An experimental comparison of standard stereo matching algorithms applied to cloud top height estimation from satellite IR images

Anna Anzalone[1] and Francesco Isgrò[2] for the JEM-EUSO collaboration

[1]INAF - IASF, Istituto di Astrofisica Spaziale e Fisica Cosmica, via U. La Malfa 153, Palermo, Italy and INFN - Sezione di Catania, Italy;
[2]DIETI, Università degli Studi di Napoli Federico II, via Claudio 21, Napoli, Italy and INFN - Sezione di Napoli, Italy

ABSTRACT

The JEM-EUSO (Japanese Experiment Module-Extreme Universe Space Observatory) telescope will measure Ultra High Energy Cosmic Ray properties by detecting the UV fluorescence light generated in the interaction between cosmic rays and the atmosphere. Cloud information is crucial for a proper interpretation of these data. The problem of recovering the cloud-top height from satellite images in infrared has struck some attention over the last few decades, as a valuable tool for the atmospheric monitoring. A number of radiative methods do exist, like C02 slicing and Split Window algorithms, using one or more infrared bands. A different way to tackle the problem is, when possible, to exploit the availability of multiple views, and recover the cloud top height through stereo imaging and triangulation. A crucial step in the 3D reconstruction is the process that attempts to match a characteristic point or features selected in one image, with one of those detected in the second image. In this article the performance of a group matching algorithms that include both area-based and global techniques, has been tested. They are applied to stereo pairs of satellite IR images with the final aim of evaluating the cloud top height. Cloudy images from SEVIRI on the geostationary Meteosat Second Generation 9 and 10 (MSG-2, MSG-3) have been selected. After having applied to the cloudy scenes the algorithms for stereo matching, the outcoming maps of disparity are transformed in depth maps according to the geometry of the reference data system. As ground truth we have used the height maps provided by the database of MODIS (Moderate Resolution Imaging Spectroradiometer) on-board Terra/Aqua polar satellites, that contains images quasi-synchronous to the imaging provided by MSG.

Keywords: Stereo-Vision-Algorithm, Cloud-Top-Height, Matching, Meteosat, Infra Red images

1. INTRODUCTION

Meteorologists and climatologists have studied clouds, for their key role in the Earth's climate system and their relevant impact on the radiative processes. Cloud Top Height (CTH) estimation in particular, provides information on cloud vertical structure, and improves the knowledge of the cloud's radiative effects. Reliable estimation of CTH is crucial, not only for meteorologists, but it is also matter of interest in very different frameworks such as the JEM-EUSO (Japanese Experiment Module-Extreme Universe Space Observatory) mission,[1–4] whose main objective is to detect, from a telescope[5] placed on the the International Space Station (ISS), Extreme Energy Cosmic Rays (EECR)[6] with energy above $5 \cdot 10^{19}$eV. These rays can be detected through the showers they produce in the atmosphere.[7,8] Aim of the JEM-EUSO telescope observations, is to record the time evolution and the intensity of the fluorescence light, produced in the spectral band 300 - 400 nm during its interaction with the atmosphere.

When an EECR particle reaches the atmosphere, it produces a nuclear interaction which leads to the generation of a cascade of billions of particles, such as electrons, photons, muons and neutrinos, which form the so-called

Further author information: (Send correspondence to Anna Anzalone)
Anna Anzalone: E-mail: anna.anzalone@ifc.inaf.it, Telephone: +39 091 6809476
Francesco Isgrò: E-mail: francesco.isgro@unina.it, Telephone: +39 081 679308

Extensive Air Shower (EAS). The electrons, continuously produced during the evolution of the cascade in the atmosphere, excite the nitrogen molecules in air and produce fluorescence light. The maximum light intensity is at 5-10 km above the Earth's surface, therefore the presence of clouds above the phenomena can affect the detection, and the influence depends on the cloud top altitude. To monitor the cloud coverage, an on purpose Atmospheric Monitoring System has been designed that will include an infrared (IR) camera,[9] and a LIght Detection And Ranging (LIDAR) device,[10] together with the support of global atmospheric models generated from the analysis of all available meteorological data provided by global weather services, such as the European Centre for Medium-range Weather Forecasts.

The design of the infrared camera is taking into account the possibility of using several different ways to retrieve the required atmospheric information, such as the most common methods based on radiative transfer and also the stereoscopic method,[11,12] where the latter can be applied exploiting the motion of the ISS (see Figure 1): two consecutive observations can be acquired at different and relatively close instants in such a way that the two images partially overlap, i.e., they are images of the same scene. In this way an stereo pair of images of the same scene from different view points is available.

Stereo methods rely on information that can be derived from the images, and on the knowledge of the main parameters of the stereo system,[13,14] therefore it can be considered purely geometric. Conversely, the more common standard radiative methods,[15–18] exploit several ancillary atmospheric data: temperature or pressure profiles, derived cloud temperatures, etc. The possibility of using stereo methods in remote sensing for CTH retrieval has been successfully explored over the last 15 years.[17,19–23] This method represents a completely different approach in comparison with the frequently used radiative algorithms. It can be applied whenever multiple views are available, and can be suitable for those cases in which a large amount of meteorological information are not available. Nevertheless, these methods do not exclude each other, in fact they can be also used in a synergistic way, as shown in.[17] Comparisons between stereo and radiative methods,[24,25] have shown good confidence in the values that can be retrieved by means of stereoscopy.

A crucial step for the reconstruction of a tridimensional scene from a stereo pair is the matching.[13] Most of the literature on CTH retrieval from stereo, adopts ad hoc algorithms. For instance, in[23,26] it is exploited the fact that the input are infrared images of cloud scenes; the geometry of the camera system adopted in the mission is exploited in the multi-camera method presented in.[20] In this paper we follow a different approach, and investigate how more classical area based stereo matching methods, ordinarily used in computer vision with standard images in the visible domain, perform with infrared satellite image for CTH estimation.

These algorithms have been tested[27] with a set of images for which it was possible to measure a ground truth for the matching. When working with satellite images of cloudy scenes, it is very difficult to obtain such a ground truth, and in a previous work[28] we run our experiment measuring the error only on ground pixels, for which an elevation map is known. In this paper we move the work a step forward, and try to test the performance of the same algorithms on the cloudy pixels; in particular we test how the retrieval of the CTH is affected by the different parameters on which these matching algorithms depend. To this end, the resulting disparity maps were transformed into final maps of CTH, according to the geometry of the reference data system. Finally, they were compared with a set of height maps obtained by other imaging sensors, that were chosen as ground truths. Real IR data from the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) on board Meteosat Second Generation (MSG) geostationary satellites, were used as images to test the algorithms, while the heights of the corresponding images from MODerate resolution Imaging Spectroradiometer (MODIS) sensor, installed on the polar satellite Terra, were used as ground truths.

The paper is structured as follows. In Section 2 a brief introduction on the most diffuse approaches used to deal with the problem of correspondence between images, is given, together with a brief description of how the stereo reconstruction works. In Section 3 the nature of the data used in this paper is briefly described. The experiments are reported and discussed in Section 4. Finally Section 5 is left to some final remarks.

