Fast Dense Disparity Estimation of Stereo Couples from Image Gradient

Charles Beumier
Signal and Image Centre, Royal Military Academy, Brussels, Belgium
beumier@elec.rma.ac.be

Abstract

This paper presents an original development of disparity estimation of stereo couples primarily based on the similarity of image gradient between both images. Each pixel of the left image is associated to right image pixels with same gradient and similar intensity, directly leading to a set of disparities equal to the horizontal displacement in the specific case of epipolar images. Disparity values are restricted to an a priori acceptable range generally known by the application. The final disparity value assigned to a left image pixel is obtained from the histogram peak of candidate disparity values associated to pixels in the vicinity of the left pixel. A dense disparity map is produced. The conception and development have followed a very fast processing objective, resulting in about 0.75 Mpixel of disparity values per second on a 2.33 GHz Intel Core 2 Duo CPU. It has been successfully applied to highlight buildings in several types of satellite and aerial images.

1. Introduction

Although image processing tasks have traditionally been considering 2D images as the most direct and easiest way to get information automatically, there is an increasing number of applications which take advantage of 3D information.

Technology breakthrough of the last decade has enabled very accurate and fast cameras for image acquisition. Some systems also integrate image processing and possibly light or LASER projection to deliver 3D measures automatically.

In the field of remote sensing, 3D measurements are of high importance when considering cartography, environmental monitoring or change detection. Although active 3D capture systems also exist (SAR, LIDAR), they remain expensive and are restricted to pure 3D information. We find stereoscopic pairs more attractive as they can also provide for radiometric information, for instance to filter out elevated vegetation when looking for buildings.

In order to recover 3D measures from a stereoscopic pair, what is called computational stereo [1], two types of information are necessary. First, the position, orientation and characteristics (focal length, optical centre and lens distortion) of the camera for each image have to be known precisely thanks to calibration. Secondly, the image coordinates of each object of interest have to be extracted in each image. The 3D position of an object point is obtained at the intersection of the two 3D lines defined by each image point and associated camera geometry.

For practical reasons, the images are often rectified by a projective transformation so that corresponding points in both images lie on a horizontal line [1]. In this geometry called epipolar, the correspondence problem is simplified to a 1-D search.

This paper presents an original correspondence algorithm which aims at very fast (horizontal) disparity estimation for epipolar images thanks to gradient similarity. Matching left and right pixels of similar gradient leads to multiple disparities which are filtered according to intensity similarity to form the set of candidate disparity values attached to the left pixel. The disparity value retained for each left pixel corresponds to the peak in a histogram filled by candidate disparities for a square area surrounding the left pixel. This aggregation phase by a square area allows pixels with no candidate disparity to receive a probable disparity value so that the approach delivers a dense disparity map.

The paper is organized as follows. Section 2 presents a discussion about disparity estimation which has guided the proposed algorithm detailed in section 3. Results are presented in section 4 in the form of disparity maps for different types of images. Section 5 presents the disparity map as a step for building verification. Finally, section 6 concludes the paper and present perspectives.

2. Disparity estimation

2.1. Generalities

The main topic of the paper is the estimation of the disparity which is, in the specific case of epipolar geometry, the horizontal signed distance in pixel between corresponding points in the left and right images.

Two major classes of approaches have originally been followed to derive disparity [2]. The first one, called area-based approach, search for left and right pixel correspondences through the comparison of pixel neighborhoods to increase matching robustness. There is a tradeoff between robustness (in favor of larger areas) and localization precision obtained for smaller windows. Methods with multiple window sizes have been developed to address this tradeoff ([3,4]). Area-based approaches deliver a disparity value at each pixel (dense disparity map), are generally simple to program but require much computational power.

The second general class of first approaches, called feature-based, aims at identifying features of interest corresponding in both images. The type of features depends on the application but usually consists of points with high contrast like corners, lines or 2D homogeneous patches. The set of extracted features from the left and right images are then matched, possibly using associated radiometric or geometrical characteristics. Computational time is generally low, thanks to the proper selection of pertinent features, but disparity values are available at the extracted feature points only, leading to a sparse disparity map. A dense map may then be derived by interpolation.

