Urban Localization based on Correspondences between Street Photographs and 2D Building GIS Layer

N. Bioret[1] G. Moreau[1] M. Servières[1]

[1] CERMA UMR 1563

Ecole d'Architecture de Nantes
Rue Massenet - F-44319 Nantes Cedex 3

{nicolas.bioret, guillaume.moreau, myriam.servieres}@ec-nantes.fr

Abstract

We propose a localization approach in urban environments based on correspondences between street photographs and a 2D GIS. Assumption of vertical facades and use of vanishing points theory allows extracting a partial geometry from the images : angles and width ratios between facades. This information is then transformed into GIS queries, which result in a layer of every possible pose for the camera. Wrong poses are then discarded based on consistency criteria, using road layer information or building layer information that has not been used before. Eventually, combination of possible poses coming from several images of the same site result in a unique pose, the true one. We show results of our experiments based on real data of Nantes city, France.

Keywords

Augmented Reality, Geographic Information Systems, Pose computation, Vanishing points, Correspondences between image and model, 3D Model.

1 Introduction

Onsite GIS applications meet an increasing demand in urban environments, for retrieving, creating or updating geographic information. Such applications always require a precise localization of the user. While geographic information is usually in 2 dimensions, 3D building models become available as well as facade textures and also street pictures if they can be registered in the GIS frame. Need for accurate and reliable 3D data becomes serious in the field of urban planning in a context of sustainable development. Moreover, tools for 3D virtual exploration of cities grow popular [1], but they are limited by 3D data low accuracy and coverage. Interaction with 3D data brings new challenges, to which Augmented Reality seems to be a promising solution. Geographic data is overlaid in real time on pictures of a scene, which demands a real time registration of a camera frame in a GIS frame, involving both 3D localization and orientation of a camera.
In our approach, such registration is obtained by matching photographs taken in streets with geographic data. Since we work with a 2D building layer and a 2D road layer, we achieve a 2D localization and 3D orientation, assuming facades are vertical. As a feedback loop, our GIS-connected Augmented Reality application will insert new 3D data in the GIS, for instance facade textures and the associated semantic information.
This article is structured as follows : part 2 presents a state of the art about outdoor localization based on computer vision. Part 3 explains how image data is matched with GIS data. Part 4 presents the results of our experiments and part 5 discusses the possible improvements and future work.

2 Existing work

Outdoor localization based on computer vision typically relies on a model on the viewed scenes. However, some applications get rid of any model and build a cloud of 3D points modeling the scene, but the registration of such a local model is still required [2, 1, 3, 4, 5].
The first type of model data used for outdoor localization is street images, which have previously been registered in the GIS frame. Such approaches rely on image matching and triangulation as presented in [6]. The size of an image database can get huge as its coverage grows, leading to non real time computation. The first way traditionally explored to overcome that limitation is reduction of computation time through an optimized indexing algorithm [7, 8, 9, 6, 10]. The second way is reducing the database size with simplified textures [11], spherical images [12] or region of interests rather than complete images [13].
Vector data, for instance 3D wireframe building models, are also used. Virtual tracking provides an accurate outdoor localization [14, 15, 16] but initialization of the camera pose is still a challenging issue. Posit algorithm [17] aims at establishing automatic correspondences between images and 3D model, but no result have been shown in an outdoor environment. Use of GPS is interesting [18, 19] but the accuracy in dense urban environment may not be enough for pose initialization. Most of the approaches mentioned above consider a 3D model of only 1 building and there-

fore, research of the right building in a GIS is not considered. [20] uses a model made of a 2D map, containing several buildings, but correspondences between images and 2D model are found manually, which implies the user knows how to establish such correspondences.
A third type of model, combination of vector data and textures, brings a better robustness and accuracy. Initialization of the camera pose is still required, and differential GPS is often considered [21, 22, 23, 24] in association to a sweeping algorithm. However, this kind of approach faces the 2 main limitations we find in approaches based only on image data or only on vector data. On the one hand, creation of an image database with large coverage and accurate registration in big cities seems currently out of reach. On the other hand, no satisfying method has been found for pose initialization using a model containing several buildings.

3 Matching GIS and images

3.1 Geometry extraction from images

Our approach is based on facade extraction from the images, based on the assumption that facades are vertical. Vanishing points associated to the facades can be extracted using for instance [25] or [26]. The vertical vanishing point is the same for every facade, and each facade is associated to 1 horizontal vanishing point. Based on our previous work [27], we can extract the width ratio between 2 neighbor facades and the angle between 2 facades even if these are not neighbor facades. Each image measurement is associated to a measurement uncertainty, which we propagate to the angles and width ratios measurements. However, it is not directly possible to extract length information from the image, because of the unknown scale factor.

