

Camera and laser robust integration in engineering and architecture applications

Pablo Rodriguez-Gonzalvez,
Diego Gonzalez-Aguilera and Javier Gomez-Lahoz
*Department of Cartographic and Land Engineering
High Polytechnic School of Avila, Spain
University of Salamanca*

1. Introduction

1.1 Motivation

The 3D modelling of objects and complex scenes constitutes a field of multi-disciplinary research full of challenges and difficulties, ranging from the accuracy and reliability of the geometry, the radiometric quality of the results up to the portability and cost of the products, without forgetting the aim of automatization of the whole procedure. To this end, a wide variety of passive and active sensors are available of which the digital cameras and the scanner laser play the main role. Even though these two types of sensors can work in a separate fashion, it is when are merged together when the best results are attained. The following table (Table 1) gives an overview of the advantages and limitations of each technology.

The comparison between the laser scanner and the digital camera (Table 1) stresses the incomplete character of the information derived from only one sensor. Therefore, we reach the conclusion that an integration of data sources and sensors must be achieved to improve the quality of procedures and results. Nevertheless, this sensor fusion poses a wide range of difficulties, derived not only from the different nature of the data (2D images and 3D scanner point clouds) but also from the different processing techniques related to the properties of each sensor. In this sense, an original sensor fusion approach is proposed and applied to the architecture and archaeology. This approach aims at achieving a high automatization level and provides high quality results all at once.

Scanner laser	Digital camera
↓ Not accurate extraction of lines	↑ High accuracy in the extraction of lines
↓ Not visible junctions	↑ Visible junctions
↓ Colour information available on low resolution.	↑ Colour information on high resolution
↑ Straightforward access to metric information	↓ Awkward and slow access to metric information
↑ High capacity and automatization in data capture	↓ Less capacity and automatization in data capture
↓ Data capture not immediate. Delays between scanning stations. Difficulties to move the equipment.	↑ Flexibility and swiftness while handling the equipment.
↑ Ability to render complex and irregular surfaces.	↓ Limitations in the renderization of complex and irregular surfaces
↓ High cost (60.000€-90.000€.)	↑ Low cost (From 100€)
↑ Not dependent on lighting conditions.	↓ Lighting conditions are demanding.
↓ 3D model is a "cloud" without structure and topology.	↑ The 3D model is accessed as a structured entity, including topology if desired.

Table 1. Comparison of advantages and drawbacks of laser scanner and digital camera.

1.2 State of the art

The sensor fusion, in particular concerning the laser scanner and the digital camera, appears as a promising possibility to improve the data acquisition and the geometric and radiometric processing of these data. According to Mitka (2009), the sensor fusion may be divided in two general approaches:

- *On-site integration*, that resides on a "physical" fusion of both sensors. This approach consists on a specific hardware structure that is previously calibrated. This solution provides a higher automatization and readiness in the data acquisition procedures but a higher dependency and a lack of flexibility in both the data acquisition and its processing. Examples of this kind of fusion are the commercial solutions of Trimble and Leica. Both are equipped with digital cameras that are housed in the inside of the device. These cameras exhibit a very poor resolution. (<1Mp). With the idea of accessing cameras with higher quality, other manufacturers present an exterior and calibrated frame to which a reflex camera can be attached. Faro Photon, Riegl LMS-Z620, Leica HDS6100 and Optech Iris-3D are some of the laser systems that have incorporated this external sensor.

Even though these approaches may lead to the idea that the sensor fusion is a straightforward question, the actual praxis is rather different, since the photos shoot time must be simultaneous to the scanning time, thus the illumination conditions as well as other conditions regarding the position of the camera or the environment may be far from the desired ones.

- *Office integration*, that consists on achieving the sensor fusion on laboratory, as the result of a processing procedure. This approach permits more flexibility in the data acquisition since it will not require neither a previously fixed and rigid framework nor a predetermined time of exposure. Nevertheless, this gain in flexibility demands the challenge of developing an automatic or semi-automatic procedure that aims at "tuning" both different data sources with different constructive fundamentals. According to Kong et al., (2007), the sensor fusion can be divided in three categories: the sensorial level (low level), the feature level (intermediate level) and the decision level (high level). In the sensorial level raw data are acquired from diverse sensors. This procedure is already solved for the *onsite integration* case but it is really complicated to afford when sensors are not calibrated to each other. In this sense, the question is to compute the rigid transformation (rotation and translation) that renders the relationship between both sensors, besides the camera model (camera calibration). The feature level merges the extraction and matching of several feature types. The procedure of feature extraction includes issues such as corners, interest points, borders and lines. These are extracted, labeled, located and matched through different algorithms. The decision level implies to take advantage of hybrid products derived from the processed data itself combined with the expert decision taking.

