IMAGE RECTIFICATION AND SYNTHESIS FOR AN IMAGE BASED AUTOMATIC GUIDANCE SYSTEM IN A 3D MODELLED SITE

VANDEPORTAELE Bertrand, CATTOEN Michel, MARTHON Philippe.

IRIT, LEN7, ENSEEIHT
2 rue Camichel, 31071 TOULOUSE, FRANCE
email: bvdp@enseeiht.fr, cattoen@enseeiht.fr,marthon@enseeiht.fr

ABSTRACT

We are developing a system to allow localization and guidance of people inside a 3D modelled site using imaging techniques.

The system operates in two phases. Firstly, a database of the site has to be built by an expert. Then a user wishing to locate himself in this site takes pictures with a portable device which includes a standard camera and an orientation sensor. These pictures are compared with the contents of the database and if a high correlation is detected, the user can be located inside the site. Finally, the desired path through the site from the actual user's location to the place where he wants to go is found, allowing the guidance.

The database of the site consists in a 3D textured model. To build this database, an omnidirectional 3D sensor has been developed.

A rotation invariant representation of locations inside the site is computed using a virtual omnidirectional sensor. Two steps are necessary to match the user's images with the 3D database. Firstly, the 3D database has to be dynamically projected by this virtual sensor onto 2D omnidirectional images using an optimized method. Secondly, the user's images have to be projected by the same virtual sensor using information from the orientation sensor.

In this paper, we focus on the following two tasks: omnidirectional image synthesis and image rectification.

1. INTRODUCTION

Localization and guidance of persons are efficiently done today thanks to the Global Positioning System. For outdoor applications, this system can be used without problems, if the signal from satellite can be received. For application inside buildings, Local Positioning System has to be used. Instead of using satellites, LPS uses local terminals. This kind of system requires the installation of terminals and suffers from radio frequency interferences.

We report the development of a system allowing localization and guidance of people inside a known site, using imaging techniques. This system does not require any special installation inside the site. The drawback is that it does require more complex sensors and computations. This functionality is planned to be integrated inside advanced cell phones with camera, the computations being carried out by a server receiving data from the phone and guiding the person by voice in real time. Cell phones will have to be improved in order to include an orientation sensor.

Many applications are possible, particularly guidance inside public places or assistance for visually handicapped persons.

Our system uses methods already applied in robotic applications [1], [2], [3], [4], where omnidirectional 2D images from catadioptric sensors [5] are matched.

Our goal is to overcome many drawbacks of existing systems. Firstly, it should be possible to use different sensors to acquire the site and to locate the user. Hence, use of low cost standard cameras for the user's localization is desirable. These sensors are smaller, cheaper and less fragile than catadioptric sensors used for the acquisition. Secondly, the system has to be able to adapt to changes inside the site easily thanks to an intelligent representation of the scene.

In this paper, we describe the system in development and focus on the image rectification and synthesis problems which are necessary to allow matching.

2. SYSTEM PRINCIPLE

A complete diagram of the system is shown in figure 1. The system operates in two phases. Firstly, a database of the site has to be constituted by the expert. Then a user wishing to locate himself in this site uses a portable device to take a picture. This picture is compared with the content of the database and if a high correlation is detected, the system can locate the user within the site. Image matching is done on 2D representations of the scene which are omnidirectional images made of two panoramic images showing upper and lower parts of the scene. These images are all oriented in the same angular reference. Data from database and the user's sensor are formatted to fit this 2D representation.

A 3D textured model of the site is used as a part of the database. We are developing a 3D omnidirectional sensor to construct this model. This model is small to store and easy to update as the expert does not require hundred of pictures to be taken each time the site changes. If a part of the site cannot be modelled in 3D, it is possible to directly acquire panoramic images thanks to a panoramic sensor [5], [6]. The

