



Urban areas change detection using DEMs generated automatically from high spatial resolution stereo satellite images

Detección de cambio en zonas urbanas utilizando MDTs generados automáticamente a partir de imágenes satelitales estereoscópicas de alta resolución espacial

Sandra Lucia Cobos Mora

Faculty of Engineering, Industry and Construction, Católica de Cuenca University, 010101, Ecuador
Alcala University, Spain
German Aerospace Center, Germany
scobosm@ucacue.edu.ec

Abstract

In the spatial-temporal modeling for vertical change detection by a mathematical difference, Digital Elevation Models (DEMs) are used to monitor changes in height values. The main aim of this paper is to improve the accuracy of the change detection technique by applying a merge between DEMs of the same characteristics and therefore achieve better quality results. The research was developed using data gathered between 2006 and 2011 in two test areas. Two robust DEMs were generated from these dates, which were subtracted to detect altered areas. The DEMs have been automatically generated using the semi global matching method (SGM) from high spatial resolution stereo images which were provided by the Ikonos and GeoEye-1 satellites. All these images were previously subjected to a block adjustment in order to correct the geometric differences. The proposed technique in comparison with the basic technique, has improved the correctness in 3D change detection by reducing the number of pixels detected as false positives in 2,03% in Test Area 1 and 3.24% in Test Area 2.

Key words: Stereo Images, Digital Elevation Models (DEM), 3D Change Detection, Urban Areas.

Resumen

En el modelado espacio-temporal para la detección de cambios verticales, los modelos digitales de elevación (MDE) se utilizan para monitorear los cambios en los valores de altura, mediante una resta entre dos MDEs de diferentes fechas, como técnica básica. El objetivo principal de este trabajo es proponer una alternativa para mejorar la precisión de la técnica básica de detección de cambios mediante la fusión de MDEs de características similares para aumentar la calidad de los mismos. Este estudio se desarrolló entre 2006 y 2011 en dos áreas de prueba. Se generaron dos MDEs robustos a partir de estas fechas, que se restaron para detectar áreas alteradas. Los MDEs se han generado automáticamente utilizando el método de correspondencia semi global (SGM) a partir de imágenes estereoscópicas de alta resolución espacial proporcionadas por los satélites Ikonos y GeoEye-1. Previamente, todas estas imágenes fueron sometidas a un ajuste de bloque para corregir las diferencias geométricas. La técnica propuesta, en comparación con la técnica básica, ha mejorado la corrección en la detección del cambio 3D al reducir el número de píxeles detectados como falsos positivos en 2,03 % en el Área de Prueba 1 y 3,24 % En el área de prueba 2.

Palabras clave: Imágenes Estereoscópicas, Modelos Digitales de Elevación (MDT), Detección de Cambio tridimensional, Áreas Urbanas.

I. INTRODUCTION

A. Stereo Sensors

IN recent years, satellite systems have been implemented with the ability to provide stereoscopic data of the earth

surface. The data consists of two images obtained with seconds of difference between them and covering the same area from different observation angles during a single sensor pass, therefore making a difference in perspective. This

technology has been commonly applied in aerial photography and achieved through the sequential photographs taken during the same flight path with an overlapping area of approximately 60% on the horizontal plane. With the recent advances of remote sensing, images from very high spatial resolution satellites have achieved the same performance but at a lower cost than aerial photography. German Modular Optoelectronic Multispectral Scanner (MOMS) which was onboard of the Russian space station named MIR, was among the first systems to introduce this stereoscopic capability. A detailed description of this sensor can be found on Lehner & Kornus [1] and Poli [2].

This stereoscopic capability is only included in some very high spatial resolution satellites such as Ikonos, operated by GeoEye, which has a rational polynomial coefficient (RPC) camera model file that provides stereo data with a coverage of 360 degrees, allowing the generation of DEMs and digital surface models from anywhere in the world with a spatial resolution of five meters or less. Ikonos offers a spatial resolution of 1 m in the panchromatic mode and 4 m in the multispectral mode. It has a horizontal and vertical precision of 25 m and 22 m respectively, without considering control points, otherwise it goes up less than 2.5 m horizontally and less than 1.5 m in the vertical [3]. The same company that operates Ikonos put the satellite system GeoEye-1 in orbit on September 6, 2008, which has the second highest spatial resolution for commercial sensors. Images from this satellite are commercially available since February 2009 with a spatial resolution of 0.41 meters for military applications, and 0.5 m in panchromatic band for the rest of applications. The stereo pairs are taken on the same orbital pass, with an elevation above 60° with a difference of 30 seconds between them [4].

Nowadays, Digital Globe owns the most sophisticated constellation of satellites with the highest spatial resolution. WorldView-1 launched on September 2007 provides and spatial resolution of 50 cm in the panchromatic band. It is located at an altitude of 496 km and has an average day revisit time of 1.7 [5]. WorldView-2, launched on October 2009, has 8 multispectral bands with a pixel size of 1.85 meters and one panchromatic with 0.46 cm but they are resampled to 2.0 and 0.5 respectively, to comply with U.S. Regulation. It is operating at 770 km of altitude with an average 1.1 days in revisiting time. These two satellites are equipped with control moment gyroscopes that improve the slew time from 60 seconds to 12 seconds to efficiently collect in-track stereo images to create highly detailed DEMs [6]. This company also operates Quick Bird which has a spatial resolution of 61 to 72 cm in the panchromatic band, depending of the nadir viewing angle that could be between 0 to 25 degrees. The scene covers 16.5 km x 19 km in the across-track direction, with a frequency of acquisition between 1 to 3.5 days [7] and used to provide stereo capability until 2006.

The SPOT-5 satellite with its HRS2 and HRS1 sensors was the first satellite of the SPOT series to provide stereo-

scopic data, obtaining the stereo pair with a difference of 90 seconds. This satellite has a spatial resolution of 5 meters length and 10 meters in width and a nadir angle of +20 and -20 degrees for HRS1 and HRS2 respectively, to cover a stereoscopic area of 600 km from a whole scene of 1200 km x 1200 km. In the same way, Cartosat-1 with its 2 cameras, F (forward) and A (after), has angles of 26 degrees and -5 degrees, respectively, and provides images with a difference of 52 seconds with 0.62 meters of spatial resolution [9] and a temporal resolution of 26 days. In Radhadevi et al.[10] is described the work carried out by the authors to compare quantitative and qualitative DEMs generated from these two sources.

Also, ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer), an imaging instrument flying on Terra satellite, offers a stereoscopic view in the near infrared bands 3N with three detector line arrays (Bands 1, 2, 3) and 3B which detects in only one band, by using two telescopes developed to reduce the backward distortion and the nadir looking. It is done with an angle of 27.7 degrees and a spatial resolution of 15 meters [11].

