

DATA ADJUSTMENT OF THE GEOGRAPHIC INFORMATION SYSTEM, GPS AND IMAGE TO CONSTRUCT A VIRTUAL REALITY

Adel FRIDHI¹, Benzarti FAOUZI¹, Amiri HAMID¹

DOI: 10.21163/GT_2017.121.04

ABSTRACT:

Visualization of spatial data in a virtual reality environment is catching the imagination of researchers across the disciplines of spatial information sciences and computer vision alike. Different approaches have been investigated into for visualization of three-dimensional objects in a virtual reality environment. An important aspect of creation of virtual world is the increasing interest shown by the research community in building virtual world database with contents of real world object. Geographic Information System, GPS and image techniques have become an important link in the process of creation of virtual worlds. In this paper we present a system for computing georeferenced positions and orientations of images of buildings. Providing such information is a mandatory step to well conditioned large scale and precise 3D reconstruction of urban areas. Our method is based on the registration of multimodal datasets, GPS measures, geolocation, and rough 3D models to create a virtual reality. The 3D modeling of an environment has long been a widely studied topic. Its importance stems from the various applications of such a modeling: virtual reality, extended reality, etc. The difficulties are: the creation of a geolocated virtual reality of an environment via Google Earth, a large-scale data acquisition, a large-scale data processing.

Key-words: Modeling, 3D, GPS, GIS, Geolocation, Virtual reality.

1. INTRODUCTION

Computation of realistic 3D models of urban environments has numerous applications in both virtual reality. A first example is the tremendous demand for high-quality 3D building models access in virtual reality applications, such as Google Earth. Unfortunately, in these kinds of virtual tours, the photo-realistic aspect of the buildings is not always obvious. Some are textured automatically using aerial photographs, leading mainly in poor resolution or misaligned images, while others (the most famous ones) are modeled manually. Another application is the use of 3D building models within forthcoming location-based systems, such as global positioning system (GPS) navigation, where one would like to be able to add additional virtual information online and not only to offline constructed maps. In both these complementary cases, the key point is the fusion—or registration—between real images and synthetic 3D models. It is the mandatory step if one wants either to enhance existing rough 3D models with ground-based images or to provide additional virtual information to urban videos.

The endeavor of this article is part of a perspective of a definition of future products, an implementation of acquisition means and processing methods to enable the field operator (e.g. laboratory LRSITI Image Signal and Information Processing at The National School of Engineers of Tunis) to visualize and interact using georeferenced data that consist in applying a system of coordinates on images and scaling them in a given system.

¹ *National Engineering School of Tunis, 1002 Tunis, Tunisia, fridhi2013@gmail.com, benzartif@yahoo.fr, hamidlamiri@gmail.com.*

We present a system for calculating positions and orientations of geo-referenced images of a location. Our approach is based on data in accordance with the real pictures taken during the construction of a virtual reality of our location and acquired with GPS measurements and a database type SIG which consists of several 3D models of virtual reality. If this type of modeling is relevant to an aerial view, it is not satisfactory for navigation in our 3D environment at the ground level. The real images used in our virtual tour and virtual reality and the basis of SIG contain information: the images contain geometric details of our virtual reality, while the SIG models give a complete geometry of the scene. GPS measurements are acquired in synchronization with the images. After having combined these different data types, first they must be put in correspondence in the same coordinate system. (Numerous 3D model information of a georeferenced environment exist in alphanumeric formats into spreadsheets or databases in meters, degrees, decimal degrees or degrees, one can use a GPS to record the locations. The expression of spatial georeferenced data depends on a projection system (UTM or Lambert), the coordinate system and spatial reference). We present, in this article, some methods to convert these GPS data, SIG and Image related to our virtual reality in the new system.

The article is structured as follows: a section that presents the work on the construction of a virtual reality. Then, the used data are described in section three and a method of their adjusting to be realized in Section 4 and some results to be introduced in Section 5. Finally, we will give some perspectives. Starting with a part of the system and an algorithm.

Our image-based model consists of a panoramic mosaic image to the data acquired. The algorithm building the model are:

- (1) Acquisition of data
 - #Acquisition image#
 - # Acquisition GIS model#
- (2) Registration data
 - # Range data filtering#
 - #Range data representation#
 - #Global registration (image to image warp)#
- (3) Creating cylindrical and spherical panorama
 - #Image Refinement#
 - #Linking and detection#
 - #Triangulation Constraint#
 - #Segments in panoramic 3D line parameter#
- (4) Data adjustment
 - #Used data: GIS, GPS#
- (5) Image adjustment
 - # Registration data to image & video#
- (6) Display

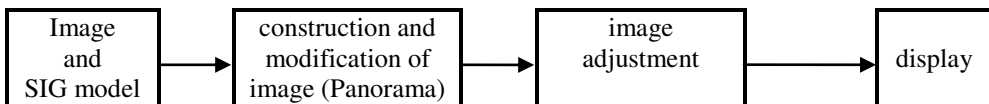


Fig. 1. Part of the system.

