Proceedings of the IEEE ICRA 2009
Workshop on People Detection and Tracking
Kobe, Japan, May 2009

Visual Receding Horizon Estimation for human presence detection

Damien Brulin
Institut PRISME/MCDS UPRES EA 4229
Ecole Nationale Supérieure d'Ingénieurs de Bourges
88, bd. Lahitolle 18020 BOURGES Cedex, France
Email: damien.brulin@ensi-bourges.fr

Estelle Courtial and Guillaume Allibert
Institut PRISME/MCDS UPRES EA 4229
Université d'Orléans - Polytech'Orléans
8, rue Léonard de Vinci, 45072 Orléans Cedex 2, France
Email: estelle.courtial@univ-orleans.fr
guillaume.allibert@univ-orleans.fr

***Abstract*—This paper deals with human presence detection by using a receding horizon estimator based on computer vision results. The visual position estimation problem is formulated into a nonlinear constrained optimization problem in the image plane. A global model combining the behavior of the human motion and the camera model is used to estimate the evolution of the visual features on a past finite horizon. The main interest of this method is the capability to easily take into account constraints. Experimentations in two different configurations highlight the efficiency of the proposed approach, especially with an image occlusion treated as a visual constraint.**

***Index Terms*—Receding Horizon Estimation, Human detection, Computer Vision**

I. Introduction

The management of energy consumption (electricity, heating...), the improvement of the autonomy of the elderly or also the automation of lighting systems are all issues of our society against the rising of energy prices, the ageing of population and environmental concerns. Besides, both industries and individuals wish to better manage their energy consumption thanks to a maximum control of the house or building equipments, like lighting or heating. This is why Domotic science has been developed to offer solutions to supervise and organize a global system that ensures comfort and security. Among the applications stem from Domotic, human presence detection holds an important place.

Several systems, from the state of the art, already offer solution more or less reliable, depending on their use and the monitored phenomenon. Passive InfraRed Detector (PIR) or hyper-frequency sensor can be cited [1][2]. All these systems are called "presence" detectors but in the majority of cases, they are simply movement detectors. The goal of CAPTHOM project is to really detect the presence of a human in an indoor environment by using a multi-sensors system. The final objective is to develop an application which will be able to detect a situation of emergency. The first stage is then to estimate correctly the position on the ground of the human target.

Among the different kinds of sensors, vision seems to be well-adapted to give information about the human presence in a given environment [3]. As an image is a rich source of information, an algorithm should be used to extract the relevant data. In [4], the authors have developed an algorithm which can detect the presence of a human in a scene and can give information about its approximate position. The position estimation of a moving object by using visual information in real time has been largely investigated in the computer vision literature [5][6][9]. However, because visual measurements are usually affected by significant noise and disturbances, for example due to lens distortion, the estimation of the position and orientation could be a difficult task. To enhance the estimation, the extended Kalman Filter (EKF) is usually chosen because it offers many advantages, e.g., accuracy of estimation, prediction capability, temporal filtering [7][8]. The limits of EKF are the conditions that have to be satisfied in order to obtain good results. Adaptative EKF has been proposed in [10] for visual applications. However, difficulties like occlusion or obstacle avoidance, which can be considered as visual constraints, can not be taken into account with EKF based approaches.

The aim of this article is thus to propose a method based on Receding Horizon Estimation (RHE) to realize the estimation of the human target's position in the image. The position estimation problem is transformed into a nonlinear optimization problem. A global model combining the camera model and the human motion model is used to estimate the visual features over a past finite horizon. The optimization algorithm minimizes the error between the features measured thanks to the computer vision algorithm and the features estimated by the RHE. The estimation horizon moves one step forward at each sampling instant and the procedure is repeated. The main advantage of RHE is the capability to easily handle constraints contrary to EKF. When an occlusion appears, the computer vision algorithm gives a false position of the target. The proposed method can bypass this problem either by considering the occlusion as a visual constraint or by considering the largest admissible movement of the human as a state constraint. The comparison between the position given by computer vision and the position estimated by the visual receding horizon estimator is then used to detect an emergency situation.