B. Digital elevation models (DEM) from stereo data

The generation of DEMs from stereo data has been investigated during the last 30 years and has become the most common method to produce these models. Initially this technique had a strong dependency on the aerial images since they were the only data to be used in stereoscopy procedures. However, since 1980 stereo data is also available from satellite sensor considering factors such as the spatial and temporal resolution, and the radiometric difference between the stereo pair; all of them with a strong influence on the final product [11]. To work with these images, a matching of the stereo pairs is required. However, considering this manual process would require a lot of time and increment the product's cost. Because of this, many techniques have been explored and proposed for automatic procedures like the Semi-Global Matching which generates DEMs based on added cost for each pixel. According to Buyuksalih et al. [12], these automatic procedures could be based on the area, sub-matrix of gray values or features. Feature based matching is used for images with a very high spatial resolution in urban areas, using points positions, edges or patches. The area based matching procedure, is better used in open areas to get a complete coverage of the whole scene with height points

These automatic algorithms can increase their accuracy with the use of ground control points (GCP). According to Toutin [11] cited in Akl [13], only four precise GCPs are theoretically necessary to obtain DEMs with good absolute accuracy (quality of accuracy in each pixel). However, with a large number of these points, the possibility of keeping the accuracy in the order of one pixel is higher. This is because they prevent the error propagation in the model. Unfortunately, the possibility of counting with control points in this phase is not always certain, in many

cases due to the difficulty to select the correct ones [14].

This situation has led different authors to investigate the best way to produce DEMs with greater accuracy by using post processing algorithms to improve precision. Some of these methods have been discussed in Reinartz et al. [9] where it is shown that the accuracy of the DEMs from satellite images with 10 meters of spatial resolution (provided by SPOT-5) and without GCP, is improved using auxiliary information such as tie points to make a bundle block adjustment in order to correct the small biases in all three coordinates (x, y, z). Also, a fusion with other DEMs with the same quality but obtained from different sources helped to raise the overall quality. After the matching phase, interpolation is necessary, which is also done using automatic functions [12].

The generation of DEMs has been performed using a variety of sensors through the use of different techniques such as the stereo pairs of by mixing images coming from the same sensor, but from different orbits, with a significant time difference. Unfortunately, in the last case, a large number of unmatched areas usually exist, due to different illuminations and atmospheric conditions, being necessary to execute a Gauss filter as previous step to get better results. Definitely, the best results are obtained when images from the same orbit are used, and the difference in time between them is the shortest possible [14].

It is important to consider that with these automatic algorithms the result is not a DEM, this is due to the fact that the match points could correspond to the top of a building or to the tree crown, therefore digital surface models (DSM) are produced. In most cases, this is not the product required, and it is necessary to convert it to a DEM by applying a filtering method which is only possible if at least a few points are located on the bare ground [12].

C. Temporal change detection using DEMs

Change detection has been one of the major operational applications of satellite images. Several techniques have been proposed in the literature, some of them specifics of the main object/cover to be identified as the one using vegetation indices to focus in vegetation changes or the one using water index to monitor the water content or water stress. There are also generic procedures that allow detecting all changes between land covers. All these techniques are discussed in detail in [15].

DEMs provide the basis for modeling and analyzing topographic information. This allows urban monitoring, change and damage detection in the third dimension (height). These changes have been ignored for the 2D techniques, where only changes related to reflectance values or texture can be detected. This information plays an important role for disaster assessment, urban area construction, destruction monitoring, etc. [16].

Change detection by using DEMs can be classified into two categories. The first category is based on the combination of stereo and multispectral images to detect changes

in 2D and 3D. The second one uses the subtraction of two of these models from different dates with similar specifications and covering the same area. This last technique discloses every change in height when the difference in elevation is higher than the combined maximum possible error of each DEM and the change occurs over an area larger than its spatial resolution, plus the horizontal error [13]. Here, the resolution and the data accuracy are determined to reduce false positive changes. This approach has been used by several authors as Zhang [17], Reinartz, et al. [18], and Akca [19].

II. MAIN AIM

Merge various DEMs generated automatically from high spatial resolution stereo images before being part of a 3D change detection procedure to analyze potential accuracy improvements in the results.

III. DATA AND METHODOLOGY

A. Data

In this study, very high spatial resolution images obtained by commercial sensors have been used. There are two GeoEye-1 and four Ikonos images, all of them covering the same area. These images were converted to XDibias format, DLR's own software. This process creates 4 images for each date; one image containing the four multispectral channels and the other one only the panchromatic band. Because they are stereo images, there is one image captured from the left side, and the other one from the right side. The images used in this study are panchromatic because they have 1 m spatial resolution compared with 4 m of the multispectral images.

The period of time studied goes from 2006 to 2011. The images obtained between these years have been used to improve the DEM of the last date. Unfortunately, there were not images available before 2006 to help the improvement of its model. Table I shows the main information about the data used.

B. Methodology

The methodological steps considering the merging of DEMs adopted to improve the basic 3D change detection technique, which is based in the subtraction of DEMs, is exposed in Figure 1.

1. Tie points generation

The initial phase in the image processing is the finding of tie points. These are stable features that describe the same geographic position in two or more images that cover the same area. They are useful in geometric image transformation, as it will be used in the present project; but they are also used on image mosaicking and three dimensional model generation. Tie points are visually recognizable in the overlap area among the images and are used to further establish relationships between them.

TABLE I
IMAGES USED IN THE STUDY

Image	Date	Sensor	Abbr.
Ge20091220	December 20th, 2009	GeoEye-1	geo2009
Ge20110502	May 2th, 2011	GeoEye-1	geo2011
Ik20060223	February 23th, 2006	Ikonos	ik2006
Ik20100112	January 12th, 2010	Ikonos	ik2010.1
Ik20100513	May 13th, 2010	Ikonos	ik2010.5
Ik20110107	January 7th, 2011	Ikonos	ik2011

These image features are invariable to the image scale, rotation, illumination or addition of noise from the stereo images and help to provide robust matching. They should be well located in spatial and frequency domains, making them harder to be interrupted by clutter or noise. The approach used in this project to generate tie points is called Scale Invariant Feature Transform (SIFT) and it was developed by Lowe [20]). Its main advantage is the generation of a large number of features that cover the whole image over the full range of scales and locations.

This algorithm is performed by considering 4 major stages, in order to find the set of tie points. First, it is

necessary to detect the scale space extrema by recognizing locations and scales that can be assigned when the image has a different angle view to the nadir. This can be done by searching all the points that are stable in the images when the scale has changed. According to Lowe [20], the best way to do this is using the scale space extrema in the difference of Gaussian function. The number of samples with different scales must be large, due to the fact that the success of the object recognition is closely linked to the quantity of correctly matched key points and not on their percentage, in relation to the total number of them.