2. STEREO MATCHING

The 3D structure of an object in a standard binocular vision system,[13,14] is inferred from the analysis of two images acquired by spatially separated cameras. Estimation of the resulting parallax effect contributes to reconstruct its depth, i.e., the distance of the object from the sensor. The parallax effect on the image is an apparent

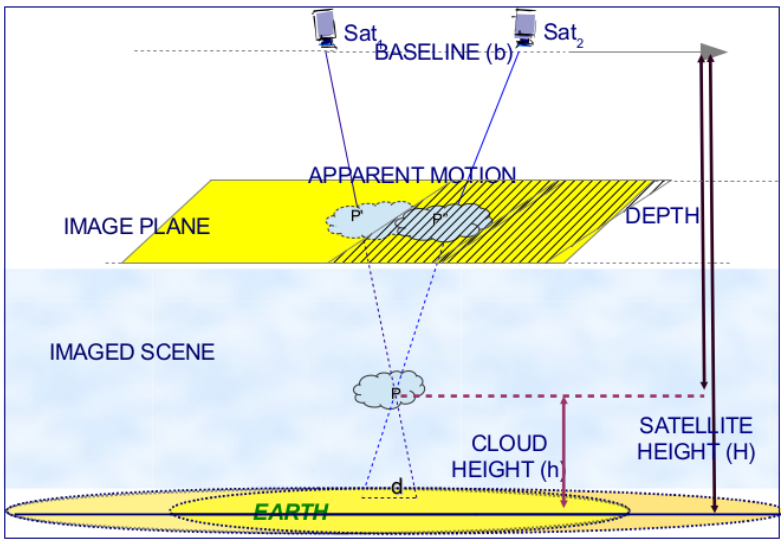


Figure 1. Stereo Reconstruction. The scheme shows the height reconstruction for the cloudy pixel P. Its projection points ($P^{'}$ and $P^{"}$) are detected by the matching step and the depth is recovered by triangulation. Finally the height is calculated by subtracting these values from the satellite altitude.

motion of the object from the first image of the pair to the second one. This motion is known as disparity, and it is constrained by the geometry of the stereo system. This is a crucial step and in this paper a set of standard known area-based techniques for disparity estimation are settled in order to find the best performance for images from space.

2.1 Approaches to stereo matching

Passive stereo remains one of the fundamental technologies for estimating 3-D information. It is desirable in applications because it requires no modifications to the scene, and because dense information (that is, at each image pixel) can nowadays be achieved at video rate on standard processors for medium-resolution images (e.g., CIF, CCIR). Large-baseline stereo systems, generating significantly different images and therefore making the matching more difficult, is particularly important as it is not always possible to position cameras close enough to achieve small baselines. There are two classes of correspondence algorithms, seeking to achieve, respectively, a sparse set of corresponding points (yielding a sparse disparity map) or a dense set (yielding a dense disparity map). Determining a sparse set of point matches among the images, is a key problem for multi-view analysis. It is usually performed as a first step to calibrate (fully or weakly) the system, when anything about the geometry of the imaging system is known. At this stage then, no constraint can be used in order to help the matching.

The algorithms in the first category select feature points independently in the two images, then match them using tree searching, relaxation, maximal clique detection or string matching.[29–31] Particular robustness is achieved in[32] through a three-step algorithm. A different algorithm is given in,[33,34] where the author present an algebraic approach, based on position and correlation measure.

Algorithms in the second category select templates in one image (usually patches with some texture information), then look for corresponding points in the other image using some similarity measures.[13,14,35,36] The algorithms in this class tend to be slower than the ones in the first class, as the search is less constrained. The search for matches between two images, is sped up enormously if the two images are warped in such a way that corresponding points lie on the same scanline in both images (in other words, the epipolar lines are the same in the two images and parallel to the horizontal image axis). This process is called rectification.[37,38] The rectified images can nearly invariably be regarded as acquired by cameras rotated with respect to the original ones. Most of the stereo algorithms reported assume rectified images.

For cloud images it is a challenge detecting cloud features suitable for the algorithms of the first category. Features normally used, such as edges and corners, are generally smoothed in cloud images and change their

appearance from different viewing angles. Therefore we focus our attention on area-based approaches. Dense stereo matching is a well-studied topic in image analysis.[13,14] An excellent review including suggestions for comparative evaluation is given in.[27] We refer the reader to this paper for an exhaustive list of known algorithms.

The output of a dense matching algorithm is a correspondence or disparity map, which associates to each pixel in one image, its displacement (disparity) vector with respect the one of the other images. As already mentioned, the matching point must satisfy geometric constraints as the epipolar[13] one, in the case of two views. However, some other constraints (physical and photometric) are imposed, including:

Similarity the image patches of corresponding pixels must be similar.

Order if two points in two images match, then matches of nearby points should maintain the same order. Notice that this constraint can fail, for instance for points on different surfaces.

Smoothness the disparities should change smoothly with the distances from the camera. This is not always valid, of course, along region boundaries.

Uniqueness each pixel cannot match more than one pixel in any of the other images.

2.2 Steps of a matching algorithm

For the purpose of the experimental evaluation we adopted a few algorithms well known among the computer vision community. According to,[27] four steps for a general stereo matching algorithm can be considered:

1. matching cost computation
2. cost aggregation
3. disparity computation and optimisation
4. disparity refinement.

Needless to say that the step sequence can be different, depending on the particular algorithm considered.

The most traditional matching cost functions are the absolute intensity difference[39] (AD), and the squared intensity difference[40] (SD). These cost functions have been used for decades in the computer vision community, and although more complex matching functions have been proposed, they are still considered for their simplicity which brings optimal performance in terms of speed. The cost functions, during the aggregation, can be sensitive to pixels with a too large intensity, therefore truncation values are sometimes used to make them more robust.[27] Various cost functions have been proposed in the more recent literature, and among them we mention the one proposed by Birchfield and Tomasi in[41] (BT), where each pixel in the first image is compared against a linearly interpolated function of the second image, instead of using shifts by integral amounts; for the comparison any of the SD or AD cost function can be used.

