Automatic and robust external camera calibration for high accuracy mobile mapping

Werner Goeman[a], Koen Douterloigne[a], Peter Bogaert[b+c], Rui Pires[b], Sidharta Gautama[a]
[a]Department of Telecommunciations and information processing (TELIN), Ghent University, St. Pietersnieuwstraat 41, B-9000, Ghent, Belgium; [b]Geodata & Mobile Mapping group, Grontmij Belgium NV, Meersstraat 138, B-9000, Ghent; [c]Department of Geography, Ghent University, Krijgslaan 281, S8, B-9000 Ghent, Belgium

ABSTRACT

A mobile mapping system (MMS) is the answer of the geoinformation community to the exponentially growing demand for various geospatial data with increasingly higher accuracies and captured by multiple sensors. As the mobile mapping technology is pushed to explore its use for various applications on water, rail, or road, the need emerges to have an external sensor calibration procedure which is portable, fast and easy to perform. This way, sensors can be mounted and demounted depending on the application requirements without the need for time consuming calibration procedures. A new methodology is presented to provide a high quality external calibration of cameras which is automatic, robust and fool proof.

The MMS uses an Applanix POSLV420, which is a tightly coupled GPS/INS positioning system. The cameras used are Point Grey color video cameras synchronized with the GPS/INS system. The method uses a portable, standard ranging pole which needs to be positioned on a known ground control point. For calibration a well studied absolute orientation problem needs to be solved. Here, a mutual information based image registration technique is studied for automatic alignment of the ranging pole. Finally, a few benchmarking tests are done under various lighting conditions which proves the methodology's robustness, by showing high absolute stereo measurement accuracies of a few centimeters.

Keywords: Mobile mapping, extrinsic calibration, external calibration, automation, image registration, mutual information

1. INTRODUCTION

In recent years, mobile mapping systems (MMS) have evolved to a point where they can claim their position as one of the standard methods for the acquisition of survey grade geospatial data. The speed and the multimodal character of the data acquisition process have made it an attractive alternative to the classical surveying techniques. Equipped with high dynamic range cameras, 3D laser scanners, tightly coupled GPS/INS systems and other sensors, the platform is capable of taking a multimodal snapshot of its immediate surroundings in a glance. As each sensor's performance improves with time, the technology is pushed to explore its use in various new application fields on water, rail, road, and even more difficult environments such as tunnels and urban canyons.

When exploring new applications using a mobile platform, the problem has to be studied in relation to the wide variety of available sensors. Modern MMS systems consist of an all-digital platform combining the complementary characteristics of different sensors. It generates highly redundant sensor data which is exploited by the customized data fusion algorithms with near real-time capabilities[2]. It gives the end-user more and more feedback about completeness and quality assurance.

However, the quality of the extracted data highly depends on the quality of the calibration of the sensors. Two sets of parameters can be identified for each sensor, intrinsic and extrinsic. Both parameters' sets can be estimated separately and have already been studied extensively. The intrinsic parameters describe the characteristics inherent to the sensor itself. For cameras, the more practical methods are described in [3] and [4] which are used as a basis for the camera calibration toolbox in Matlab. Here, the intrinsic calibration is performed based on this toolbox and the parameters are considered known before extrinsic calibration. The intrinsic parameters do not change in relation to time as long as the lens/camera combination doesn't change.

The focus of this paper is on extrinsic camera calibration which calculates the relative rotation and translation between different sensor coordinate systems. They have to be calibrated and integrated with the whole sensor system. The method used for extrinsic calibration of video cameras is often poorly documented[5]. Also in many experiments, extrinsic calibration parameters are treated as already known through pre-calibration, while few details can be found on how this is done. However, it is common knowledge that the process of extrinsic calibration is labor intensive. For this reason, off-the-shelf platforms such as [6] [7][8] try to eliminate the calibration step for the end user by providing an all in one sensor package which can be mounted on a mobile platform. By doing so, they also eliminate the flexibility of a sensor configuration by preventing sensor adaptations customized for more specialized surveying and computer vision tasks, such as road inspection, surveys on rails, offshore operations, infrared heat inspection, visual odometry etc.