3.2 Segments queries in the GIS

Then this information is transformed into GIS queries. Our GIS contains a 2D building layer, which is made of 2D polygons representing building footprints, and a 2D road layer made of polylines representing the median axis of the roads. For 2 neighbor facades F_1 and F_2, we extract an angle α and a width ratio r, associated with uncertainty measurements. Those uncertainties are given as standard deviations and are then transformed into tolerances δ_α and δ_r. Then, we query in the GIS every couple of segments (s_1, s_2) such that :

- s_1 and s_2 share a common extremity,
- s_1 and s_2 are separated by an angle belonging to $[\alpha - \delta_\alpha, \alpha + \delta_\alpha]$
- s_1 and s_2 are related by a length ratio belonging to $[r - \delta_r, r + \delta_r]$

Figure 1 details how these queries work. We have then a set of every possible couple of segments in the GIS building layer, with correspondences between s_i segments and F_i facades. Therefore we can compute a set of every possible pose. This number of possible poses can be seriously reduced using consistency criteria. If more complex structures are identified in the images such as collections of n neighbor facades, we can use more complex segments queries, which will result in fewer possible poses than with only 2 neighbor facades.

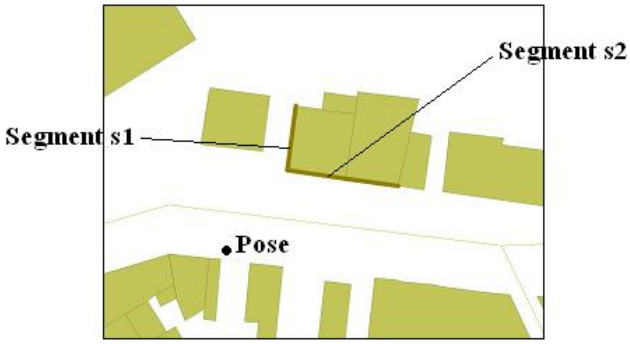

Figure 1 – *Segments queries in the GIS building layer*

3.3 Consistency queries in a GIS

First of all, if several unrelated angles and ratios are computed from an image, several sets of possible poses are generated and are then merged together. During the merging of n poses layers, we can remove every pose which do not belong to a group of n poses. In this query, as in every other consistency query, we use tolerances which prevent from deleting the true pose.

Secondly, we can tell from the image if there is a road between the pose and the viewed facades. For each computed pose and using GIS road layer, we can obtain the same information. If there is a contradiction between GIS and image measurements, the pose is discarded. Similarly, we can compute for each pose and using the building layer if there is an obstruction by a building located between the pose and the viewed facades. If such a situation occurs, the pose is discarded, which allows us to discard any pose located inside of a building. Figure 2 gives an obstruction example.

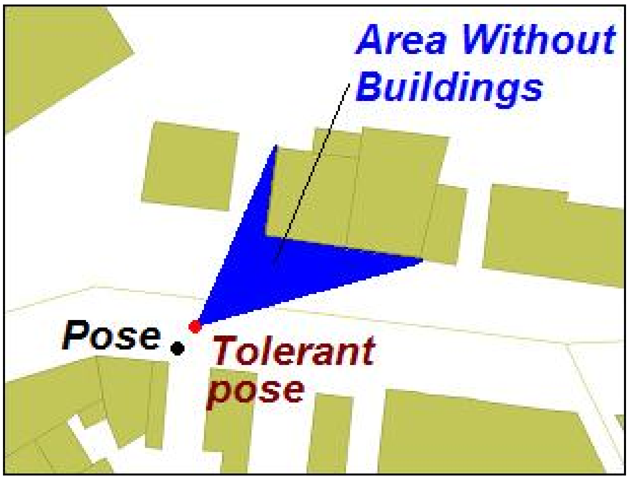

Figure 2 – *Example of obstruction query in the GIS building layer*

Thirdly, more precise measurements can be made in the image, like azimuth measurements. For each possible pose, a viewing polygon is computed, defined as follows : this includes every point in the GIS that can be viewed from the pose, in a particular angular domain and within a given range distance (in our experiments this range distance was about 100 m, which corresponds to the limit of the image resolution to reliably distinguish vertical facades). This viewing polygon contains every azimuth angle of viewed vertical edges and every orientation of vertical facades, which can also be computed from the images. For every inconsistency between GIS and image measurements, the pose is discarded. Definition of tolerances for such a query is however difficult because it must take into account GIS modeling errors. Figure 3 details how these queries work.