The aggregation of the cost function for a pixel (x, y) is traditionally performed summing or averaging over a squared neighbourhood of the pixel itself. For the two cost functions mentioned above, the standard aggregations are the sum of squared intensity differences (SSD), and the sum of absolute intensity differences (SAD). For instance, the SSD for a pixel is defined as

$$SSD(x, y, d) = \sum_{j=-w}^{w} \sum_{i=-w}^{w} (I_l(x + j, y + i) - I_r(x + d + j, y + i))^2$$

where I_l and I_r are the two images, d is the disparity for which the cost function is being estimated, and $2 * w + 1$ is the size of the neighbourhood. If a truncation value σ is used the formula above is written as

$$SSD(x, y, d) = \sum_{j=-w}^{w} \sum_{i=-w}^{w} \min((I_l(x + j, y + i) - I_r(x + d + j, y + i))^2, \sigma^2)$$

Other methods for the aggregation do exist (see[14,27] for further references), for the work presented in this paper, it is worth mentioning the shiftable windows,[42] the iterated binomial filter,[43] regular and membrane diffusion.[44]

For the third step, we can distinguish between local and global methods. Local methods compute the disparity following a winner-take-all optimisation,[27] as the core of these algorithms is in the first two steps. Therefore, they simply choose as best match the one with the highest matching score. For instance, using again the SD cost function, we have that the disparity is computed as

$$d(x, y) = \min_d SSD(x, y, d)$$

Conversely, global methods may sometimes skip the aggregation step at all. Most of the algorithms in this class are formalised as an energy minimisation problem, i.e., they aim to find a disparity function minimising some energy function. Among the various methods proposed in literature for locating the minimum for the energy function, we mention dynamic programming (DP), scan-line optimisation and graph cuts.[27]

Finally, the last step is introduced as the disparities are usually computed in a discrete space, and for some application (e.g., image based rendering, augmented reality) higher accuracy can be necessary. For this reason some algorithms do apply a further step of sub-pixel refinement, after the discrete matching has been performed. Sub-pixel disparity can be achieved in many different ways, and we refer the reader to[27] for further details.

3. DATA

The experiments discussed in this paper were run on a set of data, selected from the Meteosat/SEVIRI database. Cloudy image pairs were used to the aim of recovering the combination, that outputs the best disparity map. The quality of the recovered CTH maps has been tested against the height maps produced by MODIS.[45]

3.1 Stereo data

In this paper we have considered a geostationary stereo system given by the combination of the Meteosat/SEVIRI satellites MSG-2 and MSG-3, located on the Equatorial plane respectively at 9.5°N and 0°E at about 36,000 km of altitude. MSG-3 provides full disc imagery every 15 minutes of Europe and Africa, whereas the Rapid Scan System of MSG-2 delivers images every 5 minutes over the Northern Hemisphere. Parts of Europe, Africa and adjacent oceans are then observed from two different sights of view, and with a pixel resolution of 3 km at the Sub Satellite Point (SSP). The distance between MSG-2 and MSG-3 positions, i.e., the baseline, is $\approx$ 7000 km, that is insufficient to provide an accurate reconstruction for all cloud heights. This occurs because Meteosat satellites were not planned to be a stereo system. Following,[22] the accuracy of the CTH (σ_{CTH}) estimated from a stereo system can be expressed as:

$$\sigma_{CTH} = \frac{\sigma_d}{b/H_s} \tag{1}$$

where σ_d is the disparity estimation error at the ground, b is the baseline of the stereo system and H_s is the satellite altitude. Using Equation 1 and considering an average disparity error of no more than half a pixel, i.e., $\sigma_d = \pm 1.5$ km, the accuracy results within 7.6 km. Hence, if the disparity is estimated with an accuracy of half a pixel, in the Meteosat configuration clouds with heights $h \in$[7.5,15] km might not be distinguished between each other, as well as lower clouds might be merged to the background. The two imaging devices are quasi-synchronous, and ancillary data include also the actual satellite positions for each row of the images utilised for the CTH estimation.

We run our experiments on selected parts of several MSG-2/MSG-3 pairs. The parts of the images were selected among the cloudy ones for which corresponding data from MODIS, were available within a reasonable time delay. In this paper we report only the results relative to a stereo pair obtained from data acquired on February the 6th, 2015. The stereo pair is shown in Figure 2.

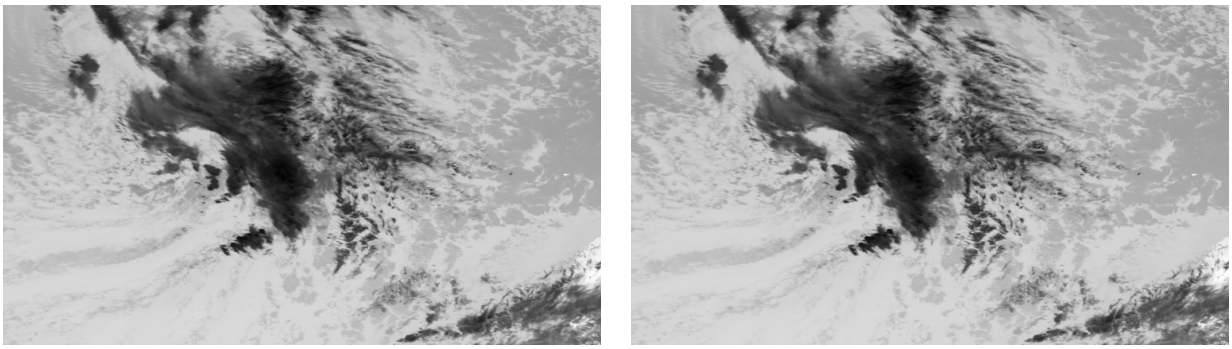


Figure 2. Stereo pair used for the experiments. On the left the MSG2 image and on the right the MSG3 image, are shown. Darker colours represent pixels with lower temperature, that is points with larger height.

3.2 Cloud top height data

The MODIS[46] images on the Terra and Aqua Earth Observing System platforms, provide the capability for globally retrieving these properties using passive solar reflectance and infrared techniques. In addition to providing measurements similar to those offered on a suite of historical operational weather platforms, such as the Advanced Very High Resolution Radiometer (AVHRR), the High-resolution Infrared Radiation Sounder (HIRS), and the Geostationary Operational Environmental Satellite (GOES), MODIS provides additional spectral and/or spatial resolution in key atmospheric bands, along with on-board calibration, to expand the capability for global cloud property retrievals.

The core MODIS operational cloud products include cloud top pressure, thermodynamic phase, optical thickness, particle size, and water path, and are derived globally at spatial resolutions of either 1 or 5 km (referred to as Level-2 or pixel-level products). In addition, the MODIS atmosphere team (collectively providing cloud, aerosol, and clear sky products) produces a combined gridded product (referred to as Level-3) aggregated to a 1° equal-angle grid, available for daily, eight-day, and monthly time periods. The wealth of information available from these products provides critical information for climate studies as well as the continuation and improved understanding of existing satellite-based cloud climatologies derived from heritage instruments. Recently,[45] the data products have been enriched with cloud macrophysical properties including cloud-top pressure, temperature and height and, cloud thermodynamic phase.

In Figure 3 it is shown the CTH map produced by MODIS relative to the stereo pair in Figure 2. It appears clear for a simple visual inspection that the area covered by the MODIS data is smaller than the one covered by the stereo pair. This is not a particular problem, and it will be taken into account when computing the quality metrics for the experimental evaluation.

It is worth pointing out that, although we are using the CTH maps as ground truth in our experiment, they are actually measures that are prone to errors. Moreover the resolution of the MODIS data is different from the one of the MSG data, therefore a re-sampling is needed for the comparison, and this, as it is well known, does have a negative effect on the quality of the maps.

4. EXPERIMENTS

In this Section, we describe the evaluation strategy and quality metrics we used for evaluating the performance of stereo correspondence algorithms. Furthermore, we describe and discuss the experiments used to evaluate the effect of the different blocks of a stereo algorithm, on the estimation of the CTH map.

