AR Mapping of GIS Information by Pattern-based Tracking with Particle Filter

Yuko Uematsu
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Japan
yu-ko@ozawa.ics.keio.ac.jp

Hideo Saito
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Japan
saito@ozawa.ics.keio.ac.jp

Myriam Servieres
Ecole Centrale de Nantes
BP 92101 F-44321 Nantes Cedex 3 FRANCE
myriam.servieres@ec-nantes.fr

Guillaume Moreau
Ecole Centrale de Nantes
BP 92101 F-44321 Nantes Cedex 3 FRANCE
guillaume.moreau@ec-nantes.fr

Abstract

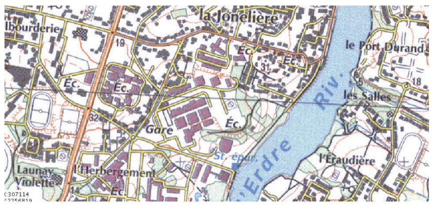
This paper presents an AR system for mapping GIS information. GIS is a kind of database system that has geographical information in layers. By using our system, when a user watches some scene from a high place with a camera phone, some information about the scene, such as a building's name or a street's name, are overiald onto the scene though the display. For the registration of the information data which is generated with CG and the camera, pattern-based camera tracking with particle filter is introduced in this paper. In this method, a top-view image of the scene which is captured from a airplane and is included in the GIS database is used for the tracking. The particle filter evaluates the parameters of the camera's position and pose by consistency of the top-view image and input images captured from the user's camera. In the experiment, we demonstrate the mapping of the geographical information onto the user's input images.

1. Introduction

Augmented Reality (AR) is a technique for overlaying virtual objects onto the real world, which is achieved by superimposing 3D CG objects onto 2D images of the real world. We can see the virtual objects as if they really exist in the real world. Therefore AR can provide users very effective views by giving some additional information [1].

Geographical Information Systems (GIS) are databases that include a spatial component. They have become unavoidable tools for the development and management of territories. They are used for diagnosis, public debate and as decision making tools. The goal of GIS is to provide information about a given space. They include many types of geographical objects organized in themes frequently represented as layers (tiers architecture) containing objects of the same type (roads, buildings). GIS data can be vectorial as shown in Fig. 1 (a) or raster data as shown in Fig. 1 (b).

(a) Vectorial data



(b) Raster data

Figure 1. GIS database

In this paper, we introduce AR mapping system of GIS data on the captured images. When a user watches the scene of the town from a high place through a camera phone, the map information about town, such as the name of the buildings, roads, rivers etc., is overlaid onto the image in the display. Therefore this system is available for the navigation or the sightseeing guide.

For overlaying such GIS map information which is generated which is generated with CG onto the input images, the position and pose of the camera has to be estimated at every frame. Such geometrical registration problem is one of the most important issues for implementing AR appli-

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

cations. In this paper, our method utilizes the framework of patricle filter [2], which is a kind of time-sequential filtering technique for estimating some parameters. The best way in terms of accuracy is full search of all parameter candidates, however, it is quite unreasonable for an on-line system from a computational perspective. Therefore the particle filter has some hypotheses (particles) which are sampled candidates. It is assumed that the correct value is probably included in the candidates. The confidence of each particle is evaluated at every frame and is propagated to the next frame. Therefore how to evaluate the confidence is the key of this method.

For evaluating the hypotheses (particles) of the parameters, this method uses the top-view image which is captured from a airplane and is included in the GIS database. By comparing the top-view image with the input image captured by a user, the camera's parameters are computed. In particular, we focus on the planar area of a building, such as a roof, and transform the top-view image to the input image by planar transformation (homography). Then the transformed top-view image is compared with the input image.

2. Camera Tracking with Particle Filter

In this section, the tracking method of the camera's rotation and translation parameters (extrinsic parameters) is explained. The algorithm using the particle filter is divided into two phases as shown in Fig. 2; initialization phase, tracking phase. Because our method focuses on the planar area of the building such as a roof, a planar homography between the top-view image and the input image is computed at the initialization. Then the initial extrinsic parameters are computed from the homography. At the tracking, the best parameters are searched around the initial parameters by using the framework of the particle filter.

The probability density function (PDF) of the particle filter consists of a set of discrete hypotheses (particles) of the extrinsic parameters and corresponding weight values. For computing the weight values, the particle which can generate the fittest pattern to the actual input image is searched from the candidates as the best particle. In particular, the top-view image are transformed to the input image by using the hypothetical parameters. Then the patterns around the feature points which are detected from the top-view image are compared to the pattern from the input image.

2.1. Initialization Step

Our method uses M feature points on the top-view image for computing weight values in the particle filtering. In the experiment, the feature points are detected by using OpenCV library (cvCornerMinEigenVal function).