With the set of key points found in the step above, it

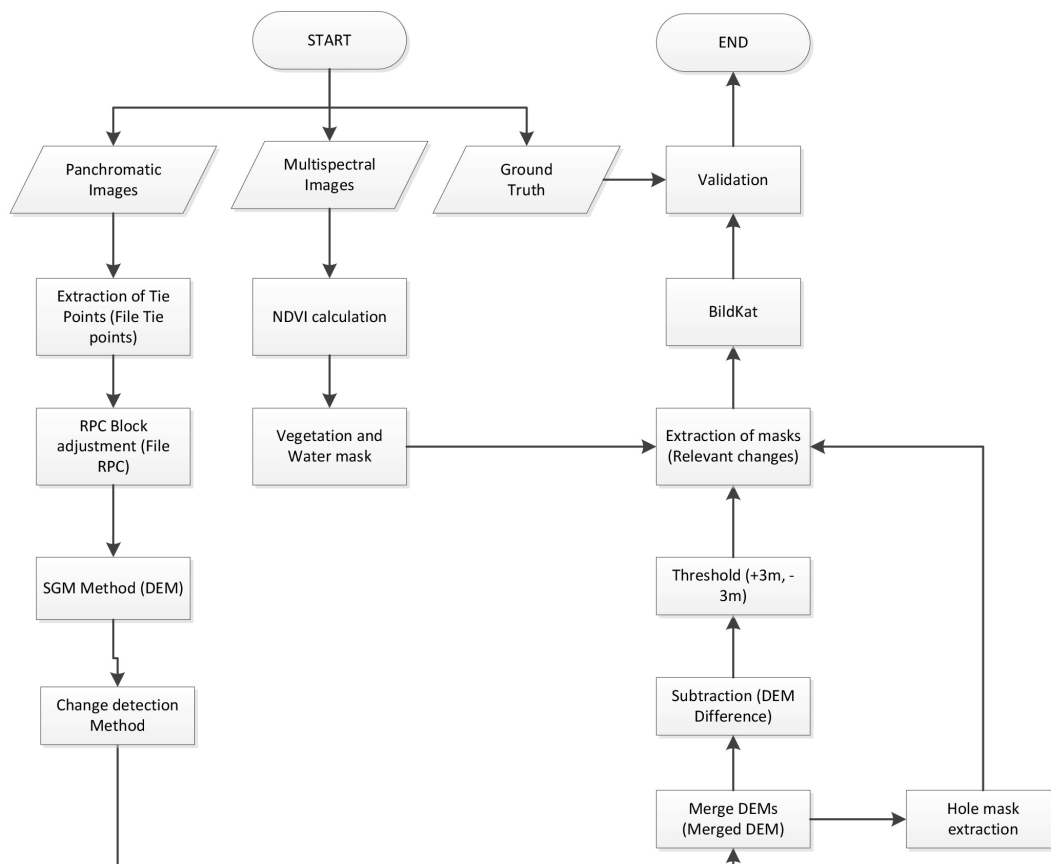


FIG. 1. Flowchart of the proposed technique to identified 3D changes in urban areas from stereopair images.

is necessary to eliminate the points with low contrast or a poor location along an edge. This is achieved by the adjustment of the nearby data for location, scale and ratio of principal curvature. The features with the higher stability measurement are assigned to a consistent orientation, scale and location based on local image properties, describing this region for a repeatable local 2D coordinate system. This leads to the last step which calculates the highlighted invariant descriptor for the local region. The algorithm used to create it is described in Lowe [20].

In this process it is important to take into account the following two facts. First, the tie points' accuracy is linked to the database size, formed with all the features extracted from the reference images that will be matched with a new image; and second, to detect small objects in cluttered areas, at least three points should be correctly matched for a trustworthy identification [20].

In this study, the tie points have been generated using four Ikonos and two GeoEye-1 images with a maximum distance for outlier measurements of 200 pixels, 500 meters for the grid size, used for thinning the resulting points and a matching ratio of 0.65. Snaveley et al. [21] propose this last value in 0.6 because higher values not only produce more matches, but also more outliers.

In this context, for each descriptor of the first image, the algorithm finds the nearest neighbor in the second image looking through a tree, formed with the features of this last image. It accepts those matches if $\frac{d_1}{d_2} < 0.65$, where d_1 and d_2 are the distances of two nearest neighbors to the second image. To make this procedure efficient, the search is limited to a maximum of 200 bins of the tree. This gives geometrically consistent matches but only between a pair of images. It is necessary to join all the keypoints that match in multiples images, in one set called a track. If one point matches two or more times in the same image, it is considered as inconsistent [21].

2. Bundle Block Adjustment

Bundle Block Adjustment is used in this methodology to correct the geometric difference between a set of images, by converting each three dimensional image into two dimensional images. Here, the outliers tie points are eliminated to obtain a correct matching. The block adjustment implemented here, has been done by using rational polynomials, mathematically simpler and numerically more stable than the traditional methods. This method is directly related to the geometric properties of the sensor. In the original paper that describes this technique, the authors consider the physical features of Ikonos sensor, but it is demonstrated that it could be used for any sensor with a stable calibration on its interior orientation, a priori corrected exterior orientation and a narrow field of view [22].

The rational polynomial coefficient (RPC) models are used to convert three-dimensional object space coordinates, determinate in latitude, longitude and height (Φ, λ, h) , into two dimensional image space coordinates (Line, Sample),

a process that is quite complicated in satellite images due to the fact that the whole image is composed by several number of lines, each one captured in a different period of time; therefore, each line has its own exposure position and its own orientation [23], which does not happen in aerial photography. For this reason, a series of mismatches in the image need to be corrected by the introduction of two parameters, the line parameter and sample parameter. These two include effects of orbit, attitude and residual interior orientation errors in line direction, for the first one, and sample direction for the second one [22].

The RPC block adjustment model is determined for a function $(\Delta p, \Delta r)$ that expresses the differences between the measured and the nominal line and sample coordinates of GCP or tie points. In this case, the generated tie points are used. In the equations below (Equation 1 and 2), a_0 includes errors on the line direction (in-track errors) and b_0 includes errors on the sample direction (cross-track). The parameters a_L and b_L are referred to temporally linear error in satellite attitude and they are only required with images that cover more than 50 km and include all small effects due to gyro drift during the image scan. Parameters a_s and b_s includes radial ephemeris error and interior orientation errors, and they have been negligible for this purpose [22].

$$\Delta p = a_0 + a_s \cdot \text{Sample} + a_L \cdot \text{Line}, \quad (1)$$

$$\Delta r = b_0 + b_s \cdot \text{Sample} + b_L \cdot \text{Line}. \quad (2)$$

Due to the fact that there are no measured GCPs, the intersection of one point in the images has still a weak geometry. Therefore, it is necessary to use an additional DEM created with the GeoEye-1 image, acquired on December 20th, 2009; to interpolate the original elevation values obtained from the intersection of the point in the images (Z) with the elevation value obtained from the auxiliary DEM (Z_1); which helps to fix these errors and decreased even more the geometric discrepancies between images, such as position and altitude. A standard deviation of 0.6 pixels has been used for the image tie points.

3. Semi-Global Matching (SGM)

The Semi Global Matching (SGM) is a new image correlated approach, used to generate dense DEMs or digital surface models (DSM) from stereo images by matching them according to a global minimum cost function. It is developed and described in Hirschmüller [24].

SGM consists in the aggregation of a pixel wise cost function. The calculation of these cost values can be based on basic mathematical operations such as absolute value or squared value of the intensity difference [24]. However, these costs are susceptible to radiometric differences. Techniques based on mutual information (MI) are more stable and they have the best performance according to Gehrke et al. [25], since they are not sensible to changes in light emission and recording techniques [24].



FIG. 2. Test Areas A) Test Area 1 on 2006 B) Test Area 2 on 2006 C) Test Area 1 on 2011 D) Test Area 2 on 2011.

The calculation of costs based on Mutual Information is defined as the value of similarity for potential pixels to be matched in two images [26], called entropy by Hirschmüller [24] and requiring an epipolar geometry for them. That is when the scanning lines of the stereo pair are epipolar lines, which occurs when two axes of the stereoscopic camera system are parallel to each other and perpendicular to the base. The shape and size of the area to be matched is important for improving the performance, knowing beforehand that the combination is more robust when it comes to large areas [24]. For a stereoscopic pair correctly recorded, the entropy union will be low since one image can be predicted by the other. Kim et al.[27] cited in [24] established the joint entropy of the images as the sum of pixels using Taylor's transformation.