4.1 Evaluation strategy and quality metrics

A quantitative evaluation of the performance of the algorithms is obtained measuring the effects of the variation of a set of parameters on the quality of the CTH map. Since we are interested in understanding the behaviour

of the stereo matching on the the cloudy pixels, we neglect pixels at the ground level. This can be easily done from an analysis of the MODIS altitude maps.

For a quantitative analysis of the results we compute two quality measures based on ground truth data:

1. Root mean squared error (RMS) between the CTH map $h_C(x, y)$ retrieved from the computed disparity map and the ground truth map $h_T(x, y)$, that is

$$RMS = \left(\frac{1}{N} \sum_{(x,y)} (h_C(x, y) - h_T(x, y))^2 \right)^{\frac{1}{2}} \tag{2}$$

   where (x, y) are cloudy pixels (i.e., $h_T(x, y) > 0$), and N is the total number of cloudy pixels. This condition also remove from the error measure all those pixels which are not included in the MODIS map. The height values are measured in km.

2. Percentage of good matching pixels, computed as the number of pixels for which the recovered height $h_C(x, y)$ is below a given threshold δ_h, that is

$$G = \frac{1}{N} \sum_{(x,y)} (|h_C(x, y) - h_T(x, y)| > \delta_h) \tag{3}$$

   where, again, the sum is restricted to the cloudy pixels and N is the amount of cloudy pixels. For the experiments reported in this paper we used the threshold values of $\delta_h = 7.5, 3.75, 2$ km. This values come from the analysis on the accuracy that can be expected on the MSG stereo pairs reported in Section 3 and based on Equation 1.

4.2 Experimental results

The experiments used for our evaluation, are described in the remainder of this Section. Following the experiments presented in,[27] we will test how the performance depend on the parameters governing some of the steps of a matching procedure discussed in Section 2. In particular, in this paper we will report results on the first two steps: matching cost computation, and cost aggregation.

It is worth reminding that the quality is measured against the CTH maps provided by MODIS, and not against a real ground truth. Therefore, we are computing the amount of agreement with the MODIS measures instead of the real error. Moreover, stereo pairs in the case of CTH estimation are rarely images taken at the same instant, but there is always a time delay that can cause some changes in the scene. These observations explain why the quality of the results obtained in this study is lower than what has been shown in similar studies in the computer vision community, where *standard* images were used.

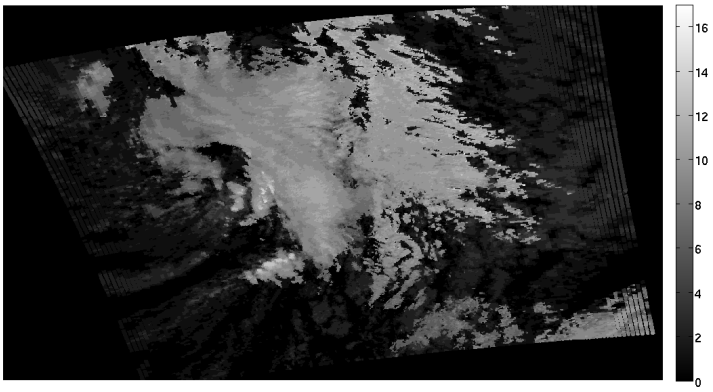


Figure 3. Cloud top height map produced by MODIS relative to the stereo pair in Figure 2. The unit measure is km.

4.2.1 Matching cost

The fist three experiments compare different matching costs discussed in Section 2, and how their performances are influenced by other components of the matching procedure.

Experiment 1 In this first experiment the matching costs compared are AD, SD, AD and BT, SD and BT, using 9×9 aggregation windows. The optimisation is the one always used is the standard winner-take-all. In the experiment different truncation values are used: 1, 2 , 5, 10, 20, 50 and 1000. The last value, being so large, means that no truncation is applied.

In Figure 4 we show the plots of the two evaluation measures RMS and G, defined in equations (2) and (3), plotted against the different truncation values used in the experiment. Both AD and SD show similar behaviours, and even using BT does not change too much the quality of the results. The plots of the RMS show a better performance with a heavy truncation, however the range of the RMS values obtained is very small (about 0.3 Km). The plot of the G quality measure, instead, makes clear that a heavy truncation (less than 10) decreases the amount of pixels with an agreement within 2 km with the MODIS data. This is in line with what obtained with other studies on natural images.[27]

Experiment 2 This second experiment is like the previous one, where the aggregation is done using a 9×9 shiftable window.[42] The results are very similar to what returned by the previous experiment, with a very small overall increase in the RMS measure. In[27] it was reported that the use of a shiftable window on standard images, can avoid the selection of the best truncation value. Experiments on these data, however, do not show evidence of this fact. This is probably due to the lower resolution of these satellite images with respect to the standard natural image in image analysis, that results in a not particularly accurate CTH map, as discussed in Section 3.

Experiment 3 This third experiment tries to determine what is the influence of the matching cost when a global optimisation approach is used. In the experiment the global optimisation paradigms used are dynamic programming (DP), scan-line optimisation (SO), and graph cuts (GC). AD was used as matching cost, with different truncation values.

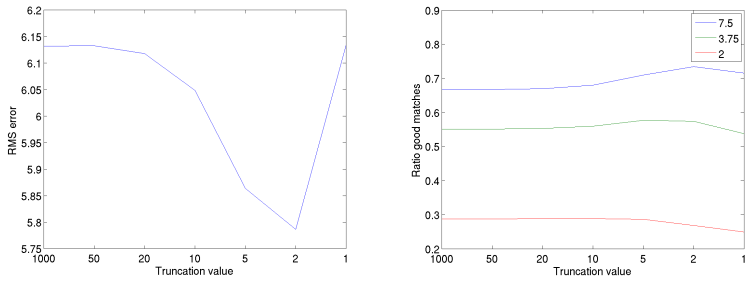
The results of this experiment are reported in Figures 6 and 7. Overall, it seems that global optimisation is not affected by the truncation value. This outcome, again, is different from what reported in,[27] and it might be due to the nature of the images used in the experiments. This calls for further studies using satellite images with a better resolution.

4.2.2 Aggregation

The last step we consider in this paper is the aggregation, which will be performed considering square neighbourhood, shiftable windows, binomial filters, regular diffusion, and membrane diffusion.

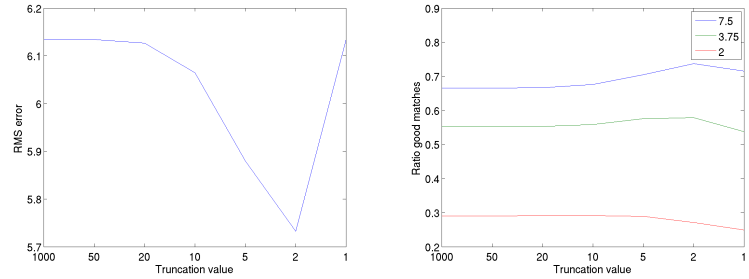
Experiment 4 The cost function considered for this experiment is, as for the last one, AD. No truncation is used this time. Following[27] we considered the following aggregation methods:

- square neighbourhood with sizes running from 3 to 29;
- shiftable square windows with sizes ranging from 3 to 29;
- iterated binomial filters with iterations ranging from 2 to 28;
- regular diffusion with iterations ranging from 10 to 150;
- membrane diffusion for 150 iterations and β ranging from 0.9 to 0.0.



