PAL Based Localization Using Pyramidal Lucas-Kanade Feature Tracker

Zoltán Vámossy, Ákos Tóth, Péter Hirschberg

Budapest Tech, John von Neumann Faculty of Informatics

Institute of Applied Informatics

Budapest, III. Nagyszombat u.19.

Phone: +36 1 3689840/178, Fax: +36 1 3689632

Email: vamossy.zoltan@nik.bmf.hu

Abstract:
The objective of the current research is to improve the localization accuracy of a mobile robot during its navigation in known environment. The robot uses GPS receiver to rough positioning. To determine more accurate location panoramic annual image of the omnidirectional imaging tool mounted on the CCExplorer robot is processed continuously. The applied software library algorithm [3] lets us compute optical flow based on the Lucas-Kanade feature tracker in real time. The optical flow computation is implemented in pyramidal fashion, from coarse to fine resolution. Consequently, the algorithm can handle large pixel flows, while keeping a relatively small window of integration, and therefore achieve high local accuracy (of the order of 0.1 pixel).

The paper summarizes the features of the PAL imaging tool, the applied computer vision method adapted to panoramic image and the results of the navigation experiment.

Keywords: *omnidirectional vision, PAL, mobile robot, navigation*

1. Theoretical basis of PAL

The main characteristics of any centric minded imaging (CMI) system [1] is that it is in the center of the coordinate system that describes the three-dimensional scene, and not on its periphery, as in the case of conventional see-through-window (STW) imaging strategy. In STW strategy, one “sees” the 3D world only through discrete chunks as if looking through a window by moving the head up and down or turning around. As a result one cannot get full 3D information in real time. CMI based systems, however, because of their being in the center of the 3D space, totally eliminate this drawback of STW imaging. The most up-to-date CMI block is the Panoramic Annular Lens (PAL) [2].

PAL considers the visual field to be cylindrical, rather than spherical. The image of object points will then be projected first onto an imaginary cylinder wall that is located at a distance equal to the prevailing vision distance, and then this panoramic image projection will be transformed onto a plane perpendicular to the axis of the cylinder.

As a result, the whole 360° visual field will appear on the plane surface as an annular image, and the image points retain the same 1:1 relation of the original object points. The width of the ring will correspond to the vertical viewing angle of the PAL, while points on concentric rings will represent different horizontal spatial angles within a given vertical viewing angle. Thus, the two-dimensional skeleton of the three-dimensional world will be encoded in this annular image.
This leads also to the coding of the appearance of depth on a two-dimensional surface as a convergence into one point, resulting in just one single vanishing point. For this 3D optical coding method, the term Flat Cylinder Perspective (FCP) imaging was introduced. Here the entire 360° panoramic space becomes