To fix these problems we present an extrinsic camera calibration methodology which is automatic, robust, portable and easy to perform. This way, extrinsic calibration can be performed without time consuming procedures. There is no need for a target site, as presented in [9][10]. Only a one dimensional model such as a standard ranging pole is needed. It can easily be up to 4m in length and is still practical and portable compared to the checkerboards used in [11]. The procedure can be performed by a non-expert and be used as a standard check for quality assurance, both reducing the operational cost. Moreover, when used for quality assurance, the methodology is more reliable than online calibration algorithms as these sometimes prove to be unpredictable [12], i.e. their performance depends on the surrounding characteristics of the scene, e.g. rural or urban.

2. CALIBRATION METHODOLOGY

The methodology consists of four phases. The first phase is a small survey around the ranging pole. The pole has to be viewed from different viewing angles and different distances. In a second phase the computer vision algorithm will search for good images useful for calibration. Third, in the selected images, a registration will be performed by which the reference model will be fitted to the image with sub pixel accuracy. Based on these registrations it is possible in the final phase to calibrate the camera without interference of an operator.

2.1 The mobile platform

The platform used here is the MMS constructed by Grontmij Belgium NV. The main advantage of this platform is the flexibility for its use in a wide range of applications. The navigation sensor is an Applanix POSLV420 tightly coupled GPS/INS system. The INS makes it possible to maintain the positional accuracy during GPS outages due to low satellite visibility, occurring in urban canyons, tunnels, etc.

Figure 1. The mobile platform from Grontmij Belgium NV mounted on car, boat and train

The platform can handle up to 6 synchronized cameras and 4 laser scanners, a ground penetrating radar, an infrared scanner, a multibeam, and various other types of sensors. Here, we focus on the automatic calibration of the CCD sensor. The cameras can be oriented conform the application requirements. For the experiments, 2 MP Scorpion Pt Grey color cameras are mounted. Lenses with focal length of 4.8 mm and 6.4 mm are used.

2.2 Sensor model and calibration

Using a camera's pinhole model, the projection from the 3D space to the image plane can be described by:

$$s\,\boldsymbol{p} = \boldsymbol{A}\langle \boldsymbol{R}|\boldsymbol{t}\rangle \boldsymbol{P}^{W} \qquad \text{or} \qquad s\,\boldsymbol{p} = \boldsymbol{A}\boldsymbol{P}^{C} \tag{1}$$

Where $\mathbf{P}^w(X^w,Y^w,Z^w,1)$ are the coordinates of a 3D point in the world coordinate space W, **p** are the coordinates of the projection point in pixels, s is a scale factor. **A** is called a camera matrix, or a matrix of intrinsic parameters, and $\langle \boldsymbol{R}|\boldsymbol{t}\rangle$ denotes a matrix gathering the extrinsic parameters (rotation and translation) of the camera. In (1), **p** is not the actual observed image point since virtually all imaging devices introduce a certain amount of nonlinear distortions. Among the nonlinear distortions, radial distortion is present and increasing along the radial direction from the center of distortion. It has been recognized to be the most severe. In this paper it is assumed that the intrinsic parameter set of the camera is known by an off-line calibration process such as the camera calibration algorithm presented in [10]. The intrinsic parameters define the projection from a point in the camera frame to the pixel coordinates in the image plane.

The transformation from world coordinate frame to the camera frame is defined by the following equations:

$$\boldsymbol{P}^{INS} = \boldsymbol{R}_{W2INS}.\boldsymbol{P}^{w} + \boldsymbol{t}_{W2INS} \tag{2}$$

$$\boldsymbol{P}^{c} = \boldsymbol{R}_{INS2C}.\boldsymbol{P}^{INS} + \boldsymbol{t}_{INS2C} \tag{3}$$

$$\boldsymbol{R}_{INS2C} = \boldsymbol{R}_1(\alpha).\boldsymbol{R}_2(\beta).\boldsymbol{R}_3(\gamma) \tag{4}$$

$$\boldsymbol{t}_{INS2C} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \tag{5}$$

