USING MOBILE SYSTEMS TO TRANSMIT LOCATION BASED INFORMATION

U. Kretschmer

Fraunhofer Institute for Computer Graphics, Fraunhoferstr. 5, 64283 Darmstadt, Germany
Ursula.Kretschmer@igd.fhg.de

Commission III, WG III/7

ABSTRACT:

This paper presents an approach for enabling a mobile system to estimate position and orientation of users walking around a city. In an urban environment, localization methods based on the cellular phone network or GPS are not accurate or not reliable enough. This is the reason why we propose to use information out of a 3D model representing the surroundings where the users are staying. A camera represents the line of sight of the users. Frames out of the video stream can then be used to be compared with views of the 3D model. To do so, image processing techniques are applied to the system to calculate the absolute orientation of the camera which the user is carrying. Reconstructions of ancient buildings can then be presented by their exact position. By using specific displays that can be hold in front of the eyes the users can watch the reality and at the same time observe the replaced buildings. This proposed method will be applied in the GEIST research project. Here, a learning game for students is being developed to transmit historical facts of the history of a town.

1. INTRODUCTION

Mobile systems and services are an important part of the liberty most of us want to rely on. They enable us to stay in contact all the time with other persons or to access data stored on remote systems. Location based services are offered from various sides for users of cell phones. Persons can get guiding support to the next restaurant or ATM, and even meetings of friends can be enhanced by these services. For these applications the cell phones can be localized by use of the cellular phone network or GPS (Global Positioning System). To overlay the reality with additional information while the users are watching it, for example by holding a display in front of their eyes, these techniques are not sufficient. There are several techniques to solve this problem. The approach being developed at Fraunhofer Institute for Computer Graphics (Fraunhofer-IGD) for application in urban areas takes into account additional information about the environment. We do not use photos of the surroundings as applied to other systems but a 3D city model. Information in the 3D model about features is compared to pictures taken by the camera, which represents the line of sight of the user. After calculating the absolute orientation of the camera, 3D reconstructions can be presented at the correct position in the display. Our solution will be part of the GEIST research project.

In section 2.1 some examples of services based on localization techniques are given. Section 2.2 introduces different methods to estimate the position of a mobile device. Section 2.3 provides an overview of different outdoor techniques for Augmented Reality. In section 2.4 the approach for a system is presented which enables to estimate position and orientation of a person in an urban environment. An application for this system is shown in chapter 3. Chapter 4 portrays the state of this approach as well as further development.

2. LOCATION BASED INFORMATION

2.1 Mobile Reality

There is a wide range of mobile information systems on the market. Laptops or pen-driven computers allow people to work on the move. Specific software enables them not only to do the same as in the office but to permit employment which would not be possible with a static system. An example is the support of work of surveyors or for people fixing machines on sight. PDAs (Personal Digital Assistants) have started to replace the use of calendars. They can be used not only to store appointments but also contacts joined with the connection by e-mail are possible or to load articles of newspapers. The increasing memory cell of PDAs leads to a dominance of this technique. But to a greater extent, cell phones govern our daily touch with mobile life. The screen of these devices is mostly smaller than the one of PDAs but advantage of these devices is still the possibility to response to changes in position in real time. Some network companies use this technique for gathering information about the position of the cell phone for special home fares. But also the transmission of specific information and services depending on the position can be obtained wherever there is data available. Common applications are broadcasting of traffic jams, route planning, city guides, hotel and restaurant pilots, localization of service stations, shopping support to indicate the next supermarket or the nearest ATM. Also the presentation of the position of friends is feasible. In Germany, most of these services are already part of the offering of different network companies.