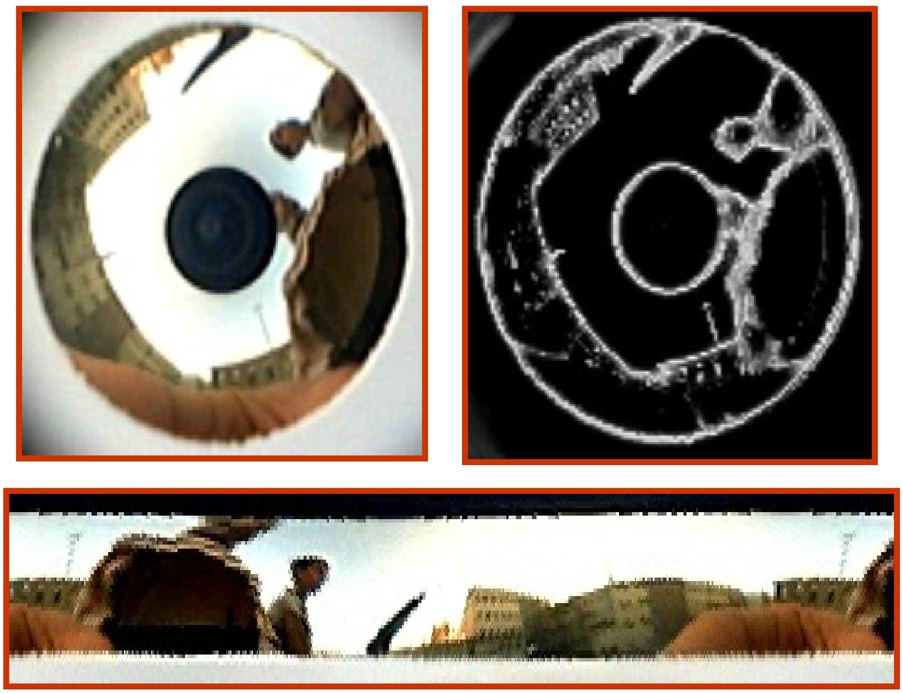

Figure 1. PAL image, segmentation results, straightened image

visible at once, however, both the usual and the inverted perspective will appear simultaneously. This is why at the first glance one has difficulties to orient himself in this type of picture. However, using appropriate software, the ring shaped image can be „straightened out", i.e., the 360° panoramic image displayed in polar coordinates can be converted into Cartesian coordinates, and the mentioned discomfort immediately disappears.

One consequence of this type of CMI is that not a real but a virtual image is formed inside the PAL. Better to say a virtual image volume, which can be regarded as a miniaturized image volume of the 3D space encircling the imaging block, and one can assume that it contains all the imaging point data from the real 3D space.

2.1 Geometrical characteristics of the lens

The omnidirectional imaging block PAL can be considered to be basically of catadioptric type, since it consists of a single piece of glass that has two refractive and two reflective surfaces.

A PAL-optic

- is almost afocal, and both a virtual and a real image are formed inside the optic;
- it renders, via a relay lens, a sharp image from right up against the lens surface out to infinity;
- its center region around the optical axis does not take part in the forming of the panoramic annular image; it serves only to ensure undisturbed passing through of the image forming rays,
- objects to the front of the optic are imaged to the interior of the annular image, and objects to the rear of the optic appear on the outer rim of the annular image;
- a collimated light beam entering the PAL-optic through its plane surface, after passing through the lens, leaves it in form of a light cylinder that evenly illuminates the surrounding. The height of the light cylinder at a given distance from the optical axis depends upon the refractive index of the material the CMI block has been made of.

3 Optical flow based feature tracking

Optical flow is defined as an apparent motion of image brightness. Let $I(x, y, t)$ be the image brightness that changes in time to provide an image sequence. Two main assumptions can be made:

1. Brightness $I(x, y, t)$ smoothly depends on coordinates x, y in greater part of the image.

2. Brightness of every point of a moving or static object does not change in time.

Let some object in the image, or some point of an object, move and after time dt the object displacement is (dx, dy). Using Taylor series for brightness $I(x, y, t)$ gives the following:

$$I(x+dx, y+dy, t+dt) = I(x, y, t) + \frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt + \ldots,$$

where "…" are higher order terms.

Then, according to Assumption 2: $I(x+dx, y+dy, t+dt) = I(x, y, t)$

and $\frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt + \ldots = 0$

Dividing the equation by dt gives an equation $\frac{\partial I}{\partial t} = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt}$

usually called optical flow constraint equation, where dx/dt and dy/dt are components of optical flow field in x and y coordinates respectively.

3.1 Pyramidal Lucas-Kanade feature tracker [3]

In this section the Pyramidal implementation of the Lucas Kanade feature tracker will be briefly summarized based on the description of [3]. Let I and J be two 2D gray scaled images. The two quantities $I(\mathbf{x}) = I(x; y)$ and $J(\mathbf{x}) = J(x; y)$ are then the grayscale value of the two images are the location $\mathbf{x} = [x\ y]^T$, where x and y are the two pixel coordinates of a generic image point $\mathbf{x}$. The image I will be referenced as the first image, and the image J as the second image.