Combining the area and feature-based approaches is one possible way to design a fast and robust sparse disparity extractor. In Beumier [5], pixels with locally maximal gradient are matched based on gradient orientation and ranked according to local correlation. The best correlation is accepted as matching left-right pair with a confidence proportional to the correlation score. This approach has the ability to estimate the disparity of edge points of very high importance, especially for remote sensing images at 0.5m or 1m resolution.

2.2. Image observation

In the search for designing a correspondence algorithm appropriate for remote sensing images (satellite or airborne), many grey profiles of left and right images were analyzed visually. I summarize here the main observations.

Intensity levels of left and right images (referring to panchromatic in the case of satellite Quickbird or Ikonos imagery and to one channel, typically green, in the case of airborne imagery) may undergo very large discrepancies due to occlusion, object motion and specular reflection. A method based on absolute intensity alone will not be adequate. In area-based approaches, intensity variation is usually addressed by the normalized cross correlation which normalizes differences relatively to the mean and standard deviation of the intensities in the window.

Gradient values are robust against intensity level changes and provide for precise location (gradient maxima). 2D gradients contain in addition local geometrical information thanks to their orientation value.

In the specific case of remote sensing imagery with resolution around 0.5m or 1m, little information is conveyed in slowly varying areas, except color when available, because shading is usually not captured and noise may be preponderant. Borders of such areas are the most reliable features on which to match left and right image parts.

2.3. Computation time

Computation time has been an important concern when developing the proposed solution initially devoted to remote sensing imagery which generally consists of 10 to 100 Mpixel images. Computer memories nowadays support such resolutions and processing the image as a whole prevents the tedious work of region selection or partitioning, and assembling the resulting parts into one image.

As said before, area-based approaches suffer from the huge number of comparisons. A feature-based approach is much more attractive in that respect.

One general way to speed up disparity estimation consists in multi-resolution, considering image versions at different resolutions. Smaller resolution images are processed first, implying a smaller search area and a reduced set of possible disparity values. Higher resolutions are then considered, taking advantage of disparity values obtained at previous resolution to confine search. Apart from the obvious computational gain, the approach also benefits from a higher robustness thanks to lower ambiguity at lower resolutions and the confined search. Last but not least, this scheme is also a good solution for noise and large uniform areas which tend to lack matching features.

3. Proposed disparity estimation

Following observations mentioned in 2.2 and computation time considerations outlined in 2.3, our research work was oriented towards finding image features which are simple to detect and to handle within a computer program, easy to compare while sufficiently spread in images to avoid too sparse disparity maps. As opposed to area-based approaches where pixel specificity for matching is brought by its neighborhood ('aggregation in [7]'), we thought that disparity votes of possible matches, even incorrect, could be a posteriori filtered. Moreover, handling aggregation once disparities are obtained is more efficient and allows to output dense disparity maps. From a previous experience with disparity thanks to gradient [5], we adopted the horizontal gradient as basic feature.

Integrating gradient into the process of computational stereo is not new, as attested by [6] which presents an overview of matching techniques based on image gradients. In this respect, the originality of the proposed approach is to avoid the time consuming of an a priori aggregation for gradient comparison used to select the winner left/right pair and to postpone the final disparity selection to an a posteriori aggregation phase by histogram of individual candidate disparity values obtained from left and right pixel comparison (see section 3). By this way, the approach combines the advantage of the feature-based and area-based approaches.

When considering the taxonomy of stereo algorithms given in [7], the proposed correspondence method mainly concerns steps 3 and 4 about disparity computation and refinement. The candidate disparity values by gradient comparison is the 'local' aspect of the approach while the aggregation by histogram makes it more a 'Global method', distinction also made in the survey of Brown [1].

To go into implementation details, our development considered horizontal gradient of each left pixel as local feature to obtain a set of matching right pixels and derive associated disparity values. This set of candidate values is filtered by the allowable disparity range related to the application, by the gradient orientation and by the similarity of grey values. The final disparity value assigned to a left pixel is obtained as the maximal occurrence of candidate disparity values associated to pixels contained in a square area surrounding the considered left pixel. Details are given in the following subsections.