2.2 Location based Services

For some of the above-mentioned information, manual input of the area code, the zip code, or the name of the city would be sufficient. It enables the system to present information about the environment of the user's position. Voice entry would be an

alternative to using buttons or a pen. The drawback for this kind of positioning technique is that the user needs to use complex input masks as well as the impossibility to use these systems for persons staying at an unfamiliar location. Asking for the best way to a specified site is impossible without knowing the current position. For these problems, an automated technique to estimate the position of a user is required. This could be the use of the terrestrial system of the cellular phone network. Among these, two techniques are being used: a network-based and a handset-based localization, depending on whether the network or the handset itself estimate the position. Here, accuracies of up to 50-100 m are possible. This accuracy is sufficient to get information about the next restaurant or shop in a defined circle around them. Advantages of this technique are the reception of the signal mostly in closed rooms, in narrow streets, or in tunnels, in case they are provided with repeaters. An application where the accuracy is not sufficient is broadcasting information about surrounding buildings while standing in the middle of an urban square. GPS (Global Positioning System) should then be applied to the system. This requires an additional hardware, which should be connected to the cell phone. With common PCMCIA cards, which can be plugged in the hardware, according to manufacturers' instructions an accuracy of up to three meters can be achieved. An improvement of the accuracy can be realized by applying Differential GPS (DGPS) to the system. Here, correction data from a fixed receiver is broadcasted to the hardware and taken into account by the system. These positioning techniques are common for a lot of systems running on cell phones or also on PDAs equipped with special cards.

2.3 Augmented Reality

One step further into the mobile reality is a system where the users can watch the current surroundings while the view is enhanced with additional information. This requires using a display which can be hold in front of the eyes.

Figure 1 shows n'vision's VB-30 hand-held display. It is supplied with a see-through mode, which facilitates watching reality and at the same time overlapping it with information depending on the position and orientation. This technique is called Augmented Reality (AR).

Figure 1: Hand-held display with orientation tracker

The position and the orientation of the user are the basic information, AR systems need to be provided with. With this information the system can evaluate which building the tourists are looking at and it can present the respective information in their line of sight. Several AR approaches exist for a solution of the outdoor orientation problem. The environment cannot be changed - for example by sensors or markers - in contrast to indoor AR systems.

Azuma et al. (Azuma et al., 1999a) evaluated that all tracking sensors such as GPS, inertial sensors, or electronic compass and tilt sensors lack accuracy or robustness. This leads to his conclusion that only a combination of tracking technologies, which he calls a hybrid tracking system, should be used. There are several systems using a GPS device and an orientation tracker, which achieve quite good results. One of the most well known is the *Touring Machine*, an information system for the campus of Columbia University in New York (Feiner et al., 1997; Julier et al., 2000) As the system broadcasts information about names of buildings and not about specific building features the accuracy is not as important in this system as in other AR approaches. A similar system also based on the unification of sensors (GPS and orientation tracker) is the system of the University of South Australia (Piekarski et al., 1999). Taking into account the errors due to local distortions of the magnetic field, Azuma et al. (Azuma et al., 1999b) presented an approach which uses corrections stored in a registered map. Applied to the system is additional calibration, which helps to reduce the drift problem, which increases with time. Further on, additional gyroscopes improve the results of the orientation problem by stabilization. The system is only established to identify real-world objects with in more than 500 meters distance from the user.

To improve these results several groups have used image-processing techniques. You et al. (You et al., 1999) developed an AR system which includes the use of a camera. The results of the inertial system are corrected by the video image. The problems still persist, i.e. accumulated gyroscope drift and compass errors provoke wrong orientation results. Behringer (Behringer, 1999) used horizon silhouettes of digital terrain maps to receive better results for the orientation problem. Chen and Shibasaki (Chen and Shibasaki, 1999) showed an approach where image sequences based on landmark lines are analyzed. Here, one image has to be geo-referenced manually and registered in a database. An additional solution for the use of further data was shown in the ArcheoGuide project by Stricker (Stricker, 2001). He used a reference image for comparison with the line of sight of the user. The results show that the system works quite well in case the users are close to the viewpoint at which the reference picture was taken.

All these approaches are either not accurate enough for the overlay of a reconstruction of the real building or they cannot be applied to persons walking in a city without the requirement of fixed positions where they should stand to receive information. This leads to an approach using a 3D model to track position and orientation of the users.

2.4 Augmented Reality in an urban environment

The use of GPS in an urban environment is risky due to problems with multipathing. Processing the video stream can control the surveying results. The approach of Fraunhofer-IGD is based on the use of a 3D model of the whole environment (Coors et al., 2000). While registering the reality by use of a video camera, the images can be compared with the model. Using of this technique AR can be applied to city tours.