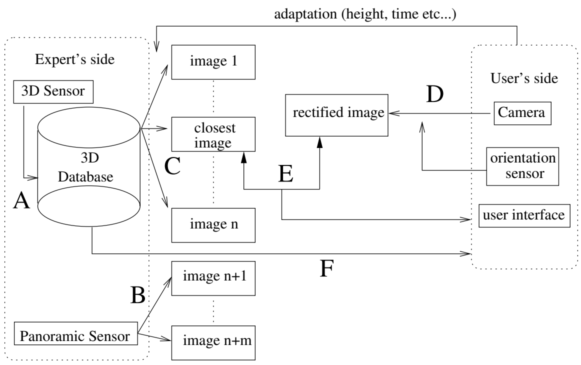

Fig. 1. system principle: A: 3D Acquisition of the site; B: Panoramic images acquisition and rectification; C: Synthesis of omnidirectional images; D: User's image rectification; E: Matching; F:Guidance

sensor used in this work is of the catadioptric type [7] for acquiring panoramic images. It is composed of a paraboloidal mirror, an orthographic lens and a digital camera. This sensor has a single viewpoint located at the focus of the paraboloid.

The database of synthesized omnidirectional images is dynamically generated from the 3D model so it can be computed adaptively according to various parameters.

The user's image is rectified to be matched with these omnidirectional images. A rotation invariant representation of the user's location is computed from the image and data from an orientation sensor mounted onto the camera.

Panoramic images acquired directly by the catadioptric sensor are also rectified to be in the same angular reference and with the same sensor parameters than the ones synthesized from the 3D database, so that all images of the database are of the same type.

Matching techniques are used to find the closest images in the database. Then, localization can be done as the image is considered as representative of one location inside the site.

Guidance is achieved simply by finding paths through the 3D model of the site.

As the prototype 3D omnidirectional sensor is still undergoing further development, we have chosen to work on a synthetic 3D site model. Results presented in this paper are thus obtained using this synthetic 3D model even if rectification works with images from a real camera.

3. OMNIDIRECTIONAL IMAGE SYNTHESIS

2D omnidirectional images have to be computed from the 3D model of the site in order to be compared with user's images. An omnidirectional image can be decomposed into two panoramic images, each one showing either the upper or lower hemisphere. The possibility to generate these images adaptively to the user's situation is potentially very advantageous. For example, the images can be rendered in accordance with the user's height, so images from the user's sensor look like database images. Other parameters can be used, for example time or date to generate database images according to the user's situation.

As the computation of a panoramic image using standard raycasting [8] is relatively time consuming, a quick rendering algorithm is required to achieve this task. We propose a method based on the decomposition of the projection of the scene to the sensor in two steps.

Figure 2 shows the synthesis principle.

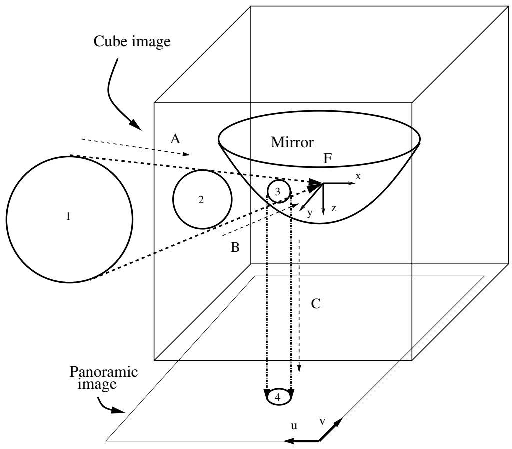

Fig. 2. Panoramic image synthesis principle. A: Projection from the scene to the cube faces. B: Projection from the cube faces to the mirror. C: Projection from the mirror to the panoramic image plane.

Firstly, the scene (Object 1) is projected onto the faces of a cube (Image 2) through its center F.

Secondly, the projection from the cube faces to a virtual catadioptric sensor with the same center of projection F is calculated via raycasting. The virtual sensor has the same parameters as the catadioptric sensor used for direct acquisition. The cube images (Image 2) are projected onto the surface of the paraboloid mirror (Image 3) and then reflected to the panoramic image plane (Image 4).

Due to this decomposition, the non-linear projection from the cube faces to the panoramic image can be stored in a lookup table. Only the linear part remains to be computed for each panoramic image.

The details of computation is next described:

Let α_u and α_v be the horizontal and vertical focal lengths in terms of pixel dimensions, and h the mirror parameter. Let u_0 and v_0 be the optical center projection onto the digital camera image plane.