Regarding the two first levels (sensorial and feature), several authors put forward the question of the fusion between the digital camera and the laser scanner through different approaches linked to different working environments. Levoy et al., (2000) in their project "Digital Michelangelo" carry on a camera pre-calibration facing integration with the laser scanner without any user interaction. In a similar context, Rocchini et al. (1999) obtain a fusion between the image and the laser model by means of an interactive selection of corresponding points. Nevertheless, both approaches are only applied to small objects such as sculptures and statues. With the idea of dealing with more complicated situations arising from complex scenes, Stamos and Allen, (2001) present an automatic fusion procedure between the laser model and the camera image. In this case, 3D lines are extracted by means of a segmentation procedure of the point clouds. After this, the 3D lines are matched with the borders extracted from the images. Some geometrical constraints, such as orthogonality and parallelism, that are common in urban scenes, are considered. In this way, this algorithm only works well in urban scenes where these conditions are met. In addition, the user must establish different kinds of thresholds in the segmentation process. All the above methodologies require the previous knowledge of the interior calibration parameters. With the aim of minimizing this drawback, Aguilera and Lahoz (2006) exploit a single view modelling to achieve an automatic fusion between a laser scanner and a not calibrated digital camera. Particularly, the question of the fusion between the two sensors is solved automatically through the search of 2D and 3D correspondences that are supported by the search of two spatial invariants: two distances and an angle. Nevertheless, some suppositions, such as the use of special targets and the presence of some geometric constraints on the image (vanishing points) are required to undertake the problem. More recently, Gonzalez-Aguilera et al. (2009) develop an automatic method to merge the digital image and the laser model by means of correspondences of the range image (laser) and the camera image. The main contribution of this approach resides on the use of a level hierarchy (pyramid) that takes advantage of the robust estimators, as well as of geometric constraints

that ensure a higher accuracy and reliability. The data are processed and tested by means of software called USALign.

Although there are many methodologies that try to progress in the fusion of both sensors taking advantage of the sensorial and the feature level, the radiometric and spectral properties of the sensors has not received enough attention. This issue is critical when the matching concerns images from different parts of the electromagnetic spectrum: visible (digital camera), near infrared (laser scanner) and medium/far infrared (thermal camera) and when the aspiration is to achieve an automatization of the whole procedure. Due to the different ways through which the pixel is formed, some methodologies developed for the visible image processing context may work in an inappropriate way or do not work at all.

On this basis, this chapter on sensor fusion presents a method that has been developed and tested for the fusion of the laser scanner, the digital camera and the thermal camera. The structure of the chapter goes as follows: In the second part, we will tackle with the generalities related with the data acquisition and their pre-processing, concerning to the laser scanner, the digital camera and the thermal camera. In the third part, we will present the specific methodology based on a semi-automatic procedure supported by techniques of close range photogrammetry and computer vision. In the fourth part, a robust registration of sensors based on a spatial resection will be presented. In the fifth part, we will show the experimental results derived from the sensor fusion. A final part will devoted to the main conclusions and the expected future developments.

2. Pre-processing of data

In this section we will expose the treatment of the input data in order to prepare them for the established workflow.

2.1 Data Acquisition

The acquisition protocol has been established with the greatest flexibility as possible in such a way that the method can be applied both to favourable and unfavourable cases. In this sense, the factors that will condition the level of difficulty will be:

- Geometric complexity, directly related to the existence of complex forms as well as to the existence of occlusions.
- Radiometric complexity, directly related to the spectral properties of each sensor, as well as to different illumination conditions of the scene.
- Spatial and angular separation between sensors. The so called baseline will be a major factor when undertaking the correspondence between the sensors. This baseline will also condition the geometric and radiometric factors mentioned above. A short baseline will lead to images with a similar perspective view and consequently, to images easier to match and merge. On the contrary, a large baseline will produce images with big variations in perspective and so, with more difficulties for the correspondence. Nevertheless, rather than the length of the baseline, the critical factor will be the angle between the camera axis and the

average scanning direction. When this angle is large, the automatic fusion procedures will become difficult to undertake.

The following picture (Fig. 1) depicts the three questions mentioned above:

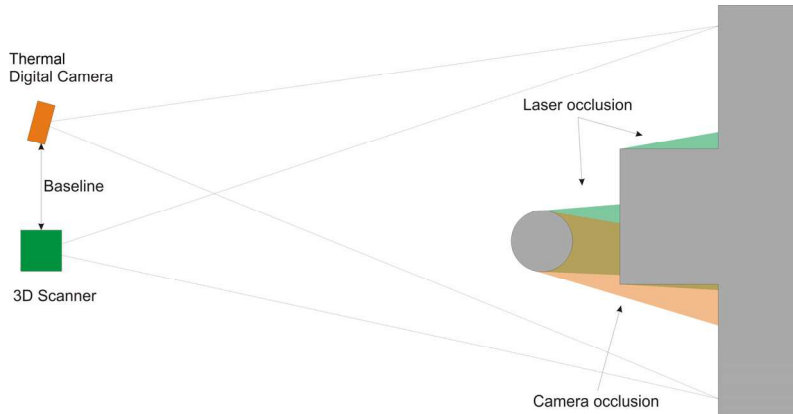


Fig. 1. Factors that influence the data acquisition with the laser scanner and the digital/thermal camera