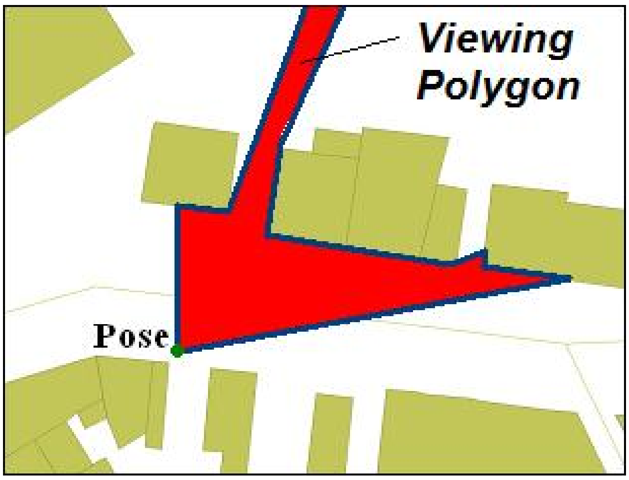

Figure 3 – *Viewing Polygon used in a visibility query*

3.4 Queries using several images

While consistency queries aim at cutting the number of possible poses down to 1, it is not always possible to achieve this goal only using 1 image. We have to combine information coming from different pictures taken at the same location, with varying orientations. One set of possible poses is generated for each image, and are merged using a distance criterion. For n images, we should observe groups of n poses located within a tolerance distance R, and every pose which do not meet this assumption are discarded.

4 Experimental results

Our experiments were based in real data from the city of Nantes, France. Our GIS data can be viewed on `http://www.geonantes.fr`, and initially contains 7138 buildings, which corresponds to a $3km^2$ area. We can remove from this GIS every segment that cannot be viewed from a street point of view, and as a result, we query a transformed building layer containing 2715 polygons. Moreover, we have taken a set of 28 street photographs, distributed on 7 different sites. On each site, the different poses are located within a circle of radius 10 m. Those photographs were taken with the same focal length, and therefore we consider that the internal camera parameters are known. Radial distortions in the image have been corrected based on OpenCV library. In our approach, vanishing points are extracted automatically but recognition of the vertical edges limiting facades, and definition of neighborhood relationship between 2 facades is done manually. One can notice that [28] gives an interesting approach for automatic segmentation of the image into facades.

On the 28 images, 4 were found correctly positioned in the GIS. 4 other images have not been located because the number of possible poses was above 1. There were 20 failures.

– 8 failures were due to a problem in GIS data, and more precisely 4 failures were due to a recent building, viewed on the image but not present in the GIS (an example is given in Figure 4), 3 errors were due to a wrong angle between 2 facades in the building layer (an example is given in Figure 5), and 1 error was due to a 3D structure which does not belong to a building but which was considered as part of the building.

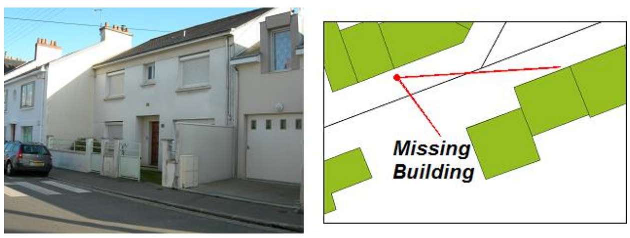

Figure 4 – *Missing building in the GIS*

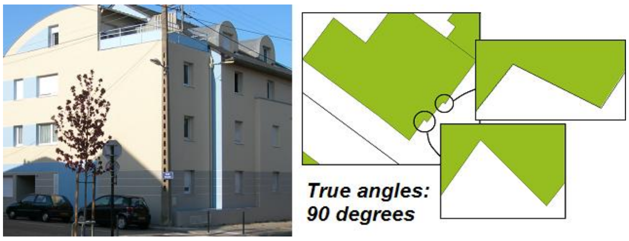

Figure 5 – *False angles in the GIS*

– 8 failures were due to image measurements from the user, and 7 of them occurred in a case of 180 degrees angle between 2 facades : the vertical edge limiting 2 facades was not evaluated at the correct location on the image. An example of this situation is shown on Figure 6, in which it is indeed difficult even for a human user, to find the true correspondences between the GIS building layer and the image.

Figure 6 – *Difficult correspondences between Images and GIS, due to 180 degrees angles*

– 3 failures were due to the fact that we do not work with the road width, but only with its median axis.
– 1 failure occurred because no entire facade was visible on an image, and therefore no pose can be computed.