3.1. Gradient as local feature

In the proposed approach, the image gradient has been retained as local feature on which to base left and right matches and collect possible disparities that will be filtered as indicated in the following subsections.

For efficiency and practical implementation, only the horizontal part of the gradient is first used. This corresponds to the sensitive information when considering epipolar images for which disparity is horizontal. Dealing with such a 1-D value allows for an efficient implementation which considers arrays of x coordinates associated to fixed horizontal gradient values. Matching a left pixel x_L is a simple lookup in the table of x_R corresponding to the horizontal gradient at x_L.

The horizontal gradient G_x at x is defined as the intensity difference $G_x(x,y) = I(x+D,y) – I(x-D,y)$, with D controlling the locality of the gradient feature. A typical value for D is 2. A lower D value results in noisier estimation while a large value tends to blur estimated disparities.

Not all gradient values need to be represented in the tables. We named L the parameter which controls the number of gradient levels. A value L=4 means that only gradient values multiple of 4 will be stored. Interpolation is used so that subpixel x coordinates are stored. A large L value reduces the number of $(x_L,x_{R)}$ pairs considered for matching at the expense of fidelity. A small L value provides more possible matches at the expense of a larger running time.

3.2. Disparity range

The allowable range for disparity may be known from the scene geometry and camera arrangement. It can also be obtained from image observation or by trial and error with the computer program on one stereo couple and maintained for other stereo couples of the same campaign. Anyway, campaign designers paid attention so that disparity is only a fraction of the image size, trying to optimize the tradeoff between precision (large disparity) and image similarity (small disparity).

The disparity program contains two limits '*Off_min*' and '*Off_max*' out of which disparity candidate values are rejected. Selecting loose limits may result in larger computation times and possibly spurious disparity values. Selecting too tight limits will reject correct values and end up with wrong disparity estimations.

3.3 Gradient orientation

To optimise implementation in terms of programming and running time, we preferred to create 1-D tables of x coordinates for some given horizontal gradient values G_x (see 3.1). The vertical component G_y of the gradient has been used as a filter.

Gradient orientation provides useful local information which is usually stable between captures. Slight orientation shifts may occur due to the different points of view of the left and right images. Rare cases involve inversion of contrast, usually due to specular reflection.

As matching points x_L and x_R share the same G_x value (3.1), the condition on gradient orientation has been simplified to a condition on G_y, avoiding the heavy computations of atan(). We use the orientation constraint:

$$k*|G_y(x_L,y) - G_y(x_R,y)| < |G_y(x_L,y)| + |G_y(x_R,y)|$$

where | | denotes the absolute value, and *k* is a parameter to set the sensitivity to gradient orientation. In the experiments *k* was set to 3, which is quite a loose constraint.

3.4. Intensity similarity

Left and right intensity levels may differ either in the form of a global shift or local differences. As presented in 2.2, local variations may be very large, principally due to occlusion, specular reflection or object motion.

In the proposed approach, gradient is given priority for matching and intensity is used as a rough filter to reduce the number of unlikely matches. This is implemented with a threshold '*Thres*' which represents the maximal value by which left and right pixels may differ in intensity. Since x_L and x_R have been determined for fixed horizontal gradients with subpixel accuracy, associated intensity values were also linearly interpolated. This interpolation is not fundamental to the method but may be advantageous in steep edges where intensity values vary a lot.

To account for a global shift of intensity levels between the left and right images, intensity comparison is corrected by the median of left and right pixel intensity differences obtained by histogram.

Most optical cameras nowadays possess several channels. The current approach uses so far only one channel such as the panchromatic band for a Quickbird or Ikonos image or the green channel for an aerial image. A direct adaptation towards colour may consist in providing a threshold based on the norm of the vector whose components are channel values. However, observation of multispectral images reveals that most edges are reflected in all channels so that the gain relatively to one channel threshold is probably limited. A better 'color' approach would consist in measuring the hue in uniform areas.