Census has been also one of the methods used to calculate the cost of combined images, proving to be robust and achieving a very good performance. The basic principle of this algorithm lies in a small processing window called kernel and a string of bits where each bit is assigned to a pixel within this window, taking the pixel a value of 1 if it has lower intensity than the central pixel. The main advantage of this transformation is its invariability to changes in the digital number, being very suitable for matching the stereo images with radiometric differences [26].

In the present study, these two methods for cost calculation have been used to determine which generates the most exact DEM with fewer gaps possible. First, they were calculated separately and then, as a sum, with 50% of

the cost calculated using mutual information, and the other 50% using census, to determine which of these three ways (only MI, only Census, sum of MI and Census) is the most adequate, considering that the purpose is to obtain the most accurate model.

If only the matching cost is analyzed, the results are ambiguous and they can easily reach a lower cost with a bad combination of images, making it necessary to include a new constraint called cost of aggregation, which penalizes the change of disparity between neighboring pixels supporting smoothness [24].

Finally, the calculation of the base image disparity and the match image, is determined in the first case by selecting for each pixel the disparity with the minimum aggregated cost, while in the second case it can be defined by traversing the epipolar line which corresponds to the image pixel. This procedure does not consider the two images in the same way, so better results are expected when a recalculation of the match and aggregation cost is done by reversing the order, the base image becomes match image and vice versa [24].

This procedure has been applied in two test areas, one located in the north part of the image, covering the urban area, where more buildings are present, with a size of 2480 columns per 3000 rows and a spatial resolution of 1 meter with the IKONOS images. The second one is located in the south of the image, covering a small constructed area, which is 1276 columns per 940 rows size. In Figure 2 are shown both test areas in 2006 and 2011. The SGM DEM calculation was carried out using the panchromatic

channel since it has the best spatial resolution of the images, but only 5 of the 6 stereo pairs allowed the generation of the corresponding DEM because the stereo pair taken on December 20th, 2009 provided for the GeoEye-1 sensor has only a difference between them of 2 seconds when the normal difference is 30 seconds, preventing to perform the SGM algorithm due the invariable difference of the point view angle.

The DEMs generated automatically by the semi-global matching method, one for each stereo pair, contained gaps in occlude areas, where the matching failed or where outliers were removed. In this study, the gaps were filled using DEMs of previous dates in case these were available as the most recent image acquired on January 7th, 2011, which was filled using the May 9th, 2011 DEM provided by GeoEye-1 with a better resolution in the panchromatic channel than the one obtained with Ikonos. A merge of more DEMs was considered for 2011 using the images obtained on January 12th, 2010 and May 2010 from Ikonos. Unfortunately, many changes were lost because during 2010 and 2011 some buildings were constructed and there were not images before 2006 that help to improve this DEM.

Merging DEMs in order to fill data gaps could be done by using the mean or the median, which are statistical parameters that best describe the data distribution in the study region. When the mean is used, the value of each pixel in the output image is calculated by averaging the elevation values from the corresponding pixels in the input images. When the median is used, the pixel values of the output images come from the middle value in ordered array of the corresponding pixels in the input images.

The mean is the optimum measurement of position when the distribution of the values is symmetrical, indicating that their tails are balanced and the existence of outliers is very unlikely. It is usually applied when the data has a normal distribution. In cases where there is no guarantee of this normal distribution, the median is a far more robust option, due to its minimum sensibility to the presence of irregular values [28]. The median has been used in this study because it has shown to better sharpen the buildings edges, key elements of this study.

4. Change detection techniques - Difference between digital elevation models

This method consists in the subtraction of two DEMs. In this case, one corresponds to the merge of DEMs generated with Ikonos and GeoEye-1 images from 2011, and the other corresponds to a single DEM from 2006. To compare the results and evaluate the improvement, the technique was repeated, but this time the subtraction was done using the same DEM from 2006 and another one acquired just by Ikonos and only from 2011.

Image subtraction is the easiest way to perform change detection but the weakest in terms of precision. It is very susceptible to false changes detection being indispensable

a pre-processing stage to refine the original images. It also requires the definition of a threshold that allows creating the binary map, where 0 means no changes and 255 means changes. Under ideal circumstances the areas of no changes would be identified by a digital value of 0, but this does not happen in reality and this threshold needs to be defined to classify those areas of significant changes and eliminate areas of no change or insignificant changes.

The threshold was decided as function of the minimum height of a buildings floor, considering that this value would represent the minimum change in a building. According to Czajkowski & Corredera [29] the minimum height of a floor is 2.70 meters plus 30 centimeters of slab and beams which gives a total height of 3 meters, adopting this values as threshold. It is necessary to determinate two thresholds, one of + 3 meters which gives the areas that increase in height, showing those new buildings existing in 2011, and another of -3 meters to determine those areas that have been decreased in height, determining those buildings that have been removed.

After applying the threshold, it is necessary to eliminate elements that are out of the main aim of this study as vegetation and water. For this purpose, the Normalized Difference Vegetation Index (NDVI) was calculated and scaled from 0 to 255 to generate a mask of the vegetation areas with all those pixels with a digital value higher than 150 and a water mask for values lower than 110 using the same index.

Using the available information, it was not possible to fill the 100% of the gaps presented in occluded areas. Therefore, it was necessary to generate a mask of all these areas to extract them from the study and avoid detecting them as false changes, especially in 2006 where no additional information was used to generate the DEM, hence more gaps are included.

Once the change areas were identified, a morphological filtering was applied to compact the result areas by using two procedures. The first consist on filling gaps in the middle of an area to form a compact one, a technique called closing. The second eliminates small details as single pixels detected as changes, called opening. Both of them were carried out using a morphological element of radius 2, since it was the one that provided the best results. A lot of single pixels were identified as changes when using small radius; constrastly when using radius which were larger than 2, the blocks of new small buildings were not identified.

5. Validation of results

After a visual or digital interpretation of an image, a verification of the quality of the results is required to estimate the precision of these results and the methodology used to generate them. It allows the user to estimate the potential risk that could be associated to decisions made based on these results. During the validation process it is necessary to compare the results with an external source that must be the faithful representation of the reality [15].

TABLE II
CONFUSION MATRIX USED IN THIS APPROACH

Change detection result	Reference data	
	Change	Non-Change
Change	TP	FP
Non-change	FN	TN

In this case, a ground truth was created manually by visual inspection of the changes between 2006 and 2011, which were digitalized and converted to a mask. Positive and negative changes were considered in different masks in both test areas. Here, the experience and skills of the analyzer plays an important role to correctly identify the changes between both images.

By using the ground truth and the results obtained from the change detection techniques, a confusion matrix was generated. In this matrix, the columns correspond to the description of the reference data while in the rows are located the results of the change detection technique. The major diagonal represents those pixels that agree in both data sets and everything off this line indicates the omission and commission errors [15].

