ISCIS & isis 2006, Joint 3rd International Conference on Soft Computing and Intelligent Systems and
7th International Symposium on Advanced Intelligent Systems, September 20-24, 2006, Tokyo, Japan

Autonomous Robotics by Perception

Ö. Ciftcioglu
Dept. of Building Technology
Delft University of Technology
Berlageweg 1, 2628 CR Delft,
The Netherlands
o.ciftcioglu@tudelft.nl

M.S. Bittermann
Dept. of Building Technology
Delft University of Technology
Berlageweg 1, 2628 CR Delft,
The Netherlands
m.s.bittermann@tudelft.nl

I.S. Sariyildiz
Dept. of Building Technology
Delft University of Technology
Berlageweg 1, 2628 CR Delft,
The Netherlands
i.s.sariyildiz@tudelft.nl

***Abstract*—Visual perception-based autonomously moving virtual agent in virtual reality as a counterpart of an actual robot moving with a given dynamics is investigated. The visual perception is mathematically modelled as a probabilistic process obtaining and interpreting visual data from an environment. The perception obtained in the form of measurements in 2D is used for perceptual robot navigation. By means of this twofold gain is obtained; while the autonomous robot is navigated, it is equipped with some human-like behaviour, thereby dealing with complexity and environmental dynamics. The visual data is processed in a multiresolutional form via wavelet transform and optimally estimated via extended Kalman filtering in each resolution level and the outcomes are fused for improved estimation of the trajectory. The perceptual robotics experiments are carried out in virtual reality for the demonstration of the feasibility of the investigations in this domain. The computer experiments are carried out with perception measurement data, and the sensor/data fusion experiments are carried out by means of simulation. The improvement on the trajectory estimation by means of sensor/data fusion is demonstrated. The research is connected to building technological robotics, where some form of perceptual intelligence, like reaction to moving objects around, is required during operation.**

I. Introduction

Visual perception is one of the important information sources playing role on human's behavior. Although the concept of perception is well established and defined, its properties and its quantification are not clearly established yet. For verbal communications and implementations by a human expert, the common concept of perception may be satisfactory. However, in tasks where precision is desired, more precise qualities of perception, like degree of perception, need to be specified. The concept of perception inherently includes some uncertainty in the act of perception. That is, it is not a definitive action but it is a circumstantial action depending on the circumstances and the environment subject to perception. With this understanding, it is natural to think of to deal with uncertainty with the existing scientific computational methodologies. That is to treat perception with the methods probability theory. Probability theoretic perception model having been established, the perception outcomes from the model is 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. Since perception is a complex intelligent human activity the subject is soft and the research is of essential interest to soft computing. The organization of the paper is as follows. Section two gives the brief description of the perception model developed in the framework of ongoing perceptual robotics research. Section three deals with the multiresolutional Kalman filtering and sensor fusion for enhanced estimation. Section four is reserved for the experimental studies. Section five gives some discussions and the conclusions.

II. Probabilistic Model of Visual Perception

We start with the basics of the perception process with a simple yet a fundamental visual geometry. This is shown in figure 1a.

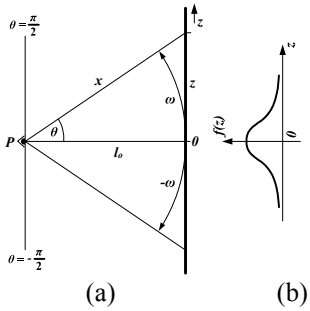

Fig. 1 The geometry of visual perception from a top view, where P represents the position of eye, looking at a vertical plane with a distance l_o to the plane; $f_z(z)$ is the probability density function in z-direction.

In figure 1, the observer is facing and looking at a vertical plane from the point denoted by P. By means of looking action the observer pays visual attention equally to all locations on the plane in the first instance. That is, the observer visually experiences all locations on the plane without any preference for one region over another. Each

point on the plane has its own distance within the observer's scope of sight which is represented as a cone. The cone has a solid angle denoted by θ. The distance of a point on the plane and the observer is denoted by x and the distance between the observer and the plane is denoted by l_o. Since visual perception is associated with distance, it is straightforward to proceed to express the distance x of visual perception in terms of θ and l_o. From figure 1, this is given by

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

Since we consider that the observer pays visual attention equally to all locations on the plane in the first instance, the probability of getting attention for each point on the plane is the same so that the associated probability density function (pdf) is uniformly distributed. This positing ensures that there is no visual bias at the beginning of visual perception as to the differential visual resolution angle $d\theta$. Assuming the scope of sight is defined by the angle $\theta = \pm \pi/2$, the pdf f_θ is given by

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

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

To find the pdf of the variable x denoted $f_x(x)$ for a given x we consider the theorem on the *function of random variable* [1] and solve the equation

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

for θ in terms of x. If $\theta_1, \theta_2, ..., \theta_n, ..$ are all its *real* roots, $x=g(\theta_1) = g(\theta_2) = ... = g(\theta_n) = ...$
Then

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

According the theorem above,

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

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

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

Using (7) in (5), we obtain

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

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

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

for the interval $l_o \le x \le \infty$. For this interval, the integration below becomes

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

as it should be as pdf. The sketch of $f_x(x)$ vs x is given in figure 2. As to (9), two observations are due. Firstly, it is interesting to note that for the plane geometry in figure 1, the visual perception is sharply concentrated close to $\theta \cong 0$, that is perpendicular direction to the plane. This striking result is in conformity with the common human experience as to visual perception.