3.5. Aggregation

Following the taxonomy of Scharstein & Szeliski [7], an aggregation phase is present in most correspondence algorithms.

In area-based approaches, the difference between pixel pairs is aggregated over a window to improve robustness. In feature-based methods, robustness is achieved by the specificity of features. In the case of geometrical features, this corresponds to an aggregation of pixels into objects. For the feature-based approach proposed in this paper, the locality of the gradient is not sufficient to resolve pairing ambiguities so that a larger scale representation is needed. Handling objects of possibly different and varying forms requires programming development that we preferred to avoid. The proposed aggregation scheme is performed a posteriori, once disparity candidate values are available. The disparity value assigned to a left pixel is derived from the distribution of candidate disparity values associated to neighboring left pixels of a square window.

More specifically, the distribution of disparities associated to pixel x_L is captured in a histogram considering a square neighborhood of size $(2*SV+1)*(2*SV+1)$ and centered on pixel x_L. For performance considerations, the histogram is initiated for each line at $x = SV$ and updated by the inclusion of column x_L+SV+1 and the exclusion of column x_L-SV. A target disparity is obtained from the highest peak of 3 consecutive bins in the histogram. This target value is refined to the highest histogram bin within +/- 1 bin. This last refinement reduces the blur effect of the aggregation.

A value of 5 for *SV* corresponding to a 11x11 aggregation window seems appropriate as a compromise between blurring and wrong estimation by lack of candidate disparity values.

If a sparse disparity map is sufficient for the application, it is possible to de-activate aggregation (SV=-1) so that only disparity values at gradient pixels are output. In this case, x_R candidates are ranked by grey similarity. This results in faster operation but very noisy estimates.

4. Results

Results of the proposed correspondence algorithm are given for different stereo couples of satellite and aerial images. Resolutions range from 10 cm to 1 m. For data containing multi-spectral information, only one channel has been used so far: panchromatic for Ikonos images and the green channel for aerial images. The intensity range has been cast to 8-bit if it had more resolution but no contrast adjustment has been made.

4.1 Sensitivity to parameters

Several parameters have been mentioned and discussed in section 3. Qualitative tests through disparity map observation were carried out with five stereo couples to find the influence and appropriate values of those parameters.

Parameter *D* (locality of gradient, 3.1) was chosen equal to 2. Although *D*=1 may result in better localization, it is also more sensitive to noise. For *D* >= 4, blur effects become important for a small gain in noise reduction. A noisy image will favor *D*=2 or 3 while a clean image could use *D*=1. Compared to *D*=1, about 5 to 10% time saving was observed for *D*=3 thanks to the reduced gradient variation implying less matching pixels.

For the 8-bit intensity data considered, the value of parameter *L* (3.1) was set to 2. A small *L* value means that more gradient levels are stored, what eventually leads to more disparity candidate values. Working with integer values for gradient G_x, *L*=1 implies the maximal number of candidate pixel pairs and the resulting disparity map offers the more coherent disparities. However, results are similar for *L*=2 or *L*=3, implying less computations as less pixels are stored in the tables. From *L*=1 to *L*=4, about 15 to 20% less global computation time was observed. The differences are mainly found in uniform areas where gradient values lack.

The a priori allowable range of disparities (*Off_min*, *Off_max*) was discussed in 3.2 and does not need practical tests as the range is defined by the acquisition campaign. However, the scene geometry of a specific image (e.g. flat terrain) may allow a tighter range which can prevent some incoherent values. Running time increases with the range of disparities.

Gradient orientation has a secondary role in our implementation and is controlled through parameter *k*. A small value for *k* will allow large orientation variations between the left and right pixels, increasing the number of pixel pairs (and candidate disparities) to be processed. Values below 2 seemed to filter few candidate pairs while values above 4 introduced disparity estimation with little confidence due to a lack of matching pixel pairs. We adopted *k*=3, keeping a confidence similar to *k*=1 while rejecting enough pixel pairs to reduce computation time (in the order of 25 to 30%).