Consider an image point $\mathbf{u} = [u_x\ u_y]^T$ on the first image I. The goal of feature tracking is to find the location $\mathbf{v} = \mathbf{u} + \mathbf{d} = [u_x + d_x\ u_y + d_y]^T$ on the second image J such as $I(\mathbf{u})$ and $J(\mathbf{v})$ are "similar". The vector $\mathbf{d} = [d_x\ d_y]^T$ is the image velocity at $\mathbf{x}$, also known as the optical flow at x. It is essential to define the notion of similarity in a 2D neighborhood sense. Let ω_x and ω_y two integers determine the half size of the so called integration window above $\mathbf{x}$. We define the image velocity $\mathbf{d}$ as being the vector that minimizes the function ε defined as follows:

$$\varepsilon(\mathbf{d}) = \varepsilon(d_x, d_y) = \sum_{x=u_x-\varpi_x}^{u_x+\varpi_x} \sum_{y=u_y-\varpi_y}^{u_y+\varpi_y} (I(x,y) - J(x+d_x, y+d_y))^2$$

To provide solution to that problem, the pyramidal implementation of the classical Lucas–Kanade algorithm is used. An iterative implementation of the Lucas-Kanade optical flow computation provides sufficient local tracking accuracy. Let is now summarize the entire tracking algorithm in a form of a pseudo-code.

Goal: Let **u** be a point on image *I*. Find its corresponding location **v** on image *J*
Build pyramid representations of *I* and *J*: $\{I^L\}_{L=0,\dots,L_m}$ and $\{J^L\}_{L=0,\dots,L_m}$

Initialization of pyramidal guess: $g^{L_m} = \left[g_x^{L_m}, g_y^{L_m}\right]^T = [0,0]^T$

for $L = L_m$ down to 0 with step of -1

Location of point **u** on image I^L: $\mathbf{u}^{\mathrm{L}} = [p_x\, p_y]^T = \mathbf{u}/2^L$
Derivative of I^L with respect to **x**: $I_x(x; y) = (I^L(x + 1; y) - I^L(x - 1; y))/2$
Derivative of I^L with respect to **y**: $I_y(x; y) = (I^L(x; y + 1) - I^L(x; y - 1))/2$
Spatial gradient matrix:

$$G = \sum_{x=p_x-\varpi_x}^{p_x+\varpi_x} \sum_{y=p_y-\varpi_y}^{p_y+\varpi_y} \begin{bmatrix} I_x^2(x,y) & I_x(x,y)I_y(x,y) \\ I_x(x,y)I_y(x,y) & I_y^2(x,y) \end{bmatrix}$$

Initialization of iterative Lucas-Kanade method: $\overline{\nu}^k = [0\ 0]^T$

for k = 1 to K with step of 1 (or until $\|\overline{\boldsymbol{\eta}}^k\| <$ accuracy threshold)

Image difference:
$$\delta I_k(x,y) = I^L(x,y) - J^L(x + g_x^l + \nu_x^{k-1}, y + g_y^l + \nu_y^{k-1})$$

Image mismatch vector:
$$\overline{\mathbf{b}}_k = \sum_{x=p_x-\varpi_x}^{p_x+\varpi_x} \sum_{y=p_y-\varpi_y}^{p_y+\varpi_y} \begin{bmatrix} \delta I_k(x,y) I_x(x,y) \\ \delta I_k(x,y) I_y(x,y) \end{bmatrix}$$

Optical flow (Lucas-Kanade):
$$\overline{\boldsymbol{\eta}}^k = G^{-1}\overline{\mathbf{b}}_k$$

Guess for next iteration:
$$\overline{\mathbf{v}}^k = \overline{\mathbf{v}}^{k-1} + \overline{\boldsymbol{\eta}}^k$$

end of for-loop on k

Final optical flow at level L: $d^L = \overline{\mathbf{v}}^k$

Guess for next level L - 1: $g^{L-1} = \left[g_x^{L-1}, g_y^{L-1}\right]^T = 2\left[g^L + d^L\right]$

end of for-loop on L

Final optical flow vector: $\mathbf{d} = \mathbf{g}^0 + \mathbf{d}^0$
Location of point on *J*: $\mathbf{v} = \mathbf{u} + \mathbf{d}$

Solution: The corresponding point is at location **v** on image *J*

3.2 Feature Selection [3]

So far the tracking procedure is described that takes care of following a point **u** on an image *I* to another location **v** on another image *J*. However it is not described what means to select the point **u** on *I* in the first place. This step is called feature selection. It is very intuitive to approach the problem of feature selection once the mathematical ground for tracking is led out. Indeed, the central step of tracking is the computation of the optical flow vector. At that step, the **G** matrix is required to be invertible, or in other words, the minimum eigenvalue of **G** must be large enough (larger than a threshold). This characterizes pixels that are easy to track. Therefore, the process of selection goes as follows:

1. Compute the **G** matrix and its minimum eigenvalue λ_m at every pixel in the image *I*.
2. Call λ_{max} the maximum value of λ_m over the whole image.
3. Retain the image pixels that have a λ_m value larger than a percentage of λ_{max}. This percentage can be 10% or 5%.
4. From those pixels, retain the local max. pixels (a pixel is kept if its λ_m value is larger than that of any other pixel in its 3 x 3 neighborhood).
5. Keep the subset of those pixels so that the minimum distance between any pair of pixels is larger than a given threshold distance (e.g. 10 or 5 pixels).

After that process, the remaining pixels are typically "good to track". They are the selected feature points that are fed to the tracker.

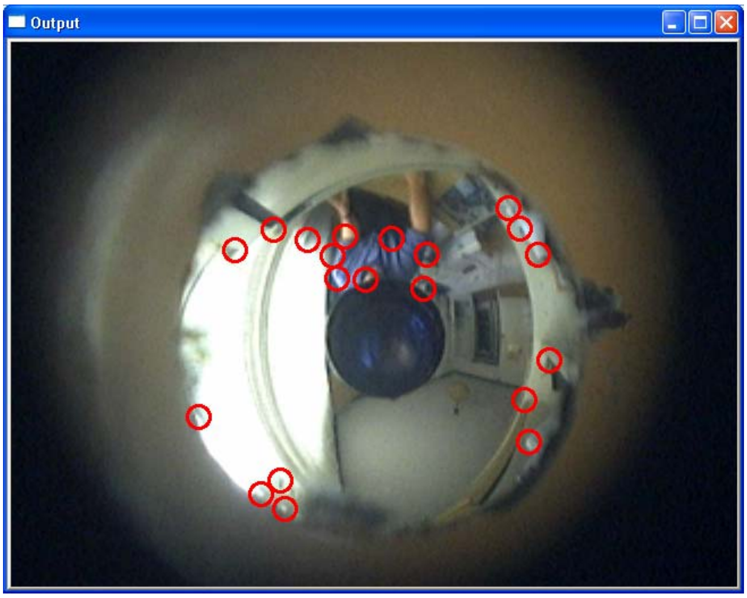

Figure 2. Good features to track – PAL image in lab environment

4 The wheeled mobile robot

The aim of the CCExplorer project [4] is to design and build a radio controlled wheeled mobile robot, based on a model car controlled and navigated by an external computer. The car should carry only the sensors, the accumulators and the sender-receiver units needed for the communication. Navigation in the macro environment is based on the Global Positioning System (GPS) and a camera using PAL for refining the location data, while obstacles in the micro environment are recognized by another monochrome camera.

The camera on the front of the robot monitors the area in front of the robot to detect and avoid obstacles. The program observes the field before the robot using the optical flow method. Two frames following each other are compared. The displacement of each pixel is straight to the distance of the object containing the pixel. Controlling commands can be determined by the displacements recognized on the left and right side of the seen image.