Through a careful planning of the acquisition framework, taking into account the issues referred before, some rules and basic principles should be stated (Mancera-Taboada et al., 2009). These could be particularised for the case studies analyzed in section 5 focussing on objects related to the architectural and archaeological field. In all of them the input data are the following:

- The **point cloud** is the input data in the case of the laser scanner and exhibits a 3D character with specific metric and radiometric properties. Particularly, the cartesian coordinates XYZ associated to each of the points are accompanied by an intensity value associated to the energy of the return of each of the laser beams. The image that is formed from the point cloud, the range image, has radiometric properties derived from the wavelength of the electromagnetic spectrum, that is, the near or the medium infrared. This image depends on factors such as: the object material, the distance between the laser scanner and the object, the incidence angle between the scanner rays and the surface normal and the illumination of the scene. Also, in some cases, this value can be extended to a visible RGB colour value associated to each of the points.
- The **visible digital** image is the input data coming from the digital camera and presents a 2D character with specific metric and radiometric properties. Firstly, it is important that its geometric resolution must be in agreement with the object size and with the scanning resolution. The ideal would be that the number of elements in the point cloud would be the same that the number of pixels in the image. In this way, a perfect correspondence could be achieved between the image and the

point cloud and we could obtain the maximum performance from both data sets. In addition, for a given field of view for each sensor, we should seek that the whole object could be covered by a single image. As far as this cannot be achieved we should rely on an image mosaic where each image (previously processed) should be registered in an individual fashion. On the other hand, from a radiometric point of view, the images obtained from the digital camera should present a homogeneous illumination, avoiding, as far as it is possible, the high contrasts and any backlighting.

- The **thermal digital** image is the input data coming from the thermal camera and presents a 2D character with specific metric and radiometric properties. From a geometric point of view they are low resolution images with the presence of high radial lens distortion. From a radiometric point of view the values distribution does not depend, as it does in the visible part of the electromagnetic spectrum, on the intensity gradient of the image that comes from part of the energy that is reflected by the object, but from the thermal gradient of the object itself as well as from the object emissivity. This represents a drawback in the fusion process.

2.2 Laser pre-processing

Aiming to extrapolate part of the approaches that had already been applied to images by the photogrammetric and the computer vision communities, one of the first stages of the laser pre-processing will concern to the transformation of the point cloud into the range image.

2.2.1 Generation of a range image

The range image generation process resides on the use of the collinearity equations (1) to project the points of the cloud over the image plane.

$$\begin{aligned} x_A &= -f \cdot \frac{r_{11} \cdot (X_A - X_S) + r_{12} \cdot (Y_A - Y_S) + r_{13} \cdot (Z_A - Z_S)}{r_{31} \cdot (X_A - X_S) + r_{32} \cdot (Y_A - Y_S) + r_{33} \cdot (Z_A - Z_S)} \\ y_A &= -f \cdot \frac{r_{21} \cdot (X_A - X_S) + r_{22} \cdot (Y_A - Y_S) + r_{23} \cdot (Z_A - Z_S)}{r_{31} \cdot (X_A - X_S) + r_{32} \cdot (Y_A - Y_S) + r_{33} \cdot (Z_A - Z_S)} \end{aligned} \quad (1)$$

To obtain the photo coordinates (x_A, y_A) of a three-dimensional point (X_A, Y_A, Z_A) the value of the exterior orientation parameters $(X_S, Y_S, Z_S, \omega, \phi, \kappa)$, must have been computed. These are the target unknowns we address when we undertake the sensors registration procedure. As this is a question that must be solved through an iterative process, it becomes necessary to provide the system of equations (1) with a set of initial solutions that will stand for the exterior orientation of the virtual camera. The registration procedure will lead to a set of corrections in such a way that the final result will be the desired position and attitude.

In this process it is necessary to define a focal length to perform the projection onto the range image. To achieve the best results and to preserve the initial configuration, the same focal length of the camera image will be chosen.

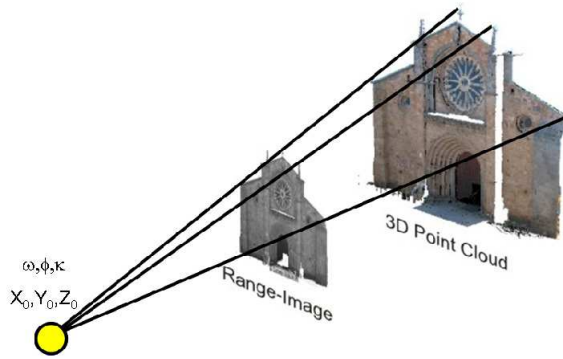


Fig. 2. Range-image generation from laser scanner point cloud

Likewise, in the procedure of generation of the *range-image* a simple algorithm of visibility (*depth correction*) should be applied since there is a high probability that two or more points of the point cloud can be projected on the same image pixel, so an incorrect discrimination of the visible and occluded parts would hinder the appropriate application of the matching procedures (section 3). This visibility algorithm consists on storing, for every pixel, the radiometric value, as well as the distance between the projected point and optic virtual camera centre (both in the laser coordinate system). In this way, every time that a point cloud receives a pair of photo-coordinates, the new pair will be received only if the former point happens to be closer to the point of view than the latter (Straßer, 1974).