The intensity similarity constraint, controlled by parameter *Thres*, has little influence on the results as long as it is not too strict. The considered image couples gave satisfying results with threshold above 8 grey levels, except for couple '1' ('Bagdad') containing large left and right intensity differences in flat roofs. Selecting a sufficiently high threshold value generally helps solving a few ambiguities (appearing as spurious disparities) and looking for the lowest acceptable threshold saves around 5 % computation time. Uniform areas are here also the most sensitive parts. *Thres*=15 was good for all couples except couple '1' (*Thres*=25).

Aggregation by histogram is controlled by the size parameter *SV*. With *SV*=1 or 2, not enough pixels are used for the histogram, resulting in many spurious and erratic values, especially in low contrast areas. *SV*=4 (9x9 window) is better. *SV*=6 seems to be right relative to the coherence of disparity values with tested images but object outlines are blurred. *SV*=6 is about 10% slower than *SV*=4. This parameter has by far the major impact on the visual results and was chosen equal to 5.

As a conclusion about parameter sensitivity, aggregation seems to have made the algorithm little dependent on the parameters except *SV*. *SV* results from the tradeoff between disparity homogeneity and fidelity.

4.2 Ikonos images

Ikonos satellite imagery is captured with 1m panchromatic and 4m multispectral images. Only the panchromatic channel has been considered due to its resolution. Multispectral bands have not been tested neither added to the panchromatic data.

One stereo pair covers a zone near Bagdad (Fig. 1). The images have little contrast (intensity in the range [10..90], enhanced for display in Fig. 1). Large discrepancies exist between left and right intensity levels, mainly for flat roofs. The disparity range is [–32..0] and the map is represented in false color to highlight small differences and similar elevation. The 1m resolution limits the detection by disparity to medium or large size structures (mainly located in the top left quadrant) although areas with houses are highlighted (right part). Elevated vegetation is also visible (wood at middle bottom and rows of trees at middle top) but with a variable quality. Erroneous and erratic disparity values are mainly found in large uniform areas.

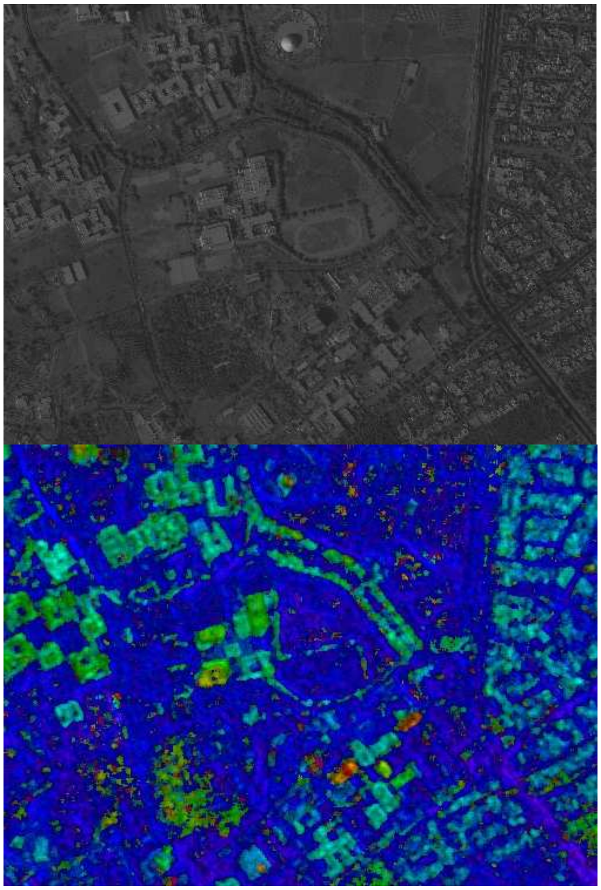

Figure 1. a) Left image of part of Ikonos image around Bagdad; b) Disparity map

A part of another Ikonos image was successfully processed, covering an area of Graz. As a 3D ground truth is available for this stereo couple, a quantitative estimation is planned.