The paper is organized as follows. In section 2, the issue of human presence detection is introduced. Difficulties due to presence detection or due to the computer vision algorithms are pointed out. In section 3, the principle of the receding horizon estimation is briefly recalled. Then, the proposed approach, called Visual Receding Horizon Estimation (VRHE), is detailed. Finally, section 4 presents experimental results in two different configurations: a first one illustrating the method without constraints and a second one showing the efficiency of the method in case of occlusion.

II. The issue of human presence detection

The current devices, for example PIR or ultrasound sensors, allow to detect the movement of a person in a room but not really its presence. If the target stops and does not move, it becomes invisible for the system and the latter answers that there is nobody in the room. The goal is to always be able to know if there is someone or not in the given environment. However, human presence detection sets out some difficulties.

A. Problems due to presence detection

The first problem which can appear is the detection of non-human targets. The system should be able to differentiate a detection brought out by the movement of an animal and the detection due to the presence of a human being. It exists two ways to solve this differentiation problem. The first approach is a technological one, simply by making a sensor positioning that does not detect movement of small entities. The second one is rather software. It consists of registering excluded areas of the scene or by defining threshold values.

Another problem of presence detection is the presence of several persons in the same room. Furthermore, a room can have more than one exit. So, one has to be able to manage the possibility that a person can enter in the room by one door and leave by a different door. Finally, because the presence of several people can happen, the system must differentiate two or more human targets and must adapt its behavior.

With computer vision algorithm, we can find solutions to deal with these problems. In [4], the authors have proposed a real time human detection based on visual information. Firstly, in order to reduce the search space of the classifier, they perform a background substraction to detect change. The program draws a box including the detected target. Then the algorithm tries to find if the detected target is a human or not. The classification between human and non-human being is done with machine learning tools. Furthermore, each box has its own identifying number. So we can also bypass the problem of multi-presence in a same environment and track each target independently.

In our application, we need to extract the coordinates of the box in order to estimate the human position in the image. However, even if the use of vision offers solutions faced with human presence detection difficulties, it possesses its own drawbacks.

B. Problems due to the computer vision algorithm

Although visual sensor gives a lot of information, many difficulties appear during its use. We will not do an exhaustive list of these difficulties but we will raise the main problems encountered during the development of our application.

In order to detect a change in images, computer vision programs store in memory a model which serves as background. Each image is then compared with the background model to detect a difference. This background is regularly updated. However, if a lightning change happens in the environment, the program will be disrupted by this sudden noise and it will not return good results.

Limitations concerning the camera placement also exist. To use a camera, we need to calibrate it in order to obtain the transformation matrix which allows to calculate the coordinates in the image reference of a point, knowing its coordinates in the environment reference. Once the computation of the matrix is done, the camera does not have to move because the coefficients of the matrix are linked to the camera position.

Another problem, that can have an impact on our position estimation, is the occlusion. If the person walks behind an obstacle, the person is partially masked and so the computer vision algorithm could encounter some difficulties to decide if this target is a human or not. In all environment, there are several obstacles like table, chair or just a box that can disturb the visual acquisition. A last problem, concerning the difficulty of human recognition, could happen if the person falls. The majority of computer vision algorithms use a database composed of images with humans standing up. So, if a person lays down or falls, it will not be recognized as a human being.

Our approach will try to propose a solution, faced with these problems, by combining a receding horizon estimation approach with visual information in order to estimate the position of a human in the image.

III. Visual Receding Horizon Estimation

A. Receding Horizon Estimation

The estimation of the position and orientation of a moving object has been largely investigated in the literature for the past few years. The estimation of the pose of the target is often required in position-based visual servoing approaches. Kalman filtering, especially the Extended Kalman Filter (EKF), offers a satisfactory rejection of disturbance or noise and an accurate estimation. In [10], the authors proposed an adaptative version of EKF for visual applications. However, the EKF may encounter difficulties for practical implementations, when state constraints have to be handled and when the process is highly nonlinear [11]. In order to overcome these problems, a receding horizon estimation (RHE) can be used. The strategy of the RHE is to formulate the constrained state estimation into an online nonlinear optimization problem. The constraints can easily be added to the optimization problem [12].