To compute the lookup table, each pixel of the panoramic image is associated with a pixel in one of the six cube images.

The surface of the lower mirror is defined as:

$$z_1 = \frac{h^2 - \left(x_1^2 + y_1^2\right)}{2.h} \tag{1}$$

The surface of the upper mirror is given by $z_2 = -z_1$.

The orthographic projection can be written as:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} -\alpha_u & 0 & 0 & u_0 \\ 0 & -\alpha_u & 0 & v_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{pmatrix} \quad (2)$$

So the center of each pixel (u,v) of the panoramic image is projected to the surface of the mirror to a point $P_1 = (x_1, y_1, z_1)^T$. For the lower mirror:

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} -\frac{u-u_0}{\alpha_u} \\ -\frac{v-v_0}{\alpha_v} \\ \frac{h^2-x_1^2-y_1^2}{2.h} \end{pmatrix} \quad (3)$$

The first intersection of the ray from $(0, 0, 0)$ to P_1 with one of the six cube faces determines which pixel of which image corresponds to (u,v) in the panoramic image.

To compute the six projections of the scene onto the cube, six perspective virtual cameras are used. These cameras have a 90 degrees field of view and are all placed at the same location F. Their orientation is set to create six different views corresponding to the faces of the cube as shown in figure 3.

Two panoramic images are computed from this set of images via simple raycasting to obtain the omnidirectional representation. Figures 4 and 5 show the computed images for the mirror surface oriented either to the top or to the bottom. We have decided to use the nearest pixel's value for each ray.

As generated panoramic images are computed from a sampling of the scene at the surface of the cube, the ray used to compute the color of the pixel ($R2$) is not exactly in the same direction as the one obtained from reflection ($R1$) as shown in figure 6. Experiments have shown that the maximum angular error between $R1$ and $R2$ is about 0.1˚for a panoramic camera with $\alpha_u = \alpha_v = 5$, $h = 32$ and image size of 512*512 pixels for both panoramic image and images at the surface of the cube.

The six projections of the scene at the surface of the cube are computed efficiently with an image synthesis API like OpenGL. The rendering is done by a 3d accelerated graphic card. Thanks to the chosen method, the rendering process time of panoramic images can be reduced by a factor of about thirty compared to pure raycasting method for a single image. As two panoramic images are computed at the same time, the effective gain is nearly multiplied by two as the six projections onto the cube can be used twice.

4. IMAGE RECTIFICATION

User's images have to be rectified in order to be matched with panoramic synthesis images generated from the 3D database. The virtual sensor used to generate the panoramic images database has a central viewpoint. So it is possible to project the image from another sensor with central viewpoint onto an image of this panoramic sensor. We have developed an orientation sensor in order to know the orientation of the user's