4.3 Aerial images

Airborne campaign is the preferred solution for the capture of a digital surface model in built-up areas. We have tested our correspondence algorithm on three types of stereo aerial imagery.

First, two black and white images corresponding to the old generation of photogrammetric data at the Belgian National Geographic Institute (IGN) were considered. These are digitized versions of ‘analog’ pictures shot at 30 cm ground resolution and scanned. They are very noisy and suffer from a large base to height ratio, implying large left/right differences from the important parallax. For this couple, we have at our disposal the building and road vector database allowing for a building verification assessment. This is presented in section 5.

Secondly, a couple of digital multispectral images in the region of Virton (Belgium) was tested. With a ground resolution of 0.5m, these images correspond to what IGN Belgium is currently acquiring for database update by restitution. The green channel was selected as intensity input.

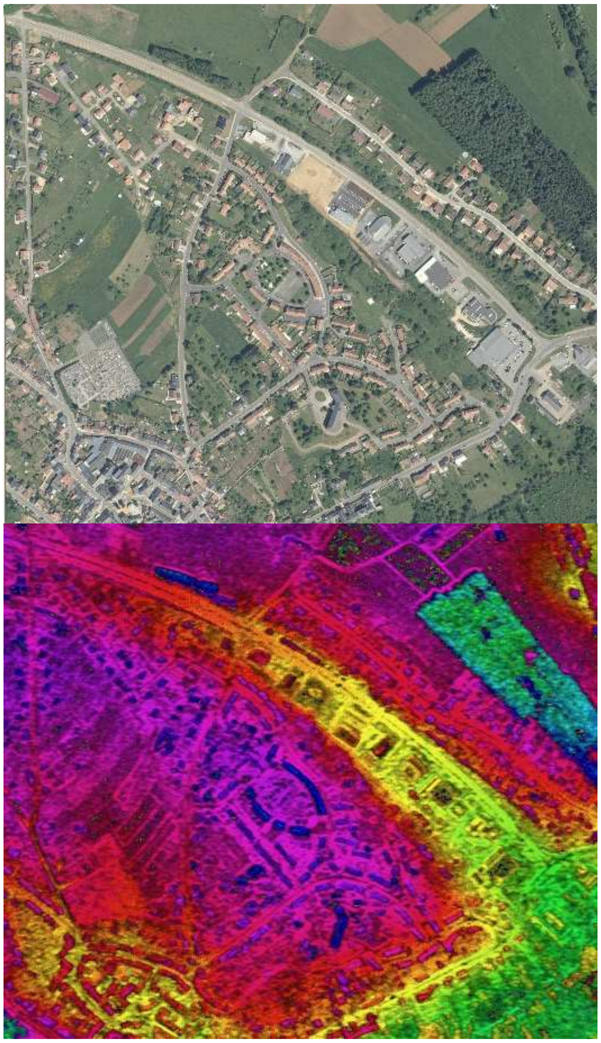

Figure 2. a) Part of the left RGB image of stereo couple around Virton; b) Disparity map

Fig. 2 shows the disparity map of an area covering about 1 km^2. With the available resolution, buildings and isolated houses clearly appears in the false color map. The figure shows the hilly aspect of the area and the proper detection of woods (top right). Incorrect values are found in poorly textured areas (fields of middle top and large buildings in the yellow/bright area around the centre).

A close-up in Fig. 3 shows the quality of the building outlines and the restitution of most isolated trees.