According to Liang [30] this matrix was classified into four variables: the true positives (TP), which are the number of pixels correctly classified as change; the false positives (FP) or pixels incorrectly classified as changes; the true negatives (TN) or the number of pixels correctly classified as no change; and the false negatives (FN), which are the number of pixels incorrectly classified as no change. All of them are represented in the confusion matrix showed in table II

With these four values it is possible to calculate the accuracy of the methodology by analyzing four parameters. Completeness (cp), which represents the producer's accuracy and it is the proportion of correctly detected changes with respect to the total number of pixels that changed in the reality, showed in the reference data (Equation 3). Correctness (ct) representing the user's accuracy and it is the proportion of correctly detected as changes with respect to the total number of pixels detected as change (Equation 4). Quality percentage (qp) which is an overall measure of the results and considers completeness and correctness (Equation 5). Overall accuracy (oa) which is the proportion of correctly classified pixels with respect to the total number of pixels in the image (Equation 6). It is important to remember that both images, the result and the mask, must have the same size [30].

$$cp = 100 \times \frac{TP}{TP + FN}, \quad (3)$$

$$ct = 100 \times \frac{TP}{TP + FP}, \quad (4)$$

$$qp = 100 \times \frac{TP}{TP + FP + FN}, \quad (5)$$

$$oa = 100 \times \frac{TP + TN}{TP + FP + FN + TN}. \quad (6)$$

IV. RESULTS AND DISCUSSION

6. Pre - processing

The block adjustment technique enhanced the geometric accuracy of the 6 stereo pair images by combining multiple physical camera model parameters into a single adjustment parameter having the same net effect on the object-image relationship, as it is explained in Grodecki & Dial [22], ranging from an image residual of 42.8 pixels to 2.1 pixels in the seventh iteration (in the eighth iteration there were not more outliers to remove) by removing 985 outlier tie points of 5290 generated. Considering the total number of generated points, only 2 tie points matched all images while the great majority did it with 2, 3 or 4 images. Most points matched with the second image of the stereo pair captured on January 7th, 2011, the newest images of the set, which comprised 84.73% of all tie points, while the first image of the same pair included 78.77%. The first image of the GeoEye pair corresponding to this year had 78.94%, while the first image of this stereo pair had the fewest number of tie points, only 56.82%. This is shown in Figure 3.

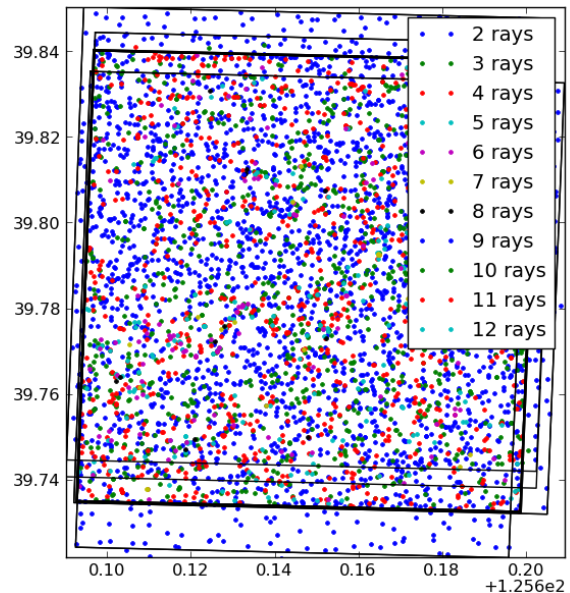


FIG. 3. Tie points between the twelve images (6 stereo pairs).

Not all the tie points were good enough to help with the correction of the images, which is the reason why some of them were removed during the different iterations of the

block adjustment procedure, improving in each iteration the image residual and the standard deviations which reduced drastically to finally obtaining an average value of 0.43 pixels. In the first iteration, the images with the lowest standard deviation were the GeoEye-1 stereo pair acquired on 2009 with a value of 6.82 pixels each one, while the ones with the higher value are the second Ikonos image of the stereo pair from 2006 and the first Ikonos image of the stereo pair from May 2010, with 67.06 and 64.65 pixels respectively. This order is not preserved in the last iteration because the smallest standard deviation still belongs to 2009 with 0.23 pixels, but the highest values correspond to the first image of 2006 and both Ikonos stereo pair of 2011 with 0.51 pixels. As expected, the same trend is observed for the mean. Minimum values ranged from 0 to 0.21 pixels in the first iteration to 0 in all images in the last iteration. Maximum values decreased from 1266.53 pixels to 2.39 pixels after the elimination of the outliers. A detailed description of each image during the first and last iteration is shown in Table III.

The distribution of the residuals finally showed a normal distribution, forming a positive asymmetric Gaussian curve as it is shown in Figure 4, going through 0 to 2.5 pixels, where most of them have residuals between 0 and 0.5 and only a small number reached residuals above this value. During the first iteration, most of the tie points were also in the range of 0 pixels but there were a group of tie points with residuals above 1000 pixels that caused a big geometric distortion in the images and were removed. The residual contrast between the remaining tie points was 2.5 pixels while in the first iteration it was 1266.53 pixels, therefore, a large reduction of the 99.8% in the residuals was achieved.

With this block adjustment, the orthorectification was performed using tie points, which is not possible to do when employing the Ikonos or GeoEye ground station, ensuring consistent quality of results because the block adjustment described by rational polynomials considers the imaging

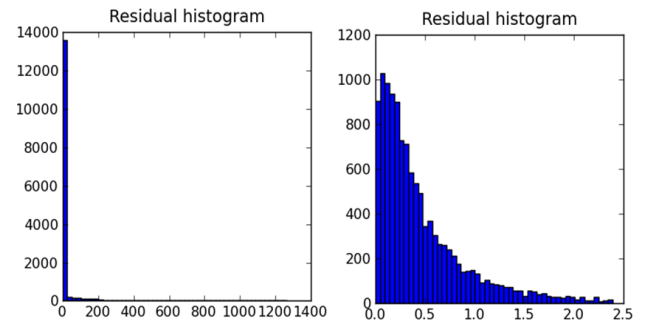


FIG. 4. Histogram of the residuals of the first iteration (left) and last iteration (right).

geometry and the satellite maneuvering modes, it does not occur with the traditional adjustment of exterior and interior orientation parameters, or the one bias compensation proposed by Grodecki & Dial [22].

7. Analysis of the changes detected with single DEMs and merged DEMs

The algorithm used to calculate the cost in the semi-global matching was Census, it provided more complete DEMs compared to Mutual information (MI) and the sum of both (MI plus Census). However, the sum of MI and Census also had a good performance, but the model generated using only MI provided a really weak result, similar to those obtained by matching images with a big difference of time and taken from different orbits (not stereo pairs) that was also tested during this study.

It was not possible to eliminate 100% of the gaps but they were filled with data of previous dates in case of 2011, what did not happen with 2006 because there were no additional images that can be used. The closest stereo pair is from 2009 but as it was mentioned in the methodology section, it was not possible to generate a DEM with those images.