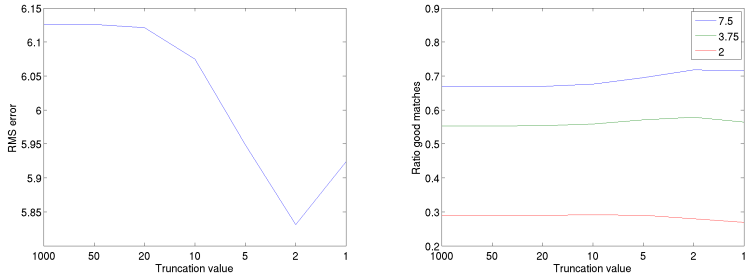
(a) SAD

(b) SAD



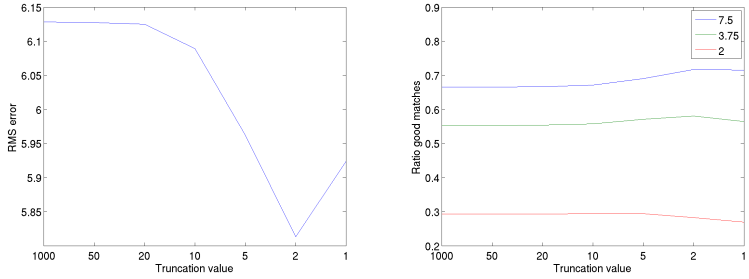
(c) SSD

(d) SSD



(e) SAD+BT

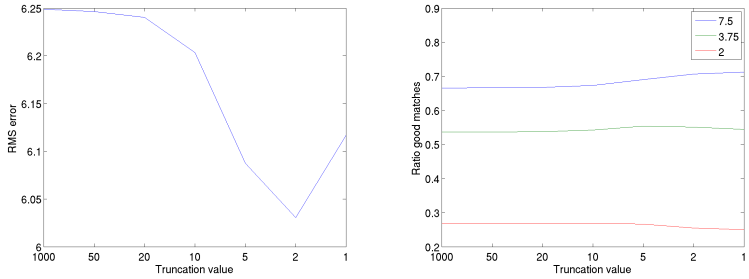
(f) SAD+BT



(g) SSD+BT

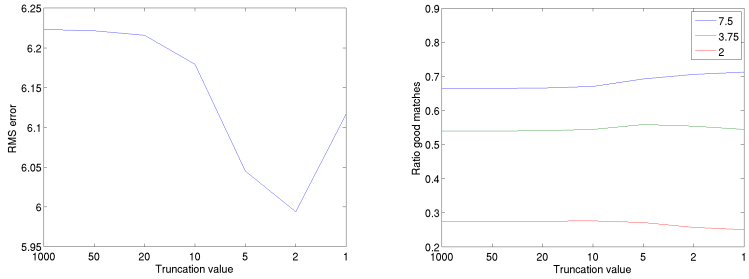
(h) SSD+BT

Figure 4. Experiment 1. Performance of different cost functions with a 9×9 aggregation window. The two error measures are plotted against different truncation values. The left column shows the graphs relative to the RMS, while the G error is displayed on the right column.



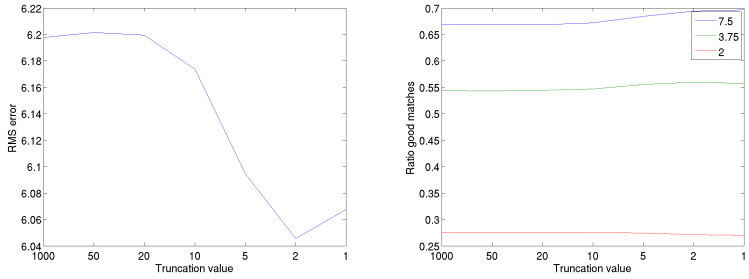
(a) SAD

(b) SAD



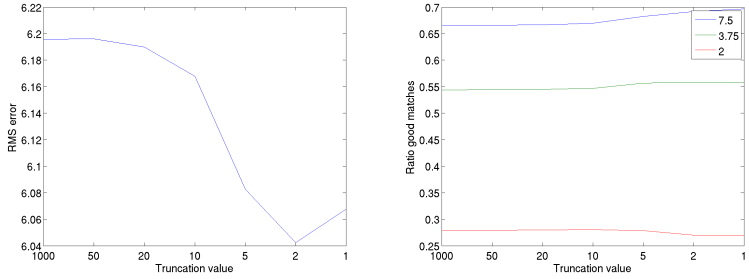
(c) SSD

(d) SSD



(e) SAD+BT

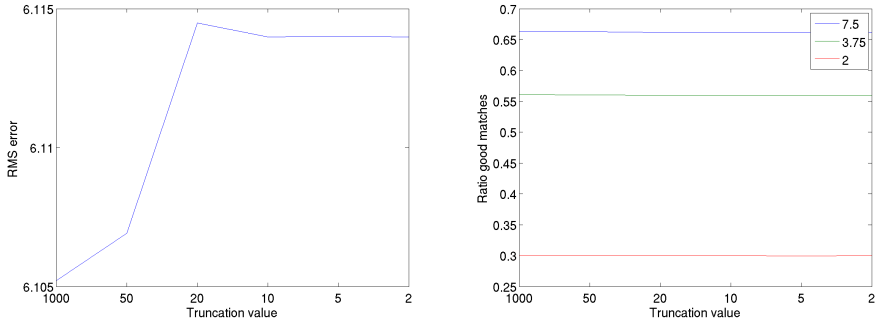
(f) SAD+BT



(g) SSD+BT

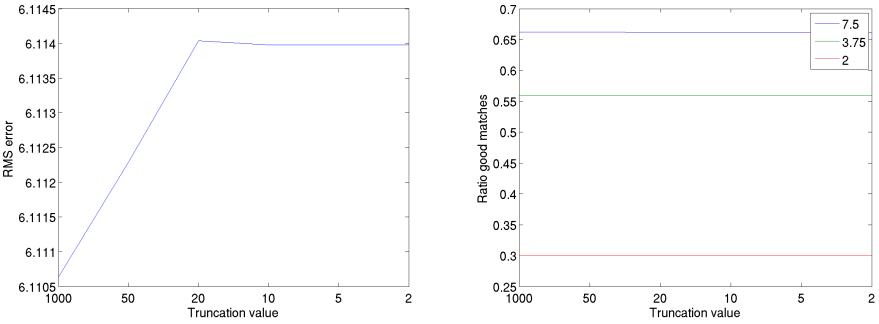
(h) SSD+BT

Figure 5. Experiment 2. Quality measures using a 9×9 shiftable window. The two error measures are plotted against different truncation values. The left column shows the graphs relative to the RMS, while the G error is displayed on the right column.