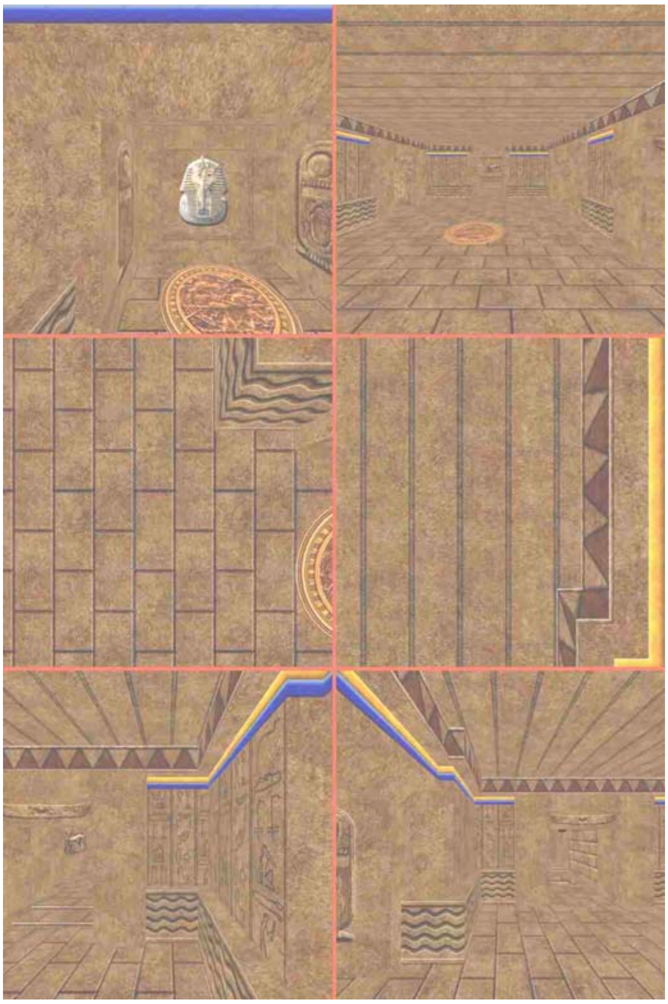

Fig. 3. Projections to the sides of the cube

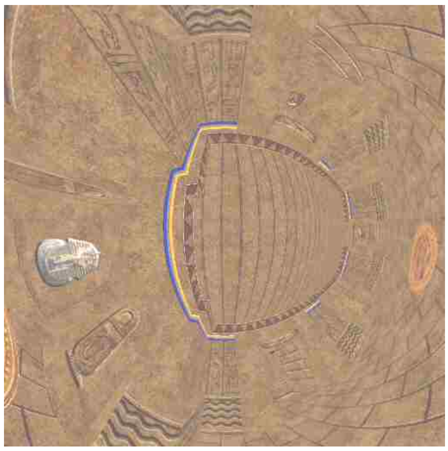

Fig. 4. Upper Panoramic Synthesis Image

camera along three axes. This sensor is composed of an electronic compass and a two axes accelerometer used as incli-

Fig. 5. lower Panoramic Synthesis Image

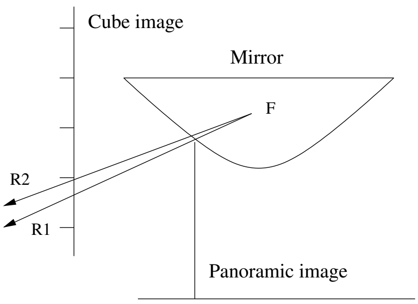

Fig. 6. Image synthesis error

nometer.

$\alpha_u, \alpha_v, h, u_0, v_0$ are parameters of the virtual catadioptric sensor with which rectified images are synthesized as defined in section 3.

The user's standard camera is modelled using the following parameters:

α_{Pu} and α_{Pv}= Focal length in terms of pixel dimensions focal distance in u and v direction respectively.

u_{P0}, v_{P0}= principal point position.

k_{P1}, k_{P2}= radial distortion parameters.

The value of each pixel of the panoramic image are computed as shown in figure 7.

The center of the pixel (u,v) is thus projected onto the mirror surface of the virtual sensor to a point $P_1 = (x_1, y_1, z_1)^T$ using equation (3).

To express P_1 in the camera coordinate system P, P_1 is multiplied by a rotation matrix which represents the rotation from C to P according to angles α, β and θ given by the orientation sensor.

$$P_p = \begin{pmatrix} x_P \\ y_P \\ z_P \end{pmatrix} = R_{(\alpha,\beta,\theta)} . \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \quad (4)$$

This point is then projected onto the surface of the camera

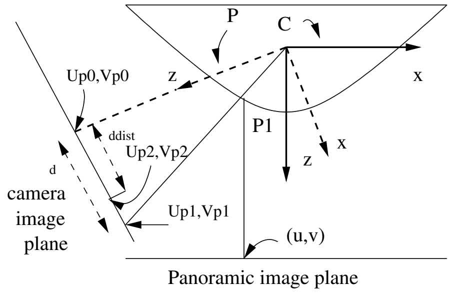

Fig. 7. Standard image rectification principle

image plane.

$$\begin{pmatrix} u_{P1} \\ v_{P1} \end{pmatrix} = \begin{pmatrix} \frac{\alpha_{Pu}.x_P}{z_P} + u_{P0} \\ \frac{\alpha_{Pv}.y_P}{z_P} + v_{P0} \end{pmatrix} \quad (5)$$