1.1 Motivations

This paper explores the two sides of the registration process between images and synthetic building models. The input data consist in a synthetic model of geo-referenced buildings (extracted from a geographic information system (GIS), storing for each building its footprint and elevation, together with images acquired at ground level within urban areas. These images are associated with GPS measures providing the approximate position of the camera in the geo-referenced coordinate frame. The key problem we wish to solve can be formulated in this single question: How can we accurately register 3D models to videos in urban areas, while having poor precision on each input data type? As a matter of fact, real datasets used in most envisioned applications are not necessarily precisely modeled, calibrated, or localized. For that reason, we wish to keep from end to-end unconstrained datasets, contrary to the majority of existing approaches on this subject. The uncertainties are here of different natures. The GIS building models have been estimated using aerial photographs. Their facades are thus approximated using simple planes, leaving the micro-structures (windows, doors, etc.) unmodeled. Second, GPS measures in urban areas usually suffer from poor accuracy, mainly due to multiple signal reflections on walls and narrow satellite coverage. Last, if we imagine a collaborative application where hundreds of users could collaborate to acquire image data and enhance building models, we cannot constrain the camera to be calibrated (in terms of its internal parameters).

1.2 Contributions

This paper presents a set of methods that allow to register 3D building models to images and also to enhance these models by synthesizing photo-realistic textures of their facades using ground level imagery. All these methods have been designed to take into account the fact that the input datasets have poor accuracy. The major contributions of this paper are the following:

1. The first contribution of this paper is the registration initialization, that is, the computation of the camera poses (position and orientation) in a georeferenced coordinate system for the first image of a panorama. Contrary to most of existing similar systems, no orientation cue is required: Only a rough estimation of the position is used as an input. Moreover, semantic information extraction is performed on the images to build correct sets of matches between images and 3D models and thus compute the initial pose of the camera in a robust way.

2. The second contribution is the exploitation of the proposed registration to enhance existing buildings 3D models. We developed a new robust hybrid texture fusion algorithm for facades textures synthesis that summarizes existing methods, which do not generally aim at solving each problem related to this concern. Pose tracking, that is, computing the pose of the camera for all images, has already been presented in (Sourimant et al, 2007). As a matter of fact, we will only remind it briefly in this paper. This pose tracking scheme is robust to partial visibility of the buildings and to occluding objects, even considering the coarse geometry of the models.

2. SYSTEM OVERVIEW AND PROPOSED APPROACH

The datasets registration principle is outlined on **Fig. 1**. The step number one of our framework consists to using GPS data together with the GIS database so as to find a first approximation of the camera localization. Rough camera position and orientation are

therefore associated with each picture of the sequence. The step number two consists in relating images and 3D model so as to get in output accurate poses of the sequence camera, for each picture in the video. The digital camera pose being initialized with the positions given by the GPS, the projection of the model is registered with the images by modifying the position and orientation of the virtual camera.

The scientific issue related to georeferenced virtual reality (You have an aerial photograph or a satellite image of a region, the coordinates (X, Y) and a virtual reality but no geographical reference) has as objective to: create a virtual reality and make the coupling between the latter and the topographic map.

The looming up of the concept of virtual reality illustrates the dynamism of interdisciplinary dialogues between computer graphics, computer-aided design, simulation, remote operation and audiovisual (Beer & Guez, 2013).

With this definition, it operates a change of viewpoint that the study of the acquired aerial and satellite images of the ground proceeds to the study of the virtual objects that are displayed. The advances in the field of construction of a 3D model made the possibility of the establishment of several more complex virtual reality applications intended for georeferenced environments.

2.1. Aerial Setting up

The works realized on the 3D modeling of an environment generally recourse to the use of a set of aerial or satellite images (panorama) **Fig. 2**. Two steps are of important: an environment is detected using a segmentation or extraction basis **Fig. 3** and a connection of different lines (Tupin & Roux, 2003). Then virtual reality is effectively reconstructed using a mapping between the set of images (panorama) and 3D combined primitives (Horna et al., 2009). These methods generally provide 3D models that are not rich geometrically, even if one has a good estimation of the overall shape (dimensions, angles), no data on the texture or small geometric details of the facades can be extracted **Fig. 3**.

Table 1.
Points of geographic coordinate on Google Earth.