We propose to extend the receding horizon estimation to visual estimation.

B. Visual Receding Horizon Estimation (VRHE)

The estimation problem of the human position is formulated into a nonlinear optimization problem in the image plane over a past receding horizon N_e. The difference between the measured features in the image plane denoted y_{imag} and the estimated features denoted y_{mod} defines the cost function J. The estimated features are obtained by using a global model combining the human motion model and the camera model. The cost function is to be minimized with respect to the human position $\underline{p} = (x_h, y_h)$ at time $k - N_e$. The position estimation at the current time k is computed thanks to the human motion model and $\underline{p}_{k-N_e}$.
At each sampling time, the past finite estimation horizon moves one step forward and the procedure is then repeated to ensure the robustness of the approach in regard to disturbances and model mismatches.

The cost function can be written in discrete-time as:

$$J(p) = \sum_{j=k-N_e}^{k} [y_{imag}(j) - y_{mod}(j)]^T Q \; [y_{imag}(j) - y_{mod}(j)] \quad (1)$$

Q is a symmetric definite positive matrix. The mathematical formulation of the Visual RHE is then given by:

$$\min_{\underline{p}_{k-N_e}} \; J(p) \quad (2)$$

subject to the nonlinear global model describing the dynamics :

$$\begin{cases} p(k+1) = f(p(k), \Delta u(k)) \\ y_{mod}(k) = h(p(k)) \end{cases} \quad (3)$$

The Figure 1 shows the scheme of the VRHE.
One of the main advantages of VRHE is the capability to explicitly take into account constraints, contrary to EKF. Numerous constrained optimization routines are available in software libraries. A drawback of the RHE strategy is the computational time required for the resolution of the nonlinear constrained optimization problem. However this computational burden is not a strong limitation for real time application due to the increase of PC power.

C. Global overview of our method

The receding horizon estimation algorithm represents the keystone of our approach as we can see in the Figure 2.

The first step consists in positioning and calibrating the camera so as to compute the model of the camera. Moreover, the transformation matrix is required for the change from image to environment reference. In [13], the authors use a planar homography matrix to compute the position of a flame front on the ground in the world reference from its position in the image.

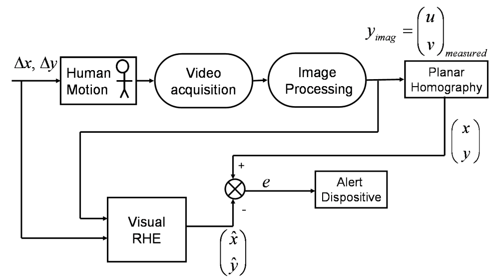


Figure 1. Scheme of the VRHE

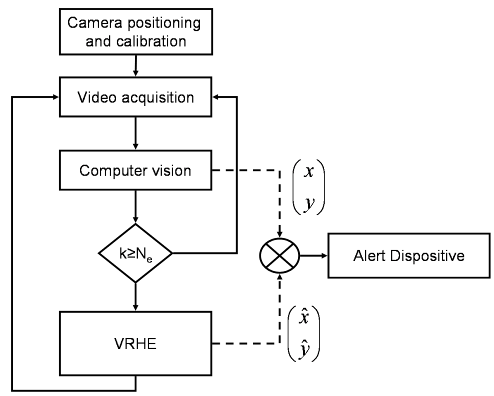


Figure 2. Overview of the VRHE

The computer vision algorithm gives the measure of the coordinates (u and v) of the middle point of the box's bottom side. Thanks to the homography matrix, we calculate the coordinates of feet on the ground (x and y) and also, for each step, the distance with the previous step (Δx and Δy). Once we have enough measures, it means when we have reached the size of the estimation horizon N_e, we can run the VRHE procedure.