2.4.1 Feature detection

For this comparison identical features in the 3D model on the one hand and in the frame of the video stream on the other hand should be detected. For these values only a part of the data within the 3D model is compared with the data in the video stream. We are using a DGPS receiver and an orientation tracker to get rough information about position and orientation at the beginning. We assume that the person is standing close to this position and is looking approximately in the direction received from the orientation tracker. One approach would be to generate views out of the 3D model and compare them to the picture of the video stream. This requires feature detection algorithms applied to both kinds of pictures. The advantage of this method is that both pictures are handled in the same way. It is a good starting point for the matching algorithm which is required to find identical features in both pictures.

For feature detection we merge the results of two operators. The edge detection algorithm is based on the Canny-operator (Encarnação et al., 1997). For a pixel belonging to an edge not only the coordinates but also the orientation is stored. These values are calculated by use of the first derivation and then by the gradient. The latter can be estimated according to the following formula:

$$\text{Gradient } \mathbf{D}\,(x,y) = \mathbf{D}_x\,(x,y)\,\mathbf{e}_x + \mathbf{D}_y\,(x,y)\,\mathbf{e}_y \qquad (1)$$

where $\qquad$ $\mathbf{D}$ = derivation
$\qquad\qquad\quad$ $\mathbf{e}$ = unit vectors

The strength of the gradient is a quantum for the strength of an edge. In the following non-maxima-suppression-method information about the direction and the size of the gradient are used to set all points without a local maximum in direction of the gradient to zero. This is done by comparing the values to adjacent gradients which are located at two different sides of the examined point. In a next step gaps in the edges caused by the previous step are closed.

To detect corners in the video stream the Susan operator (Ristow, 2001) is used. Here, a circular mask is placed over a point. Then, all the points with diverging intensity values to the central points are accumulated:

$$C\,(\mathbf{r}, \mathbf{r}_0) = \begin{cases} 1 \text{ if } |\mathbf{I}\,(\mathbf{r}) - \mathbf{I}\,(\mathbf{r}_0)| \le t \\ 0 \text{ else} \end{cases} \qquad (2)$$

where $\qquad$ $\mathbf{r}_0$ = position of the central point
$\qquad\qquad\quad$ $\mathbf{r}$ = position of any other point
$\qquad\qquad\quad$ $\mathbf{I}\,(\mathbf{r})$ = brightness of any pixel
$\qquad\qquad\quad$ t = brightness difference threshold
$\qquad\qquad\quad$ C = output of the comparison

If more than a predefined amount of the points differ, these points are potential candidates for a corner.

A drawback for the approach by examining 2D raster data is that the generation of views out of the 3D model requires much computer power. The information of the features gathered by image processing techniques is already used by creating the 3D pattern and stored within the data. To avoid the described step, visible vectors and points within the clipped part of the model should be gathered and then compared with the results of the feature detection algorithm applied only to the picture of the video stream.

2.4.2 Feature matching

The features found in the video stream and in the part of the 3D model should be matched in the next step. To do so, the matching algorithm of Gros et al. (Gros et al., 1995) is applied. It comprises the search of so-called segment configurations. They consist of two lines with a common endpoint (see Figure 2).
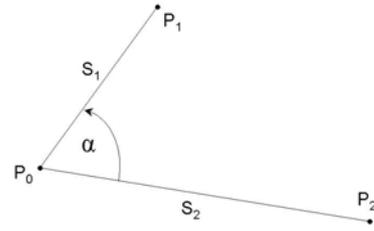

Figure 2: Segment configuration

By applying a similarity to these segment configurations there are two values which do not change: the angle between the lines and the length ratio. They are called invariants. Because a similarity is only an approximation of the correct transformation these values are called quasi-invariants. After finding segment configurations similar in both the image and the view, the transformation parameters of the similarity are calculated for each match. These parameters should be similar in case of the correct transformation. A point in R4 can represent them. Here, an area as small as possible is detected where most of the points are positioned. They represent the correct results of the transformation and lead to correspondent features.