At the initialization, corresponding points on the initial

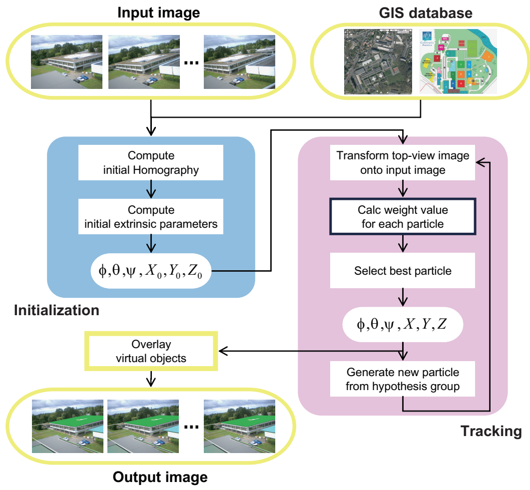


Figure 2. Overview of our tracking method.

frame of the input image are specified by manual. Then, a planar homography $\boldsymbol{H}$ between the top-view image and the input image is computed by the corresponding points. From the homography, extrinsic and intrinsic parameters of the camera can be computed as follows [3].

The extrinsic parameters consist of a rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$, and the intrinsic parameters consist of a intrinsic matrix $\boldsymbol{A}$. Therefore the homography can be represented by a combination matrix of $\boldsymbol{A}$, $\boldsymbol{R}$ and $\boldsymbol{t}$ without the third vector $\boldsymbol{r_3}$ of $\boldsymbol{R}$.

$$\boldsymbol{A}\,[\boldsymbol{R} \mid \boldsymbol{t}] = \boldsymbol{A}\,[\boldsymbol{r_1}\ \boldsymbol{r_2}\ \boldsymbol{r_3}\ \boldsymbol{t}] \tag{1}$$

$$\boldsymbol{A}\,[\boldsymbol{r_1}\ \boldsymbol{r_2}\ \boldsymbol{t}] = \boldsymbol{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \tag{2}$$

By using this theory, the extrinsic parameters are computed from the homography as an initial extrinsic parameters. At the tracking step, the best parameters are searched around the initial parameters. The process will be described in the next section.

2.2. Tracking Step

In the tracking step, the extrinsic parameters which represent the position and pose of the camera are esimated. This method uses the framework of the particle filter and searches the best parameters around the initial parameters which are computed in the initialization step.

The particle filter estimates the extrinsic parameters in the t th image frame as 6 dimentional vector $\boldsymbol{p}_t = (\phi_t, \theta_t, \psi_t, X_t, Y_t, Z_t)^\top$ in a 3D state space $\boldsymbol{S}$ where t represents the frame number.

In the framework of the particle filter, the probability density function (PDF) is represented as a set of N discrete hypotheses (particles) $\{\boldsymbol{s}_t^i\}$ in the 6D state space $\boldsymbol{S}$ and the

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

corresponding weight values π_t^i $(i = 1 \cdots N)$. This sample set can approximate an arbitrary PDF.

At the beginning of the tracking, we generate N new samples as $\boldsymbol{s}_t^0$ in neighborhood of the initial values which are rotation and translation parameters obtained at the initialization step. Then a constant value $\pi_0^{(i)} = 1/N$ is given to every particle. In this way the initial values of a set of particles are decided.

After the initial frame, the tracking is performed based on the previous particles and the current input image. The particle set in t th frame $(\boldsymbol{s}_t^i; \pi_t^i)$ is estimated based on the previous assumption set $(\boldsymbol{s}_{t-1}^i; \pi_{t-1}^i)$ and a motion model as following equation.

$$\boldsymbol{s}_t^i = \boldsymbol{s}_{t-1}^i + v_{t-1} + \mu \qquad (3)$$

where, v_{t-1} is velocity of the camera and represents the distance from t-2 th frame to t-1 th frame. μ is random noise. In our method, each particle is moved from previous sample based on the concept that the camera moves with uniform motion v_{t-1} and is diffused by adding random noise μ to become the particle in t th frame.

After obtaining N new particles $\{\boldsymbol{s}_t^i\}$, we compute the corresponding weight values π_t^i for $\boldsymbol{s}_t^i$ by evaluation based on the current image frame. The weight values mean the level of confidence for the corresponding particles. Therefore we give larger value to the particle which is closer to the truth. In our method we use the contour of the marker and evaluate how far the contour in the input image from the projected contour by each particle. In particular, we compute distances from the sampled K points on the contour to the edge in the input image. The detail will be described in the next section.

Finally we consider the particle $\boldsymbol{s}_t^i$ which has the maximum weight value as the camera pose $\boldsymbol{p}_t = (\phi_t, \theta_t, \psi_t)^\top$ in t th image.