IV. Experimental Results

The feasibility and the performance of the proposed visual position estimation algorithm have been experimentally tested using a single camera.

A. The camera and the computer vision algorithm

The camera has been calibrated by using a least square method. The resolution of the camera is 640 x 480 pixels but in the display result of the computer vision algorithm, the image size is reduced to 320 x 240 pixels. The sample time used is the minimum time allowed by the camera frame rate, T_e=0.07s. To conclude with camera's characteristics, it was placed at a height of 1.98m. The scene viewed by the camera is illustrated in Figure 3.

To compute the homography matrix, we used a reference, which can be seen in Figure 3, measuring 0.92 x 0.59 m. This

Figure 3. Camera's view of the scene and image reference

matrix permits to compute the position on the ground from the position in the image. We need a third matrix of transformation because the environment reference used for the homography and the environment reference used for the calibration is not the same. We have computed the transition matrix between these two references. In brief, the three different transition matrices are :

$$M_{intr} = \begin{pmatrix} 328.17 & 0 & 170.88 \\ 0 & -327.89 & 103.51 \\ 0 & 0 & 1 \end{pmatrix}$$

$$M_{hom} = \begin{pmatrix} 0.0065 & 0.0035 & -1.4892 \\ -0.0003 & 0.0175 & -3.7361 \\ 0.0003 & 0.0033 & -0.4043 \end{pmatrix}$$

$$M_{trans} = \begin{pmatrix} 1 & 0 & 0 & -1.02 \\ 0 & 0 & 1 & 1.98 \\ 0 & -1 & 0 & 5.11 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where M_{intr}, M_{hom} and M_{trans} are respectively the intrinsic parameter matrix, the homography matrix and the transition matrix between environment reference of homography and environment reference of calibration. The Figure 4 shows the two different environment references, one used for the homography matrix and the other one used for calibration.

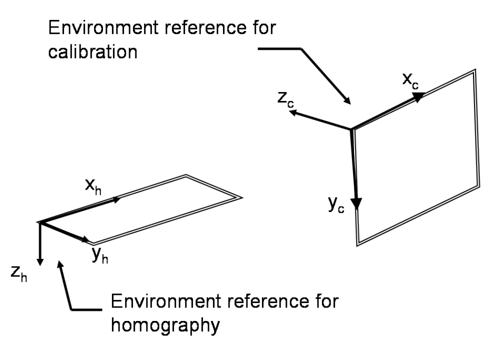
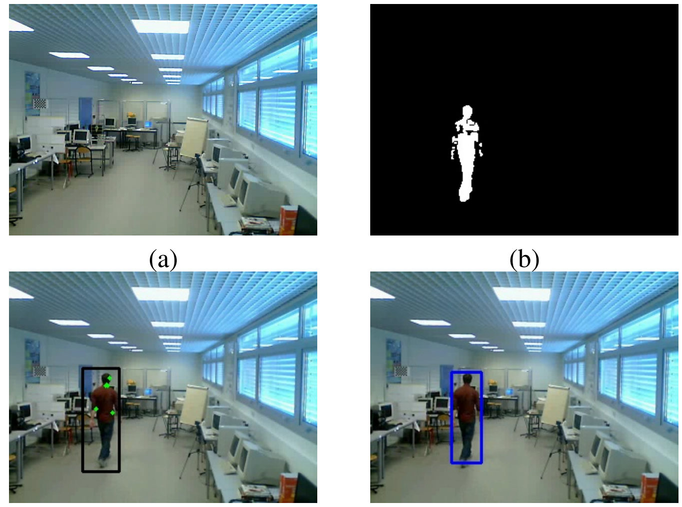


Figure 4. Environment references

Due to the knowledge of the three transition matrices, the



(a) (b) (c) (d)

Figure 5. Representation of the different steps of the computer vision algorithm

model camera can be written as :

$$y_{mod} = \begin{pmatrix} u_{mod} \\ v_{mod} \end{pmatrix} = \begin{pmatrix} \alpha_u \frac{X_c}{Z_c} + u_0 \\ \alpha_v \frac{Y_c}{Z_c} + v_0 \end{pmatrix} \tag{4}$$

$$and \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = M_{trans} \begin{pmatrix} x_h \\ y_h \\ z_h \\ 1 \end{pmatrix} \tag{5}$$

We suppose that the person walks on a flat environment (z_h is constant).