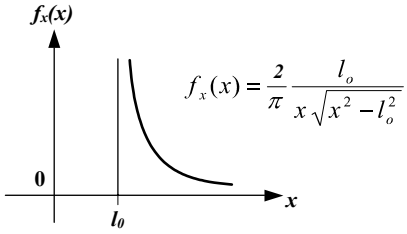

Fig. 2. Sketch explaining the relative importance of the viewing direction in visual attention and perception.

Namely, for this geometry the visual perception is strongest along the axis of the cone of sight relative to the side directions. This is simply due to the fact that, for the same differential visual resolution angle $d\theta$, one can perceive visually more details on the infinite plane in the perpendicular direction as this is sketched in figure 3.

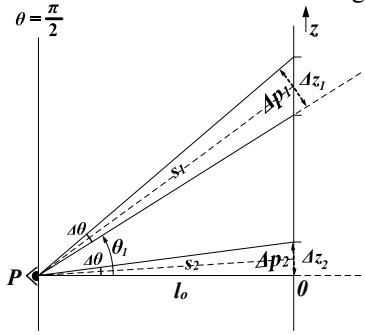

Fig. 3. Sketch explaining the relative importance of the viewing direction in visual attention and perception.

Secondly, the visual perception is given via a probability density at a point; if we consider the stimulus of perception is due to the light photons, it is the relative number of photons as stimulus at infinitesimally small interval, per unit length. Integration of these photons within a certain length gives the intensity of the stimulus, which is a measure of perception. This implies that, perception is a probabilistic concept and the measure of perception is probability, meaning that it is different than "seeing", which is goal-oriented and therefore definitive.

By means of pdf calculation similar to obtain $f_x(x)$ one can obtain $f_z(z)$ as [2].

$$f_z(z) = \frac{l_o}{\pi(l_o^2 + z^2)} \tag{11}$$

for the interval $-\infty \le z \le \infty$. For this interval

$\int_{-\infty}^{+\infty} f_z(z)\,dz = \frac{l_o}{\pi}\int_{-\infty}^{+\infty}\frac{dz}{z^2+l_o^2} = 1$ as it should be.

The variation of pdf $f_z(z)$ is shown in figure 1b.

The perception implementation of this research is made in the virtual reality where a virtual agent plays the role of humanoid robot. Considering the geometry, the derivation of perception in more general cases is presented in another research [3]. All perceptual robot behaviour is simulated in this environment to exercise the research outcome extensively without hardware or environmental limitations. However, the transfer of the perception technology being developed to the robotics is the final goal. A typical visual robotic perception measurement with the virtual agent in real-time is shown in figure 4 where the vision beams underlying the measurements together with the plot of real-time measurement outcomes are clearly seen. The implementation of the vision to virtual agent is accomplished as follows.

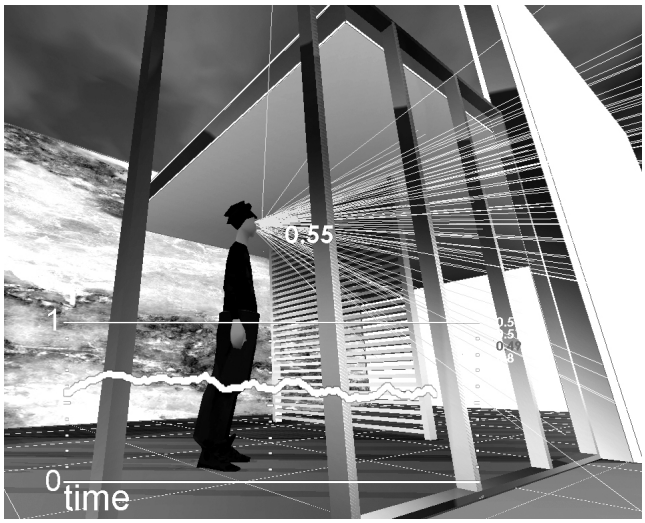

Fig. 4. Visual openness perception measurements via an agent in virtual reality.

The perception data subject to decomposition and information fusion are obtained by means a visual agent in the virtual reality environment. The virtual human senses its spatial environment by sending rays from its eyes and measuring the length of each of ray as they hit shapes, which form the environment. In the virtual reality, for the computational convenience, the vision of the virtual agent can be modelled by means of Gaussian shaped form. However, the theoretical considerations given above indicated that the vision model presents a Gaussian-like shape as given in (11). Therefore it is necessary to establish the equivalence of (11) to a corresponding Gaussian. For this aim first we expand the probability density expression given in (11), which yields

$$\frac{l_o}{\pi(l_o^2+z^2)} = \frac{1}{\pi l_0}\frac{1}{1+\left(\frac{z}{l_0}\right)^2} = \frac{1}{\pi l_0}\left(1-\left(\frac{z}{l_0}\right)^2+\left(\frac{z}{l_0}\right)^4-\left(\frac{z}{l_0}\right)^6+\ldots\right) \quad (12)$$

Second, we expand a Gaussian probability density with variance σ^2, which yields

$$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}\sigma}\left(1-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2+\frac{1}{2}\left(\frac{1}{2\sigma^2}\right)^2 z^4-\ldots\right) \quad (13)$$

Due to the equivalency we wish to have we equate the respective coefficients of the powers of the variable z.