$\boldsymbol{P}^c$, $\boldsymbol{P}^{INS}$ and $\boldsymbol{P}^w$are the coordinates of the same point in respectively the camera frame, the GPS/INS frame and the world frame. Each rotation is defined by 3 parameters as shown in (4), and each translation is defined by three parameters as shown in (5). $\boldsymbol{R}_{W2INS}$ is the rotation from world frame to GPS/INS frame. This rotation is known by the system. Also translation $\boldsymbol{t}_{W2INS}$ is known. Thus, the position and orientation of the camera frame is defined by the six remaining parameters (t_x,t_y,t_z,α,β,γ) of translation $\boldsymbol{t}_{INS2C}$and rotation $\boldsymbol{R}_{INS2C}$. These parameters can be estimated using 3D-2D correspondences. For this, a set of object points, their corresponding image projections, the camera matrix and the distortion parameters of the camera are required[15].

2.3 Data collection

The procedure to collect data for calibration has to be easy to perform and practical. Therefore, a standard ranging pole is used as a reference. The pole is set up and vertically leveled at a known ground control point. With the mobile platform a small survey around the pole is performed. This results in a set of georeferenced images in which the pole is visible from different viewing angles and at different distances from the camera. The ranging pole can be seen as a one dimensional sequence of colored segments in the image as is shown in Figure 2. Each transition in color of the ranging pole can be used as a reference point of which the world coordinates are known. The automatic frame selection will try to detect the colored segments as a first estimate of where the ranging pole can be found in the image before the registration technique is used for segment extraction with sub pixel accuracy.

Figure 2. One dimensional reference model

2.4 Automatic frame selection

The process of selecting the right frame is important for two main reasons. First, the calculation of the calibration parameters is sensitive to outliers. False positives should be prevented even if that limits the fraction of true positives. On the other hand, frames need to be collected from different viewing angles to the ranging pole. This way, a bias is prevented on the wide baseline stereo measures after calibration. The process of frame selection is presented in Figure 3. Each frame is first undistorted using the intrinsic calibration parameter set of the camera. This way the distortion of straight lines projected in the image plane such as the reference model will be corrected.

In the next phase, the interesting features of the reference model (i.e. color and shape) are used to search for its projection in the image. Using samples of different appearances of the pole, a color region in the HSV color space is created, focusing on the orange color segment in this case. The conversion to the HSV color space is performed because it allows us to model the color region of interest as a 2D polygon in the HS plane, dropping the value component of the color coordinate.

A second filter is the ridge detection filter. Lines in an image can be seen as narrow valleys or ridges in the intensity surface if one views the image as a terrain model. To determine pixels belonging to the pole, one can use several approaches like the 'facet model' based approach [13] or Steger's method [14]. In our work, the range pole detection is performed using a differential ridge detector based on Steger's method. This method describes the scale-space analysis of a bar-shaped intensity profile, which can be a pole like object, by convolving the image with a Gaussian or one of its derivatives. The main characteristic of a ridge pixel is a large eigenvalue λ_1 of the Hessian matrix, with corresponding eigenvector perpendicular to the road direction.