P	X	Y
01	602256	4076725
02	602325	4076758
03	602342	4076752
04	602325	4076734
05	602348	4076694
06	602296	4076620
07	602288	4076652
08	602261	4076666
09	602261	4076684
10	602271	4076692
11	602273	4076695
12	602279	4076680
13	602285	4076689
14	602387	4076772
15	602313	4076711



Fig. 2. Area of study (15 point of control) *Data source: LRSITI laboratory and its environment.*

2.2. Ground Setting up

The implementation on the ground refers to 3D modeling using data acquired from our work environment (still images). This type of approach allows the extraction of geometric and photometric data.

The methods for modeling 3D scenes from the panorama have been well studied by computer either the structure via motion that allows to find calibration and setting up of the camera as well as the scene structure (Hartley & Zisserman, 2004), or in the modeling based on the set of images (panorama) that allows to create the virtual reality (e.g. **Fig. 3**) that was used to model our laboratory LRSITI of the National School of Engineers of Tunis. Various works opt for a geometric setting up (Fraundorfer et al, 2006) while other projects target a large-scale setting up of our work environment from images of the ground. Cylindrical calibrated images are used to extract corresponding plans, which are textured and geometrically refined using software and hardware material for pattern recognition and computer vision. This project is designed to create 3D models (virtual reality) varying with time via a set of images taken from many different places.

The serious issue with these approaches is that they are sometimes internal or they require complex acquisitions and processing times that can become difficult during the large-scale setting up of virtual reality.

3. CREATING CYLINDRICAL AND SPHERICAL PANORAMA

We use similar machinery as for spherical panoramas. Within the shared, the θ and ψ viewport coordinates are interpreted as yaw (ψ) and pitch (θ) in the range $[-\pi, +\pi]$ and $[-\pi/2, +\pi/2]$. The desired image ray described by the unit vector \hat{r} ,

$$\hat{r} = (\cos \theta \cos \psi, \sin \theta, \cos \theta \sin \psi) \quad (1)$$

This is transferred into the frame of reference of the key frame using the virtual camera to key frame rotation matrix, R_{kc} , which is uploaded as a parameter to the shared. Finally, the camera intrinsic matrix can be used to map this to key frame image-space coordinates. Given this correspondence, we proceed as with the spherical panorama.

A sequence of images is taken on a tripod with a high resolution digital camera. **Fig. 3** shows two overlapping cylindrical images, Notice how horizontal lines become curved. We plot the coordinates $p=(X; Y; Z)$ to 2D cylindrical screen coordinates $(\theta; \psi)$ using [6].

$$\theta = \tan^{-1}\left(\frac{X}{Z}\right), \psi = Y/\sqrt{X^2 + Z^2} \quad (2)$$

Where θ is the angle of all the pictures by which we build our panorama so we can map into 2D spherical coordinates $(\theta; \psi)$ using

$$\theta = \tan^{-1}\left(\frac{X}{Z}\right), \psi = Y/\sqrt{X^2 + Z^2} \quad (3)$$

After having distorted each input image, all photos that built our panorama have become a very big problem for mere understanding. Ideally, to build a cylindrical or spherical panorama from a horizontal panning sequence, the unknown panning angles need to be recovered. In practice, small vertical translations are needed to compensate for vertical jitter and optical twist (Shum & Szeliski, 2002). In fact, tiny vertical translations are required to balance the vertical jitter and optical twist. Actually in a simultaneous tx horizontal and the vertical translation is estimated for each input photo.

To minimize discontinuities in intensity and color between the pictures' laying, we apply a layout algorithm of clear flag, that means to shade pixels on each photo in proportion to their distance from the edge (and thus their distance is invisible to the nearest pixel) (Szeliski, 1996). Creating panoramas in cylindrical or spherical coordinates has several limits. Once registration is finished, we can clip the ends (and optionally the top and bottom), and write out a single panoramic photo.

At first, we only can handle the facile case of panning motion. Secondly, it is possible to convert a photo at 2D spherical or cylindrical coordinates for a tilting angle, as ill-sampling causes big saving errors, and thirdly, it requires to knowing the focal lengths. As the latter can be calibrated (Stein, 1995), therefore estimating the focal length of a lens by saving many photos is not very accurate.

In **Fig. 3** construction of a cylindrical panorama: composited from a sequence of photos. The method we use to stitch the frames is by minimizing the gradient error between the frames, and uses the equation,

$$E(\Delta x, \Delta y) = \sum_{xy} [G_1(x' + \Delta x, y' + \Delta y) - G_0(x, y)]^2 \quad (4)$$

Where \mathbf{G} is gradient operator, $(\Delta x, \Delta y)$ is the translation which is the same for all pixels, $(x'+\Delta x, y'+\Delta y)$ and (x,y) are corresponding points in two images, and \mathbf{E} is error term. Lowe for aligning images several approaches exist for different types of cases.