The principle of the computer vision algorithm, used in our application is depicted in the Figure 5. At the beginning, the program needs to take some pictures of a clear scene in order to have a fixed background (a). Then, each picture is compared with the background.

In [4], the authors first detect changes by computing the Mahalanobis' distance between pixels of the current image and the background model(b), composed of the mean of the three RGB components and of the co-occurrence matrix. This first step is done in order to reduce the search space of the classifier. Then, the preceding detected objects are observed by using a tracking of point of interest(c). The final step consists in determining the nature of the tracked object(d). Authors built a cascade of boosted classifiers based on Haar-like filters and on a boosting method to discriminate humans and non humans entities. We can then define the measured features :

$$y_{imag} = \begin{pmatrix} u_{mes} \\ v_{mes} \end{pmatrix} \tag{6}$$

coordinates of the middle of the box's bottom side.

Two different cases are considered. The first aims at illustrating the feasibility of our method with videos where there is no occlusion of the human. The second aims at studying the capability of the proposed method to deal with large path variations due to occlusion. For each case, several videos have been tested to vary movements in the observed scene. No occlusion model is used. The occlusion is treated as a visual constraint in the estimation procedure.

B. The model of the human movement

Human motion model is necessary to estimate the position of a human. However, it is difficult to model and describe precisely the movement of a man. Indeed, one can not predict where the target will be at the next step because it does not follow precise rules. The motion of a human can be described by nonholonomic model [14]. To prove the feasibility of our method, we have just chosen a simple model of the human motion. By observing the human displacements on video, we have observed that the movement can be modeled by a single integrator.

With a first order discretization, the model of human motion can be written as :

$$\begin{cases} x_h(k+1) = x_h(k) + \Delta x \\ y_h(k+1) = y_h(k) + \Delta y \end{cases} \tag{7}$$

where $\Delta x = x_h(k) - x_h(k-1), \Delta y = y_h(k) - y_h(k-1)$ are the displacements respectively in x and y.

The global model is then composed of the camera model (4), the transition matrix (5) and the human motion model (7).

C. Simulations without any constraints

For all experimentations, the size of the estimation horizon is fixed to 5 ($N_e = 5$), the matrix Q is the identity matrix. The VRHE algorithm has been implemented in Matlab software and the computer vision algorithm in Visual C++.

In this case study, scenarios have been realized in a room without any occlusion possibilities. The aim was to verify the feasibility of our method with a simple case. The human just goes to the far end of the room, stays in position during a short time and goes back. The Figures 6 and 7 illustrate respectively the estimation of the position with VRHE according to u and v in the image reference. For both figures, the dashed line represents the estimation, result of VRHE, and the solid line indicates the measures obtained by the computer vision algorithm.

The first five points are at zero because the VRHE begins to run as soon as the program has reached the estimation horizon N_e and has sufficient information. As we can see, the estimates are closed to the measures in both directions.

Same results have been obtained with several videos and proved the feasibility of the proposed approach.

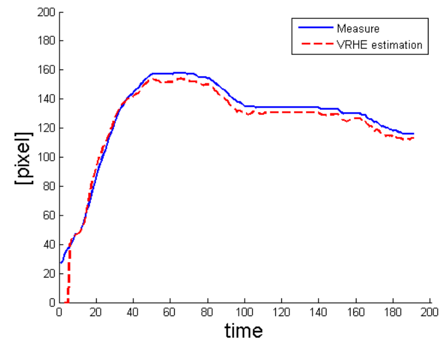


Figure 6. Time history of the measured trajectory and the estimated trajectory according to u