We also noticed that in the case there were no failures, position accuracy was less than 1 meter, which is very interesting because this is enough for pose initialization using visual tracking.

5 Discussion

5.1 Solving failures

Some improvements can be made to solve the failures listed in paragraph 4. First of all, use of multi-hypotheses approach could help find the right correspondences between GIS and images in ambiguous cases, although it could generate much more possible poses. In case of facades separated by a 180 degrees angle, the vertical edges limiting 2 neighbor facades are difficult to find because they cannot be found using vanishing points, only by texture analysis. In case of a wrong GIS modeling, wrong segment orientation or over segmentation of a polygon, multi-hypotheses approach could also help because such modeling errors occur in very special case that could be predicted. Secondly, one could also increase tolerances used in queries so as to take into account error modeling in GIS. Thirdly, we could also take advantage of a width modeling for the roads.

5.2 Reducing the number of possible poses

We can also explore several ways to reduce the number of possible poses. It depends on the number of considered polygons in the GIS building layer, but also on the relative orientations of the facades. For instance, unusual angles in a city are easier to find than angles of 90 or 180 degrees, which are extremely frequent in a city like Nantes. Use of an accurate GPS would help reduce the number of polygons to query, although GPS signals are hardly available in dense city centers. Use of absolute orientation sensors, even with a low precision, will help filter out many wrong poses. However, this will really be useful if the distribution of facade orientations in the city is isotropic.

5.3 Using 3D data

Use of 3D data allows highly selective pose filtering, but we face the problems of availability, accuracy and reliability of such data. Indeed, modeling of 3D data is much more difficult that modeling of 2D data. On the one hand, correspondences between images and 3D model will be ambiguous in some cases, which implies that human interaction will be mandatory to establish such correspondences. On the other hand, correspondences between images and 2D model are also ambiguous because we do not take into account building height in inter visibility computation. As a result, we plan for the future to work on the use of 3D data using interactive approach.

6 Conclusion

We have proposed in this article an outdoor localization approach for street images, based on GIS building layer. Automatic correspondences between GIS and geometry extracted from the images result in a layer of every possible pose. Wrong poses are then filtered out using consistency criteria with building and road layer and also merging the possible poses coming from several images. Experimental results have shown that this approach is relevant, but limitations appear like the ambiguity in correspondences between images and 2D GIS. Many ways can be explored to overcome these limitations, and chiefly multi-hypotheses approach, use of 3D data, and multi view localization. Moreover, it seems that interactive localization is more robust than automatic one. Therefore our future work also includes a better integration of the user role during the localization process.

Références

[1] Noah Snavely, Steven M. Seitz, et Richard Szeliski. Photo tourism : exploring photo collections in 3d. *ACM Transaction on Graphics*, 25(3) :835–846, July 2006.

[2] Konrad Schindler et Joachim Bauer. Towards feature-based building reconstruction from images. Dans *Computer Graphics, Visualization and Computer Vision, Plzen, Czech Republic, February 3-7*, 2003.

[3] Bogumil Bartczak, Kevin Koeser, Felix Woelk, et Reinhard Koch. Extraction of 3D freeform surfaces as visual landmarks for real-time tracking. *Journal of Real-Time Image Processing*, 2(2-3) :81–101, November 2007.

[4] Chieh-Chih Wang, Chuck Thorpe, Martial Hebert, Sebastian Thrun, et H. Durrant-Whyte. Simultaneous localization, mapping and moving object tracking. *The International Journal of Robotics Research*, 26(6) :889–916, June 2007.

[5] Oussama Moslah, Vincent Guitteny, et Serge Couvet. Fitting uncalibrated photographs to geo-referenced urban 3d models. Dans *Journees de l AFRV, Bordeaux, France, 30-31 october*, pages 57–61, 2008.

[6] Wei Zhang et Jana Kosecka. Image based localization in urban environments. Dans *International Symposium on 3D Data Processing, Visualization and Transmission, North Carolina, Chapel Hill, June 14-16*, pages 33–40, 2006.

[7] Hao Shao, Tomas Svoboda, Tinne Tuytelaars, et Luc Van Gool. Hpat indexing for fast object/scene recognition based on local appearance. Dans *International Conference on Image and Video Retrieval, University of Illinois at Urbana-Champaign, USA, July 24-25*, pages 71–80, 2003.

[8] Toon Goedeme, Tinne Tuytelaars, et Luc Van Gool. Fast wide baseline matching for visual navigation. Dans *CVPR, Washington, DC, USA, June 27 - July 2*, pages 24–29, 2004.