For the first term we write:

$\frac{1}{\pi l_0} = \frac{1}{\sqrt{2\pi}\,\sigma}$, which yields $\pi l_0 = \sqrt{2\pi}\,\sigma$, so that

$$\sigma = \sqrt{\frac{\pi}{2}}l_0 \quad (\sigma_1 \approx 1.213\,l_0 \text{ or } l_{01} \approx 0.800\,\sigma_1) \quad (14)$$

For the second term we write:

$\frac{1}{\pi l_0}\frac{1}{l_o^2} = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{2\sigma^2}$, which yields $\pi l_0^3 = 2\sqrt{2\pi}\sigma^3$, so that

$$\sigma^3 = \frac{\pi}{2\sqrt{2\pi}}l_0^3 \quad \sigma^3 = \sqrt{\frac{\pi}{8}}l_o^3 \quad (\sigma_2 \approx 0.856\,l_0 \text{ or } l_{02} \approx 1.168\,\sigma_2) \quad (15)$$

For the third term we write:

$\frac{1}{\pi l_0}\frac{1}{l_0^4} = \frac{1}{\sqrt{2\pi}\,\sigma}\frac{1}{2}\left(\frac{1}{2\sigma^2}\right)^2$, which yields $\pi l_o^5 = 8\sqrt{2\pi}\,\sigma^5$, so that

$$\sigma = \sqrt[10]{\frac{\pi}{128}}l_0 \quad (\sigma_3 \approx 0.690 l_0 \text{ or } l_{03} \approx 1.449\,\sigma_3) \quad (16)$$

If we consider the weighted average of the standard deviation we have to identify the weights as follows. Consider the coefficient $1/\pi\ l_o$ in (12). As to this coefficient, taking l_{o1} from (14), for the first term, we obtain

$$w_1 = \frac{1}{\pi l_{o1}} = \frac{1}{\pi\,0.8} = 0.400$$

In the same way, taking l_{o2} from (15), for the second term, we obtain

$$w_2 = \frac{1}{\pi l_{o2}} = \frac{1}{\pi 1.168} = 0.272$$

and for the third term, taking l_{o3} from (16), for the third term, we obtain

$$w_3 = \frac{1}{\pi l_{o3}} = \frac{1}{\pi 1.445} = 0.220\ .$$

Noting that $w_1+w_2+w_3=0.897$, the normalized weight coefficients become

$$w_{1n} = 0.45; \quad w_{2n} = 0.30; \quad w_{3n} = 0.25$$

and finally from (14), (15), and (16) we obtain the standard deviation in terms of l_0 by three term approximation in series expansion as

$$\begin{aligned}\sigma &= w_1\,1.25\,l_0 + w_2\,0.86\,l_0 + w_3\,0.69\,l_0 \\ &= 0.56\,l_0 + 0.26\,l_0 + 0.17\,l_0 \approx 1.00\,l_0 \\ \sigma &\approx l_0 \end{aligned} \quad (17)$$

To model forward vision in 3D-space implemented via a virtual agent three Gaussian functions are used, which each delivers independently a random number for three orthogonal coordinate axes. These numbers are used as the components of the direction vectors for the vision rays. The formation of simulated vision in conformity to (11) is quite important for modelling the perception accurately. 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 figure 5; that means the intensity of vision rays are within the cone of vision as shown in figure 2 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 cone as 2θ, as shown in figure 5. From figure 5, we can write

$$s_z = \frac{1}{tg\theta} s_x \tag{18}$$

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 to be have a minimal value or larger all the time within the probability limit designated as s_x in the figure.

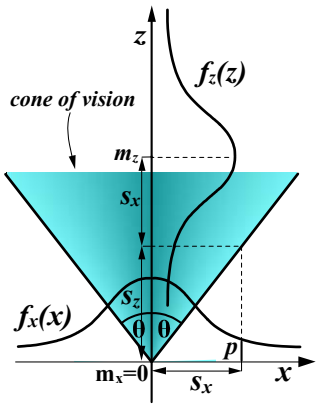

Fig. 5 The Gaussian visual perception in forward direction with a cone of angle 2θ, where z is the forward direction.