TABLE III
STATISTICAL RESULTS OF THE 12 IMAGES AFTER THE FIRST AND THE LAST ITERATION

Observation	First iteration,(Pixel Units)				Last iteration,(Pixel Units)			
	mean	Std	Min	max	mean	Std	min	max
ge2009_1	2.61	6.82	0	86.71	0.29	0.23	0	1.57
ge2009_2	2.61	6.82	0	86.71	0.29	0.23	0	1.57
ge2011_1	13.01	36.9	0.01	287.33	0.46	0.46	0	2.39
ge2011_2	12.02	35.45	0.01	232.71	0.37	0.37	0	2.34
ik2006_1	14.05	36.82	0.21	477.35	0.53	0.51	0	2.39
ik2006_2	17.81	67.06	0.01	1266.53	0.5	0.5	0	2.36
ik2010.1_1	14.58	40.79	0	239.89	0.45	0.43	0	2.38
ik2010.1_2	13.06	37.54	0	263.33	0.45	0.46	0	2.28
ik2010.5_1	16.36	64.65	0	1233.33	0.48	0.45	0	2.39
ik2010.5_2	14.22	43.25	0	473.59	0.47	0.46	0	2.38
ik2011_1	12.36	33.68	0.01	260.39	0.65	0.51	0	2.35
ik2011_2	10.27	32.36	0.01	240.91	0.59	0.51	0	2.39

Merging two years of information to get a complete DEM allowed to improve the completeness of the DEM by filling the gaps where the occlusion of certain areas did not allow a good matching. In both test areas, these results are much more complete than the one obtained only using images from 2011. Unfortunately, this merge uses a statistical value to fill the gaps that is calculated for every single pixel of the DEM, including those pixels that already have information and should be unchangeable and stable. With this procedure it does not occur because it is changed to a new value calculated as the median of the common pixels in all images to be merged, losing information in those areas where new buildings were constructed between 2010 and 2011, causing that some major changes get lost, that is why a merge of only images from 2011 will be used to avoid the introduction of omission errors.

For both test areas, the improvement between the proposed merge technique and the simple difference was more evident in the positive than in the negative changes due to the fact that the latest were less frequent.

The positive and negative changes would be analyzed separately considering that two masks were created for each test area. For test area 1, the total number of analyzed pixels is 7440000 and for test area 2 it is 1199440. All of those distributed in correct detections, in commission errors and omission errors as showed in the confusion matrix. For both test areas, the improvement between the proposed merge technique and the simple difference is most notorious in the positive changes than in the negative changes due to the fact that there are not so many of these changes between this period of time.

In test area 1, positive changes occurred between both dates (Figure 5), comparing the proposed technique vs. the not merge subtraction, the completeness decreased 18.84%. Because the total number of pixels detected as changes decrease from 3.36% to 1.06% with the new technique, the number of correct changes detected also decreased.

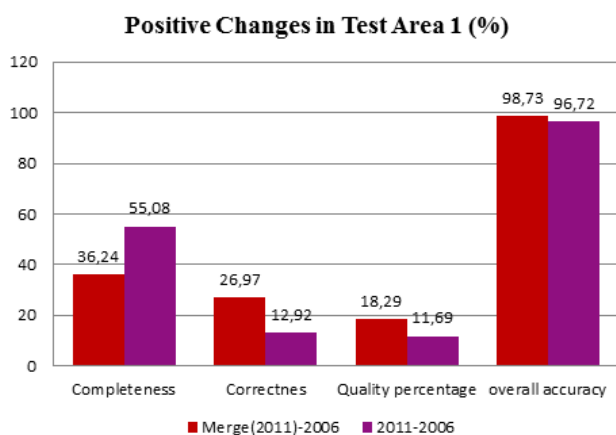


FIG. 5. Comparison between the merging difference technique and simple difference of the positive changes in test area 1.

The correctness, quality percentages and overall accuracy increased 14.05%, 6.6% and 2.01% respectively due the fact that there was a reduction of 14% in the number of pixels incorrectly detected as changes with respect to the number of pixels detected as change (Figure 6 A and B). In the negative changes of test area 1, the improvement was not as visible as in the positive changes as is showed in Figure 7. Completeness increased 2.12%, while correctness, quality percentage and overall accuracy stayed almost the same with an imperceptible rise of 0.37%, 0.37% and 0.13% respectively. These results were not detected visually (Figure 6 C and D). The number of pixels correctly detected as well as the number of pixels incorrectly detected remained almost the same.

In test area 2, the positive changes had a similar behavior to the positive changes in test area 1, because the completeness also decreased 11.71% due the fact that there was a reduction from 5.16% to 1.83% in the number of pixels detected as change with respect to the total number of pixels. The correctness, quality percentage and overall accuracy showed in Figure 9, increased 4.58%, 0.84% and 3.24% respectively. In this case, the reduction of the pixels incorrectly detected as changed ranged from 96.21% to 91.62% with respect to the total detected as change (Figure 8 A and B). The lost buildings have not been detected by neither of the two methods since they are of small height and the established threshold did not permit to detect them as negative changes; it provided a completeness and correctness of zero. This is shown in Figures 9 and 10. However, the commission error in the suggested alternative using the merge technique, is lower than the error obtained with no merging (Figure 8 part C and D)

After a visual analysis of the change detection results, it can say that most of the commission errors (incorrect changes detected) correspond to shadowed areas, causing a correctness and completeness in both methods below 50% as shown in the statistics. Also, there were some vegetation areas that could not be removed properly with the vegetation mask because the spectral behavior of the roofs of some buildings was similar to this land cover preventing the possibility of applying an appropriate threshold based on NDVI, so it was necessary to establish an additional condition which allowed to mask buildings in conflict but also some areas with sparse vegetation. These last zones and the shadows finally increase the number of false positive pixels.

However, The technique used in this study, demonstrated to better handle these areas reducing considerable the number of wrong detected positive changes in test area 1 which ranged from 3.3% of incorrect detected pixels without merging DEMs to 1.27% in the proposed merging technique. In test area 2, the reduction of incorrect positive changes is even bigger going from 5.13% to 1.89%.

Areas incorrectly detected as changes correspond to small areas below the spatial resolution of the DEM or when the difference in elevation is lower than 3 meters,

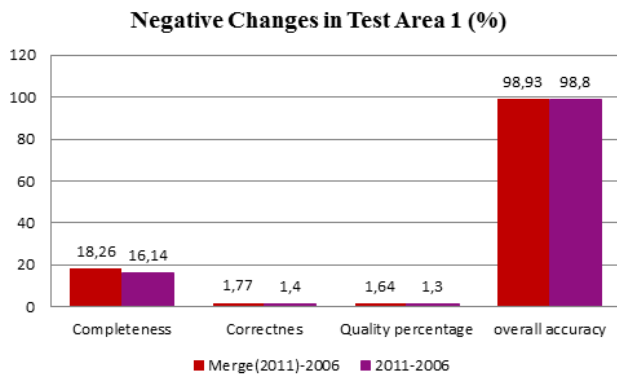


FIG. 7. Comparison between the merging difference technique and simple difference of the negative changes in test area 1.

values that has been taken as threshold in the methodology. When using smaller threshold values, the correctness decreases considerably because the number of wrongfully detected changes both increase. In contrast, if the threshold values are higher, the completeness decreases since there are many important changes that go undetected.

Unfortunately, despite the overall accuracy is high, it is not sufficient to assess the accuracy of the results because

it takes into account the total number of pixels correctly classified and it do not consider the good performance in individual classes. In this study, we have two classes, changes class and no-changes class. Although the changes class do not show the expected precision, the no-changes class does it. It can be demonstrated with the percentage of pixels detected as no changes in positive and negative changes in both test areas, all this with respect of the total number of pixels that according to the ground true, must be detected as no changes. In test area 1, where the positive changes were evaluated, the improvement went from 97.05% to 99.22% and with the negative changes, it went from 98.89% to 99%. In test area 2, the improvement went from 98.01% to 98.31% for positive changes and 98.1% to 99.2% for negative changes.