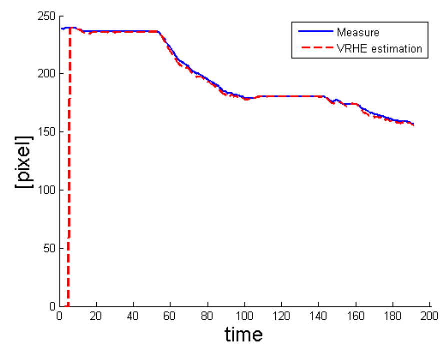


Figure 7. Time history of the measured trajectory and the estimated trajectory according to v

D. Simulations with occlusion

If the person walks behind an obstacle, a problem of occlusion will appear and the size of the box determined by the computer vision algorithm will suddenly change, leading to deviant measures. The Figure 8 illustrates the problem of occlusion and shows how the including box dimension changes when the target is behind an obstacle. In (a), we can see a representation of the scene viewed by the camera. Before the person walks behind the obstacle, the box correctly includes the person (b). When the person is partially masked by the obstacle (c), we clearly see that the box is two times smaller than the previous one. Once the person is no longer hidden by the obstacle, the box recovers its original size (d). We can remark that the human reflect on the window has also been detected by the computer vision algorithm. It is one of the practical difficulties.

The Figures 9 and 10 represent respectively the position estimation according to u and v axis. We clearly see, in the Figure 10, the two moments when the person has been hidden by the obstacle. During these two events, the position estimation according to v axis does not follow the measure. A constraint on the admissible displacement of the human has been taken into consideration. The displacement computed from computer vision is aberrant. So, based on the past movements, a constrained admissible displacement has been

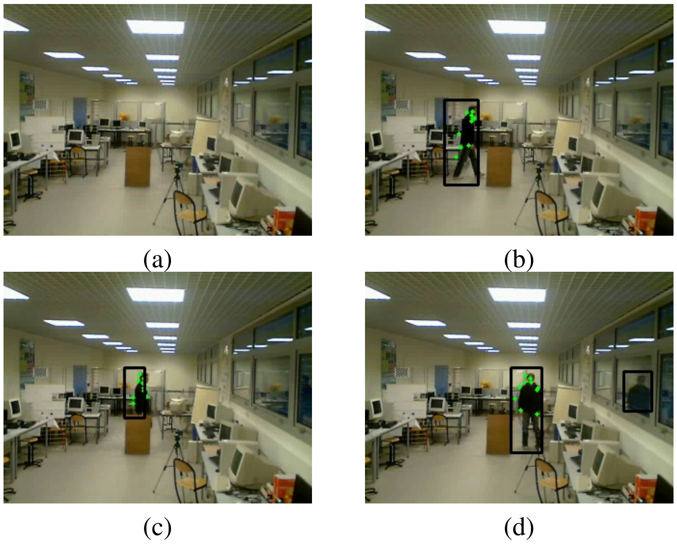


Figure 8. Representation of the problem of occlusion

applied to the global model, especially to the human motion model. Another strategy is to determine, by image processing, obstacle dimensions and to treat it as a visual constraint in RHE. However, this last approach needs more computational time.

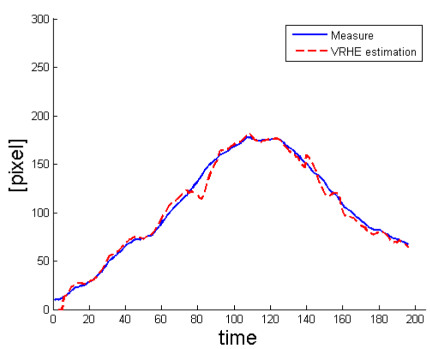


Figure 9. Time history of the measured trajectory and the estimated trajectory according to u