2.2.2 Texture Regeneration

It is very common that the range image exhibits empty or white pixels because the object shape may lead to a non homogeneous distribution of the points in the cloud. Due to this, the perspective ray for a specific pixel may not intersect with a point in the cloud and consequently, it may happen that not all the points in the cloud have a correspondent point in the image. This lack of homogeneity in the range image texture drops the quality of the results in the matching processes because these are designed to work with the original conditions of real images. To overcome this drawback, the empty value of the pixel will be replaced by the value of some neighbouring pixels following an interpolation, based on distances (IDW - Inverse Distance Weighted) (Shepard, 1968). This method performs better than others because of its simplicity, efficiency and flexibility to adapt to swift changes in the data set. Its mathematical expression is

$$Z_k = \frac{\sum_{i=1}^n Z_i w_i}{\sum_{i=1}^n w_i} \tag{2}$$

where Z_k is the digital level of the empty pixel, Z_i are the digital values of the neighbouring pixels, w_i is the weighting factor and n is the number of points that are involved in the interpolation. Specifically this weighting factor is defined as the inverse of the square of distance between the pixel k and the i -th neighbouring pixel

$$w_i = \frac{1}{d_{k,i}^2} \quad (3)$$

The neighbouring area is defined as a standard mask of 3x3 pixels, although this size may change depending on the image conditions. In this way, we ensure a correct interpolation within the empty pixels of the image according to its specific circumstances.

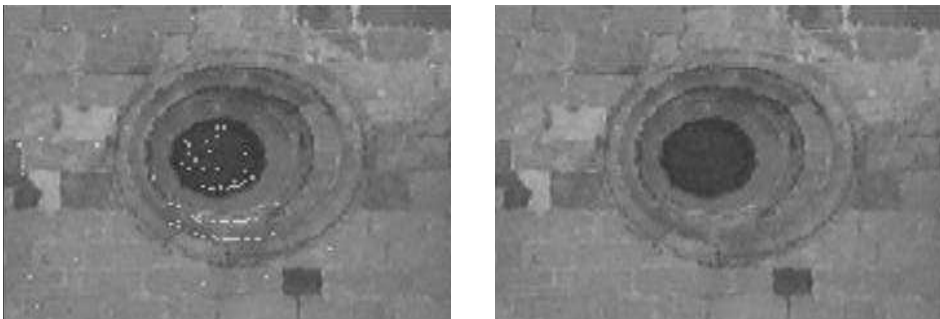


Fig. 3. Before (Left) and after (Right) the texture regeneration

In (2) only one radiometric channel is addressed because the original point cloud data has only one channel (this is common for the intensity data) or because these data have been transformed from a RGB distribution in the original camera attached to the laser scanner to a single luminance value. At last, together with the creation of the range image, an equal size matrix is generated which stores the object coordinates corresponding to the point cloud. This matrix will be used in the sensor fusion procedure.

2.3 Pre-processing of the image

The target of the pre-processing of the image that comes from the digital camera and/or from the thermal camera is to provide high quality radiometric information for the point cloud. Nevertheless, before reaching this point, it is necessary to pre-process the original image in order to make it in tune with the range image in the further procedures. In the following lines we will present the steps in this pre-processing task.

2.3.1 Determination and correction of the radial distortion

One of the largest sources of error in the image is the existence of radial distortion. Particularly, in the context of the sensor fusion, the importance of the accurate determination and correction of the radial lens distortion resides in the fact that if this is not accurately corrected, we can expect that large displacements occur at the image edges, so this could lead to inadmissible errors in the matching process (Fig. 4).

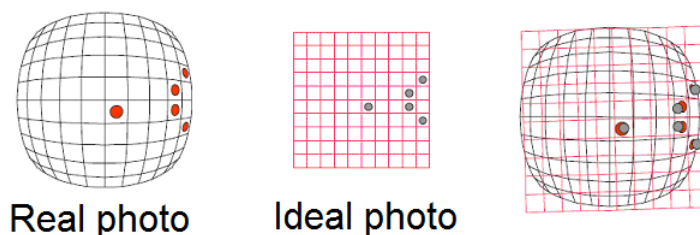


Fig. 4. Displacement due to the radial distortion (Right): Real photo (Left), Ideal Photo (Centre)

Please note in Fig. 4, how if the camera optical elements would be free from the radial distortion effects, the relationship between the image (2D) and the object (3D) would be linear, but such distortions are rather more complicated than this linear model and so, the transformation between image and object needs to account for this question.

On the other hand, in the determination of the radial distortion we find that its modelling is far from being simple because, first of all, there is little agreement at the scientific community on the standard model to render this phenomenon and this leads to difficulties in the comparison and interpretation of the different models and so, it is not easy to assess the accuracy of the methodology. As a result empirical approaches are rather used (Sánchez et al., 2004).