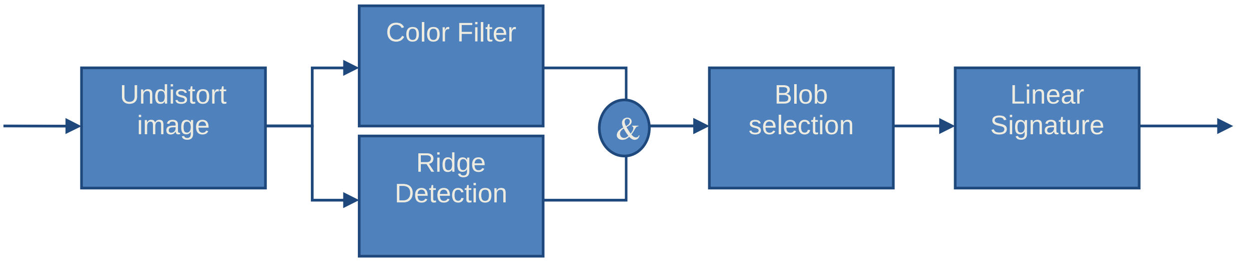


Figure 3. The frame selection process

In order to calculate the eigenvalues in a certain pixel, the image I is convolved with the second derivatives of a Gaussian G_{xx}, G_{xy}, G_{yy}. The convolution kernels have dimension $(2w+1)$ by $(2w+1)$, w denoting half the kernel size. These convolutions give rise to smoothed and differentiated images. We will refer to the values of a pixel in these images with R_{xx}, R_{xy} and R_{yy} respectively. The eigenvalues and eigenvectors can now easily be calculated using the Hessian H for each pixel:

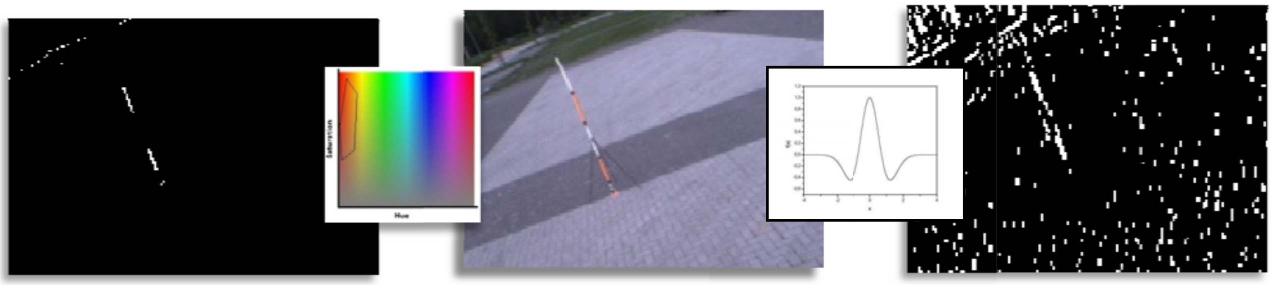
$$H = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{xy} & R_{yy} \end{bmatrix}$$

If we focus our attention on vertical ridges, only the second partial derivative R_{xx} (i.e. the raw image convolved with G_{xx}) is important for determining the ridge saliency:

$$R_{yy} \approx R_{xy} \approx 0 \Rightarrow \lambda_1 \approx |R_{xx}| \ ,$$

wherein λ_+ is the largest eigenvalue of the Hessian. If, for a certain pixel, this eigenvalue is higher than a certain threshold, there is a great possibility a ridge is detected. Of course, determining the best threshold is crucial for good detection of the ranging pole and is calculated here based on the distance to the camera. In Figure 4 a result of this process is given.

The two masks are combined to extract visual cues. This extraction based on color and ridge information is prone to noise. Therefore, each segment in the combined mask is analyzed. Based on some characteristics the segment is kept or eliminated from the mask. The criteria used here are: linearity, size and second color at beginning and ending of each segment. In this case, segments which are big linear blobs of orange, starting and ending with black pixels, are selected as possible parts of the ranging pole. It is clear that the starting and ending image coordinates of the segments do not correspond to a perfect segmentation of the segments of the ranging pole. Because of illumination effects, chromatic aberration and jpeg compression, the extraction of these visual cues will only be used as a first estimate to initialize the model registration.



(a) (b) (c)

Figure 4. The HS color filter creates a mask (a) for the original image (b) leaving only the colors present in the reference model. The ridge detector creates a mask (c) only leaving the pixels for which the eigenvalue λ_1 is higher than the threshold