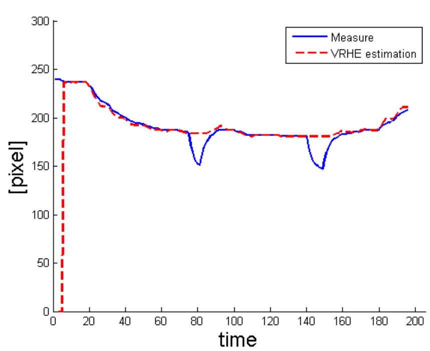


Figure 10. Time history of the measured trajectory and the estimated trajectory according to v

V. Conclusion

In this paper, a method for the visual estimation of the position of a moving human has been proposed. The approach is based on the extension of RHE principle to visual estimation. The main advantage is its capability to take into account constraints. The minimization of the cost function is performed in the image plane. The visual estimates are obtained by the knowledge of a global model combining the human motion model and the camera one. The experimental results confirm the feasibility and the efficiency of the proposed approach. A first approach to avoid problem of deviant measures due to occlusion has been presented. In future works, the residual generation, error between the measures and the estimates, will be used to generate an alert signal.

VI. Acknowledgement

We especially thank all our partners involved in the CAPTHOM project. This work was realized with the financial help of the French Industry Ministry and local collectivities, within the framework of the CAPTHOM project of the Competitiveness Pole S^2E^2, www.s2e2.fr.

References

[1] I. Masri, T. Boudet and A. Guillot, *Hyperfrequency detection method and detector using said method*, patent nb. EP1818684, August 2007.
[2] C.F. Tsai and M.S. Young, *Pyroelectric infrared sensor-based thermometer for monitoring indoor objects*, Review of Scientific Instruments 74 (12), pp. 5267-5273, 2003,doi:10.1063/1.1626005.
[3] N.A. Ogale, *A Survey Of Techniques For Human Detection From Video*, scholarly papers for the Master's of Science degree in Computer Science of the University of Maryland, http://www.cs.umd.edu/Grad/scholarlypapers/papers/neetiPaper.pdf.
[4] Y. Benezeth, B. Emile, H. Laurent and C. Rosenberger, *A Real Time Human Detection System Based on Far Infrared Vision*, ICISP 2008, Elmoataz et al. (Eds.), LNCS 5099, Springer, pp. 76-84, 2008.
[5] T. Broida and R. Chellappa, *Estimation of object motion parameters from noisy images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1, pp. 90-99, 1986.
[6] C. Harris, *Tracking with rigid models*, in A. Blake and A. Yuille (Eds.), Active vision, Cambridge, MA: MIT Press, pp. 57-73, 1992.
[7] S. Lee and Y. Kay, *An accurate estimation of 3-D position and orientation of a moving object for robot stereo vision: Kalman filter approach*, proceedings of 1990 IEEE international conference on robotics and automation, pp. 414-419, 1990.
[8] J. Wang and W.J. Wilson, *3D relative position and orientation estimation using Kalman filter for robot control*, proceedings of 1992 IEEE international conference on robotics and automation, pp. 2638-2645, 1992.
[9] B. Leibe, K. Schindler and L. Van Gool, *Coupled Detection and Trajectory Estimation for Multi-Object Tracking*, Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on Volume , Issue , 14-21, pp. 1-8.
[10] V. Lippiello, B. Siciliano and L. Villani, *Visual motion estimation of 3D objects: an adaptive extended Kalman filter approach*, Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Volume 1, pp. 957-962, 2004.
[11] E.L. Haseltine and J.B. Rawlings, *Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation*, Industrial and Engineering Chemistry Research, 44 (8), pp. 2451-2460, 2005.
[12] C.V. Rao, J.B. Rawlings and D.Q. Mayne, *Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations*, IEEE Trans. Auto. Cont., 48(2):246-258, February 2003.
[13] E. Pastor et al., *Computing the rate of spread of linear flame fronts by thermal image processing*, Fire safety Journal 41, pp. 569-579, 2006.
[14] G. Arechavaleta, J.P. Laumond, H. Hicheur and A. Berthoz, *The nonholonomic nature of human locomotion : a modeling study*. In 1st IEEE/RAS-EMBS Int. Conf. on Biomedical Robotics and Biomechatronics. Pisa, Italy, 2006.