* **2** (2005), 91–104.
[20] J.M. Findlay and I.D. Gilchrist, *Active Vision*, Oxford University Press, 2003.
[21] J. Foster, *The Nature of Perception*, Oxford University Press, 2000.
[22] K. Ghosh, S. Sarkar and K. Bhaumik, A possible explanation of the low-level brightness-contrast illusions in the light of an extended classical receptive field model of retinal ganglion cells, *Biological Cybernetics* **94** (2006), 89–96.
[23] R. Hecht-Nielsen, *The mechanism of thought*, Proc. IEEE World Congress on Computational Intelligence WCCI 2006, Int. Joint Conf. on Neural Networks, July 16–21, 2006, Vancouver, Canada, 1146–1153.
[24] P. Herrero, C. Greenhalgh and A. De Antonio, Modelling the sensory abilities of intelligent virtual agents, *Autonomous Agents and Multi-Agent Systems* **11** (2005), 361–385.
[25] D.H. Hubel, *Eye, brain, and vision*, Scientific American Library, 1988.
[26] D.H. Hubel, Exploration of the primary visual cortex, 1955–1978 (Nobel Lecture), *Nature* **299** (1982), 515–524.
[27] L. Itti, C. Koch and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans On Pattern Analysis and Machine Intelligence* **20**(11) (1998), 1254–1259.
[28] B. Julesz, Binocular depth perception without familiarity cues, *Science* **145** (1964), 356–362.
[29] J.T. Kajiya, The rendering equation, in: *Computer Graphics* (*Proc. of ACM SIGGRAPH*), (Vol. 20), D.C. Evans and R.J. Athay, eds, 1986, pp. 143–150.
[30] M.W. Levine and J.M. Shefner, *Fundamentals of Sensation and Perception*, London: Addison-Wesley, 1981.
[31] D. Marr, *Vision*, Freeman, 1982.
[32] J.K. O'Regan, H. Deubel, J.J. Clark and R.A. Rensink, Picture changes during blinks: looking without seeing and seeing without looking, *Visual Cognition* **7** (2000), 191–211.
[33] G. Oriolio, G. Ulivi and M. Vendittelli, Real-time map building and navigation for autonomous robots in unknown environments, *IEEE Trans. on Systems, Man and Cybernetics – Part B: Cybernetics* **28**(3) (1998), 316–333.
[34] S.E. Palmer, *Vision Science*, Cambridge, MIT Press, 1999.
[35] A. Papoulis, *Random Variables and Stochastic Processes*, McGraw-Hill, New York, 1965.
[36] J.M. Plumert, J.K. Kearney, J.F. Cremer and K. Recker, Distance perception in real and virtual environments, *ACM Trans Appl Percept* **2**(3) (2005), 216–233.
[37] T.A Poggio, V. Torre and C. Koch, Computational vision and regularization theory, *Nature* **317**(26) (1985), 314–319.
[38] M.I. Posner and S.E. Petersen, The attention system of the human brain, *Annual Review of Neuroscience* **13** (1990), 25–39.
[39] S.J.D. Prince, A.D. Pointon, B.G. Cumming and A.J. Parker, Quantitative analysis of the responses of V1 Neurons to horizontal disparity in dynamic random-dot stereograms, *J Neurophysiology* **87** (2002), 191–208.
[40] A. Rabinovich, T. Lange, J.M. Buhmann and S. Belongie, *Model Order Selection and Cue Combination for Image Segmentation*, Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition CVPR'06, June 17–22, 2006, New York, USA, 1130–1137.
[41] R.A. Rensink, J.K. O'Regan and J.J. Clark, To see or not to see: The need for attention to perceive changes in scenes, *Psychological Science* **8** (1997), 368–373.
[42] H. Robinson, *Perception*, Routledge, 2001.
[43] D. Smith, *The Problem of Perception*, Harvard University Press, 2001.
[44] J.G. Taylor, *Towards an autonomous computationally intelligent system*, (Tutorial), IEEE World Congress on Computational Intelligence WCCI 2006, July 16–21, 2006, Vancouver, Canada.
[45] A.M. Treisman, *How the deployment of attention determines what we see*, Visual Cognition 14, 4–8, -8/August–December 2006, 411–443.
[46] A.M. Treisman and G. Gelade, A feature-integration theory of attention, *Cognitive Psychology* **12** (1980), 97–136.

[47] J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y. Lai, N. Davis and F. Nuflo, Modeling visual attention via selective tuning, *Artificial Intelligence* **78** (1995), 507–545.

[48] I. Wald, T. Ize, A. Kensler, A. Knoll and S.G. Parker, Ray tracing animated scenes using coherent grid traversal, *ACM Transactions on Graphics – Proc. of ACM SIGGRAPH 2006* **25**(3) (2006), Boston, USA, 485–493.

[49] M. Wang and J.N.K. Liu, Online path searching for autonomous robot navigation. Proc. IEEE Conf. on Robotics, in: *Automation and Mechatronics*, (Vol. 2), Singapore, December 1–3, 2004, pp. 746–751.

[50] M. Wertheimer, Untersuchungen zur Lehre der Gestalt, II, *Psychologische Forschung* **4** (1923), 301–350. Translation published as Laws of organization in perceptual forms, in: *A source book of Gestalt psychology*, W. Ellis, London: Routeledge & Kegan Paul, 71–88.

[51] B. Widrow, Cognitive memory (invited talk) at IEEE World Congress on Computational Intelligence WCCI 2006, July 16–21, 2006, Vancouver, Canada, also in: *Computational Intelligence: Principles and Practice*, G.Y. Yen and C.B. Fogel, eds, WCCI 2006, IEEE Computational Intelligence Society, 2006, pp. 17–46.

[52] T.N. Wiesel, Postnatal development of the visual cortex and the influence of environment (Nobel Lecture), *Nature* **299** (1982), 583–591.

[53] P. Willemsen and A.A. Gooch, *Perceived egocentric distances in real, image-based, and traditional virtual environments*, Proc. IEEE Virtual Reality Conference 2002, March 24–28, 2002, 275–276.

[54] P. Wonka, M. Wimmer, K. Zhou, S. Maierhofer, G. Hesina and A. Reshetov, Guided visibility sampling, *ACM Transactions on Graphics – Proc. of ACM SIGGRAPH* **25**(3) (2006), Boston, USA, 494–502.

[55] A.L. Yarbus, *Eye Movements and Vision*, New York, Plenum Press, 1967.

[56] W. Zhang, L. Wenyin and K. Zhang, Symbol recognition with kernel density matching, *IEEE Trans. Pattern Analysis and Machine Intelligence* **28**(12) (2006), 2020–2024.