To apply the two-parameter radial distortion, the distance from the principal point is computed:

$$d = \sqrt{\left((u_{P1} - u_{P0})^2 + (v_{P1} - v_{P0})^2\right)} \quad (6)$$

A new distance is calculated using k_{P1} and k_{P2}:

$$d_{dist} = d.\left(1 + k_{P1}.d^2 + k_{P2}.d^4\right) \quad (7)$$

The rectified position of the point is given by this new distance, keeping the same direction:

$$\begin{pmatrix} u_{P2} \\ v_{P2} \end{pmatrix} = \begin{pmatrix} u_{P0} + \frac{d_{dist}}{d}.(u_{P1} - u_{P0}) \\ v_{P0} + \frac{d_{dist}}{d}.(v_{P1} - v_{P0}) \end{pmatrix} \quad (8)$$

The color of the point is computed using the nearest pixel. Figure 8 shows a synthetic image generated with a virtual pinhole camera. In this case, there is no optical distortion ($k_{P1} = k_{P2} = 0$). Figure 9 shows a rectified image from the data of figure 8.

Fig. 8. Image from a synthetic pinhole camera

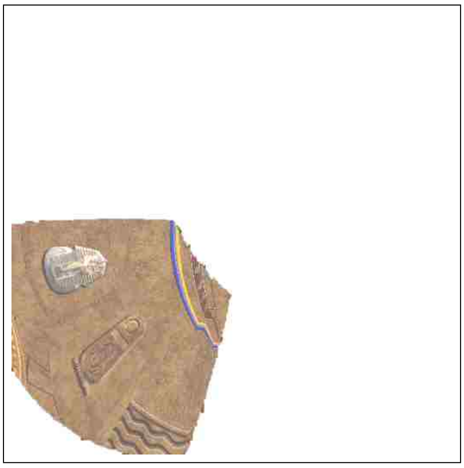

Fig. 9. Rectified image (upper panoramic part)

Calibration of the user's sensor is necessary. Firstly, parameters of the user's camera are estimated. Then, mechanical calibration is performed in order to align the orientation sensor axes with the camera.

Note that it is possible to integrate data from multiple images onto the same panoramic image if the camera viewpoint does not move, and due to this, a bigger part of the scene can be seen on the image (in figure 10). This leads to a more robust localization as the user's environment is more constrained.

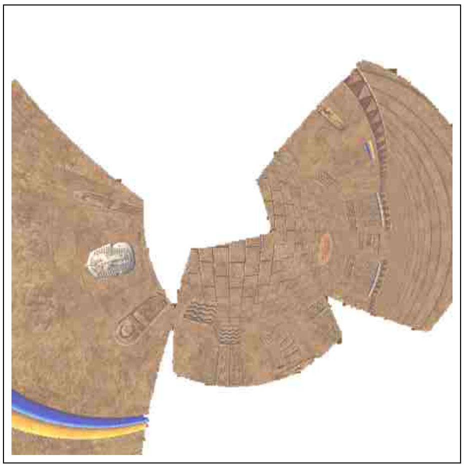

Fig. 10. Rectified combined image (downer panoramic part)

The same rectification principle can be applied to images acquired with a panoramic sensor.

$\alpha_{Cu}, \alpha_{Cv}, h_C, u_{C0}, v_{C0}$ are parameters of the real catadioptric sensor used for panoramic images acquisition as defined is section 3. The model of the catadioptric sensor is improved in order to take in account radial distortion due to the camera. k_{C1}, k_{C2} are radial distortion parameters.

The value of each pixel of the rectified panoramic image is computed as shown in figure 11. The center of the pixel

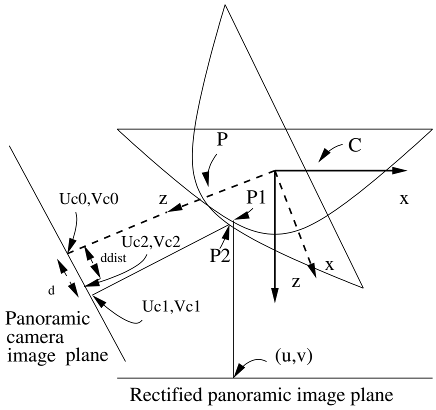

Fig. 11. Panoramic image rectification principle

(u,v) is next projected onto the mirror surface of the virtual sensor to a point $P_1 = (x_1, y_1, z_1)^T$ using equation (3).

