Motion Detection and Tracking by Autonomous Mobile robot in Indoor Environment

Zeghlache SAMIR , Razibaouene Amir, Mohand Saïd DJOUADI

Laboratoire robotique & productique,
Ecole militaire polytechnique, Bp : 17 Bordj El Bahri 16111 Alger, Algérie.
zeghlache_samir@yahoo.fr, amirrazi506@yahoo.fr , msdjouadi@gmail.com

Abstract— **In this current work, we are concerned with the robust detection of moving objects in videos taken from an mobile robot. The main task is to compensate the motion of the observer by estimating the Homography between two successive frames and detect the independent motion of the moving ground target in the image. To determine the 3-D position of the target using the stereovision and applied the classical controller to track the target (human), finally, we give some results of our detecting algorithms in the indoor environment using the fast frame differencing method. This result shows the effectiveness of the vision system.**

***Keywords*— mobile robot, stereovision, vision based control, motion detection, tracking, particular filtering**

I. INTRODUCTION

In the recent years, efforts have been made to give autonomy to a single mobile robot by using different sensors to collect information from the surroundings and react to the changes of its immediate environment. Computer vision is one of the most popular perception sensors employed for autonomous robots. In many task such surveillance and grasping, visual path tracking, visual tracking of a moving object. The application of vision based tracking control design for robotic application has been an active area of robotic research, in the two papers [1], [2] present study of the visual servoing approaches to tracking control of nonholonomic mobile robots, in [3] present study of visual navigation mobile robots, this work subdivided into the visual navigation in indoor environment and the visual navigation in outdoor environment. A method and illustration of the extraction the vision data, in order to allow a robot to operate in at a known and dynamic environment, in [4] the authors have presented a vision-based scheme for driving a non-holonomic mobile robot to intercept a moving target on the lower level, the pan-tilt platform which carries the on-board camera is controlled so as to keep the target at the centre of the image plane. On the higher level, the robot operates under the assumption that the camera system achieves perfect tracking. In particular, the relative position of the ball is retrieved from the pan/tilt angles through simple geometry, and used to compute a control law driving the robot to the target. Various possible choices are discussed for the high-level robot controller. In [5] the robust visual tracking controller is proposed for tracking control of a mobile robot in image plan, this work based on the proposed error-state model, the visual tracking control problem is transformed into the stability problem. The robust control law is then proposed to guarantee that the visual tracking system satisfies the necessary stability condition based on LYAPUNOV theory. In [6] described the position control of autonomous mobile robot using combination of kalman filter and fuzzy logic technique. Both techniques have been used to fuse information from internal and external sensors to navigate the mobile robot in unknown environment. An obstacle avoidance algorithm using stereovision technique has been implemented for obstacle detection. In [7] presented a strategy for a nonholonomic mobile robot to autonomously follow a target based on vision information from an onboard pan camera unit. Homography-based techniques are used to obtain relative position and orientation information from the monocular camera images. The proposed kinematic controller, based on the Lyapunov method, achieves uniform ultimately bounded tracking. In [8] and [9] the authors have proposed the state feedback control using the fuzzy logic controller to realize a leader and follower mobile robot based the laser and infrared sensors. Tracking a moving ground objects in aerial video has a variety of real world applications, and presents a major interest for the civilian and military ones. These include aerial recognition, remote surveillance, traffic monitoring.
Detecting motion of external objects from a moving robot is the subject of active research [10,11,12,13,14,15]
This is a challenging task as target sizes are small and they must be acquired and tracked through changing environment.

We propose an efficiently engineered system which reliably locates and tracks object by using a modular approach. The approach involves four modules which include:

- Features detection and tracking
- Ego motion estimation
- Features selection.
- Robuste Indépendant motion détection
- 2D Target Localization by particle filter
-

The monocular vision based tracking control suffers to obtain the 3D target position. However, in this paper the stereovision system has been used to solve this problem, we can at each instant have a pair of images, from two geometrical defined cameras, that allow us to have four image coordinates. The triangulation equation is used to estimate the relative 3D position (tracker-target). We want to examine the target motion detection and tracking using the nonholonomc mobile robot by measuring the direction and the depth of the target.

First we will describe the kinematical model of a mobile robot (Section II), and the camera model (section III), section IV present feature detection and tracking, section V we present the ego motion estimation, section VI

we present the ego motion estimation, section VI present the feature selection, section VII present the independent motion detect, section VIII present the target tracking using the particular filter, section XI present the 3D target localisation the triangulation equations and the parameters of the cameras, section X presents the mobile robot's visual control development. Finally we arrive to the conclusion of the whole work.

II. MOBILE ROBOT MODEL

In this work is considered the unicycle mobile robot, the navigation is controlled by the speed on either side of the robot. This kind of robot has non-holonomic constraints, which should be considered during path planning. The kinematical scheme of a mobile robot can be depicted as in Fig. 1, where v is the velocity of the robot, v_l is the velocity of the left wheel, v_r is the velocity of the right wheel, r is the radius of each wheel, l is the distance between the left and the right wheels, x and y are the position of the mobile robot, and ϕ is the orientation of the robot.

This type of robot can be described by the following kinematics equations:

$$\begin{cases} \dot{x} = v\cos\theta \\ \dot{y} = v\sin\theta \\ \dot{\theta} = \omega \end{cases} \quad (1)$$

The non-holonomic restriction for model (1) is

$$\dot{y}\cos\theta - \dot{x}\sin\theta = 0 \quad (2)$$

According to the motion principle of rigid body kinematics, the motion of a mobile robot can be described using equations (1) and (2), where ω_l and ω_r are the angular velocities of the left and right wheels respectively, and ω is the angular velocity.