The important values to be considered in the evaluation of the change class, are the correctness and completeness. In both methods these values are really low (Figures: 5, 7, 9, 10) to evidence the success of one technique over the other. The change detection technique based in the difference between digital elevation models from 2011 and 2006 has an overall week performance even thought a geometric correction was done before processing them and a threshold was set to keep only the relevant changes that occurring in these areas. However, if the desirable in the application is

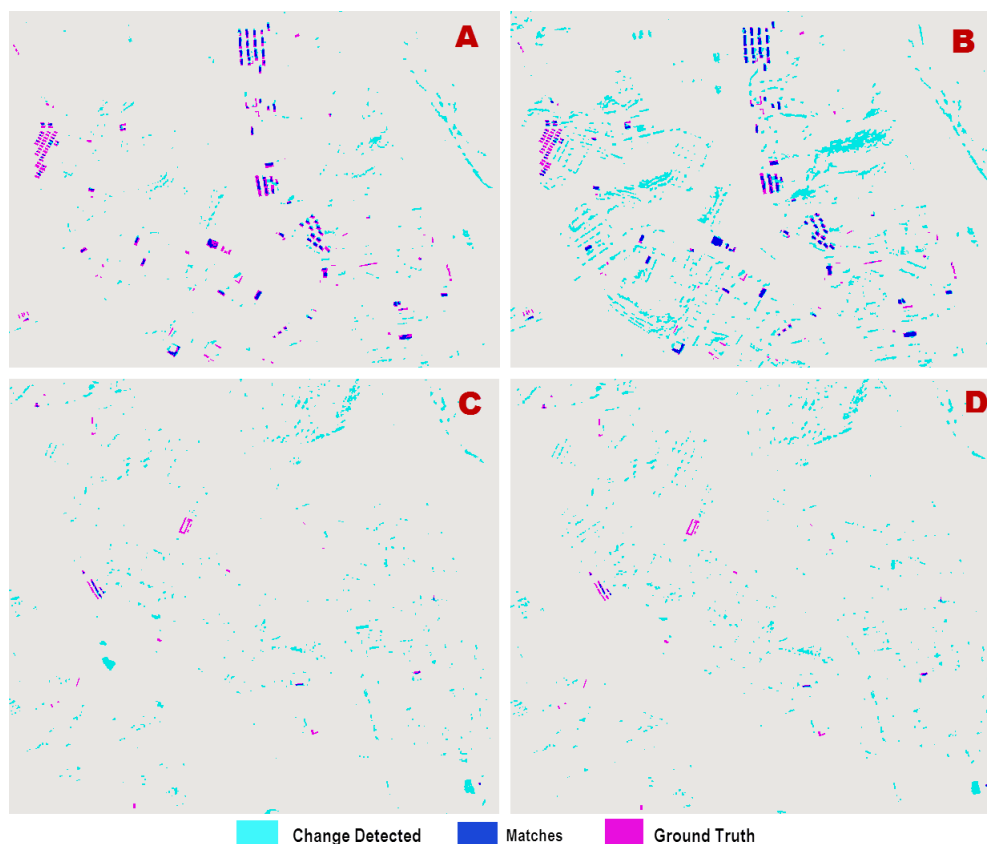


FIG. 6. Changes detected in test area 1: A) positive changes using the proposed technique, B) positive changes using the simple difference, C) negative changes using the proposed technique, B) negative changes using the simple difference.

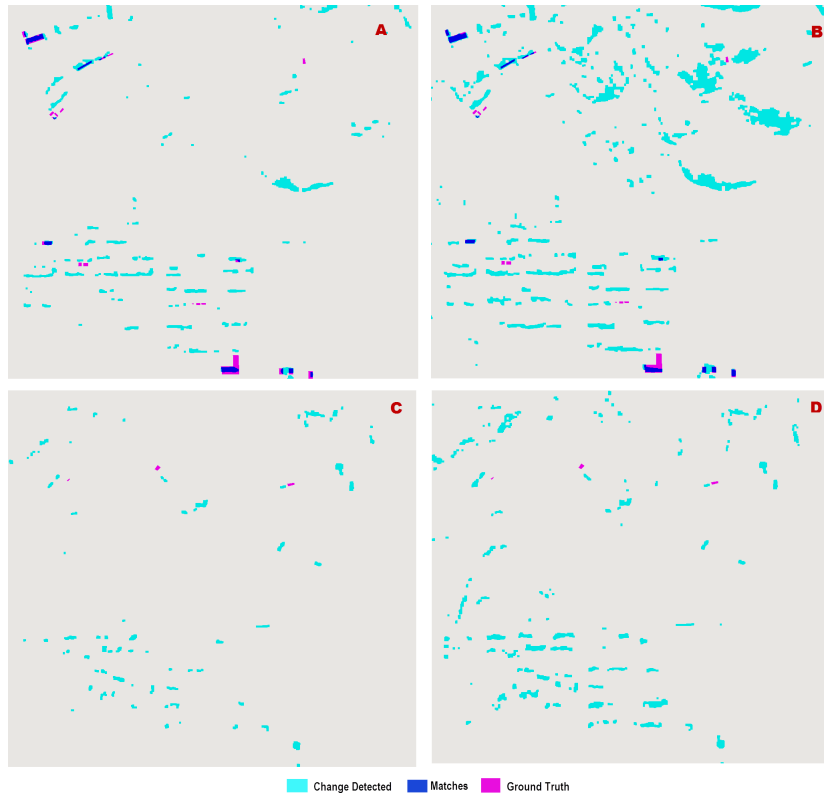


FIG. 8. Changes detected in test area 1: A) positive changes using the proposed technique, B) positive changes using the simple difference, C) negative changes using the proposed technique, B) negative changes using the simple difference.

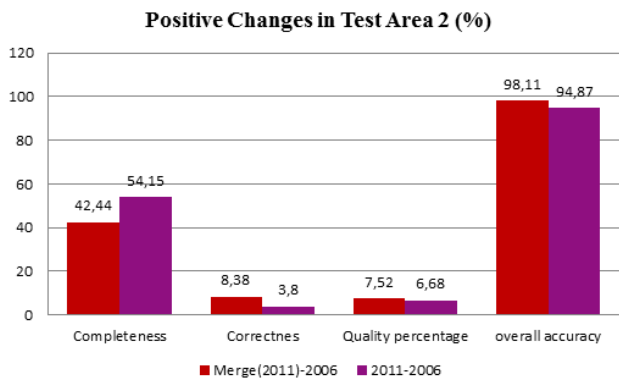


FIG. 9. Comparison between the merging difference technique and simple difference of the positive changes in test area 2.

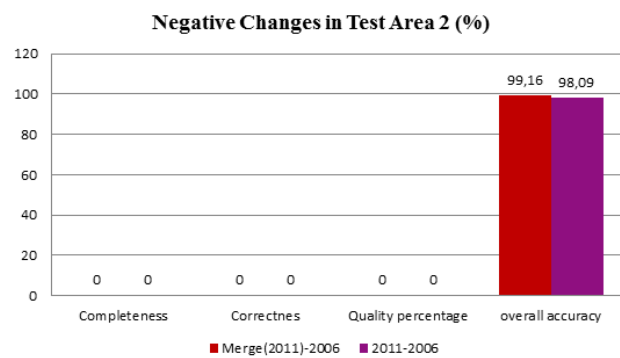


FIG. 10. Comparison between the merging difference technique and simple difference of the negative changes in test area 2.

to improve the correctness, the proposed technique shows to have a good performance.