[9] Gerald Fritz, Christin Seifert, et Lucas Paletta. Urban object recognition from informative local features. Dans *International Conference on Robotics and Automation, Barcelona, Spain, April 18-22*, pages 131–137, 2005.

[10] Ulrich Steinhoff, Dusan Omercevic, Roland Perko, Bernt Schiele, et Ales Leonardis. How computer vision can help in outdoor positionning. Dans *Ambient Intelligence, Darmstadt, Germany, November 7-10*, pages 124–141, 2007.

[11] Björn Johansson et Roberto Cipolla. A system for automatic pose-estimation from a single image in a city scene. Dans *Signal Processing, Pattern Recognition, and Applications, Crete, Greece, June 25-28*, 2002.

[12] Jong Weon Lee, Suya You, et Ulrich Neumann. Tracking with omni-directional vision for outdoor ar systems. Dans *ISMAR, Darmstadt, Germany, September 30-October 1*, pages 47–58, 2002.

[13] Duncan Robertson et Roberto Cipolla. An image-based system for urban navigation. Dans *BMVC, London, UK, September 7-9*, 2004.

[14] Reinhold Behringer, Jun Park, et Venkataraman Sundareswaran. Model-based visual tracking for outdoor augmented reality applications. Dans *ISMAR, Darmstadt, Germany, September 30-October 1*, pages 277–278, 2002.

[15] Bolan Jiang, Ulrich Neumann, et Suya You. A robust hybrid tracking system for outdoor augmented reality. Dans *IEEE Virtual Reality, Chicago, Illinois, USA, March 27-31*, pages 3–10, 2004.

[16] Miguel Ribo, Peter Lang, Harald Ganster, Markus Brandner, Christoph Stock, et Axel Pinz. Hybrid tracking for outdoor augmented reality applications. *IEEE Computer Graphics and Applications*, 22(6) :54–63, November 2002.

[17] Philip David, Daniel DeMenthon, Ramani Duraiswami, et Hanan Samet. Softposit : Simultaneous pose and correspondence determination. *IJCV*, 59(3) :259–284, September 2004.

[18] Norbert Haala et Jan Böhm. A multi-sensor system for positioning in urban environments. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58 :31–42, January 2003.

[19] Gaël Sourimant, Luce Morin, et Kadi Bouatouch. Gps, gis and video fusion for urban modeling. Dans *Computer Graphics International, Petropolis, Brazil, May 30-June 2*, pages 79–88, 2007.

[20] Duncan. P. Robertson et Roberto Cipolla. Building architectural models from many views using map constraints. Dans *7th European Conference on Computer Vision, Part II, Copenhagen, Denmark, May 28-31*, 2002.

[21] Gehrard Reitmayr et Tom W. Drummond. Going out : Robust model-based tracking for outdoor augmented reality. Dans *ISMAR, Santa Barbara, California, USA, October 22-25*, pages 109–118, 2006.

[22] Gehrard Reitmayr et Tom W. Drummond. Initialisation for visual tracking in urban environments. Dans *ISMAR, Nara, Japan, November 13-16*, pages 1–9, 2007.

[23] Cindy Cappelle, Maan El Badaoui El Najjar, Denis Pomorski, et Francois Charpillet. Localisation in urban environment using GPS and INS aided by monocular vision system and 3D geographical model. Dans *IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, June 13-15*, pages 811 – 816, 2002.

[24] Volker Coors, Tassilo Huch, et Ursula Kretschmer. Matching buildings : Pose estimation in an urban environment. Dans *IEEE and ACM International Symposium on Augmented Reality, Munich, Germany, October 5-6*, pages 89–92, 2000.

[25] Mahzad Kalantari, Franck Jung, Jean-Pierre Guédon, et Nicolas Paparoditis.. Détection automatique des points de fuite et calcul de leur incertitude à l'aide de la géometrie projective. Dans *RFIA, Amiens, France, January 22-25*, 2008.

[26] Jana Kosecka et Wei Zhang. Video compass. Dans *7th European Conference on Computer Vision, Part IV, Copenhagen, Denmark, May 28-31*, pages 476–490, 2002.

[27] Nicolas Bioret, Guillaume Moreau, et Myriam Servières. Localization approach based on image/GIS correspondence using a simple 2d building layer. Dans *REGARD, Québec, Canada, August 28-29*, 2008.

[28] B. Micusik, H. Wildenauer, et J. Kosecka. Detection and matching of rectilinear structures. Dans *IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, June 24-26*, 2008.