We call this value as bias along the axis z, which is designated as $m_z = s_x + s_z$ in figure 8. 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 (18). This implies that, the z component of the Gaussian should have a bias given by

$$m_z = s_z + s_x = [1+1/tg(\theta)]\, s_x \tag{19}$$

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

$$\frac{1}{2\pi\sigma} e^{\frac{1}{2}\left(\frac{s_x - m}{\sigma}\right)^2} = p \tag{20}$$

From here we obtain

$$s_x = \sqrt{-2\ln\left[\sqrt{2\pi}\sigma\, p\right]} \tag{21}$$

Using this value in (19) we obtain

$$m_z = \left(1 + \frac{1}{tg\theta}\right)\sqrt{-2\ln\left[\sqrt{2\pi}\sigma\, p\right]} \tag{22}$$

One notes that σ above is given by (17) and to have a solution in (22) σ and p must have the condition

$$\sigma^2 p^2 < \frac{1}{2\pi} \tag{23}$$

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

$$\sigma^2 p^2 << \frac{1}{2\pi} \tag{24}$$

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 (24) 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.

The visual perception is computed via exponential averaging [4] of the distances associated with the backscattered beams. A typical measurement is shown in figure 6, where the delay of the measurement due to the time constant (τ) of the exponential averaging is clearly seen.

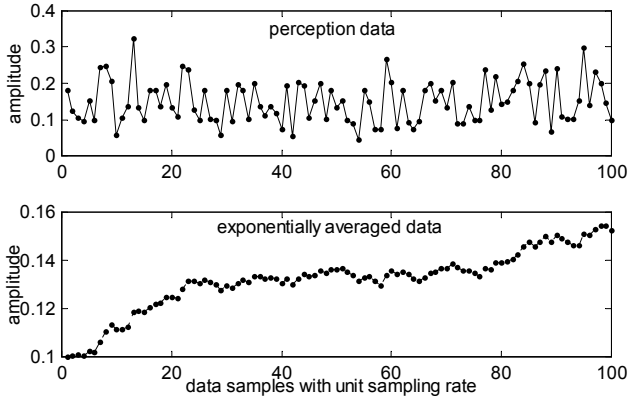

Fig. 6 The perception data (upper) and visual perception measurement outcome (lower) obtained by exponential averaging.

The perception data subject to decomposition and information fusion are obtained by means a visual agent in the virtual reality environment. The rays stemming from the agent's eye interact with the environment and return back as result of backscattering. 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. The autonomous movement of the agent or avatar/virtual-robot is modelled by multiresolutional filtering.

III. MULTIRESOLUTIONAL FILTERING USING WAVELET TRANSFORM

In this section the multiresolutional filtering (MF), as proposed by Hong [5] will be briefly explained and presented. However, since the Kalman filter underlies the MF algorithm firstly, a brief description of the Kalman filter is presented.

A. The Kalman Filter

Consider a linear stochastic system to describe the propagation in time of a state vector X_t:

$$\begin{aligned} X_{t_k} &= \mathbf{\Phi}(t_k, t_{k-1}) X_{t_{k-1}} + B(t_k) u_{t_k} + G(t_k) W_{t_k}, k = 1,2,.. \\ X_{t_o} &= X_o \end{aligned} \quad (25)$$

where X_{t_k} is an n-vector state process, $\mathbf{\Phi}(t_k, t_{k-1})$ is n×n system dynamics matrix, $B(t_k)$ is an n×r input matrix , u_{t_k} ia an r-vector deterministic input, $G(t_k)$ ia an $n \times p$ noise input matrix and w_{t_k} is a p-vector white Gaussian noise process.

$$\left\{ W_{t_k} W_{t_k}^T \right\} = Q(k) \quad (26)$$

Measurements are available at times points $t_1, t_2,$ ..and are modelled by

$$Z_{t_k} = \mathbf{C}(t_k)\, X(t_k) + V_{t_k} \quad (27)$$

where Z_{t_k} is the c-vector measurement process, $\mathbf{C}(t_k)$ is c×n measurement matrix and V_{t_k} is an c-vector white Gaussian noise process with statistics

$$\left\{ V_{t_k} V_{t_k}^T \right\} = R(k) \quad (28)$$

The optimal state estimate is propagated from measurement time t_{k-1} to measurement time t_k by the equations

$$\begin{aligned} X(k \mid k-1) &= \mathbf{\Phi}(t_k, t_{k-1}) X(k-1 \mid k-1) + B(t_k) u_{t_k} \\ P(k \mid k-1) &= \mathbf{\Phi}(t_k, t_{k-1}) P(k-1 \mid k-1) \mathbf{\Phi}(t_k, t_{k-1})^T + \\ &\qquad G(t_k) Q(k)\, G(t_k)^T \end{aligned} \quad (29)$$

where P is the covariance matrix. At measurement time t_k, the measurement Z_{t_k} becomes available. The estimate is updated by the equation

$$\begin{aligned} X(k \mid k) &= X(k \mid k-1) + K(k)\left[Z_{t_k} - C(t_k) X(k \mid k-1) \right] \\ P(k \mid k) &= P(k \mid k-1) - K(k)\, C(t_k)\, P(k \mid k-1) \\ K(k) &= P(k \mid k-1)\, C(t_k)^T \left[C(t_k)\, P(k \mid k-1)\, M(t_k)^T + R(k) \right]^{-1} \end{aligned} \quad (30)$$