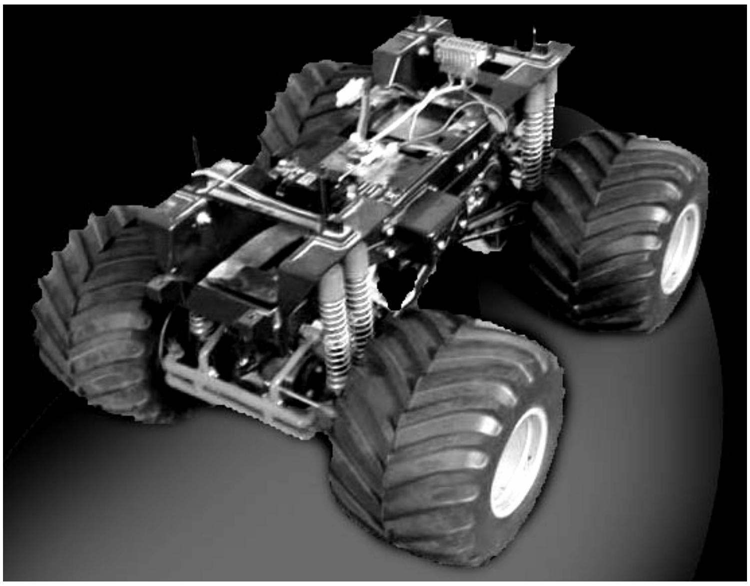

Figure 3. CCExplorer mobile robot

5 Localization during the navigation

The navigation is based on the analysis of the pictures taken about the environment and the data from the global positioning system (GPS) receiver. The image and the GPS data are transmitted via radio connection. The transmitted video signal is connected through a TV tuner to the PC. For getting the geographical position of the robot, the data from the built-in GPS receiver is used. The camera on top of the robot has a vertical-axis, mounted with PAL-optics and produces a 360 degree omnidirectional image of the outer world. With this system, the position of the robot can be calculated more accurately in a known environment on the above described way.

As the first step the operator manually selects the small regions of interest, where the program tries to find good features to track. These regions should be lay where the contrast is high, the intensity changes suddenly: for example in the case of outdoor navigation near the corners of the roof or the gables, in the case of indoor navigation the corner points of the ceiling and the walls, or the intersection of the jambs, or the corners of the cupboard. From the tracked points with known spatial coordinates and from their positions on the PAL image the location of the mobile robot can be calculated solving non linear equation system built up triangulation method.

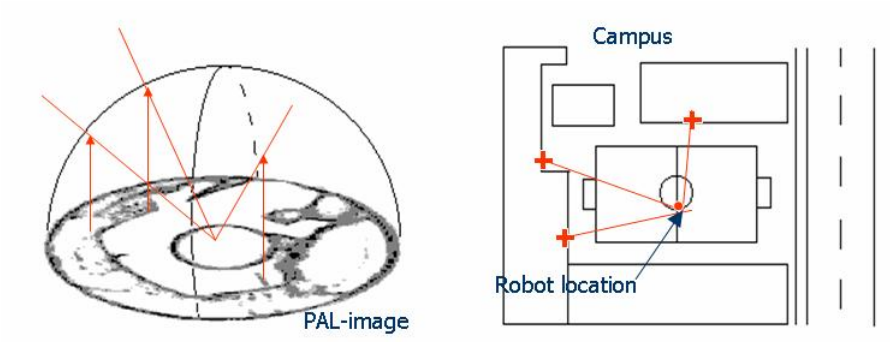

Figure 4. The localization method

Results and conclusions

The developed system was tested in the sport field (40 × 58 m) of the campus located among the buildings of the John von Neumann Faculty. In this case the accuracy of the geographical data received from the GPS was 5-15 m depending on the weather conditions and the DOP (dilution of precision) parameters. However collecting three or more good features points and solving the triangulation based equations the accuracy of the location was reduced 2–3 m depending on the position of the robot.

In indoor environment the GPS is not usable. The PAL based image processing system was tested in lab (4 × 7.8 m) and corridor (3.4 × 13 m) environment. The average failure of the localization in this case was less than 0.6 m.

References

[1] Veress, M., Greguss, P. (1998), Centric Minded Imaging in Space Research, In: Proc. of 7th International Workshop on RAAD'98 (Slovak Academy of Sciences), (K. Dobrovodsky (Ed.)), pp. 121 - 126. Bratislava, Slovakia.

[2] Hungarian Patent 192125, USA Patent 4566763, Japan Patent 1962784, German Patent 3402847.

[3] Bouguet, J. Y.: Pyramidal Implementation of the Lucas Kanade Feature Tracker, Intel Corporation, Microprocessor Research Labs, 2000, http://www.intel.com/research/mrl/research/opencv/

[4] Vámossy, Z., Molnár, A., Hirschberg, P., Tóth, Á., Máthé, B. (2003) Mobile Robot Navigation Projects at BMF NIK, In: International Conference in Memoriam John von Neumann, December 12, 2003., Budapest, pp. 209–219.