2.5 Automatic model registration

When accurately registering the ranging pole model to the image, we have to keep in mind that the ranging pole can move in 3D in the world frame, which is different from a simple translation or rotation. Under certain conditions, the visual appearance of the ranging pole can change, while its four corner points in the image stay fixed. In Figure 5, a person or camera in the $x = 0$ plane, observes a ranging pole $\boldsymbol{ab}$, with $\boldsymbol{a} = (a_y, a_z)$ and $\boldsymbol{b} = (b_y, b_z)$. Without loss of generality we can say that $a_y = 0$, $a_z = 1$ and $b_z = 1$. The projection of a point $\boldsymbol{e} \in \boldsymbol{ab}$ onto the viewing plane, seen by the observer or camera, is given by $\frac{e_y}{e_z}$. We now consider a longer ranging pole $\boldsymbol{ac}$, which stands tilted. In practice the pole will always be perfectly vertical, but the camera can tilt, resulting in the same effect. To the viewer, this new pole has the same size as the $\boldsymbol{ab}$ pole. This is because $\boldsymbol{c}$ lies on the line $\boldsymbol{0b}$, so we can write $b_y = \frac{c_y}{c_z}$. While the observed size stays the same in both cases, the observed ratio does not. A point $\boldsymbol{d}$, taken halfway on $\boldsymbol{ac}$, will not be projected onto a point halfway between $\boldsymbol{a}$ and $\boldsymbol{b}$, but on a point $\boldsymbol{e}$ instead. To get an expression for $\boldsymbol{e}$ in function of $\boldsymbol{d}$, we write down their relations to $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ first,

$$d = (1-t)\boldsymbol{a} + t\boldsymbol{c}$$
$$e = (1-u)\boldsymbol{a} + u\boldsymbol{b},$$

with $t, u \in [0,1]$. If $\boldsymbol{d}$ and $\boldsymbol{e}$ have the same projection, then $\frac{d_y}{d_z} = \frac{e_y}{e_z}$. Expanding this gives

$$\frac{(1-t)a_y + tc_y}{(1-t)a_z + tc_z} = \frac{(1-u)a_y + ub_y}{(1-u)a_z + ub_z}.$$

Filling in the known coordinates and the expression for b_y, we get

$$\frac{tc_y}{(1-t) + tc_z} = u\frac{c_y}{c_z},$$

which gives u as

$$u = \frac{c_z t}{1 + (c_z - 1)t}.$$

This expression depends only on the unknown c_z, with the $\boldsymbol{ac}$ pole being closer to the observer when $c_z \in [0,1[$ and being further away when $c_z \in]1, \infty]$. To get a more natural parameter describing the deformation of the observation of the pole due to the depth-effect, we put $c_z = e^{\varphi}$. Now negative φ values correspond to a negative depth deformation, and vice versa. Furthermore, the deformation is now symmetrical, a useful property for transformations, as deforming first with φ and then with $-\varphi$ gives the identity transformation.

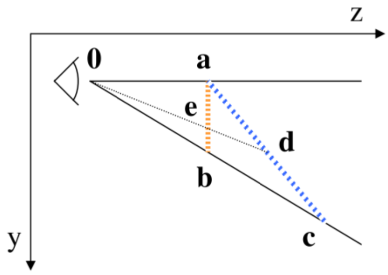
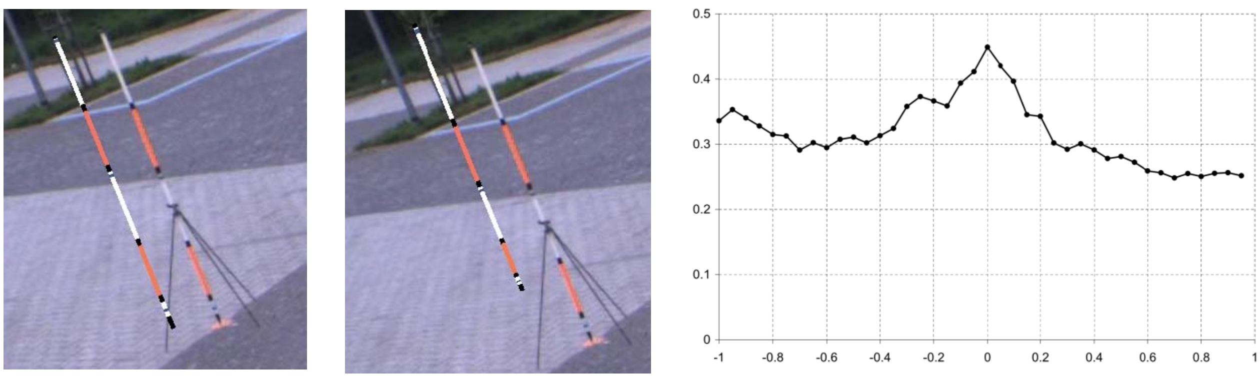