K is the filter gain. Since the avatar's movement trajectory is non-linear, a linear model does not provide a valid description. Therefore, we consider a non-linear stochastic system

$$\begin{aligned} X_{t_k} &= \mathbf{\Phi}(X_{t_k}, t_k, t_{k-1}) + B(t_k) u_{t_k} + G(t_k) W_{t_k}, k = 1,2,.. \\ X_{t_o} &= X_o \end{aligned} \quad (31)$$

where $\mathbf{\Phi}(X_{t_k}, t_k, t_{k-1})$ is an n-vector describing the system dynamics. The measurements are modelled by the non-linear equation

$$Z_{t_k} = \mathbf{T}(X_{t_k}, t_k) + V_{t_k} \quad (32)$$

where $\mathbf{T}(X_{t_k}, t_k)$ is a vector describing the relation between the state and the measurements. For a reference trajectory x_{t_k} the state equation (25) can be linearized by Taylor expansion around the point X_{t_k}, t_k, t_{k-1}, so that it yields

$$\begin{aligned} X_{t_k} &= \mathbf{\Psi}(X_{t_{k-1}}, t_k, t_{k-1}) X_{t_{k-1}} - \mathbf{\Psi}(x_{t_{k-1}}, t_k, t_{k-1}) x_{t_{k-1}} \\ &\quad + \mathbf{\Phi}(x_{t_{k-1}}, t_k, t_{k-1}) + B(t_k) u_{t_k} + G(t_k) W_{t_k}, \end{aligned} \quad (33)$$

$$Z_{t_k} = \mathbf{M}(x_{t_k}, t_k)\, X_{t_k} - \mathbf{M}(x_{t_k}, t_k) x_{t_k} + \mathrm{T}(x_{t_k}, t_{k)+V_{t_k}} \quad (34)$$

where

$$\left[\mathbf{\Psi}(x_{t_{k-1}, t_{k-1}, t_k}) \right]_{ij} = \frac{\partial \left(\mathbf{\Phi}(x_{t_{k-1}}, t_{k-1}, t_k) \right)_i}{\partial (x_{t_k})_j} \quad (35)$$

$$\left[\mathbf{M}(x_{t_k}, t_k) \right]_{ij} = \frac{\left(\partial (\mathbf{M}(x_{t_k}, t_k) \right)_i}{\partial (x_{t_k})_j} \quad (36)$$

and the approprimate linear observation equation

$$Z_{t_k} = \mathbf{M}(x_{t_k}, t_k) X_{t_k} - \mathbf{M}(x_{t_k}, t_k) x_{t_k} + \mathbf{T}(x_{t_k}, t_k) + V_{t_k} \quad (37)$$

Given the linearized model described, the standard Kalman filter is used to obtain the estimate of the state X_{t_k} and it covariance matrix. For the reference trajectory, the obvious choice is

$$x_{t_k} = \mathbf{\Phi}(x_{t_{k-1}}, t_{k-1}, t_k) + B(t_k) u_{t_k} \quad (38)$$

so that the reference trajectory is completely determined by the prior estimate of the state. This estimator is called the linearized Kalman filter. More information about the Kalman filter, can be found in [5-10].

B. The Wavelet Transform

The functions given by discrete data of the form $f(x_i)=d_i$, $i=1,2,...m$, can be represented in multiresolutional form in a dyadic structure as a counterpart of the continuous wavelet transformed time-frequency representation. The multiresolution theory can be conveniently described by the theory of function spaces. A function space is made of embedded subspaces V_m the limit of their union is $\boldsymbol{L^2(R)}$ where for each function $f(x) \in V_m$ we can write that

$$\begin{aligned} & f(x) \in V_m \Leftrightarrow f(2x) \in V_{m-1} \\ & .... \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset ..... \end{aligned}$$

where the functions

$$\phi_{m,n}(x) = 2^{-m/2}\phi(2^{-m}x - n) \tag{39}$$

form an orthonormal basis for V_m. These are called scaling functions and for m=0, we basically write

$$\phi_{0,n}(x) = \phi(x - n) \tag{40}$$

The function *f(x)* in each subspace can be expressed by these orthogonal base functions as approximation in such a way that $f_m(x) \in V_m$ and

$$f(x) = \lim_{m\to\infty} f_m(x) \tag{41}$$

All functions in V_m can be represented by using linear combinations of the scaling functions. In other words, $f_m(x)$ is an orthogonal projection of *f(x)* onto V_m,

$$f_m(x) = \sum_n < \phi_{m,n}(x), f(x) > \phi_{m,n}(x)$$
$$= \sum_n c_{m,n}\,\phi_{m,n}(x) \tag{42}$$

where $< \phi_{m,n}(x), f(x) >$ is the inner product:

$$< \phi_{m,n}(x), f(x) > = \int_{-\infty}^{\infty} \phi_{m,n}(x)\, f(x)\, dx \tag{43}$$