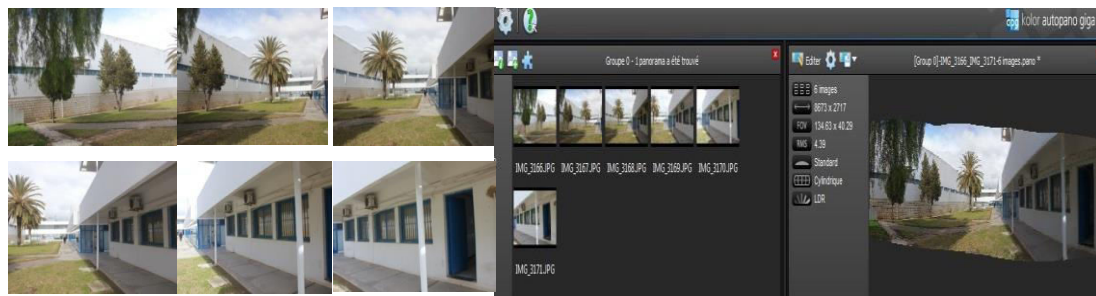


Fig. 3. Shows a part of cylindrical panorama composed from a sequence of images.

Most of the existing approaches are for panoramic imaging or for creating mosaics photo to work offline. The search for key elements is used in several offline approaches (Wagner et al., 2010). The discussion of a method in which the successive photos of a digital camera’s view finder are aligned online and in real-time (Adams et al, 2008).The viewfinder algorithm is used for tracking the camera’s digital motion to create a high resolution panoramic photography which requires an offline processing. The approach described in this (Xiong & Pulli, 2010).

3.1. Image Refinement

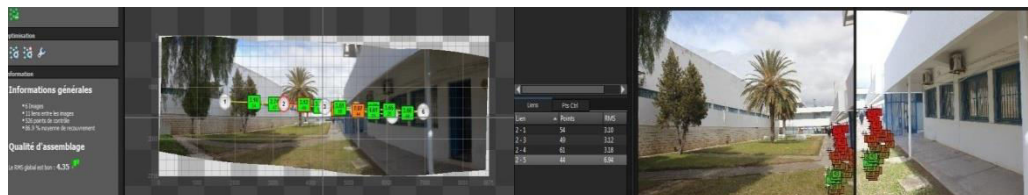


Fig. 4 Homogenous areas due to diverging (Moving the camera away from the light source again brightens the input photo in an un-proportional way).

Fig. 5 Reducing the brightness with a modified brightness correction version reduces the differences in brightness between former digital camera photos.



Fig. 6 Offset Correction.

A significant problem occurs while taking panoramic photos in real-time which is the changing exposure time of the camera. When moving the camera towards a light source the

exposure is reduced, this significantly darkens the input photo. Moving the camera away from the light source again brightens the input photo in an un-proportional way, **Fig. 4**.

However, using a mapping approach we can directly employ blending effects while the panoramic photo is recorded. No additional photo information has to be saved on the device.

A manual correction the differences in brightness values of the digital camera image we seek to find matching points in **Fig. 5**, in the panoramic photo and the digital camera image then calculating their brightness difference from the color data is thus forwarded and considered in the mapping process.

We calculate the offset for the feature points found by the tracker **Fig. 6**. The advantage of this approach is that it can be performed at almost no additional computational overhead, since the tracker provider and the actual pixel values are comparable. The virtual environments are digitally rendered spaces that blur the distinction between reality and its virtual representation (Hoyt et al, 2003).

As mentioned earlier, we used real pictures for our project which we will later change in **Fig. 3**. The observer feels totally immersed in a virtual world and each object can be seen around with a definite angle. The virtual laboratory and the environment are composed of areas of static landscapes. The observer has a definite angle by which he can see the whole scene in a combination of real and virtual, in all directions. Each area can be constructed by a photo with a projection process. After synthesis of all areas, with modifications, a panoramic image is obtained.

3.2 Concluding approaches.

Frueh et al (2005) project two cameras for data acquisition. A camera, placed vertically, measures a dense cloud of points of the study location while the camera is used for texturing. A camera, placed horizontally, used to estimate the position, during the acquisition, is matched with an aerial map of the location contours to ensure overall consistency of the different 3D models. The main drawback of this approach is the complexity of the acquisition system and the complex post-processing of the acquired data, mainly concerning the extremely dense 3D cloud of dots”.