In our case, the radial distortion has been estimated by means of the so called Gaussian model as proposed by Brown (Brown, 1971). This model represents a “raw” determination of the radial distortion distribution and does not account for any constraint to render the correlation between the focal length and such distribution(4).

$$dr = k_1 r^3 + k_2 r^5 \quad (4)$$

For the majority of the lenses and applications this polynomial can be reduced to the first term without a significant loss in accuracy.

Particularly, the parameters k_1 and k_2 in the Gaussian model have been estimated by means of the software sv3DVision Aguilera and Lahoz (2006), which enables to estimate these parameters from a single image. To achieve this, it takes advantage of the existence of diverse geometrical constraints such as straight lines and vanishing points. In those cases of study, such as archaeological cases, in which these elements are scarce, the radial distortion parameters have been computed with the aid of the open-source software Faucal (Douskos et al., 2009).

Finally, it is important to state that the radial distortion parameters will require a constant updating, especially for consumer-grade compact cameras, since a lack of robustness and stability in their design will affect to the focal length stability. A detailed analysis of this question is developed by Sanz (2009) in its Ph.D thesis. Particularly, Sanz analysis as factors

of unsteadiness of the radial lens distortion modelling the following: switching on and off, use of zooming and focus and setting of the diaphragm aperture.

Once the camera calibration parameters are known, they must be applied to correct the radial distortion effects. Nevertheless, the direct application of these parameters may produce some voids in the final image since the pixels are defined as entire numbers (Fig. 5), that is, neighbouring pixels in the original image may not maintain this condition after applying the distortion correction.

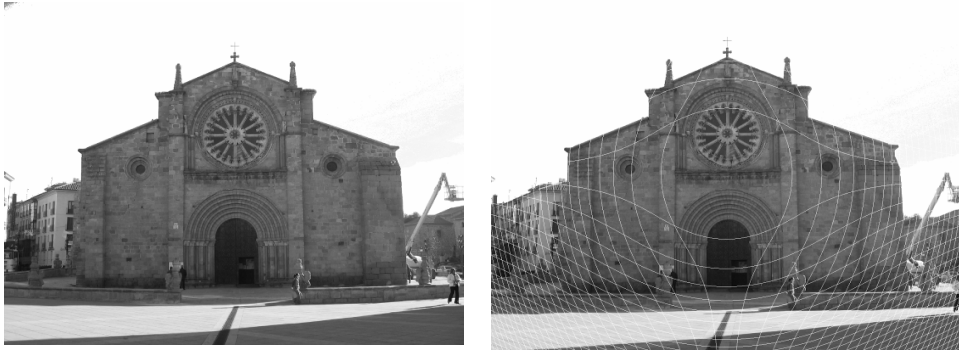


Fig. 5. Left: original image with radial distortion. Right: image without radial distortion, corrected by the direct method.

Trying to avoid this situation as well as applying an interpolation technique which would increase the computing time considerably, an indirect method based on Newton-Raphson (Süli, 2003) has been adapted in order to correct images of radial lens distortion. Particularly, the corrected image matrix will be consider as the input data, so for every target position on such matrix (x_u, y_u) , the corresponding position on the original image (x_d, y_d) will be computed.

2.3.2 Radiometric correction of the images

With the correction of the radial distortion, many of the problems of the image processing are solved but it is advisable to also correct radiometric problems such as:

Treatment of background of the visible image. Usually when we acquire an image, some additional information of the scene background that is not related with object of study is recorded. On the counterpart, the range image has as a main feature that there is any information at all corresponding to the background (by default, this information is white) because it has been defined from the distances to the object. This disagreement has an impact on the matching quality between the elements that are placed at the objects edges, since their neighbourhood and the radiometric parameters related with them become modified by scene background.

From all the elements that may appear at the background of an image shut at the outside (which is the case of the facades of architecture) the most common is the sky. This situation cannot be extrapolated to those elements in the inside or to those situations in which the illumination conditions are uniform (cloudy days), so this background correction would not be necessary. Nevertheless, for the remaining cases in which the atmosphere appears clear, the background radiometry will be close to blue and, consequently it will turn necessary to proceed to its rejection. This is achieved thanks to its particular radiometric qualities. (Fig. 6).



Fig. 6. Before (Left) and after (Right) of rejecting the sky from camera image.

The easiest and automatic way is to compute the blue channel in the original image, that is, to obtain an image whose digital levels are the third coordinate in the RGB space and to filter it depending on this value. The sky radiometry exhibits the largest values of blue component within the image (close to a digital level of 1, ranging from 0 to 1), and far away from the blue channel values that may present the facades of buildings (whose digital level usually spans from 0.4 to 0.8). Thus, we just have to implement a conditional instruction by which all pixels whose blue channel value is higher than a certain threshold, (this threshold being controlled by the user), will be substituted by white.