The difference spaces can be represented by W_m and are defined as the orthogonal complement of the spaces V_m with respect to V_{m-1},

$$V_{m-1} = V_m \oplus W_m$$

where V_m and W_m are orthogonal to each other. Now, let $\psi(x) = \psi_{0,0}(x)$ be a basis function of W_o. Note that $\psi_{0,0}(x) \in W_0 \subset V_{-1}$ and therefore can expressed in terms of basis functions $\phi_{-1,n}(x)$ and therefore, we can also define functions $\psi_{m,n}(x)$ that are shifted and dilated versions of one prototype function $\psi(x)$ of the form

$$\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x - n) \tag{44}$$

The functions $\psi_{m,n}(x)$ are identical to the wavelets described before, after the discretization. There are strong relations between $\phi(x)$ and $\psi(x)$. The introduction of the wavelet functions enables us to write any function *f(x)* in $\boldsymbol{L^2(R)}$ as a sum of projections on W_j, $j \in R$, of the form

$$f(x) = \sum_{j=-\infty}^{\infty} w_j(x) \tag{45}$$

where

$$w_j(x) = \sum_k < \psi_{j,k}(x), f(x) > \Psi_{j,k}(x) \tag{46}$$

Considering a certain scale *m*, the function *f(x)* can be written as the sum of a low resolution part $f_m(x) \in V_m$ and the detail part which is constituted by the wavelets $w_j(x) \in W_j$ so that

$$f(x) = f_m(x) + \sum_{j=-\infty}^{m} w_j(x) \qquad = \sum_n < \phi_{m,n}(x), f(x) > \phi_{m,n}(x) +$$

$$\sum_{j=-\infty}^{k}\sum_{k}^{m} < \psi_{j,k}(x), f(x) > \psi_{j,k}(x) \tag{47}$$

which can be expressed as

$$f(x) = \sum_n c_{m,n}\phi_{m,n}(x) + \sum_{j=-\infty}^{m}\sum_k d_{j,k}\psi_{j,k}(x) \tag{48}$$

Above, the coefficients $d_{j,k}$ are known as the *wavelet coefficients*. From the preceding equation multiresolution decomposition is represented by an approximation i.e., the first term with $\phi_{m,n}(x)$ functions, and the detail part i.e., the second term with the $\psi_{j,k}(x)$ functions. The variable *m* indicates the scale and is called *scale factor* or *scale level*. If the scale level *m* is high, it indicates that the function in V_m is a coarse approximation of *f(x)*, so the details are neglected. On the contrary, if the scale level is low, a detailed approximation of *f(x)* is achieved. More information about the wavelet transform, can be found in [11-13].

C. Decomposition and Reconstruction Using Haar Wavelets

The time series signal first decomposed to lower resolution levels using Haar wavelet. Haar wavelet is a two-tap high pass filter given by

$$g_{haar} = [g_1 \;\; g_2] = \left[\frac{1}{2}\sqrt{2} \quad -\frac{1}{2}\sqrt{2}\right] \tag{49}$$

The two-tap Haar low pass filter coefficients are

$$h_{haar} = [h_1 \;\; h_2] = \left[\frac{1}{2}\sqrt{2} \quad \frac{1}{2}\sqrt{2}\right] \tag{50}$$

The signals at the lower levels constitute the respective measurements and at each level extended Kalman filter is applied. Note that, these are *calculated* measurements and they contain less information than the original measurements. However, they can better capture certain information at lower resolutions as result of low-pass filtering during decomposition. The decomposition of state variables at higher resolution level to lower resolution levels and fusion of the estimations are shown in figure 7. The simulated measurements are within a sliding window which contains measurements at different resolution levels. The window width accommodates two samples at the highest resolution level.

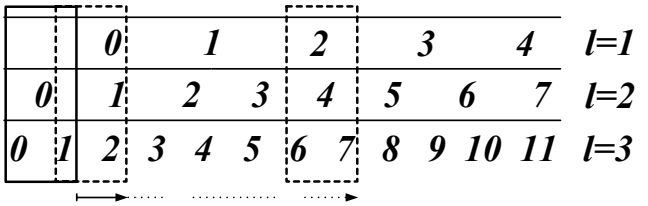

Fig. 7. Measurements at different resolution levels i for i=1,2,3.

The basic scheme for dynamic multiresolutional filtering is shown in figure 8.

Once, within the moving window, the sequences of updated state variables and error covariances

$X_{m+1|m+1}^{[N,i]}$ and $P_{m+1|m+1}^{[N,i]}$ for $I=1,2,..,N,$ are determined, they must be fused to generate an optimal $X_{m+1|m+1}^{[NF]}$ and $X_{m+1|m+1}^{[NF]}$. $X_{m+1|m+1}^{[NF]}$ and $X_{m+1|m+1}^{[NF]}$. For the minimum fusion error covariance $X_{m+1|m+1}^{[NF]}$, the fused estimate $X_{m+1|m+1}^{[NF]}$ is calculated as

$$X_{m+1|m+1}^{[NF]} = P_{m+1|m+1}^{[NF]}\left[\sum_{i=1}^{N}\left(P_{m+1|m+1}^{[N,i]}\right)X_{m+1|m}^{[N,i]} - (N-1)\left(P_{m+1|m}^{[N]}\right)^{-1}X_{m+1|m}^{[N]}\right] \quad (51)$$

where the minimum fusion error covariance $P_{m+1|m+1}^{[NF]}$ becomes

$$\left(P_{m+1|m+1}^{[NF]}\right)^{-1} = \sum_{i=1}^{N}\left(P_{m+1|m+1}^{[N,i]}\right)^{-1} - (N-1)\left(P_{m+1|m}^{[N]}\right)^{-1}. \quad (52)$$

The fused estimate $X_{m+1|m+1}^{[NF]}$ is a weighted summation of both predicted $X_{m+1|m}^{[N]}$ and updated $X_{m+1|m+1}^{[N,i]}$, for $I=1,2,..,N$. The sum of the weight factors equal to the identity I. This can be seen by substitution of $P_{m+1|m+1}^{[NF]}$ given above into the expression of $X_{m+1|m+1}^{[NF]}$ in (52).