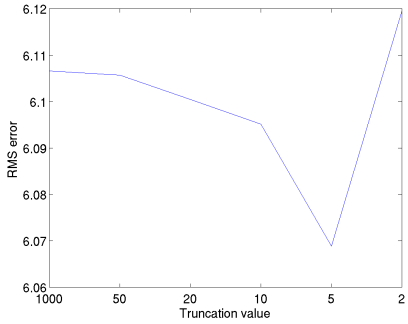
(a) DP

(b) DP

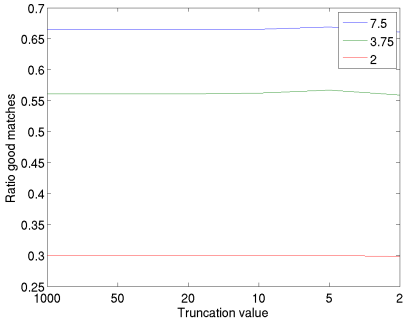


(c) DP+BT

(d) DP+BT

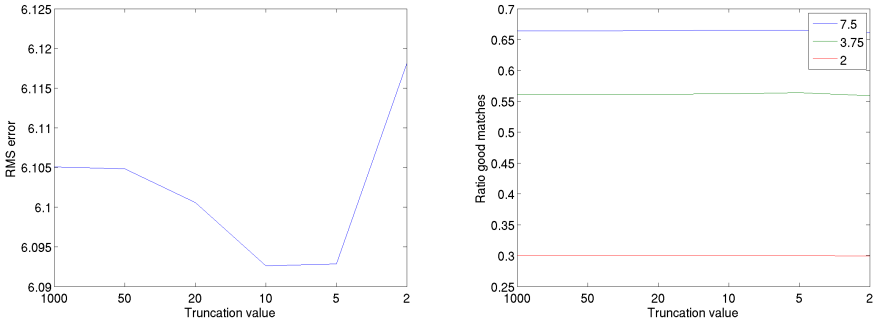


(e) SO



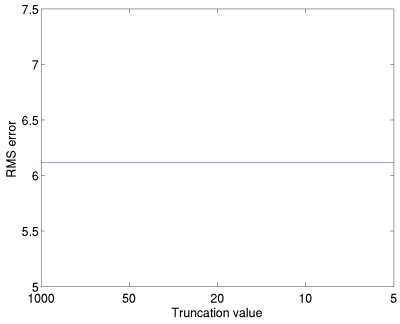
(f) SO

Figure 6. Experiment 3. Quality measures of different cost functions when global optimisations are used. The two error measures are plotted against different truncation values. The left column shows the graphs relative to the RMS, while the G error is displayed on the right column.

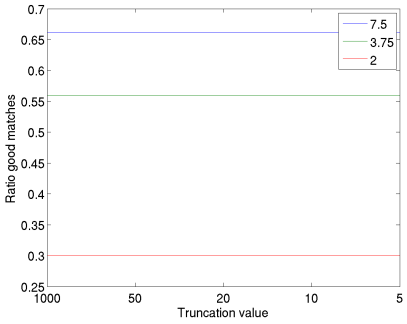


(a) SO+BT

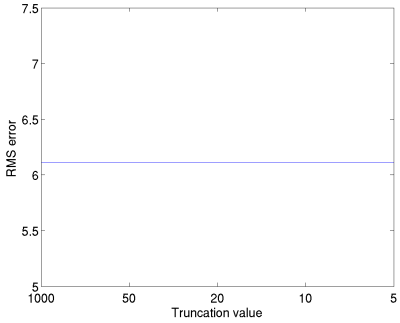
(b) SO+BT



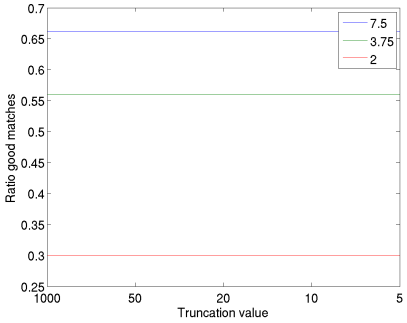
(c) GC



(d) GC



(e) GC+BT



(f) GC+BT

Figure 7. Experiment 3. Quality measures of different cost functions when global optimisations are used. The two error measures are plotted against different truncation values. The left column shows the graphs relative to the RMS, while the G error is displayed on the right column.

The results of this experiment are shown in Figures 8 and 9. Overall, the graphs show that an heavier aggregation brings better performance, as the value of the G measure increases with the aggregation; this behaviour is particularly visible for the G curve relative to the lowest threshold. The RMS shows more clearly that there is a saturation point after which the larger aggregation does not change the performance too much.

The outcome of these experiments can be justified by the fact that satellites infrared images of cloudy scene are almost textureless, that is have large areas where the information is almost constant, therefore a small aggregation do not use enough information to discriminate between the different parts. On the other hand, it is well known that large aggregation windows reduce the matching performance near the borders of the different regions in the image, as shown in.[27] This is particularly important for natural images, which are usually rich of details, but it is probably not a particular problem for the case we are studying.

5. CONCLUSIONS

In this paper we presented an experimental study on the performance of standard stereo matching algorithms on satellite infrared images of cloudy scenes. This work, although motivated by the JEM-EUSO mission, it is of general interest, as it addresses the popular problem of the Cloud Top Height estimation. Here, the problem is tackled using a stereo approach, and the main contribution of the paper is to study how some of the parameters that drive a stereo matching algorithm, influence the performance of the matching.

This work builds up on previous work[27] by other authors who made a complete study on the performance of stereo matching for standard natural images. Here we show that, due to the peculiarity of the satellite infrared images, the conclusion of the previous study cannot be applied blindly.

The experiments have been run on stereo pairs obtained from data of the MSG-2 and MSG3 satellites. Since, differently from the case of natural images, it is impossible to have a real ground truth for the cloud top height, we decided to compare the results obtained with a map provided by other sources, MODIS in this case.

The work needs to move forward in different directions. As a first thing, we need to complete the experimental study considering also the last two steps of a matching algorithm (optimisation and refinement), as this work mainly focused on the cost function and the aggregation. Since we have no real ground truth we might use, for the experimental comparison, more than just one CTH map: this would make the results more robust. As a last thing we must use in this study images from other satellites, such as AATSR, which have a ground resolution different from the MSG data.

Acknowledgements

This work has been partially supported by the Italian Ministry of Foreign Affairs and International Cooperation, and by Istituto Nazionale di Fisica Nucleare (INFN). The authors also wish to thank the JEM-EUSO collaboration for many fruitful internal discussions.

The authors acknowledge the MODIS Science team for the Science Algorithms, the Processing Team for producing MODIS data, and the GES DAAC MODIS Data Support Team for making MODIS data available to the user community. We also thank EUMESAT for providing Meteosat data.

For the experiments reported in this paper we used the publicly available code from http://vision.middlebury.edu/stereo/.