Conversion of colour models: RGB->YUV. At this stage the RGB space radiometric information is transform into a scalar value of luminance. To achieve this, the YUV colour space is used because one of its main characteristics is that it is the model which renders more closely the human eye behaviour. This is done because the retina is more sensitive to the light intensity (luminance) than to the chromatic information. According to this, this space is defined by the three following components: Y (luminance component), U and V (chromatic components). The equation that relates the luminance of the YUV space with the coordinates of the RGB space is:

$$Y = 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B \quad (5)$$

Texture extraction. With the target of accomplishing a radiometric uniformity that supports the efficient treatment of the images (range, visible and thermal) in its intensity values, a region based texture extraction has been applied. The texture information extraction for purposes of image fusion has been scarcely treated in the scientific literature but some

experiments show that it could yield interesting results in those cases of low quality images (Rousseau et al., 2000; Jarc et al., 2007). The fusion procedure that has been developed will require, in particular, the texture extraction of thermal and range images. Usually, two filters are used for this type of task: Gabor (1946) or Laws (1980). In our case, we will use the Laws filter. Laws developed a set 2D convolution kernel filters which are composed by combinations of four one dimensional scalar filters. Each of these one dimensional filters will extract a particular feature from the image texture. These features are: level (L), edge (E), spot (S) and ripple (R). The one-dimensional kernels are as follows:

$$\begin{aligned}
 L5 &= [1 \quad 4 \quad 6 \quad 4 \quad 1] \\
 E5 &= [-1 \quad -2 \quad 0 \quad 2 \quad 1] \\
 S5 &= [-1 \quad 0 \quad 2 \quad 0 \quad -1] \\
 R5 &= [1 \quad -4 \quad 6 \quad -4 \quad 1]
 \end{aligned}
 \tag{6}$$

By the convolution of these kernels we get a set of 5x5 convolution kernels:

$$\begin{aligned}
 &L5L5 \quad E5L5 \quad S5L5 \quad R5L5 \\
 &L5E5 \quad E5E5 \quad S5E5 \quad R5E5 \\
 &L5S5 \quad E5S5 \quad S5S5 \quad R5S5 \\
 &L5R5 \quad E5R5 \quad S5R5 \quad R5R5
 \end{aligned}
 \tag{7}$$

The combination of these kernels gives 16 different filters. From them, and according to (Jarc, 2007), the more useful are E5L5, S5E5, S5L5 and their transposed. Particularly, considering that our cases of study related with thermal camera correspond to architectural buildings, the filters E5L5 and L5E5 have been applied in order to extract horizontal and vertical textures, respectively.

Finally, each of the images filtered by the convolution kernels, were scaled for the range 0-255 and processed by histogram equalization and a contrast enhancement.

2.3.3 Image resizing

In the last pre-processing images step, it is necessary to bring the existing images (range, visible and thermal) to a common frame to make them agreeable. Particularly, the visible image that comes from the digital camera usually will have a large size (7-10Mp), while the range image and the thermal image will have smaller sizes. The size of the range image will depend on the points of the laser cloud while the size of the thermal image depends on the low resolution of this sensor. Consequently, it is necessary to resize images in order to have a similar size (width and height) because, on the contrary, the matching algorithms would not be successful.

An apparent solution would be to create a range and/or thermal image of the same size as the visible image. This solution presents an important drawback since in the case of the

range image this would demand to increase the number of points in the laser cloud and, in the case of the thermal image, the number of thermal pixels. Both solutions would require new data acquisition procedures that would rely on the increasing of the scanning resolution in the case of the range image and, in the case of the thermal image, on the generation of a mosaic from the original images. Both approaches have been disregarded for this work because they are not flexible enough for our purposes. We have chosen to resize all the images after they have been acquired and pre-processed seeking to achieve a balance between the number of pixels of the image with highest resolution (visible), the image with lowest resolution (thermal) and the number of laser points. The equation to render this sizing transformation is the 2D affine transformation (8).

$$\begin{aligned} \mathbf{R}_{\text{img}} &= \mathbf{C}_{\text{img}} \times \mathbf{A}_1 \\ \mathbf{R}_{\text{img}} &= \mathbf{T}_{\text{img}} \times \mathbf{A}_2 \end{aligned} \quad (8)$$

$$\mathbf{A}_{1,2} = \begin{bmatrix} a & b & c \\ d & e & d \\ 0 & 0 & 1 \end{bmatrix}$$

where \mathbf{A}_1 contains the affine transformation between range image and camera image, \mathbf{A}_2 contains the affine transformation between range image and thermal image, and \mathbf{R}_{img} , \mathbf{C}_{img} y \mathbf{T}_{img} are the matrices of range image, the visible image and the thermal image, respectively.

After the resizing of the images we are prepared to start the sensor fusion.

3. Sensor fusion

One of the main targets of the fusion sensor strategy that we propose is the flexibility to use multiple sensors, so that the laser point cloud can be rendered with radiometric information and, vice versa, that the images can be enriched by the metric information provided by de laser scanner. Under this point of view, the sensor fusion processing that we will describe in the following pages will require extraction and matching approaches that can ensure: accuracy, reliability and unicity in the results.

3.1 Feature extraction

The feature extraction that will be applied over the visible, range and thermal images must yield high quality in the results with a high level of automatization, so a good approximation for the matching process can be established. More specifically, the approach must ensure the robustness of the procedure in the case of repetitive radiometric patterns, which is usually the case when dealing with buildings. Even more, we must aim at guarantying the efficient feature extraction from images from different parts of the electromagnetic spectrum. To achieve this, we will use an interest point detector that

remains invariant to rotations and scale changes and an edge-line detector invariant to intensity variations on the images.