3.3 Proposed Approach

We will solve the problem with an approach type refinement, and starting with the 3D implemented model that is geometrically simple and non-textured, but expressed in a geo-referenced landmark. Then, we add geometric and photometric data from images (panorama) using algorithms derived from the vision. Unlike solutions based on aerial or satellite data, which give only a raw structure of the location, our system starts with these models and aims to refine using information extracted from the images of that space. Finally, our approach targets a large-scale setting up using both air and ground information with recourse to a camera and a GPS.

4. DATA ADJUSTMENT

4.1 Used data

In this section we use a SIG database that provides the raw 3D models of georeferenced locations, images from which they are extracted for texturing as well as luminance images for extracting and monitoring points. This database also supplies GPS

measures that are registered with the flow of images and that provide a proxy for the geolocation of different images.

4.1.1 GIS data

GIS data are designed to hold layers of any type of geographic information (hydrographic maps, road maps, etc.). We use a specific layer defining simple 3D models of buildings. These are specified in the UTM geo-referenced frame, using closed 2D point lists defining the buildings footprint and their altitude above sea level, together with the buildings height (both expressed in meters). These permits to generate simple polyhedral models representing large urban areas. Several limitations arise from such a representation. First, buildings facades are defined as an extrusion of their footprint. As such they are perfectly planar and do not model geometric details such as doors or windows. Moreover, curved buildings are often badly modeled due to lack of points to describe the corresponding footprint. Second, holes-like topologies, such as arches, are not modeled. Last, the footprint being expressed in the horizontal UTM frame, buildings constructed on non globally planar ground suffer from erroneous footprints.

4.1.2 Videos and associated camera

Since one of the targeted applications is collaborative modeling, the registration framework should not be restricted to specific user acquisition devices. As a consequence, the less hypotheses are made on the camera, the better. The classical pinhole projective model is chosen, and the basic camera internal parameters are supposed to be available: focal of the lens (in meters), image dimensions (in pixels), and charge-coupled device captor dimensions (in meters). This information is sufficient to get an approximation of the camera focal lengths f_x and f_y . The central point (u_0, v_0) is arbitrarily set at the image center, since errors on the central point estimation have a negligible influence in our context (Gomez et al, 2005).

The first step is an initial adjustment of the camera, using the GPS data that provides an estimation of the position and orientation of the camera in the geo-referenced SIG landmark for each image. The second step consists in the refinement of this estimation using image data. For this, original images are put in correspondence with 3D models of SIG. We then search for the camera parameters that allow the realization of the superposition of the SIG model when projected on the images of virtual reality.

4.1.3 GPS Initialization

The acquisition of GPS data allows a geographical position expressed in a geographical landmark. "This position is first converted into coordinates (X, Y) in the geo-referenced UTM system or Lambert (You have an aerial photograph or a satellite image of a region, the coordinates (X, Y) of SIG by equations from (Snyder, 1987). The obtained linear coordinates (X, Y) provide the horizontal position of the camera for every moment corresponding to an image of the virtual reality. The SIG data does not contain the ground surface. The latter is modeled by Delaunay triangulation on the heights of the ground location. We obtain, accordingly, for each image in the sequence, a position $pt = (X_t, Y_t)$. Finally, the camera's orientation is also estimated via the position data provided by the GPS, with the assumption that the axis of sight is parallel to the trajectory of the camera at the time t , the camera orientation is given by $(p_1(t+1) - p_t)$.

4.1.4 GPS measures

GPS measures are acquired jointly to videos so as to localize images with respect to the 3D model. However, GPS does not provide an exact position location. Inaccuracies arise mainly from occluding environment in urban areas, where less satellite is visible and where the signal can be delayed with multiple reflections on buildings. To estimate this imprecision, we measured fixed GPS positions (i.e., without moving the device) either with clear view or occluding environment 1 during 30 min. These measures are summarized in **Table 1**. Position accuracy at ground level is measured as the ellipse area which axes are the position standard deviations σ_X and σ_Y . We observe two facts from these results. First, we can expect ground precision to go from less than a meter to a little more than 10 m. Second, vertical precision is much poorer than horizontal one. As a consequence, one should not take vertical measures into account if an alternative is available. Because GPS measures are considered in our framework as approximate initial camera positions, we do not synchronize them with the video frames with hardware tools. The two devices are stopped at the (approximate) same time so that GPS and video timings can be roughly aligned.

5. IMAGE AND SIG ADJUSTMENT

The trajectory obtained by GPS is then refined using the image data. The adjustment between image data and SIG model consists in determining the camera parameters that allow the superposition of SIG model with images of the actual locations. For the first image of the virtual reality **Fig. 7**, the adjustment is carried out using a semi-automatic procedure. For the following images, the adjustment is made automatically by extracting and monitoring points of interest and virtual visual control.