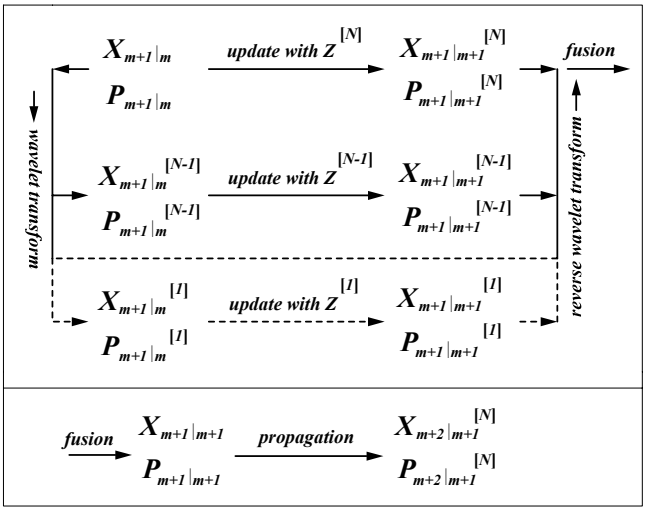

Fig. 8 Multiresolutinal decomposition during filtering

IV. Experiments for Fusion of Perceptions

Presently, the experiments have been done with the 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 times. Three bending modes are seen in figure 9 with the complete trajectory of the perceptual agent. The state variables vector is given by

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

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.

In details, there are three lines plotted in figure 9. The green line represents the measurement data set. The black line is the extended Kalman filtering estimation at the highest resolution of the perception measurement data. The outcome of the multiresolutional fusion process is given with the blue line. The true trajectory is indicated in red. In this figure they cannot be explicitly distinguished. For explicit illustration of the experimental outcomes the same figure with a different zooming range and the zooming power are given in figures 10 and 11 for bending mode and 12 for a straight-ahead case. 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.

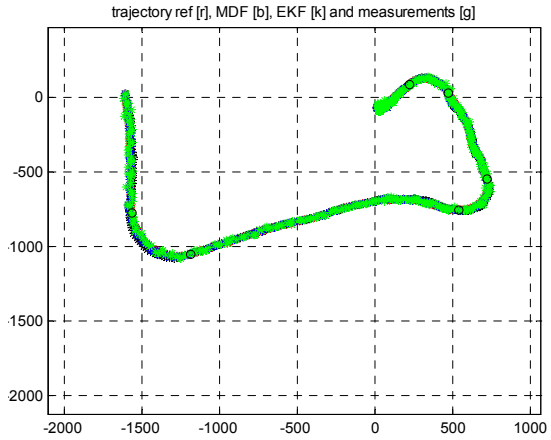

Fig. 9. Robot trajectory, measurement, Kalman filtering and multiresolutional filtering estimation.

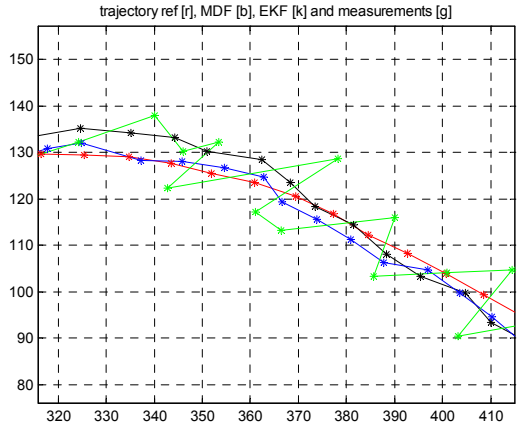

Fig. 10. Enlarged Robot trajectory, measurement Kalman filtering and multiresolutional filtering estimation, in bending mode (light grey is for measurement, the smooth line is the trajectory).

It is noteworthy to mention that, the multiresolutional approach presented here uses *calculated* measurements in the lower resolutions. In general case, each sub-resolution can have separate perception measurement from its own

dedicated perceptual vision system for more accurate executions.