V. CONCLUSIONS

The results obtained are strongly linked to the accuracy of the automatic algorithm used to generate DEMs. The gaps presented in this product due the difficult to match the two stereo images in areas with the presence of shadow or in places where its vision is hampered, will be finally detected as changes. In case the gaps in the first date are

not removed, the positive changes would increase, while the negative changes would increase if the gaps in the newest date are not removed. To eliminate positive or negative commission errors, these occluded areas should be masked out.

The merging between DEMs is a good alternative to complete them because the gaps are filled with an approximate value provided from the auxiliary data. The quality and precision of this extra data plays an important role, providing best results if the difference in time between them is the lowest possible and no major changes occur between

them, otherwise the resulting DEM would lose precision instead of improve it and the omission error would increase. A good alternative would be to use information from the same year or closer years depending on the urban growth rate, if this is going fast, even merging digital elevation models from the same year would introduce omission errors.

The technique proposed was only proved in the newest date because there was no information to merge with the oldest one. To improve the results would be appropriate to test merging the DEM from 2006, being necessary to select the auxiliary images adequately considering the closer date to this one to eliminate the possibility of loss important changes during the merge process. It is important to remember that the number of DEM used in the merge procedure does not have a limit, so more than one stereo pair could be used as additional information for each year.

For a better performance of this technique, an appropriate masking of the shadows generated by the buildings inside of urban areas is required. These would decrease the number of commission errors.

For a better performance of this technique, an appropriate management of the shadows generated by the buildings inside of urban areas is required. A possibility would be the extraction of them by a mask as it was done with the vegetation, water and gaps areas. These would decrease the number of commission errors.

Further investigations should include testing the proposed methodology with DEMs from other data sources such as the SRTM mission, which has a complete high resolution digital topographic database of the Earth. Also the implementation of other change detection techniques could be tested in order to see if the completeness and correctness could be improved.

REFERENCES

- [1] M. Lehner and W. Kornus, "Band to band registration for the german pushbroom scanner moms-2p," in *Proceedings of ISPRS Joint Workshop on "Sensors and Mapping from Space "1999*, pp. 27–30, 1999.
- [2] D. Poli *et al.*, "Georeferencing of moms-02 and misr stereoimages with strict sensor model," in *International Society for Photogrammetry and Remote Sensing, Workshop "High resolution mapping from space*, Citeseer, 2003.
- [3] S. I. C. (SIG), "Ikonos satellite images and sensor specifications," 2011.
- [4] K. Kliparchuk, A. Gisp, D. Collins, *et al.*, "Evaluation of stereoscopic geoeye-1 satellite imagery to assess landscape and stand level characteristics," 2010.
- [5] DigitalGlobe, "Worldview-1," 2011.
- [6] DigitalGlobe, "Worldview-2," 2011.
- [7] P. Cheng and C. Chaapel, "Dem generation using quickbird stereo data without ground controls using tie points only," *Geoinformatics*, vol. 9, no. 2, pp. 36–38, 2006.
- [8] H. Raggam, "Accuracy analysis and surface mapping using spot 5 stereo data," in *International Archives of Photogrammetry and Remote Sensing, XXth ISPRS congress, Istanbul*, pp. 12–23, 2004.
- [9] P. Reinartz, M. Lehner, R. Müller, and M. Schroeder, "Accuracy analysis for dem and orthoimages derived from spot-hrs stereo data without using gcp," *International Archives of Photogrammetry and Remote Sensing*, vol. 34, pp. 433–438, 2004.
- [10] P. Radhadevi, K. Jacobsen, V. Nagasubramanian, and M. Jyothi, "Comparison of digital elevation models generated from spot-5 hrs stereo data and cartosat-1 stereo data.," in *ISPRS Hannover Workshop*, pp. 2–5, 2009.
- [11] T. Toutin, "Three-dimensional topographic mapping with aster stereo data in rugged topography," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 40, no. 10, pp. 2241–2247, 2002.
- [12] G. Buyuksalih, K. Jacobsen, and I. Baz, "Dem generation based on optical space images," 2008.
- [13] P. G. Akl, "Temporal change detection using aster and usgs digital elevation models," 2005.
- [14] K. Jacobsen, "Dem generation from satellite data," *EARSeL Ghent*, pp. 273–276, 2003.
- [15] E. Chuvieco, "Teledetección ambiental (primera edición en presentación actualizada.)," *España: Editorial Ariel*, 2010.
- [16] J. Tian, H. Chaabouni-Chouayakh, P. Reinartz, T. Krauß, and P. d'Angelo, *Automatic 3D change detection based on optical satellite stereo imagery*. na, 2010.
- [17] L. Zhang, "Automatic digital surface model(dsm) generation from linear array images," *Mitteilungen- Institut für Geodäsie und Photogrammetrie an der Eidgenössischen Technischen Hochschule Zurich*, 2005.
- [18] P. Reinartz, R. Müller, M. Lehner, and M. Schroeder, "Accuracy analysis for dsm and orthoimages derived from spot hrs stereo data using direct georeferencing," *ISPRS journal of photogrammetry and remote sensing*, vol. 60, no. 3, pp. 160–169, 2006.
- [19] D. Akca, A. Grün, *et al.*, *Least Squares 3D surface matching*. Inst. für Geodäsie und Photogrammetrie, 2007.
- [20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [21] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *International Journal of Computer Vision*, vol. 80, no. 2, pp. 189–210, 2008.
- [22] J. Grodecki and G. Dial, "Block adjustment of high-resolution satellite images described by rational polynomials," *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 1, pp. 59–68, 2003.
- [23] J. Grodecki, "Ikonos stereo feature extraction-rpc approach," in *ASPRS annual conference St. Louis*, 2001.

- [24] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 328–341, 2008.
- [25] S. Gehrke, K. Morin, M. Downey, N. Boehrer, and T. Fuchs, "Semi-global matching: An alternative to lidar for dsm generation," in *Proceedings of the 2010 Canadian Geomatics Conference and Symposium of Commission I*, 2010.
- [26] P. d'Angelo and P. Reinartz, "Semiglobal matching results on the isprs stereo matching benchmark," in *ISPRS Hannover Workshop 2011: High-Resolution Earth Imaging for Geospatial Information*, 2011.
- [27] J. Kim, V. Kolmogorov, and R. Zabih, "Visual correspondence using energy minimization and mutual information," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 1033–1040, IEEE, 2003.
- [28] M. Carbonero, J. Ramírez, C. Hervás, and D. Ortiz, "Estimadores robustos de estadísticos de posición," 2005.
- [29] J. Czajkowski and C. Corredera, "Ahorro de energía en refrigeración de edificios para viviendas y propuesta de indicadores de eficiencia y valores admisibles," *Avances en energía renovable y medio ambiente*, vol. 10, 2006.
- [30] W. Liang, *Change detection for reconstruction monitoring using very high resolution optical data*. PhD thesis, DLR, 2010.

Recibido: 13 de enero de 2017

Aceptado: 03 de marzo de 2017

Sandra Cobos: Computer Science Engineer with a master degree in Geographical Information Systems. Cobos is working as full time professor and researcher at Universidad Católica de Cuenca. She also was group leader of the Spatial Data Infrastructure department in the Ecuadorian Space Center and research intern at German Aerospace Center.