The left and a right velocity of robot:

$$v_r = r.\omega_r \qquad v_l = r.\omega_l. \quad (3)$$

$$\omega = \frac{v_r - v_l}{l} \qquad v = \frac{v_r + v_l}{2} \quad (4)$$

Combining (2) with (3) we can obtain:

$$\omega = \frac{r}{l}(\omega_r - \omega_l) \qquad v = \frac{r}{2}(\omega_r - \omega_l) \quad (5)$$

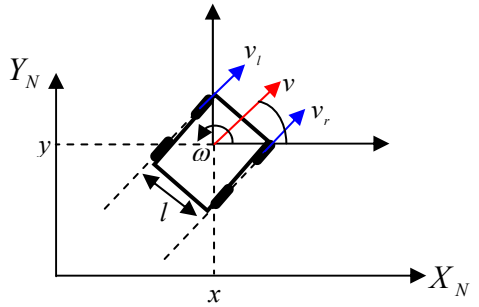


Fig. 1. Geometric description of the mobile robot

III. CAMERA MODEL

Calibration is a heavily worked on area in vision because it is necessary to estimate 3D distance information contained in an image. It allows to model mathematically the relationship between the 3D coordinates of an object in a scene and its 2D coordinates in the image [16].

The parameters of the camera are classified in two categories**, internal parameters** which define the properties of the geometrical optics and the **external parameters** which define position and orientation of the camera. More specifically, the camera calibration consists of determining the intrinsic parameters and the extrinsic parameters [17,18]. The model of the camera is presented in fig.2.

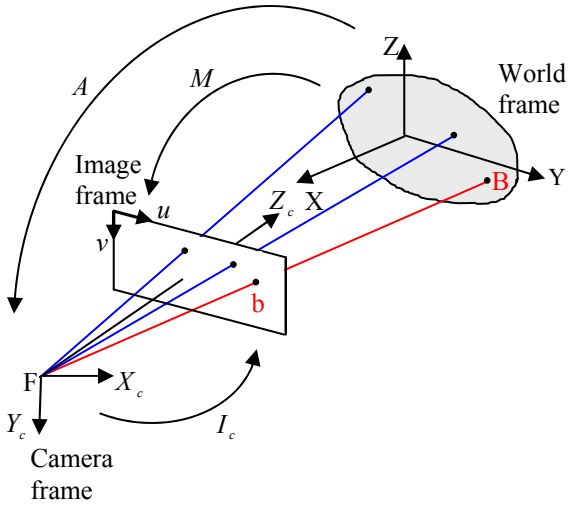


Fig. 2. Camera modele

A. Intrinsic parameters

Intrinsic parameters of the camera define the scale factors and the image centre.

$$I_c = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \begin{cases} \alpha_u = f\,K_u \\ \alpha_v = f\,K_v \\ v_0 \\ u_0 \end{cases} \quad (6)$$

K_u, K_v represent the horizontal and vertical scale factor, f represent the focal length and u_0, v_0 represent the image centre.

B. Extrinsic parameters

Which define the homogenous transformation from the world to the camera frame given by the matrix A.

$$A = \begin{pmatrix} r_{11} & r_{12} & r_{12} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \quad (7)$$

The matrix A is a combination of rotation matrix *R* and translation matrix *T* from to the world frame to the camera frame.
The transformation from the world to the image frame is given by the matrix *M*.

$$M = I_c \; A \tag{8}$$

We can write:

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{9}$$

In this equation *(X,Y,Z)* are the coordinates of a point B in the world frame and *(u,v)* are the image coordinate of the projected point B .

IV. Features Detection and Tracking

The features found and tracked by this algorithm are the Haaris corners [19]. We have also implement the technique of the "Good Features to Track" developed by [20]. These two detectors are based on the correlation matrix computation $\boldsymbol{C_w}$ in the window *w* of the whole image.

$$C_w = \begin{bmatrix} \sum_w \left(\frac{\partial I(x,y)}{\partial x} \right)^2 & \sum_w \frac{\partial I^2(x,y)}{\partial x \partial y} \\ \sum_w \frac{\partial I^2(x,y)}{\partial x \partial y} & \sum_w \left(\frac{\partial I(x,y)}{\partial y} \right)^2 \end{bmatrix} \tag{10}$$

The features f_i^{t-1} found using corner detection algorithm in the image I^{t-1} are used to estimate ego-motion. However, once these features are detected, they are tracked using a pyramidal implementation of the Lukas Kanade optical flow method [21] to find the corresponding features f_i^t location (u_i, v_i) in the image I^t. The goal of feature tracking is to minimize the residual function ε defined as follows:

$$\varepsilon(u_i, v_i) = \sum_{p=-w_x}^{w_x} \sum_{q=-w_y}^{w_y} \left[I^t(x+p, y+q) - I^{t+1}(x+p+u_i, y+q+v_i) \right]^2 \tag{11}$$

This algorithm has two major important benefits:

- Robust to fairly large displacement due to the pyramidal structure.
- Faster than a standard optical flow, because it begins to process the small image than the bigger.

The detection-tracking of the feature are done between two successive images.

V. Ego Motion Estimation

We have studied two different models, affine given by eq (12) and perspective models given by eq (13). For the first one, the compensation of changing scale factor was impossible; on the other hand, by using the second transformation, we were able to compensate this change.

$$s \cdot \begin{bmatrix} x_i^t \\ y_i^t \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_i^{t-1} \\ y_i^{t-1} \\ 1 \end{bmatrix} \tag{12}$$

$$s \cdot \begin{bmatrix} x_i^t \\ y_i^t \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \cdot \begin{bmatrix} x_i^{t-1} \\ y_i^{t-1} \\ 1 \end{bmatrix} \tag{13}$$

One the matching between features ($\boldsymbol{f}^{t-1}$, $\boldsymbol{f}^{t}$) is done, the parameters $\boldsymbol{h_{ij}}$ of the transformation model is estimating by least square or SVD method.
The Levenberg-Marquardt and iterative Gauss-Newton optimisation are used for the non-linear transformation [22]. Nevertheless, that technique could be biased [23] when *(1)* the method can deal neither with outliers (mismatched points) nor with nonrigid scenes (scenes that contain both static and moving objects), and *(2)* the method minimizes an algebraic distance and hence it gives poor results for badly conditioned data.