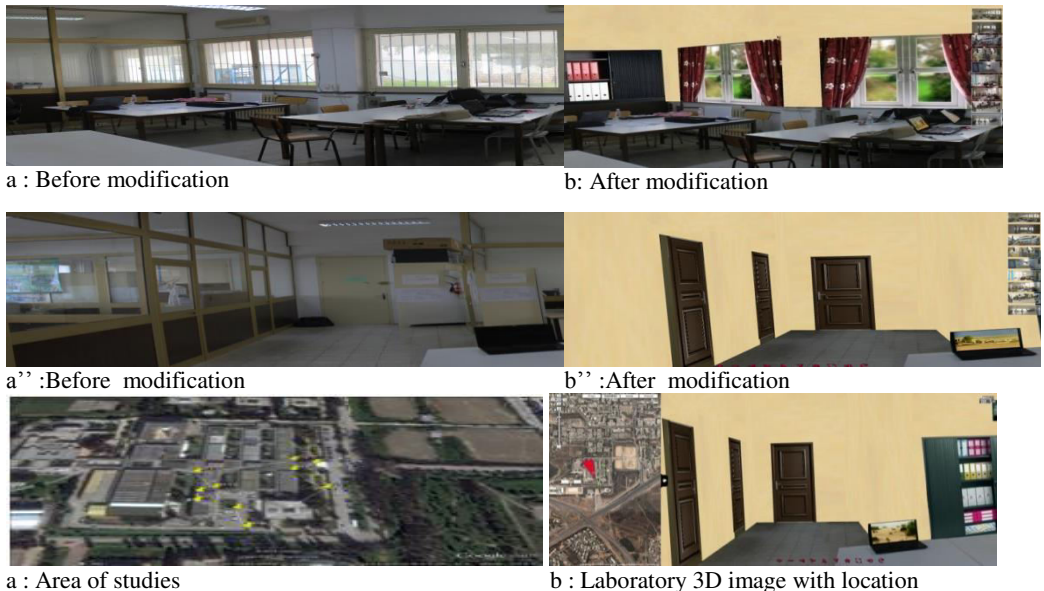


Fig. 7 Sky view of the location of the laboratory LRSITI Before and after modification.

5.1. Theoretical foundations

A coordinate of the 3D point \mathbf{P} is projected in the image with homogeneous coordinates 2D point \mathbf{p} given by:

This section exposes the issue of calculating an observation point on a simple example. To illustrate the problem, we considered the scene model as uniquely composed of points. The problem of calculating the pose, however, generalizes with any type of primitives (points, lines, contours, etc.). From a formal observation point, we consider a camera with perspective projection model, A point $\mathbf{P}=(X, Y, Z, 1)^T$ in which coordinates are expressed in point of reference of the scene \mathbf{R}_0 is thus projected on the plan image $\mathbf{p} = (x, y, 1)^T$ with:

$$\mathbf{p} = {}^c\mathbf{M}_0 \mathbf{P} \mathbf{K}$$

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \alpha \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} {}^c\mathbf{R}_0 & {}^c\mathbf{T}_0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (5)$$

$$\text{With } \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \mathbf{K} \text{ and } \begin{bmatrix} {}^c\mathbf{R}_0 & {}^c\mathbf{T}_0 \\ 0 & 1 \end{bmatrix} = {}^c\mathbf{M}_0 \quad (6)$$

or ${}^c\mathbf{T}_0$ and ${}^c\mathbf{R}_0$ is the translation vector and the rotation matrix defining the camera position relative to the scene (${}^c\mathbf{R}_0$ being a rotation matrix must obviously respect the orthogonality properties) and \mathbf{K} is the perspective projection matrix. The objective of the pose estimation, in the case of primitive type point, is to estimate the relative pose ${}^c\mathbf{M}_0$ knowing the 3D model, ${}^0\mathbf{P}$, and the measures in the image \mathbf{p} . This is in itself a non-linear problem that amounts to minimize the error of adjustment between the image measures \mathbf{p}_i and the projection model $\mathbf{K} {}^c\mathbf{M}_0 {}^c\mathbf{P}_i$

$$\hat{{}^c\mathbf{M}_0} = \operatorname{argmin} \Delta \text{ with } \Delta = \sum_{i=1}^N (\mathbf{p}_i \mathbf{K} {}^c\mathbf{M}_0 {}^c\mathbf{P}_i) \quad (7)$$

where N is the number of points considered in the model.

6. RESULTS OF ADJUSTMENT

The localization error is expressed in coordinates (x and y). Therefore, we studied the difference in x and in y between Google Earth coordinates and those of the reference. In this case, we have recourse to GPS coordinates taken in real time (RTK) with a precision of no more than 1.5 cm. The geostatistical processing is done by means of the R software.