REFERENCES

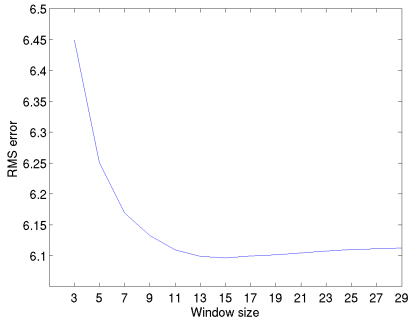
[1] Ebisuzaki, T. et al., "The JEM-EUSO mission," *Advances in Space Research* **53**(10), 1499–1505 (2014).

[2] Takahashi, Y. et al., "The JEM-EUSO mission," *New Journal of Physics* **11**(6), 065009 (2009).

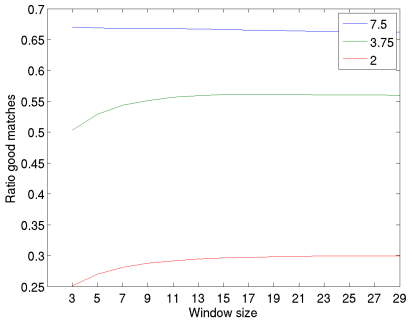
[3] Scarsi, L., Catalano, O., Maccarone, M. C., and Sacco, B., "Euso-extreme universe space observatory," in [*International Cosmic Ray Conference*], **2**, 839–842 (2001).

[4] Adams, J. H. et al., "The JEM-EUSO mission: An introduction," *Experimental Astronomy* **40**(1), 3–17 (2015).

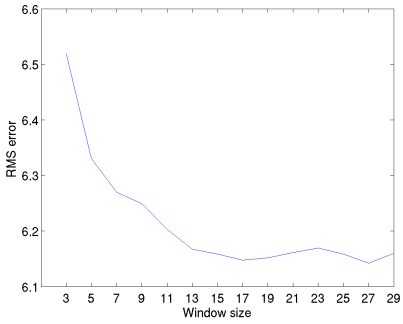
[5] Adams, J. H. et al., "The JEM-EUSO instrument," *Experimental Astronomy* **40**, 19–44 (November 2015).



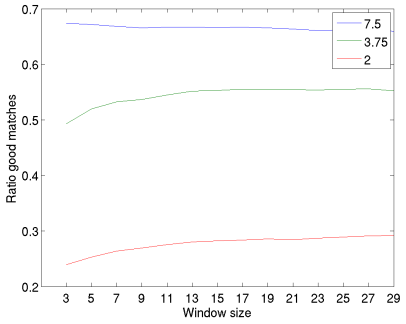
(a) Square window



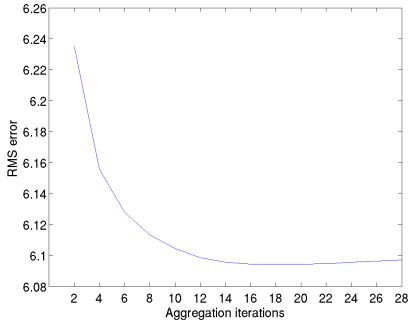
(b) Square window



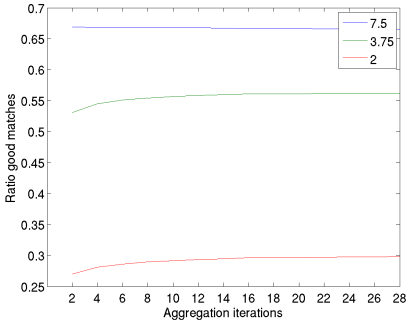
(c) Shiftable window



(d) Shiftable window

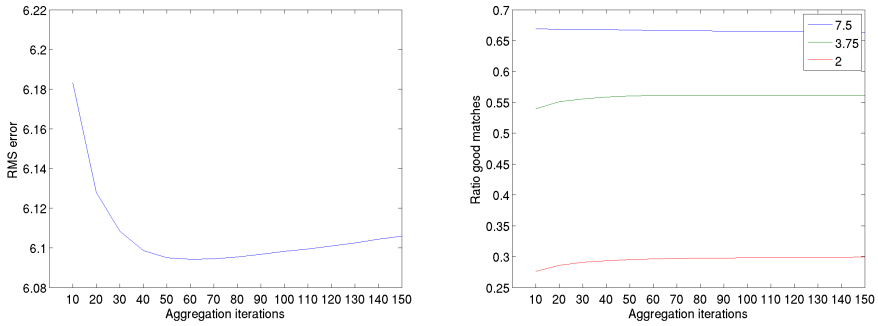


(e) Binomial filter

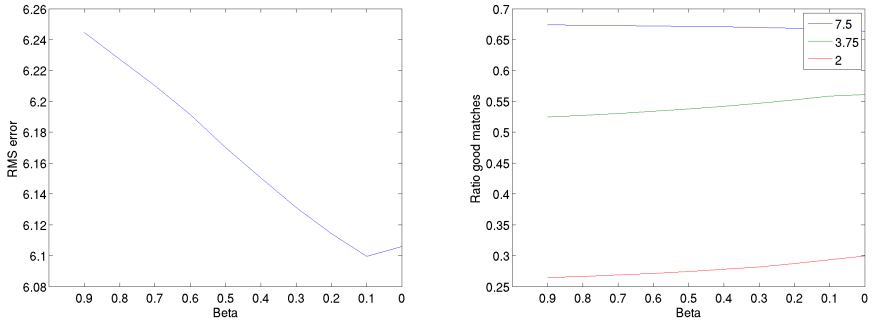


(f) Binomial filter

Figure 8. Experiment 4. Quality measures for different aggregation measures. The left column shows the graphs relative to the RMS, while the G error is displayed on the right column. In all the graphs the aggregation increases for left to right.



(a) Regular diffusion (b) Regular diffusion



(c) Membrane diffusion (d) Membrane diffusion

Figure 9. Experiment 4. Quality measures for different aggregation measures. The left column shows the graphs relative to the RMS, while the G error is displayed on the right column. In all the graphs the aggregation increases for left to right.

[6] Linsley, J., "Evidence for a primary cosmic-ray particle with energy 10^{20} ev," *Phys. Rev. Lett.* **10**, 146–148 (Feb 1963).

[7] Bird, D., Corbato, S., Dai, H., Dawson, B., Elbert, J., Emerson, B., Green, K., Huang, M., Kieda, D. B., Luo, M., et al., "The cosmic-ray energy spectrum observed by the fly's eye," *The Astrophysical Journal* **424**, 491–502 (1994).

[8] Abraham, J. et al., "Trigger and aperture of the surface detector array of the Pierre Auger Observatory," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **A613**, 29–39 (2010).

[9] Adams, J. et al., "The infrared camera onboard JEM-EUSO," *Experimental Astronomy* **40**, 61–89 (November 2015).

[10] Frias Rodriguez, M.D., Toscano, S., Bozzo, E., del Peral, L., Neronov, A., and Wada, S., "The atmospheric monitoring system of the jem-euso space mission," *EPJ Web of Conferences* **89**, 02007 (2015).