3.1.1 Extraction of interest points

In the case of range and visible images two different interest points detectors, Harris (Harris y Stephen, 1988) and Förstner (Förstner and Guelch, 1987), have been considered since there is not an universal algorithm that provide ideal results for each situation. Obviously, the user always will have the opportunity to choose the interest point detector that considers more adequate.

Harris operator provides stable and invariant spatial features that represent a good support for the matching process. This operator shows the following advantages when compared with other alternatives: high accuracy and reliability in the localization of interest points and invariance in the presence of noise. The threshold of the detector to assess the behaviour of the interest point is fixed as the relation between the eigenvector of the autocorrelation function of the kernel(9) and the standard deviation of the gaussian kernel. In addition, a non maximum elimination is applied to get the interest points:

$$R = \lambda_1 \lambda_2 - k (\lambda_1 + \lambda_2) = |\mathbf{M}| - k \cdot \text{trace}(\mathbf{M})^2 \quad (9)$$

where R is the response parameter of the interest point, λ_1 y λ_2 are the eigenvectors of \mathbf{M} , k is an empiric value and \mathbf{M} is the auto-correlation matrix. If R is negative, the point is labeled as edge, if R is small is labeled as a planar region and if it is positive, the point is labeled as interest point.

On the other hand, Förstner algorithm is one of the most widespread detectors in the field of terrestrial photogrammetry. Its performance (10) is based on analyzing the Hessian matrix and classifies the points as a point of interest based on the following parameters:

- The average precision of the point (w)
- The direction of the major axis of the ellipse (ϕ)
- The form of the confidence ellipse (q)

$$q = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 = \frac{4 \cdot |\mathbf{N}|}{\text{tr}^2(\mathbf{N})} \quad w = \frac{|\mathbf{N}|}{\frac{1}{2} \cdot \text{tr}(\mathbf{N})} \quad (10)$$

where q is the ellipse circularity parameter, λ_1 and λ_2 are the eigenvalues of \mathbf{N} , w the point weight and \mathbf{N} the Hessian matrix. The use of q -parameter allows us to avoid the edges which are not suitable for the purposes of the present approach

The recommended application of the selection criteria is as follows: firstly, remove those edges with a parameter (q) close to zero; next, check that the average precision of the point

(*w*) does not exceed the tolerance imposed by the user; finally, apply a non-maximum suppression to ensure that the confidence ellipse is the smallest in the neighbourhood.

3.1.2 Extraction of edges and lines

The extraction of edges and lines is oriented to the fusion of the thermal and range images which present a large variation in their radiometry due to their spectral nature: near infrared or visible (green) for the laser scanner and far infrared for the thermal camera. In this sense, the edge and lines procedure follows a multiphase and hierarchical flux based on the Canny algorithm (Canny, 1986) and the latter segmentation of such edges by means of the Burns algorithm (Burns et al., 1986).

Edge detection: Filter of Canny. The Canny edge detector is the most appropriate for the edge detection in images where there is a presence of regular elements because it meets three conditions that are determinant for our purposes:

- Accuracy in the location of edge ensuring the largest closeness between the extracted edges and actual edges.
- Reliability in the detection of the points in the edge, minimizing the probability of detecting false edges because of the presence of noise and, consequently, minimizing the loss of actual edges.
- Unicity in the obtaining of a unique edge, ensuring edges with a maximum width of one pixel.

Mainly, the Canny edge detector filter consists of a multi-phase procedure in which the user must choose three parameters: a standard deviation and two threshold levels. The result will be a binary image in which the black pixels will indicate the edges while the rest of the pixels will be white.

Line segmentation: Burns. The linear segments of an image represent one of the most important features of the digital processing since they support the interpretation in three dimensions of the scene. Nevertheless, the segmentation procedure is not straightforward because the noise and the radial distortion will complicate its accomplishment. Achieving a high quality segmentation will demand to extract as limit points of the segment those that best define the line that can be adjusted to the edge. To meet so, the segmentation procedure that has been developed is, once more, structured on a multi-phase fashion in which a series of stages are chained pursuing to obtain a set of segments (1D) defined by their limit points coordinates. The processing time of the segmentation phase will linearly depend on the number of pixels that have been labeled as edge pixels in the previous phase. From here, it becomes crucial the choice of the three Canny parameters, described above.

The segmentation phase starts with the scanning of the edge image (from up to down and from left to right) seeking for candidate pixels to be labeled as belonging to the same line. The basic idea is to group the edge pixels according to similar gradient values, being this step similar to the Burns method. In this way, every pixel will be compared with its eight neighbours for each of the gradient directions. The pixels that show a similar orientation

will be labeled as belonging to the same edge: from here we will obtain a first clustering of the edges, according to their gradient.