Fig.3. presents the warping results (blue quadrangle) by applying the perspective transformation. The black rectangle represents the first image before warping, and the blue one represent the warped image.

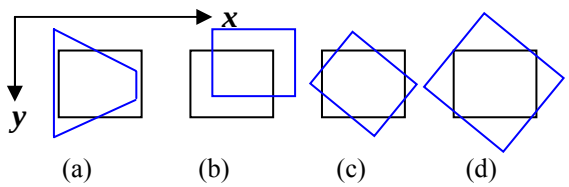


Fig.3. The main image-to-image homography transformation: (a) horizontal linear scale factor changing, (b) horizontal and vertical translation, (c) rotation, (d) constant scale factor changing +rotation.

Therefore, our motion estimation scheme must be robust enough to estimate the correct motion.
For the experiments reported in this paper the model of the Homography is represented by 3x3 matrix defined up to a scale factor.
This Homography $\boldsymbol{H}$ (planar transformation) performs a feature to feature mapping between the homogeneous coordinates of the image $\boldsymbol{x_2}$, $\boldsymbol{x_1}$, such that $\boldsymbol{x_2} = \boldsymbol{H} \cdot \boldsymbol{x_1}$, Fig.4. shows the accuracy and degree of freedom of the ego-motion changing.
This transformation between two images planes has eight degrees of freedom ($\boldsymbol{h_{33}}$=1), hence it can compensate a good number of a camera motion.

For $\boldsymbol{k}$ features points detection, we have a system of 2k linear equations:

$$\begin{bmatrix} \begin{Bmatrix} -x_1 & -y_1 & -1 & 0 & 0 & 0 & x_2x_1 & x_2y_1 & x_2 \\ 0 & 0 & 0 & -x_1 & -y_1 & -1 & y_2x_1 & y_2y_1 & y_2 \end{Bmatrix}_{f_1} \\ \vdots \\ \begin{Bmatrix} -x_1 & -y_1 & -1 & 0 & 0 & 0 & x_2x_1 & x_2y_1 & x_2 \\ 0 & 0 & 0 & -x_1 & -y_1 & -1 & y_2x_1 & y_2y_1 & y_2 \end{Bmatrix}_{f_k} \end{bmatrix} \cdot h = \begin{bmatrix} \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \\ \vdots \\ \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \end{bmatrix} \quad (14)$$

$$A_{(2k)x9} \cdot h_{9x1} = 0_{(2k)x1}$$

$$h = \begin{pmatrix} h_{11} & h_{12} & h_{13} & h_{21} & h_{22} & h_{23} & h_{31} & h_{32} & h_{33} \end{pmatrix}^T$$

With (x_1, y_1) are the first coordinate of the feature in the image ***I(t-1)***, and (x_2, y_2) the coordinate of tracked feature in the image ***I(t),*** k is the total number of features.

At least, the estimation of ***H*** requires four best features to solve (1), in the next section we will describe the manner to select them.

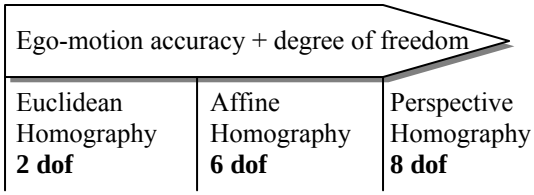


Fig. 4. Accuracy increases with the complexity of the transformation model.

VI. FEATURES SELECTION

The optimal transformation contains inherent error since some of the feature correspondences used to estimate the transformation lie on moving objects with motion independent from the camera.

Those features (outliers) should be eliminated from the feature set before the final Homgraphy is computed, we apply:

1. Compute the initial Homography using the full feature set S
2. Apply the following criteria :

$$\begin{cases} f_i \in S_{in} & \left\| f_i^{\,t} - \left(H_{t-1}^{\,t}\right) f_i^{\,t-1} \right\| \le \varepsilon \\ f_i \in S_{out} & otherwise \end{cases} \quad (15)$$

Re-compute the new Homography using only the set S_i

For the partioned features set, we paint the set S_{in} by green, and S_{out} by red. TABLE I (*a*)

Therefore, the ego-motion of the vehicle can be re-estimated accurately and robustly by using this selection method.

Other robust selection techniques are investigated to filter out matching outliers. A variant of least-median-square (LMedS) [24] and random sample consensus (RANSAC) [12, 13], both method are based on the outlier rejection strategy. More description of these algorithms can be found in [24].

VII. INDEPENDENT MOTION DETECT

To detect motion of the moving target, such as cars, trucks, the system uses the fast frame differencing method. Image ***I(t-1)*** is converted using the Homography before being subtracted by the image ***I(t)***.so for each pixel (x, y):

$$I^{\,t}{}_{comp}\ (x, y) = I^{\,t-1}\left(H_{t-1}^{\,t}(x, y)\right) \quad (16)$$

The difference image between two consecutive frames is performed by:

$$I^{\,t}{}_{diff}\ (x, y) = \left\| I^{\,t}(x, y) - I^{\,t}{}_{comp}\ (x, y) \right\| \quad (17)$$

In reality the result image I_{diff}, is noisy by a salt-and-pepper noise, in order to eliminate it, images are convolving with 3x3 Gaussian mask.