2.4.3 Relative and absolute orientation

Epipolar geometry can be used to express the transformation between the two camera systems: the camera representing the line of sight of the user and the camera to get the cutout of the 3D model. The transformation can also be applied to identical features in both systems ($\mathbf{P}_{cam1}$ and $\mathbf{P}_{cam2}$). This transformation can be expressed by a rotation $\mathbf{R}$ and a translation $\mathbf{t}$:

$$\mathbf{P}_{cam1} = \mathbf{R} \cdot \mathbf{P}_{cam2} + \mathbf{t} \qquad (3)$$

After applying a vector product and finding the lowest common denominator, it can be shown that $\mathbf{P}_{cam1}$, $\mathbf{P}_{cam2}$ and $\mathbf{t}$ are placed in one plane: the epipolar plane. Then, the following equation can be derived:

$$\mathbf{t} \times \mathbf{R} = \mathbf{T} \cdot \mathbf{R} \qquad (4)$$

Here, **T** is built by use of components of **t**.

With $\mathbf{E} = \mathbf{T} \cdot \mathbf{R}$, this leads to the algebraic presentation of the epipolar geometry:

$$\mathbf{P}^{T}_{cam1} \cdot \mathbf{E} \cdot \mathbf{P}_{cam2} = 0 \qquad (5)$$

To estimate the relative orientation of the camera, the external camera parameters concealed in the matrix **E** are calculated. **E** is a 3x3 matrix but only dependent on five parameters: three rotation parameters and two translation parameters. A scale can be applied to the equation without leading to incorrect results. This phenomenon is called scale invariance.

To estimate the matrix **E** a least median squares method is applied to the system (Torr and Murray, 1993). Here, the calculation process with eight corresponding points is realized. Afterwards, the median is calculated. This is done until a predefined calculated amount of iterations is reached (at least one subset without outliers should be withdrawn). The best results are used to calculate the values by the least squares method.

After solving the relative orientation of the camera, the absolute orientation of the camera, which is carried by the user, can be estimated as the parameters of the viewing plane of the 3D model are known.

3. APPLICATION: THE PROJECT ‚GEIST'

This video-based tracking method will be used in the GEIST project. It is financed by the German Federal Ministry of Education and Research (01 IRA 12A). In this project, a mobile system will be developed which allows experiencing history within an urban location (Kretschmer et al., 2001). The system is addressed to students and should help them to learn historical facts via a thrilling, individual story. It allows them to come into contact with history in an environment where the past actually took place. Our application area is Heidelberg. The story is related to information about the 30 years war. 3D objects which are managed in a Java3D scene graph can be positioned at the places they were or still are.

On squares defined in advance, the story continues and all functionality is available there. Here, accurate tracking is required. As soon as the students leave this area only the position is tracked. Now, they can get information by use of a PDA. This information is only position dependent but not orientation-dependent. When leaving the Old Town of Heidelberg a lost-the-way sensor is started guiding the students back in the game.

This scenario requires a lot of data, which are stored in databases. Here, the 3D objects are collected for use of accurate tracking or for the presentation of an ancient environment in the display. Also, historical facts and fictional elements are gathered for access by the story-engine which creates the story for the game. Other data important for tracking is geoinformation. It is used, among others, to let the system know when a user enters one of the predefined squares. Then, the story continues and accurate tracking is required.

4. OUTLOOK

For the first prototype only tracking by use of DGPS and an orientation tracker is provided. The next generation of the system, which is due in spring 2003, is based on analyzing pictures of the video stream and comparing them with data in the 3D model. In a first approach we implemented the edge detection algorithm. All the edges which should be used for the matching algorithm are then stored as vectors. As the information out of the 3D model is at the moment captured by generating views, this raster data has to be analyzed as well.
Figure 3 presents a photo as well as a generated view out of the model of the city center in Darmstadt and the result of the edge detection algorithm for both.
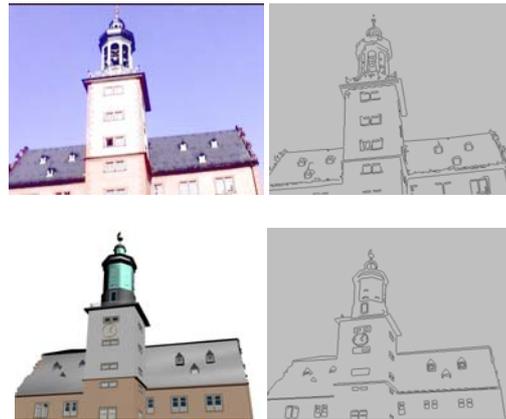

Figure 3: Application of the edge detection algorithm to a picture and the view

After the edge detection algorithm, which is programmed in C++, the matching algorithm by Gros has been implemented in the system. It is developed in Java. Due to the missing corner detection algorithm the matching leads only to satisfactory results in case the scene represented in the frame of the video stream and in the view of the 3D model is similar.