[11] Saez-Cano, G., Morales de los Rios, J.A., del Peral, L., Neronov, A., Wada, S., and Frias Rodriguez, M.D., "Thin and thick cloud top height retrieval algorithm with the infrared camera and LIDAR of the JEM-EUSO space mission," *EPJ Web of Conferences* **89**, 03001 (2015).

[12] Anzalone, A., Bertaina, M., Briz, S., Cassardo, C., Cremonini, R., de Castro, A. J., Ferrarese, S., Isgrò, F., López, F., and Tabone, I., "Methods to retrieve the cloud top height in the frame of the JEM-EUSO mission," *Submitted* (2016).

[13] Trucco, E. and Verri, A., [*Introductory Techniques for 3-D Computer Vision*], Prentice Hall, Upper Saddler River, New Jersey 07458 (1998).

[14] Szeliski, R., [*Computer Vision Algorithms and Applications*], Springer (2010).

[15] Fischer, J. and Grassl, H., "Detection of cloud-top height from backscattered radiances within the oxygen a band. part 1: Theoretical study," *Journal of Applied Meteorology* **30**(9), 1245–1259 (1991).

[16] Kuze, A. and Chance, K. V., "Analysis of cloud top height and cloud coverage from satellites using the o 2 a and b bands," *Journal of geophysical research* **99**(D7), 14481–14491 (1994).

[17] Fisher, D., Paulsen, C., Thomas, G., and Muller, J.-P., "Synergy of stereo cloud top height and ORAC optimal estimation cloud retrieval: evaluation and application to AATSR," *Atmospheric Measurement Techniques* **9**, 909–928 (2016).

[18] Briz, Susana, de Castro, Antonio J., Fernandez-Gomez, Isabel, Rodriguez, Irene, and Lopez, Fernando, "Retrieval of optically thin cloud emissivity from brightness temperatures provided by ir camera of JEM-EUSO mission," *EPJ Web of Conferences* **89**, 03011 (2015).

[19] Prata, A. and Turner, P., "Cloud-top height determination using atsr data," *Remote sensing of environment* **59**(1), 1–13 (1997).

[20] Moroney, C., Davies, R., and Muller, J.-P., "Operational retrieval of cloud-top heights using MISR data," *IEEE Transactions on Geoscience and Remote Sensing* **40**(7), 1532–1540 (2002).

[21] Muller, J.-P., Denis, M.-A., Dundas, R., Mitchell, K., Naud, C., and Mannstein, H., "Stereo cloud-top heights and cloud fraction retrieval from ATSR-2," *International Journal of Remote Sensing* **28**, 19211938 (May 2007).

[22] Seiz, G., Tjemkes, S., and Watts, P., "Multiview cloud-top height and wind retrieval with photogrammetric methods: Application to meteosat-8 HRV observations," *Journal of Applied Meteorology and Climatology* **46**(8), 1182–1195 (2007).

[23] Anzalone, A. and Isgrò, F., "A multi-spectral stereo method to retrieve cloud top height applied to geostationary satellite IR images," in [*Proceedings of the International Conference on Computer Systems and Technologies*], (2016).

[24] Naud, C., Muller, J.-P., Haeffelin, M., Morille, Y., and Deleval, A., "Assessment of MISR and MODIS cloud top heights through inter-comparison with a back-scattering lidar at SIRTA," *Geophysical Research Letters* **31** (2004).

[25] Naud, C., Mitchell, K., Muller, J., Clothiaux, E., Albert, P., Preusker, R., Fischer, J., and Hogan, R., "Comparison between ATSR-2 stereo, MOS O2-A band and ground-based cloud top heights," *International Journal of Remote Sensing* **28**(9), 1969–1987 (2007).

[26] Manizade, K. F., Spinhirne, J. D., and Lancaster, R. S., "Stereo cloud heights from multispectral ir imagery via region-of-interest segmentation," *IEEE Transactions on Geoscience and Remote Sensing* **44**, 2481–2491 (Sept 2006).

[27] Scharstein, D. and Szeliski, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision* **47**, 7–42 (Apr 2002).

[28] Anzalone, A., Isgrò, F., and Tegolo, D., "Stereo matching techniques for cloud-top height retrieval," in [*Remote Sensing*], 636507–636507, International Society for Optics and Photonics (2006).

[29] Cheng, J. and Huang, T. S., "Image registration by matching relational structures," *Pattern Recognition* **17**(1), 149–159 (1984).

[30] Horaud, R. and Skordas, T., "Stereo correspondence through feature grouping and maximal cliques," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(11), 1168–1180 (1989).

[31] Tell, D. and Carlsson, S., "Combining appearance and topology for wide baseline matching," in [*European Conference on Computer Vision*], 68–81, Springer (2002).

[32] Zhang, Z., Deriche, R., Faugeras, O., and Luong, Q.-T., "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial intelligence* **78**(1), 87–119 (1995).

[33] Pilu, M., "A direct method for stereo correspondence based on singular value decomposition," in [*Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*], 261–266, IEEE (1997).

[34] Delponte, E., Isgrò, F., Odone, F., and Verri, A., "SVD-matching using SIFT features," *Graphical models* **68**(5), 415–431 (2006).

[35] Goshtasby, A., Gage, S. H., and Bartholic, J. F., "A two-stage cross correlation approach to template matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(3), 374–378 (1984).

[36] Chou, C.-H. and Chen, Y.-C., "Moment-preserving pattern matching," *Pattern Recognition* **23**(5), 461–474 (1990).

[37] Isgrò, F. and Trucco, E., "On robust rectification for uncalibrated images," in [*Proceedings of the International conference on Image Analysis and Processing*], 297–302, IEEE (1999).

[38] Isgrò, F. and Trucco, E., "On projective rectification," in [*Proceedings of the Seventh International Conference on Image Processing and Its Applications)*], **1**, 42–46, IET (1999).

[39] Kanade, T., Kano, H., Kimura, S., Yoshida, A., and Oda, K., "Development of a video-rate stereo machine," in [*Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on*], **3**, 95–100 vol.3 (Aug 1995).

[40] Anandan, P., "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision* **2**(3), 283–310 (1989).

[41] Birchfield, S. and Tomasi, C., "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, 401–406 (Apr 1998).

[42] Bobick, A. F. and Intille, S. S., "Large occlusion stereo," *International Journal of Computer Vision* **33**(3), 181–200 (1999).

[43] Burt, P. and Adelson, E., "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications* **31**, 532–540 (Apr 1983).

[44] Scharstein, D. and Szeliski, R., "Stereo matching with nonlinear diffusion," *International Journal of Computer Vision* **28**(2), 155–174 (1998).

[45] Baum, B. A., Menzel, W. P., Frey, R. A., Tobin, D. C., Holz, R. E., Ackerman, S. A., Heidinger, A. K., and Yang, P., "MODIS cloud-top property refinements for collection 6," *Journal of Applied Meteorology and Climatology* **51**(6), 1145–1163 (2012).

[46] Baum, B. A. and Platnick, S., [*Introduction to MODIS cloud products*], 74–91, Springer (2006).