Finally, aiming at depurating and adapting the segmentation to our purposes, the edges resulting from the labeling stage will be filtered by means of the edge least length parameter. In our case, we want to extract only those most relevant lines to describe the object in order to find the most favourable features to support the matching process. To do so, the length of the labeled edges is computed and compared with a threshold length set by the user. If this length is larger than the threshold value the edge will be turned into a segment which will receive as limit coordinates the coordinates of the centre of the first and last pixel in the edge. On the contrary, if the length is smaller than the threshold level, the edge will be rejected (Fig. 7).

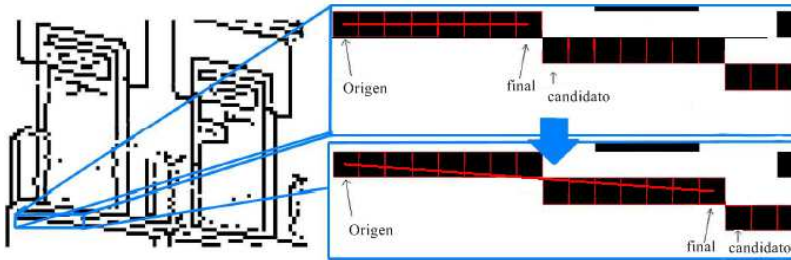


Fig. 7. Edge and line extraction with the Canny and Burns operators.

3.2 Matching

Taking into account that the images present in the fusion problem (range, visible and thermal) are very different in their radiometry, we must undertake a robust strategy to ensure a unique solution. To this end, we will deal with two feature based matching strategies: the interest point based matching strategy (Li and Zouh, 1995; Lowe, 2005) and the edge and lines based matching strategy (Dana and Anandan, 1993; Keller and Averbuch, 2006), both integrated on a hierarchical and pyramidal procedure.

3.2.1 Interest Points Based Matching

The interest point based matching will be used for the range and visible images fusion. To accomplish so, we have implemented a hierarchical matching strategy combining: correlation measures, matching, thresholding and geometrical constraints. Particularly, area-based and feature-based matching techniques have been used following the coarse-to-fine direction of the pyramid in such a way that the extracted interest points are matched among them according to their degree of similarity. At the lower levels of the pyramid, the matching task is developed through the closeness and grey level similarity within the neighbourhood. The area-based matching and cross-correlation coefficients are used as indicator (11).

$$\rho = \frac{\sigma_{HR}}{\sigma_H \sigma_R} \quad (11)$$

where p is the cross-correlation coefficient, σ_{HR} is the covariance between the windows of the visible image and the range image; σ_H is the standard deviation of the visible image and σ_R is the standard deviation of the range image. The interest point based matching relies on closeness and similarity measures of the grey levels within the neighbourhood.

Later, at the last level of the pyramid, in which the image is processed at its real resolution the strategy will be based on the least squares matching (Grün, 1985). For this task, the initial approximations will be taken from the results of the area based matching applied on the previous levels. The localization and shape of the matching window is estimated from the initial values and recomputed until the differences between the grey levels comes to a minimum(12),

$$v = F(\bar{x}, \bar{y}) - G(ax_0 + by_0 + \Delta x + cx_0 + dy_0 + \Delta y)r_1 + r_0 \rightarrow \min \quad (12)$$

where **F** and **G** represent the reference and the matching window respectively, $a, b, c, d, \Delta x, \Delta y$ are the geometric parameters of an affine transformation while r_1 and r_0 are the radiometric parameters of a linear transformation, more precisely, the gain and the offset, respectively.

Finally, even though the matching strategy has been applied in a hierarchical fashion, the particular radiometric properties of both images, especially the range image, may lead to many mismatches that would affect the results of the sensor fusion. Hence, the proposed approach has been reinforced including geometric constraints relying on the epipolar lines (Zhang et al., 1995; Han and Park, 2000). Particularly and taking into account the case of the laser scanner and the digital camera, given a 3D point in object space P , and being p_r and p_v its projections on the range and visible images, respectively and being O_l and O_c the origin of the laser scanner and the digital camera, respectively, we have that the plane defined by P , O_l and O_c is named the epipolar plane. The intersections of the epipolar plane with the range and visible image define the epipolar lines l_r and l_v . The location of an interest point on the range image p_r , that matches a point p_v on the visible image matching is constrained to lay at the epipolar line l_r of the range image (Fig. 8). To compute these epipolarity constraints, the Fundamental Matrix is used (Hartley, 1997) using eight homologous points as input (Longuet-Higgins, 1981). In this way, once we have computed the Fundamental Matrix we can build the epipolar geometry and limit the search space for the matching points to one dimension: the epipolar line. As long as this strategy is an iterative process, the threshold levels to be applied in the matching task will vary in an adaptative way until we have reduced the search as much as possible and reach the maximum accuracy and reliability.

Thank You for previewing this eBook

You can read the full version of this eBook in different formats:

- HTML (Free /Available to everyone)
- PDF / TXT (Available to V.I.P. members. Free Standard members can access up to 5 PDF/TXT eBooks per month each month)
- Epub & Mobipocket (Exclusive to V.I.P. members)

To download this full book, simply select the format you desire below