These errors vary between -9 m and 7 m to the X axis and between -7 m and 7 m to the Y axis.

Table 2.

Summary (error).

dx	dy
Min. :-9.00	Min. :-7.00
Median : 3.50	Median : 3.00
Mean : 0.90	Mean : 0.60
Max. : 7.00	Max. : 6.00
Count : 15	count : 15

Dx : The standard deviation and Dy : the error deviation value in relation to the average value represented by the red line on the graph:

> a = variance (dx)

> a

[1] 34.32222

> b=sqrt(a)

> b

[1] 5.858517

> a=variance (dy)

> a

[1] 28.26667

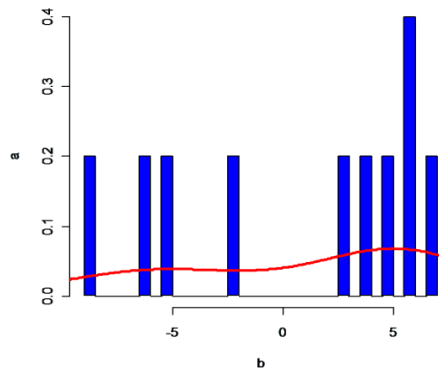
> b=sqrt(a)

> b

[1] 5.316641

The standard deviation of dx is 5.858517 and 5.316641 of dy, this value expresses the error deviation value in relation to the average value represented by the red line on the graph (for dx = 0.90 and dy= 0.60).

Dx



Dy

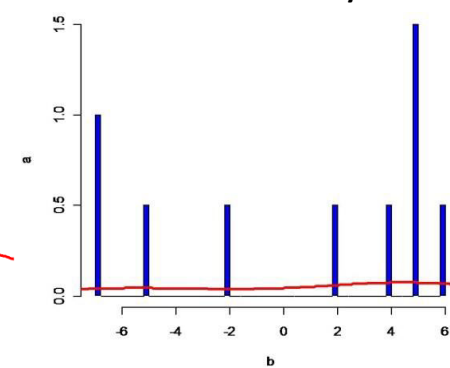


Fig. 8. Dx: probability of occurrence.

Histogram of data Dx and Dy: probability of occurrence: Histogram of data Dy. These Histogram show the probability of error for dx and dy. The DX Histogram shows that the highest probability of error is between 5 and 6 m, whereas for the error DY, it is essentially limited to a value of between -6 m and -7 m and more to 5 m.

The QQ plot is a correlation between the theoretical distribution of errors and the distribution of values studied. The distribution shows a distribution that follows the normal line with some outliers for both graphics with better distribution for the X axis.

In **Fig. 8** we can see two textures extracted from a sequence of images for which the adjustment is known. We can see that the blackout objects are removed, the image resolution is preserved, and that secularities in windows are also absent from the final texture. And later we reached the same conclusion for the actual images **Fig. 5**.

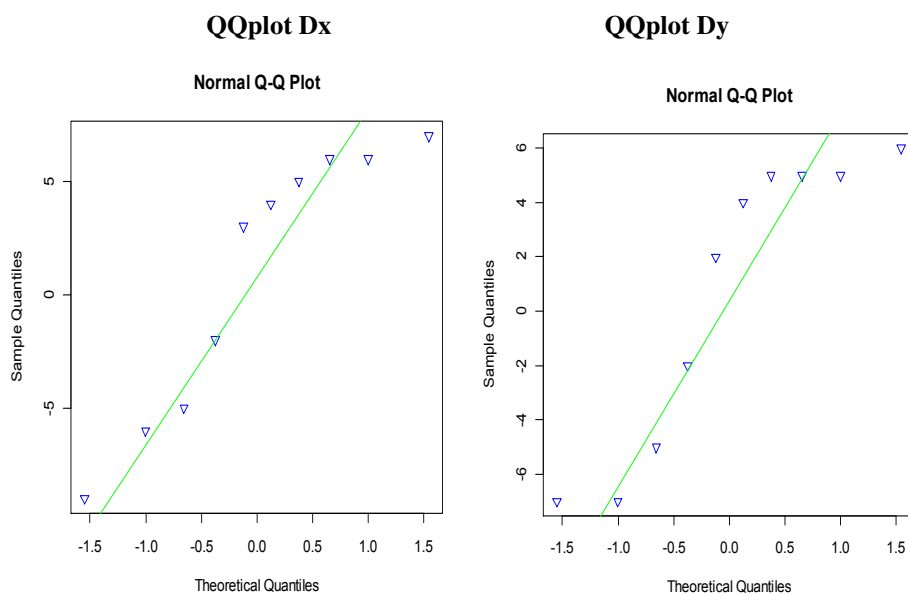


Fig. 9. Correlation between the theoretical distribution of errors and the distribution of values studied.