Further work will be to implement the corner detection to improve the results of the matching algorithm and to enhance the steps for the orientation of the camera. Here, the implementation has not been finished yet. To use the feature information of the 3D model directly, some work has to be done to gather the edges and corners visible to a person at a specific position and direction. If this step is completed, direct matching of the data of the 3D model and the features detected in the video picture will be possible.

5. REFERENCES

Azuma R., Hof B., Neely III H., Sarfaty R., Daily M., Bishop G., Chi V., Welch G., Neumann U., You S., Nichols R., and Cannon J., 1999a. Making Augmented Reality Work Outdoors Requires Hybrid Tracking. In: *Augmented Reality: placing artificial objects in real scenes: proceedings of IWAR'98,* San Francisco, Calif., pp. 219-224.

Azuma R., Hoff, B., Neely III H., and Sarfaty R., 1999b. A Motion-Stabilized Outdoor AR System. *Proceedings of IEEE Virtual Reality 1999*, Houston, Texas, USA.

Behringer R., 1999. Registration for Outdoor Augmented Reality Applications Using Computer Vision Techniques and

Hybrid Sensor. *Proceedings of IEEE Virtual Reality 1999*, Houston, Texas, USA.

Chen T. and Shibasaki, R., 1999. A Versatile AR Type Mobile GIS Based on Image Navigation Technology. *IEEE SMC '99 Conference Proceedings,* Tokyo, Japan.

Coors V., Huch T., and Kretschmer U., 2000. Matching Buildings: Pose Estimation in an Urban Environment. *Proceedings IEEE and ACM International Symposium on Augmented Reality,* Munich, Germany.

Encarnação J., Straßer W., and Klein R., 1997. *Graphische Datenverarbeitung 2.* R. Oldenbourg Verlag, München, pp. 313-318.

Feiner S., MacIntyre, B., Höllerer, T., and Webster A., 1997. A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. *Proceedings of the International Symposium on Wearable Computing*, Cambridge, MA

Gros P., Bournez, O., and Boyer E., 1995. Utilisation des Quais-invariants geometriques pour L'appariement et la modelisation des images de segments de droite. *Rapport de recherche*, Institut national de recherche en informatique et en automatique, Grenoble, France.

Julier S., Lanzagorta, M., Baillot, Y., Rosenblum, L., Feiner, S., and Höllerer T., 2000. Information Filtering for Mobile Augmented Reality. *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000),* Munich, Germany.

Kretschmer, U., Coors, V., Spierling, U, Grasbon, D., Schneider, K., Rojas, I., Malaka R., 2001. Meeting the Spirit of History. *Proceedings of VAST 2001 – Virtual Reality, Archaeology, and Cultural Heritage.* Glyfada, Greece.

Piekarski W., Gunther, B., and Thomas, B., 1999. Integrating Virtual and Augmented Realities in an Outdoor Application. *Proceedings of International Workshop on Augmented Reality 1999*

Ristow B., 2001. *Gewinnung semantischer 3D-Modelle aus Fotografien.* Shaker, Aachen, pp. 33-34.

Stricker D., 2001. Tracking with Reference Images: A Real-Time and Markerless Tracking Solution for Out-Door Augmented Reality Applications. *Proceedings of VAST 2001 – Virtual Reality, Archaeology, and Cultural Heritage.* Glyfada, Greece.

Torr P. and Murray D. 1993. Outlier Detection and Motion Segmentation. In: *Proceedings Sensor Fusion VI*, Boston MA, SPIE volume 2059, pp. 432-443.

You S., Neumann, U., and Azuma, R., 1999. Orientation Tracking for Outdoor AR Registration. *Proceedings of IEEE Virtual Reality 1999*, Houston, Texas, USA.