The point is then rotated using equation (4).

The projection to the catadioptric sensor is relatively more complicated than those of the standard camera. Point $P_p = (x_p, y_p, z_p)^T$ is projected onto the surface of the real sensor mirror through the focus $(0, 0, 0)^T$ to a point P_2:

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \\ 1 \end{pmatrix} = \begin{bmatrix} \frac{h_C . x_p}{\|P_p\| + z_p} \\ \frac{h_C . y_p}{\|P_p\| + z_p} \\ \frac{h_C^2 - x_2^2 - y_2^2}{2.h_C} \\ 1 \end{bmatrix} \tag{9}$$

Orthographic projection of P_2 to panoramic image plane is computed using equation (2) by substituting P_1 with P_2.

Radial distortion is applied as in a standard camera image using equations (6-8), substituting k_{P1} with k_{C1} and k_{P2} with k_{C2}.

Figure 12 shows an image acquired with our panoramic sensor while figure 13 shows the rectified image.

5. CONCLUSION

A new localization and guidance system using an original 2D omnidirectional representation of a location has been introduced. The advantage of this representation, is the possibility of matching data from the database to data from the user. As we are actually working on synthetic data, the matching is achieved by correlation computation between pixels. The system is able to match image from figure 9 with figure 4 and image from figure 10 with figure 5.

User's and expert's images are rectified using self developed hand held sensors composed of standard or panoramic cameras and orientation sensors.

Images are synthesized from database via an optimized synthesis method, allowing the system to work in real time.

Fig. 12. Panoramic image from sensor

Fig. 13. Rectified panoramic image

This system is operational on synthetic data but different parts of the system will have to be improved in order to work with real data, mainly:

- The 3D omnidirectional sensor will be improved and validated so it will be possible to acquire the 3D model of a real site.
- Optimized synthesis image method will be even accelerated. The drawback of our implementation is that we need to gather images from the memory of the video card, and the associated bandwidth, hence, this process is time consuming.

  More than 60% of the total time of this computing is consecrated to retrieve these images from the video memory. The new video card standard, PCI Express, is expected to solve this problem, due to its increased bandwidth and this should allow synthesis of omnidirectional images with higher speed. The OpenGL fragment program extension can be also used to compute the panoramic image directly in the video card memory, hence, avoiding the transfer of data from video card to central memory.
- Matching methods shall be investigated to achieve localization in real sites.

6. REFERENCES

[1] T.Pajdla., V.Hlavac, "Image-based self-localization by means of zero phase representation in panoramic images", In Proceedings of the 2nd International Conference on Advanced Pattern Recognition, volume 2013 of Lecture Notes in Computer Science, pages 24- 33, Heidelberg, Germany, March 2001. IAPR, Springer-Verlag

[2] J.J. Gonzalez-Barbosa et S. Lacroix, "Localisation d'un Robot Mobile par Indexation d'Images Panoramiques", 16emes Journées des Jeunes Chercheurs en Robotique. Lyon France 2002

[3] A. Argyros, K. Bekris and S. Orphanoudakis, "Robot homing based on corner tracking in a sequence of panoramic images", Computer Vision and Pattern Recognition, 2001. Pages: II-3-II-10 vol.2

[4] L. Paletta, S. Frintrop, J. Hertzberg, "Robust localisation using context in omnidirectional imaging", Proc IEEE International Conference on Robotics and Automation, ICRA 2001. Pages 2072-2077

[5] R. Benosman and S. Bing Kang, "Panoramic Vision, Sensors Theory and Applications", Ed Springer 2001

[6] T. Svoboda, T. Pajdla, V. Hlavac, "Central panoramic cameras: Geometry and design", Research report, Czech Technical University, Faculty of Electrical Engineering, 1997

[7] C. Geyer and K. Daniilidis, "Catadioptric Projective Geometry", International Journal of Computer Vision 45(3), 2001 pp. 223-243

[8] A. S. Glassner, A. Arvo, R. L. Cook, E. Haines, P. Hanrahan, P. S. Heckbert, and P. Hanrahan. "An Introduction to Ray Tracing", Harcourt Brace Jovanovich Published: 1989