VIII. TARGET TRACKING USING THE PARTICLE FILTER

The main feature of Particle Filter is Bayesian inference, which recursively estimates a posterior density of the object's state:

$$P\left(X^t \middle| I_{diff}^t\right) \propto P\left(I_{diff}^t \middle| X^t\right) P\left(X^t \middle| I_{diff}^{t-1}\right) \quad (18)$$

Where $P\left(I_{diff}^t \middle| X^t\right)$ is the likehood and $P\left(X^t \middle| I_{diff}^{t-1}\right)$ is the prior density derived from previous posterior density $P\left(X^{t-1} \middle| I_{diff}^{t-1}\right)$ and a dynamical model $P\left(X^t \middle| X^{t-1}\right)$

$$P\left(X^t \middle| I_{diff}^{t-1}\right) = \int P\left(X^t \middle| X^{t-1}\right) P\left(X^{t-1} \middle| I_{diff}^{t-1}\right) dX^{t-1} \quad (19)$$

The particle filter generate a set of weighted particles at time t, $\left\{s_i^t, \pi_i^t\right\}, i = 1..N$, where s_k^i represents the ith observation (given by eq. 20) of the object state at time k, π_i^t is the probability (the importance weight) for s_i^t to be the moving object and N is the maximum number of particles.

$$s_i^t = \begin{bmatrix} x_i^t & y_i^t & \dot{x}_i^t & \dot{y}_i^t \end{bmatrix}^T \quad (20)$$

The motion model is defined as:

$$\begin{bmatrix} x_i^t \\ y_i^t \\ \dot{x}_i^t \\ \dot{y}_i^t \end{bmatrix} = \begin{bmatrix} x_i^{t-1} + T_s \cdot \dot{x}_i^{t-1} \\ y_i^{t-1} + T_s \cdot \dot{y}_i^{t-1} \\ \dot{x}_i^{t-1} \\ \dot{y}_i^{t-1} \end{bmatrix} \quad (21)$$

Where T_s is a time interval.

$$d_i^t = \frac{1}{m \times n} \sum_{i=-m/2}^{m/2} \sum_{j=-n/2}^{n/2} I_{diff}\left(x_i^t - i, y_i^t - j\right) \quad (22)$$

The $m \times n$ mask should be big enough so that salt-and-pepper noise is eliminated.

$$\pi_i^t \propto \exp\left(-\frac{d_i^t}{\sigma^2}\right) \quad (23)$$

As shown in eq (22) and (23) only the position information of the motion data is used to evaluate particles. A model of these forms a design variable: σ. the choice of these variables determines the sensitivity of the filter to the measurements. The stapes of the particulate filter are

presented in tab 1, Fig 5 shows the output of the particle filter in indoor environment.

$$\left[\left\{x_k^{(i)}, w_k^{(i)}\right\}\right]_{i=1}^{N} = Condensation\left(\left[\left\{x_k^{(i)}, w_k^{(i)}\right\}\right]_{i=1}^{N}, z_k\right)$$

//- Initialization -//

1. ***if*** *k= 0 (Initialization)* ***than***
2. *to sample* $s_0^{(1)},\ldots,s_0^{(N)}$ *according to* $P(S_0)$, *and to pose* $w_0^{(i)} = \frac{1}{N}, i = 1,\ldots,N$
3. ***end if***

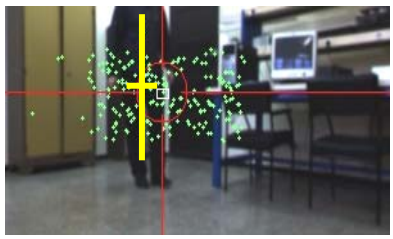
//- Propagation and weighting-//

4. ***if*** $k \geq 1$ ***than***
5. ***for*** $i = 1,\ldots,N$ ***do***
6. *to propagate the particle* $s_{k-1}^{(i)}$ *while calculating :*
$$s_k^{(i)} = P\left(s_k \middle| s_{k-1}^{(i)}\right)$$
7. *to update the weight* $w_k^{(i)}$ *according to the equation :*
$$w_k^{(i)} \propto w_{k-1}^{(i)}\, P\left(Z_k \middle| s_k^{(i)}\right)$$
Before of the step ensuring the normalization
$$\sum_{i=1,\ldots,N} w_k^{(i)} = 1$$
8. ***end for***

//- Ré-échantillonage -//

9. *Ré-échantillonner* $\left\{s_k^{(i)}, w_k^{(i)}\right\}$ *according to* $P\left(\tilde{s}_k^{(i)} = s_k^{(i)}\right) = w_k^{(i)}$, *what leads to the together of balanced particles* $\left\{\tilde{s}_k^{(i)}, \frac{1}{N}\right\}$ *such as* $\sum w_k^{(i)} \delta\left(s_k - s_k^{(i)}\right)$ *and* $\frac{1}{N}\sum_{i=1}^{N} \delta\left(s_k - \tilde{s}_k^{(i)}\right)$ *approximate* $P\left(s_k \middle| z_{1,\ldots,k}\right)$; *affecter* $x_k^{(i)}$ *et* $w_k^{(i)}$ *avec* $\tilde{x}_k^{(i)}$ *et* $\frac{1}{N}$
10. ***Fin Si***

Table 1 particular filter algorithm



(a) Indoor environnement

Fig 5 Particle filter tracking: The positions of particles are represented by small green dots, and the Yellow Cross shows the truth position of the moving target.

IX. 3D TARGET LOCALIZATION

Depth calculates from a pair of stereoscopic images it is necessary to find the matched corresponding points between the left and right images.