A completely virtual view calculating using a single panorama is presented in **Fig. 9**.

In our tests, we could not determine which method the area between measuring distance or angle was best for keeping the spatial resolution.

We recommend the first for simplicity and efficiency of implementation.



Fig. 10. Menu with various search parameters.

7. CONCLUSION

We presented a methodology for registering multimodal data, as a mandatory step to large-scale city modeling, by interpreting GPS measures with regards to a GIS database to get a coarse estimation of the camera pose, and then by refining these estimates using suitable visual virtual serving algorithms. We have then computed georeferenced poses of the camera, which provide us with useful information for future geometric refinement of the GIS 3D models, using directly the registered image sequences. However, there is still room for improvement for this method.

First, we would want to suppress the manual part of the pose initialization process, by developing a fully automatic procedure to perform this computation. Moreover we could use such automatic procedure to reduce drift introduced during the tracking phase. Such a procedure is currently studied. In the near future, we plan to take advantage of this method by using the images registered with the GIS database to enhance the coarse polyhedral 3D models, and more precisely compute their local geometric details and real texture information.

Throughout this paper, we have presented a methodology for creating a virtual reality to match different types of information for the large-scale construction of a location (laboratory LRSITI) by interpreting GPS measurements in relation to a SIG database. This is to give an approximation of the position of the camera after acquiring data on the ground. We then calculate a georeferenced orientation for each image of virtual reality.

We also show the way such an adjustment may be used to automatically calculate the textures of virtual reality from images acquired from a camera. In a more specific framework, where we have greater confidence in the accuracy of GPS measurements, we can get rid of the constraint of supervision in order to get a fully automatic adjustment method of SIG and image.

REFERENCES

- Adams, A., Gelfand, N. & Pulli, K. (2008) Viewfinder alignment. *Computer Graphics Forum*, 27 (2), 597–606.
- Beer, S. & Guez, J. (2013) Ideas about VR&AR as a new genre in fine arts'. *VRIC '13 Proceedings of the Virtual Reality International Conference: Laval Virtual*, Article 16, New York.
- Fraundorfer, F., Schindler, K. & Bischof, H. (2006) Piecewise planar scene reconstruction from sparse correspondences. *Image Vision Comput.*, 24 (4), 395–406.
- Frueh, C., Jain, S. & Zakhor, A. (2005) Data processing algorithms for generating textured 3D building facade meshes from laser scans and camera images. *International Journal of Computer Vision*, 61, 159–184.
- Gomez, J.-F.V., Simon, G. & Berger, M.-O. (2005) Calibration errors in augmented reality: A practical study. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 154–163.
- Hartley, R. & Zisserman, A. (2004) *Multiple view geometry in computer vision*, Cambridge University Press: Cambridge, UK.
- Horna, S., Meneveaux, D., Damiand, G. & Bertrand, Y (2009) Consistency constraints and 3D building reconstruction. *Computer-Aided Design*, 41(1), 13-27.
- Hoyt, C. L, Blascovich, J & Swinth, K. R. (2003) Social inhibition in immersive virtual environments. *Presence Teleoperators Virtual Environ*, 12, 183–195.
- Shum, H-Y. & Szeliski, R. (2002) Construction of panoramic image mosaics with global and local alignment. *Int. Journal of Computer Vision*, 48(2), 151-152.
- Snyder, J. P. (1987). *Map projections—a working manual*. U.S. geological survey professional paper 1395. Washington, DC: United States Government Printing Office.
- Sourimant, G., Morin, L. & Bouatouch K. (2007) GPS, GIS, and video registration for building reconstruction. In *ICIP 2007, 14th IEEE International Conference on Image Processing*, San Antonio, USA.
- Stein, G. (1995) Accurate internal camera calibration using rotation, with analysis of sources of error. In *Proc. 5th International Conference on Computer Vision*, Cambridge, Massachusetts, 230–236.
- Szeliski, R (1996) Video mosaics for virtual environments. *IEEE Computer Graphics and Applications*, 22–30.
- Tupin, F. & Roux, M. (2003) Detection of building outlines based on the fusion of SAR and optical features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58, (1), 71–82.
- Wagner, D., Mulloni, A., Langlotz, T. & Schmalstieg, D. (2010) Realtime panoramic mapping and tracking on mobile phones. In *Proceedings of the IEEE virtual reality conference (VR'10)*, Washington, DC, USA, 211-218.
- Xiong, Y. & Pulli, K (2010) Color and luminance compensation for mobile panorama construction. In *Proceedings of the 18th ACM international conference on Multimedia (MM '10)*, New York, NY, USA, 261-270.