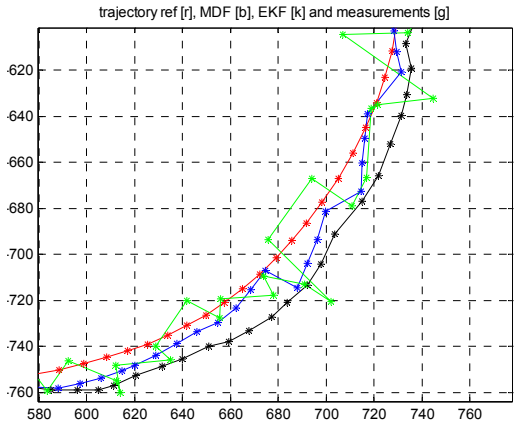

Fig. 11. Enlarged Robot trajectory, measurement Kalman filtering and multiresolutional filtering estimation, in bending mode (light grey is for measurement, the smooth line is the trajectory).

The multiresolutional fusion can still be improved by the use of different data acquisition provisions which play the role of different sensors at each resolution level and to obtain independent information subject to fusion.

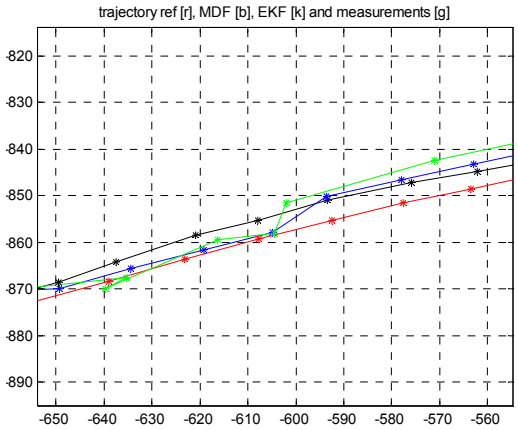

Fig. 12. Enlarged Robot trajectory, measurement Kalman filtering and multiresolutional filtering estimation in straight-ahead mode (light grey is for measurement, the smooth line is the trajectory).

V. Discussions and Conclusions

Fusion of perceptions is investigated for perception measurements where accurate estimations are aimed for autonomous robotics with perception driven navigation. The research has been accomplished in the virtual reality with a perceptual avatar-robot and using simulated perception measurement data. For this aim, several measurements in different resolutional levels are considered where each outcome at each level is combined with the others and a final outcome is obtained. This is commonly referred to as data/sensor fusion. The resolution levels are obtained by wavelet transform using Haar wavelets with two coefficients. This is generally preferred due to real time requirements in autonomous robotics. That is, after two measurements at the highest resolution level, the decomposition is executed. However, if some delay due to measurements can be afforded, then more coefficients at the wavelet filters can be used. The estimations are performed by means of extended Kalman filtering due to nonlinear trajectory of the perceptual avatar/robot.

The perceptual approach for autonomous movement in robotics is important in several respects. On one side it is very appropriate in a dynamic environment. In this case 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 this matter it is the forward direction of the robotic movement. In this way the robot navigation is better coordinated compared to alternative approaches, that are essentially non probabilistic.

Multiresolutional sensor/data fusion is especially important in autonomous perception-driven intelligent robotics since the estimations are based on soft data, like perceptual data in the present case, where measurements can vary drastically compared to non-perceptual navigation. The research has important implications for non-autonomous robotics in general. Further, the implication of this research extends to autonomous robotics for the accurate estimation from sensory information as well as perceptual robotics from the intelligent robotics viewpoints. As to intelligent robotics, soft computing methodologies are particularly important for dealing with soft aspects of the intended implementation and it is of general concern to compare these outcomes with those obtained otherwise.

References

[1] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, 1965.

[2] Ö. Ciftcioglu, M.S. Bittermann and I.S. Sariyildiz, "Studies on visual perception for perceptual robotics", Proc. *ICINCO 2006 3rd Int. Conference on Informatics in Control, Automation and Robotics*, August 1 - 5, 2006, Setubal, Portugal.

[3] Ö. Ciftcioglu, M.S. Bittermann and I.S. Sariyildiz, "Towards computer-based perception by modeling visual perception: a probabilistic theory," in *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, October 8-11, 2006, Taipei, Taiwan.

[4] T.T.J.M. Peeters and Ö. Ciftcioglu, "Statistics on exponential averaging of periodigrams," in *IEEE Trans. on Signal Processing*,vol.43, no.7, 1995, pp. 1631-1636.

[5] L. Hong, "Multiresolutional filtering using wavelet transform", *IEEE Transactions on Aerospace and Electronic Systems*, vol.29, no.4, pp.1244-1251, 1993.

[6] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York 1970.

[7] P.S. Maybeck, *Stochastic Models, Estimation and Control*, Vol I, Academic Press, New York, 1979.

[8] P.S. Maybeck, *Stochastic Models, Estimation and Control*, Vol II, Academic Press, New York, 1982.

[9] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prectice-Hall, Englewood Cliffs, New Jersey, 1979.\

[10] R.G. Brown, *Introduction to Random Signal Analysis and Kalman Filtering*, John Wiley & Sons, New York 1983.

[11] S. Mallat, *A Wavelet Tour of Signal Processing,* Associated Press, New York, 1999.

[12] S.G. Mallat, "A theory for multiresolution signal decomposition:the wavelet representation", *IEEE trans. Pattern Analysis and Machine Intelligence*, vol.11, isuue 7, pp.674-693, July 1989.

[13] D.B. Percival and A.T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge Univ. Press, 2000.