A. MATCHED CORRESPONDING

The goal is to put in correspondence two pixels 2D (right-hand side and left) corresponding to the same point 3D. To guarantee good performances, the rectification and correction of the distortions are essential for a taking into account of the horizontal epipolar lines. We present in this part the criteria to be implemented to decide that pair of primitives left/right is correct or not. We will develop three types of constraints. In the first constraint, they will be the epipolar constraints. A second type of constraints will enable us to validate compatibility between two pairings satisfying the first type of constraints. They are the constraints of order, and unicity. And the last type of constraint is the constraint of maximum disparity. The Fig 6 illustrates well the process adopted for pairing. The correspondent of a pixel (X_L, Y_L) in the rectified left image, and a pixel $(X_R, Y_R = Y_L)$ being on the same line (epipolar constraint) in the rectified right image. The value of the component X_R is found by calculating the values of correlation (eq 24) on the same horizontal line epipolar $(Y_L = Y_R)$ delimited by $X_L + D_{MAX}$ and X_L which are the pixel of a point being on a maximum distance from the stereoscopic bench and the pixel in the left image.

$$ZNCC(P_1,P_2) = \frac{\sum_{i=-w_x}^{w_x}\sum_{j=-w_x}^{w_x}\left[I_1(x_1+i,y_1+j)-\bar{I}_1(x_1,y_1)\right]\cdot\left[I_2(x_2+i,y_2+j)-\bar{I}_2(x_2,y_2)\right]}{\sqrt{\sum_{i=-w_x}^{w_x}\sum_{j=-w_x}^{w_x}\left[I_1(x_1+i,y_1+j)-\bar{I}_1(x_1,y_1)\right]\cdot\sum_{i=-w_x}^{w_x}\sum_{j=-w_x}^{w_x}\left[I_2(x_2+i,y_2+j)-\bar{I}_2(x_2,y_2)\right]}} \quad (24)$$

Once the values of correlation are calculated by using (eq 24), we fixes a threshold and we record the values of the pixels X_R whose correlation is higher than the fixed threshold. Finally, the point corresponding X_R is calculated by making the average.

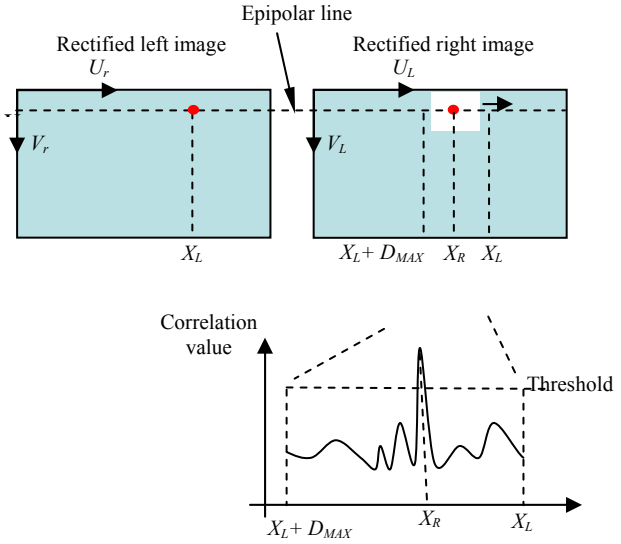


Fig 6 The principle of matched corepending

B. TRIANGULATION

If we places in the case of the three dimensional rebuilding, and if the point ${}^{l}P(x_l, y_l)$ of the left image at summer put in correspondence with the point ${}^{r}P(x_r, y_r)$ of the right image, using the eq (9) we have:

$$\begin{cases} x_l = \dfrac{m^l_{11}X + m^l_{12}\ Y + m^l_{13}\ Z + m^l_{14}}{m^l_{31}X + m^l_{32}\ Y + m^l_{33}\ Z + m^l_{34}} \\ y_l = \dfrac{m^l_{21}X + m^l_{22}\ Y + m^l_{23}\ Z + m^l_{24}}{m^l_{31}X + m^l_{32}\ Y + m^l_{33}\ Z + m^l_{34}} \\ x_r = \dfrac{m^r_{11}X + m^r_{12}\ Y + m^r_{13}\ Z + m^r_{14}}{m^r_{31}X + m^r_{32}\ Y + m^r_{33}\ Z + m^r_{34}} \\ y_r = \dfrac{m^r_{21}X + m^r_{22}\ Y + m^r_{23}\ Z + m^r_{24}}{m^r_{31}X + m^r_{32}\ Y + m^r_{33}\ Z + m^r_{34}} \end{cases} \quad (25)$$

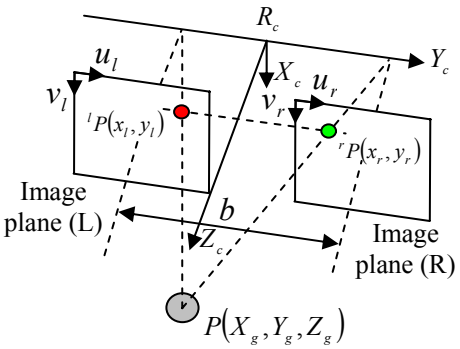


Fig.7. geometry of the stereoscopic camera

The coordinate X,Y and Z of the point P rebuilt in the calibration frame, are calculated by solving a linear system of four equations. The equation (25) can be written by:

$$\begin{pmatrix} m^l_{11} - x_l\, m^l_{31} & m^l_{12} - x_l\, m^l_{32} & m^l_{13} - x_l\, m^l_{33} \\ m^l_{21} - y_l\, m^l_{31} & m^l_{22} - y_l\, m^l_{32} & m^l_{23} - y_l\, m^l_{33} \\ m^r_{11} - x_r\, m^r_{31} & m^r_{12} - x_r\, m^r_{32} & m^r_{13} - x_r\, m^r_{33} \\ m^r_{21} - y_r\, m^r_{31} & m^r_{22} - y_r\, m^r_{32} & m^r_{23} - y_r\, m^r_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} -m^l_{14} - x_l\, m^l_{34} \\ -m^l_{24} - y_l\, m^l_{34} \\ -m^r_{14} - x_r\, m^r_{34} \\ -m^r_{24} - y_r\, m^r_{34} \end{pmatrix} \quad (26)$$

The system of equation (26) can be rewritten in the form:

$$E\ P = W \quad (27)$$

We can solve the equation (27) using the least squares method:

$$P = \left(E^t\ E\right) E^t\ W \quad (28)$$

We can also rebuilt the point P in the left camera frame using the intrinsic parameters matrix of the left and right camera I_{cl}, I_{cr}. The coordinates X_g, Y_g and Z_g of the point P are given by [25]:

$$\begin{cases} Z_g = \dfrac{b}{y_1 - y_2} \\ X_g = x_1\ Z_g \\ Y_g = y_1\ Z_g \end{cases} \quad (29)$$