Figure 5. Schematic representation of a viewer at position $\mathbf{0}$ observing two ranging poles $\boldsymbol{ab}$ and $\boldsymbol{ac}$

One detected segment is insufficient to uniquely identify the position of the ranging pole in 3D. We solve this problem by evaluating a wide range of possible φ parameters. For every φ we project the model onto the image, and compute a similarity measure between both. In this work we have used mutual information as the similarity measure, given its robustness to illumination changes [16]. The correct φ is then taken to be the φ for which the mutual information has a maximum. This is clarified in Figure 6.



(a) $\varphi = 0$ (b) $\varphi = -0.45$ (c)

Figure 6. (a)(b) The model ranging pole superimposed over the image for different φ. Notice that the position of the upper segment, of which the location was determined before, stays fixed. (c) The mutual information of the superimposed ranging pole in function of φ. The correct registration corresponds to the maximum mutual information.

3. EXPERIMENTAL RESULTS

To test the presented methodology, several calibration surveys as described above, were performed on different locations in good GPS conditions. The internal calibration of the 2MP Pt Grey color camera was done using the checkerboard method. In Table 1, the results of the frame selection procedure are shown. The true positives present the frames in which the ranging pole was in the field of view of the camera and at least one orange segment was identified correctly. The false positives correspond to detected orange segments that are incorrectly considered as part of the ranging pole. The true negatives are the frames in which the pole is not visible and not detected. False negatives are frames in which the ranging pole is visible but not detected.

Frames	*TP*	*FP*	*TN*	*FN*
(a) 205	27	2	175	1
(b) 71	20	9	40	2
(c) 141	14	2	118	7

Table 1 Results of the frame selection procedure for three calibration surveys

Three surveys were analyzed. Survey A was performed in the best circumstances, i.e. good lighting conditions and no disturbing background objects. This results in a low FP and FN value. Survey B was performed in a more interfering environment, i.e. various orange objects in the background. This results in more false positives. Survey C was performed in less environment light which results in more false negatives.

All frames which were considered positive in the frame selection procedure, are now used in the registration process. The registration should be robust against mistakes in the segment detection process. Also, it has to be able to determine which segment of the ranging pole was detected when only one segment was found. With the ranging pole used here, this means two registration trials have to be made when only one segment was found. When two segments were found, only one registration is performed.

To obtain this, we use several decision steps during registration:

1. First of all, if the registration gives a part of the ranging pole as being outside the image, then the entire registration is rejected. Even though the registration might be correct, we need the ranging pole to be completely visible for the later calibration.
2. After this we put a simple threshold on the mutual information value. When its value at the optimal φ is too low, then we will classify the registration as incorrect. The threshold should not be too strict, so as not to give too many false negatives.
3. Finally we put a threshold on the mean squared error between the model pole and the image, to remove the remaining false positives. Here we can also use the information that over a sequence of images the camera tilt will be approximately constant, meaning the φ will be constant. Images where φ is too far from the average are also discarded.

We have tested our registration evaluation method on a set of 109 registration results. Manually annotated ground truth gives us 29 correct segmentations, and 80 incorrect segmentations. Of the 109 registrations, 25 lie partly outside the image, so we discard these based on the first evaluation criterion. The remaining 84 are thresholded based on their mutual information and mean squared error values. This is shown in Figure 7. When we choose thresholds of 0.4 for the MI and 8000 for the MSE we get 1 remaining false positive and 5 false negatives. Stricter thresholds result in no remaining false positives.

4. CONCLUSION

In this paper, a new methodology was presented to automate the extrinsic calibration process of a 2 MP Scorpion Pt Grey color camera mounted on a high accuracy mobile mapping system. The methodology proves to be easy to perform and robust. To achieve this, a set of georeferenced images was collected in which a standard ranging pole is visible. Using a color and shape based method, a set of images is automatically selected from the calibration survey. Using the location of the reference object, i.e. the ranging pole, and its projected image coordinates in the image, the extrinsic camera parameters can be calculated. The mutual information based registration procedure is used to extract the image coordinates at subpixel level and is able to identify the false positives from the frame selection procedure. The results show that the automated procedure can achieve the same calibration quality compared to a manual calibration, even when jpeg compression, chromatic aberrations and less than ideal lighting conditions are present.