3. Computation of Weight Value

3.1. Overview

The weight values represent the level of confidence for the corresponding particles. Therefore we give a big value to the particle which seems to be close to the truth; give a smaller value to the particle which seems to be far from the truth. In order to obtain the weight value for particle i, this method utilizes a patter-based evaluation. The weight values $w_{patt}^{(i)}$ are normalized between -1 and 1 as $c_t^{(i)}$. After computing $c_t^{(i)}$ for all N particles, a weight value $\pi_t^{(i)}$ for each particle i is computed by gaussian function as following equation.

$$\pi_t^{(i)} \propto e^{-\frac{\left(1-c_t^{(i)}\right)^2}{2\sigma^2}} \qquad (4)$$

where, σ is standard deviation of Gaussian function. In our experiment, we let $\sigma = 3.0$. Each $\pi_t^{(i)}$ is normalized so that the sum of all the $\pi_t^{(i)}$ become 1.0. Therefore the particle which has closer to the true will obtain bigger weight value $\pi_t^{(i)}$.

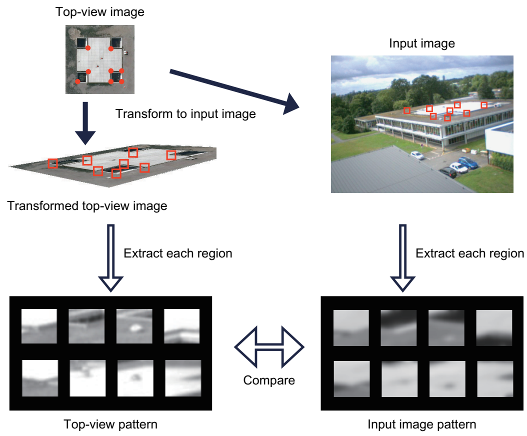


Figure 3. Computation of pattern-based weight value.

In this way, the particle which has the largest weight value is selected as the extrinsic parameters in the current frame, after computing the weight values for all the particles. Because of evaluating parameters by checking how fit each particle is to the actual input image, we can always search the best parameters at every frame. The detail of the computation will be described in the next section.

3.2. Pattern-based Evaluation

Fig. 3 shows the overview of computing pattern-based weight value. First, the top-view image is transformed to the input image by using the parameters of each particle. As described in section 2, M feature points are detected from the top-view image. Therefore the regions around the feature points on the transformed top-view image are extracted. The feature points are also projected onto the input image, and then the regions around the projected points on the input image are extracted to be compared as shown in Fig. 3.

If the parameters of the selected particle are completely correct, the regions around the projected points on the input image are consistent with the region around the feature points on the top-view image. Therefore, the difference between the regions becomes the weight value for the particle. The difference value for each feature point j between the top-view image's region and the input image's region is computed based on NCC as f_j $(1 \leq j \leq M)$. Since each f_j has the value between -1 and 1, the sum of f_j for every feature point obtained by using the particle i is normalized between -1 and 1 as follows.

$$w_{patt}^{(i)} = \frac{\sum_{j=1}^{M} f_j}{M} \qquad (-1 \leq w_{patt}^{(i)} \leq 1) \qquad (5)$$

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

Figure 4. Overlaid image sequence of map information.

Figure 5. Resulting frames that fail in registration of map information.

The score value $w_{patt}^{(i)}$ of the particle which has closer to the true parameters is closer to 1, conversely, the score value of the particle which is far away from the truth is close to -1.

4. Experimental Results

In this section, the experimental results are shown. We apply our method to the image sequence of the campus of Ecole Centrale de Nantes which is captured from a high building. Then a map information of the campus is overlaid onto the image sequence as shown in Fig. 6.

(a) Input image sequence

(b) Map of ECN

Figure 6. Input images and overlaid map information.

First, 8 feature points are detected from the top-view image, and then the corresponding points are specified on the first frame of the input images. By using the feature points, the initial homography and the initial parameters of the camera are computed. By generating the hypotheses from the initial parameters, the particle filter starts to estimate the best parameters from the hypotheses.

Fig. 4 shows the resulting image of overlaying the map information. The map information is almost overlaid onto the correct positions of the buildings' roofs. Because this method transforms the map information by using planar-transformation (homography), the information of the buildings which are not the same height as the center building can not be overlaid onto the correct position in the present algorithm. If we can use the information about the buildings' height, we can solve the problem. Fig. 5 shows another resulting images which fail in the registration. This is method evaluates the particles using only pattern matching (NCC). However, such a local matching does not work when the object pattern does not have distinctive texture or features. Therefore, we will also utilize histogram of the pattern to evaluate the particles in the future work.

5. Conclusion

In this paper, we have proposed the prototype of the AR mapping system of GIS information. For the registration of the map information onto the camera, we have proposed a pattern-based camera tracking method with particle filter. In the present method, the homography from which the initial parameters are estiamted is computed by the corresponding points specified by manual. As a future work, therefore, we will consider the way to obtain the intial parameters (ex. GPS) as well as using the histogram for evaluation process.

References

[1] R. T. Azuma. A survey of augmented reality. *Presence*, pages 355–385, 1997. 1

[2] M. Isard and A. Blake. Condensation-conditional density propagation for visual tracking. *Int. J. of Computer Vision*, 29(1):5–28, 1998. 2

[3] G. Simon and M. Berger. Reconstructing while registering: a novel approach for markerless augmented reality. In *Proc. of the ISMAR*, pages 285–294, 2002. 2

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278