With:

$$\begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = I^{-1}_{cl} \begin{pmatrix} x_l \\ y_l \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = I^{-1}_{cr} \begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix}$$

By making a translation along the Y by $\frac{b}{2}$, the 3D coordinate of the target centre can be calculated in the frame located between the two camera left and right as is shown in fig.7. Thus the equation (29) becomes:

$$\begin{cases} Z_g = \dfrac{b}{y_1 - y_2} \\ X_g = x_1\ Z_g \\ Y_g = y_1\ Z_g - \dfrac{b}{2} \end{cases} \quad (30)$$

Thus we can calculate the distance which separates the mobile robot and the target, which are given by the following equation:

$$d = \sqrt{{Y_g}^2 + {Z_g}^2} \quad (31)$$

The angle of deviation is given by:

$$\varphi = \tan g^{-1}\left(\frac{y_g}{Z_g}\right) \quad (32)$$

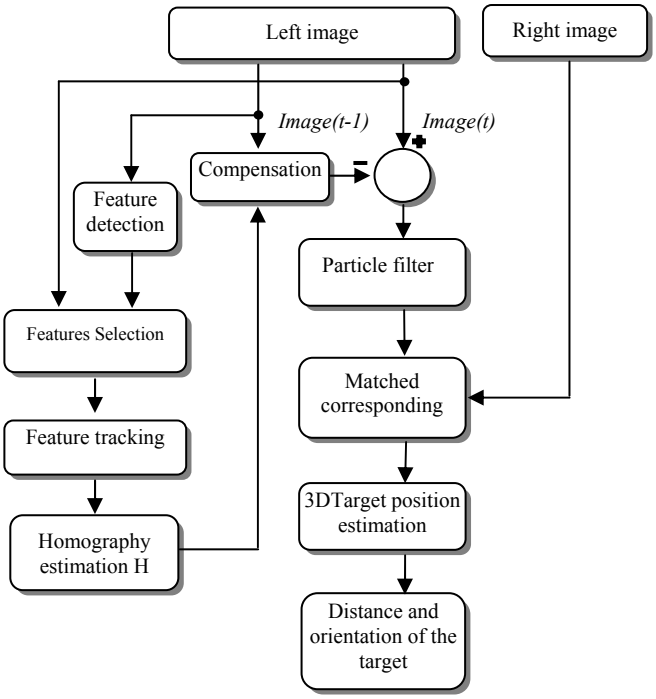


Fig 8. Global Moving human detection and localization algorithm.

X. VISUAL CONTROL OF THE MOBILE ROBOT

In this section we describe the four different robot controllers, that have been integrated in our tracking control, the following figure illustrate the case of our application

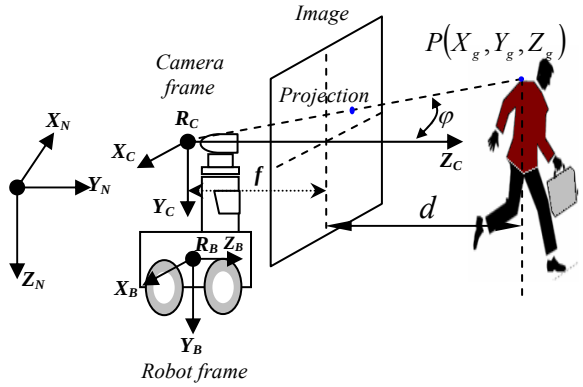


Fig 9. Placement the frames of the mobile robot equipped with a stereoscopic camera

A. Complete system design and hardware

One Pioneer wheeled robot was used in our experiment, we mounted a Bumblebee2 stereoscopic camera, which was operated at a resolution of 320×240 pixels, and it contains embedded computer with a CPU of 1.6 GHZ, the system was implemented in C++ with the openCv library of image processing [26] and the ARIA software development environment [27] running in windows XP operating system. It operated in real-time, with a calculation period of (0.08-0.09s).

Classical proportional controller

The control law proposed is defined as follows:

- The forward velocity v is calculated by:

$$v = k_1 \left(d_d - d\right) \tag{29}$$

d_d Is the desired distance between the leader and the follower mobile robot.

- The angular velocity is proportional to φ angle:

$$\omega = k_2 \ \varphi \tag{30}$$

The resulting closed-loop system is then described by the following equation:

$$\begin{cases} \dot{x} = k_1 \left(d_d - d\right)\cos\phi \\ \dot{y} = k_1 \left(d_d - d\right)\sin\phi \\ \dot{\phi} = k_2 \ \varphi \end{cases} \tag{31}$$

Fig 10 Pioneer 3AT mobile robot used in our Experimentations.