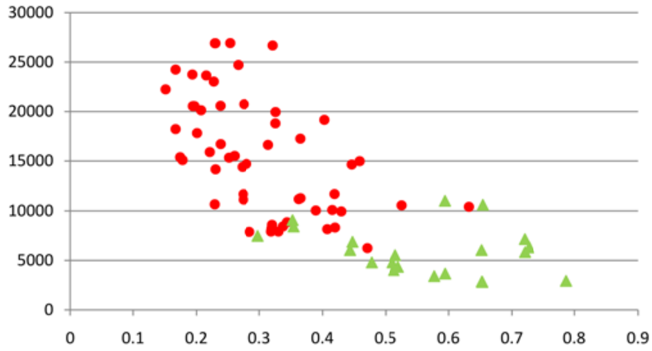


Figure 7. The mean squared error (y-axis) vs the mutual information (x-axis) for 75 segmentation and registration results. The green triangles are the correct segmentations, the red circles are the incorrect ones, based on manually created ground truth. A threshold of 0.4 for the MI and 8000 for the MSE gives a result of 1 false positive and 5 false negatives.

REFERENCES

[1] Schwarz, K. P., El-Sheimy, N., "Mobile Mapping Systems – State of The Art and Future Trends", Proc. ISPRS, (2004)

[2] Brzezinska, D., "Tutorial: Mobile mapping technology: paradigm shift and future trends," 7th International Syposium on Mobile Mapping Technology (2011)

[3] Zhang, Z.,"Flexible Camera Calibration By Viewing a Plane From Unknown Orientations", ICCV (1999)

[4] Heikkilä, J.,Silvén,O., "A Four-step Camera Calibration Procedure with Implicit Image Correction," CVPR (1997)

[5] Huing,Z.,Yuzhong,C.,Shibasaki,R.," An efficient extrinsic calib of a multiple laser scanners and cameras sensor system on mobile platform", IEEE Intelligent Vehicles Symposium, 422-427 (2007)

[6] MDL, "Dynascan 3D mobile mapper," http://mdl-laser.com/en/dynascan-3d-mobile-mapper--14739 (11 July 2012)

[7] TOPCON, "IP-S2 HD Mobile Mapping System," http://www.topcon-ositioning.eu/35/Mobile_Mapping 14739 (11 July 2012)

[8] GeoVISAT, "Data acquisition," http://www.geovisat.eu/menu031.html (11 July 2012)

[9] Hassan,T., El-Sheimy, N., "System Calibration of Land-Based Mobile Mapping Systems", ISPRS (2008)

[10] Scheller, S., Westfeld, P., Ebersback, D., "Calibration of a mobile mapping camera system with photogrammetric method," Proc ISPRS Vol XXXVI, 5th International conference on mobile mapping technology (2007)

[11] Geiger, A., Moosman, F., Car, O., Schutser, B., "Automatic camera and range sensor calibration using a single shot," IEEE International Conf on Robotics and Automation, p3936-3943 (2012)

[12] Cannelle, B., Paparoditis, N., Pierrot-Deseilligny, M., Papelard, J.P., "Offline vs online calibration of a panoramic-based mobile mapping system, " ASPRS (2012)

[13] R.Haralick, L.Watson and T.Laffey, "The topographic primal sketch", International Journal of Robotics Research, 2(1), p50–72 (1983)

[14] C.Steger, "An unbiased detector of curvilinear structures",IEEE Transactions on Pattern Analysis Machine Intelligence, 20(2), p113–125 (1998)

[15] OpenCV, opencv. willowgarage.com/documentation/ camera_calibration_ and_3d_reconstruction .html

[16] Thévenaz, P. , Unser, M., "Optimization of mutual information for multiresolution image registration," IEEE Transactions on Image Processing 9, 2083–2099 (2000)