EXPERIMENTAL RESULTS

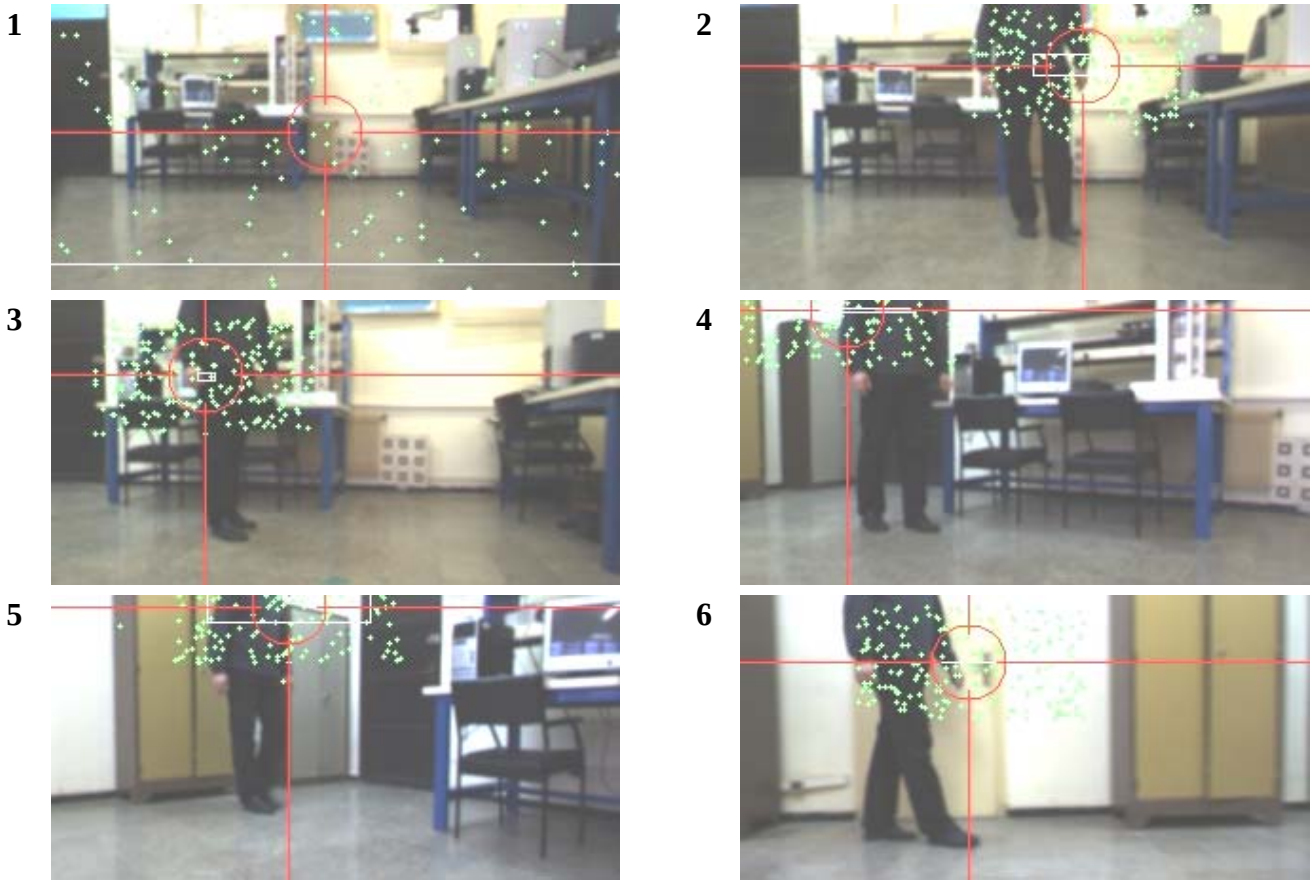


Fig 11 Moving object tracking from mobile robot in indoor environment

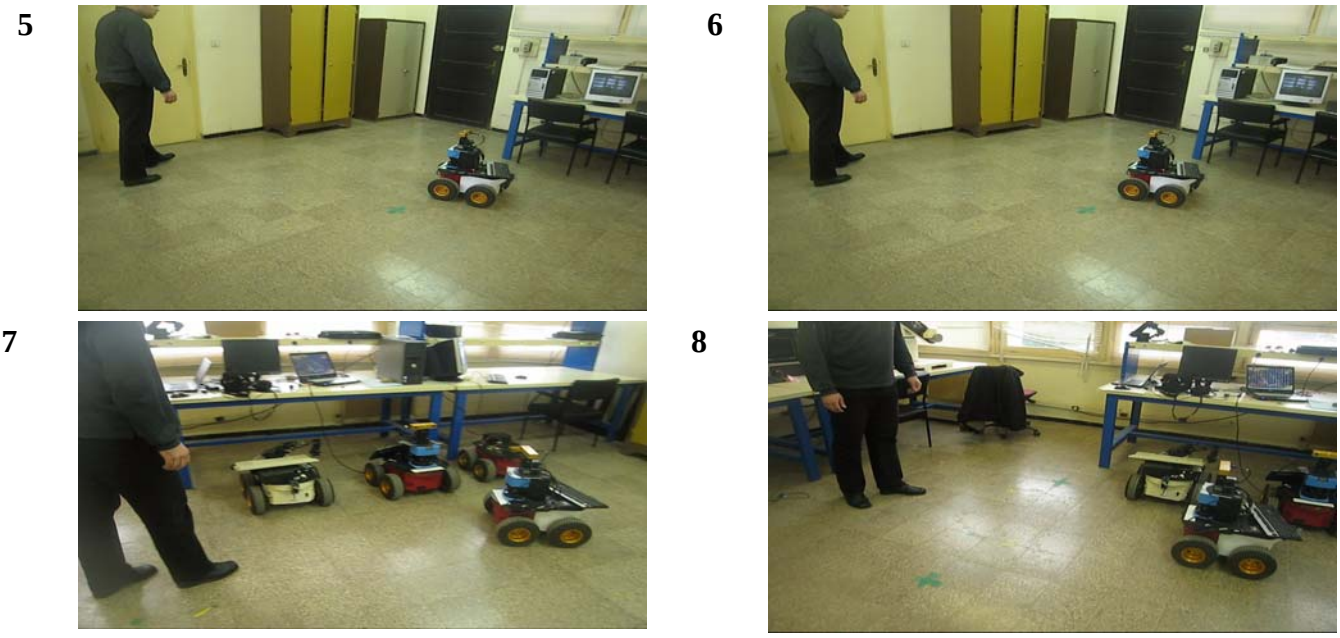


Fig 12 Snapshots of a pioneer3AT robot following a human

XI Discussion and conclusion

To illustrate the efficiency of the proposed tracking controller, we note that the relative distance between the target and the follower robot is maintained constant with the desired distance (dd=2000mm), the control actions such as the linear and the angular velocities are calculated by our classical controller and sent to the follower robot.

The performance of the tracking algorithm was evaluated by comparing with the positions of manually tracked objects. For each sequence of frames, the region of moving objects were marked manually (Yellow cross), the position of each particle is marked with green dots and used as ground truth. Fig 5 shows this evaluation process, the process of tracking the person in the indoor environment is shown in fig (11).

The figure 12, shows The snapshots of the robot following a person successfully during the visual tracking.

Bibliography

[1] D. Burschka, J. Geiman, and G. Hager, "Optimal landmark configuration for vision-based control of mobile robot," Proceedings of the 2003 IEEE International Conference *on Rob. and Auto. ,* Taipei, Taiwan, 2003.

[2] N. J. Cowan and D. E. Koditschek, "Planar image based visual servoing as a navigation problem," Proceedings of the 1999 IEEE International Conference. *Conf. on Rob. and Auto.,* Detroit, Michigan, 1999.

[3] G. N. DeSouza and A. C. Kak, "Vision for Mobile Robot Navigation: A Survey," IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 2, FEBRUARY 2002.

[4] L. Freda, and G.Oriolo, "Vision based interception of a moving target with a nonholonomic mobile robot ," Science direct on Robotics and Autonomous Systems 55 (2007) 419–432.

[5] C. Tsai, K.Song, "Robust visual tracking control of mobile Robot Based an error model in Image Plane," Proceedings of the 2005 IEEE International Conference on Mechatronics & Autoation, July, April 2005.

[6] R. Choomuang & N. Afzulpurkar "Hybrid Kalman Filter/Fuzzy Logic based Position Control of Autonomous Mobile Robot," *International Journal of Advanced Robotic Systems, Volume 2, Number 3 (2005), ISSN 1729-8806*

[7] H. Kannan, V.Chitrakaran, D. M. Dawson and T. Burg "Vision-based Leader/Follower tracking for Nonholonomic mobile robots," Proceedings of the 2007 American control conference, New York City, USA, July 11-13, 2007

[8] T. S. Li, S.J Chang, and Wei Tong "Fuzzy Target Tracking Control of Autonomous Mobile Robots by Using Infrared Sensors," IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 12, NO. 4, AUGUST 2004

[9] M. Sisto D. Gu "A Fuzzy Leader-Follower Approach to Formation Control of Multiple Mobile Robots," Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 9 - 15, 2006, Beijing, China.

[10] Boyoon Jung and Gaurav S. Sukhatme, "Detecting Moving Objects using Single Camera on a Mobile Robot in an Outdoor Envirement" in the 5th Conference on Intelligent Autonomous Systems pp.980-987,Amsterdam,The Netherlands,Marsh 10-13, 2004.

[11] Jiangjian Xiao, Changjiang Yang, Feng Han, and Hui Cheng," Vehicle and Person Tracking in UAV Videos",2007.

[12] Gérard Medioni, Isaac Cohen, Francois Bremond, Somboon Hongeng, and Ramakant Nevatia," Event Detection and Analysis from Video Streams", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 23, NO. 8, AUGUST 2001

[13] Eun-Young Kang, Isaac Cohen and Gérard Medioni ,"Robust Affine Motion Estimation in Joint Image Space using Tensor Voting", International Conference on Pattern Recognition Quebec City, Canada, August 2002

[14]Isaac Cohen and Gérard Medioni , "Detecting and Tracking Moving Objects for Video Surveillance", IEEE Proc. Computer Vision and Pattern Recognition, Jun. 23-25, 1999. Fort Collins CO.

[15]Jinman Kang, Isaac Cohen, G´erard Medioni, Chang Yuan, "Detection and Tracking of Moving Objects from a Moving Platform in Presence of Strong Parallax", Proceedings of the IEEE International Conference on Computer Vision ICCV'05 Beijing, China. October 17-20, 2005.

[16] TRUCCO E., VERRI A. "Introductory techniques for 3-D computer vision," chapter 6, Prentice Hall, 1998.

[17] TOSCANI G. "Système de Calibration et perception du mouvement en vision artificielle," Ph.D. Thesis - Université Paris Sud – 15 dec. 1987.

[18] LENZ R. K. and TSAI R. Y. "Technique for calibration of scale factor and image center for high accuracy 3D machine vision metrologie,". IEEE Transactions on Pattern Analysis and Machine Intelligence – 1988.

[19] C. Harris and M. Stephens, "*A combined corner and edge detector*", in Proceedings Alvey Conference, Manchester, UK, August 1988, pp. 189–192.

[20] Jianbo Shi and Carlo Tomasi, "*Good Features to Track*", IEEE Conference on Computer, Vision and Pattern Recognition, CVPR94 Seattle, June 1994.

[21] B. D. Lucas and T. Kanade, "*An iterative image registration technique with an application to stereo vision*", IJCAI, 1981.

[22] Boyoon Jung and Gaurav S. Sukhatme, "*Detecting Moving Objects using Single Camera on a Mobile Robot in an Outdoor Envirement*" in the 5th Conference on Intelligent Autonomous Systems pp.980-987,Amsterdam,The Netherlands,Marsh 10-13, 2004.

[23] David Demirdjian and Radu Horaud "*Motion–Egomotion Discrimination and Motion Segmentation from Image-Pair Streams*", Computer Vision and Image Understanding **78,** 53–68 (2000), available online at http://www.idealibrary.com.

[24] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim, "*Robust regression methods for computer vision: A review*", Int. J. Comput. Vision 6(1), 1991, 59–70.

[25] R Horaud et Olivier Monga "*Vision Par Ordinateur*," Edition Hermès, deuxieme Edition, France 1995.

[26] User guide of the OpenCv library http://opencvlibrary.sourceforgr.net

[27] Active media

[28] K. M.Passino "Fuzzy control," Departement of Electrical Engineering the ohio sate university 1998