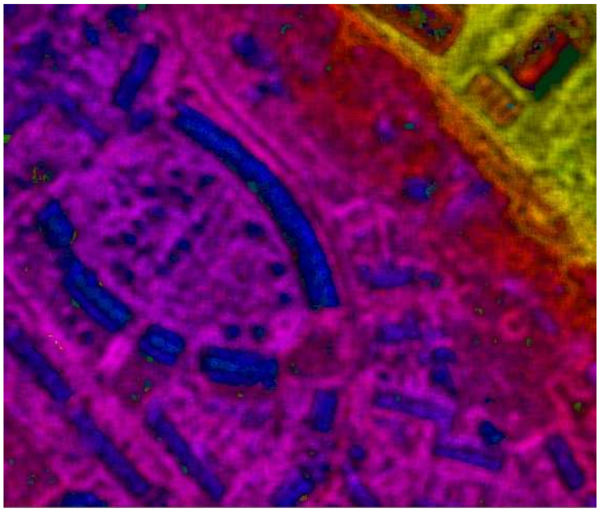

Figure 3. Close-up from central area of Fig. 2

Finally, the correspondence algorithm has been applied on a couple of very high resolution (0.1m) of multispectral images. This resolution is most of the time considered to capture the digital surface model of dense urban areas. It was ordered by the CIRB (Brussels Regional Informatics Centre) on the Brussels region.

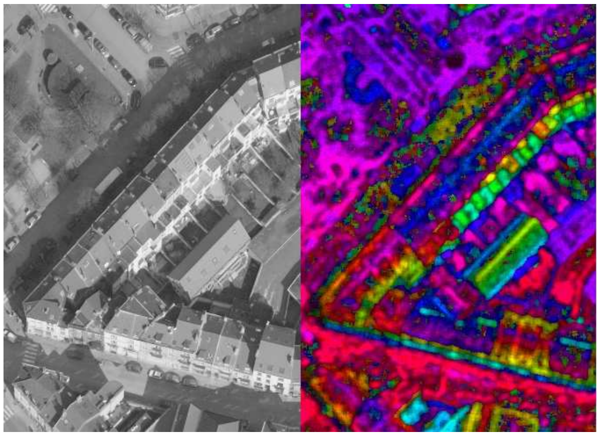

Figure 4. Small area of the left image of Brussels at 0.1m and disparity map

Fig. 4 presents a small area in Brussels captured at the 0.1m resolution and the associated disparity map. With a range of more than 60 pixels, the disparities are also better seen in false color, although in this case the 32 colormap entries were used several times, meaning that each color may represent two different disparities (separated by 32 pixels). The higher resolution enables the distinction of several details like the varying levels of gable roofs and the vehicles parked in the streets. Erroneous disparity values are found in poorly textured areas (middle of roads, large roofs and trees with no leaves in the square).

4.4 Results summary

Table 1 lists image characteristics and running times for the different experiments with the set of parameters mentioned in 4.1 (*D*=2, *L*=2, *k*=3, *SV*=5).

Table 1. Summary of data and running times

ID	Name	Image	Size	Range	Time
1	Bagdad	Ikonos, 1m	3.8 Mp	-32..0	4.3s
2	Graz	Ikonos, 1m	3.2 Mp	10..72	4.5s
3	Malle	Aerial 0.3m	12 Mp	-72..-32	14.7s
4	Virton	Aerial 0.5m	12 Mp	-32..32	16.5s
5	Brussels	Aerial 0.1m	12 Mp	-64..8	16.1s

Running times for these experiments range from 1.12s to 1.41s by Mpixel.

Wrong disparity estimations with the approach are commonly met in poorly textured areas, where candidate disparities lack. The aggregation may solve this problem to a limited extent because too large SV values results in blurred disparities. A general solution for large uniform areas consists in adopting a multi-resolution scheme, as explained in 2.3.

5. Building verification

The main application of the presented work in disparity estimation concerns change detection of buildings. As detailed in [5], we first considered the disparity of contour points so relevant for man-made structures as a way to assess the presence of buildings. In an attempt to go further in precision, we believe that a dense disparity map even with some imperfections will simplify the accurate estimation of disparity by model-based approaches. The step of acquiring the digital terrain model necessary in [5] can also be obtained very quickly with the proposed correspondence algorithm.

In order to assess the presence of buildings specified in the vector database, there is no need to reconstruct the 3D map since disparity values suffice to highlight elevated buildings. The approach detailed in [5] can be followed, replacing the sparse disparity map by the dense map of the present work.

Fig. 5 shows the potential for building verification thanks to the superposition of the vector database (polygons of the buildings and roads layers) on the disparity map for stereo couple '3' (Malle, Belgium).

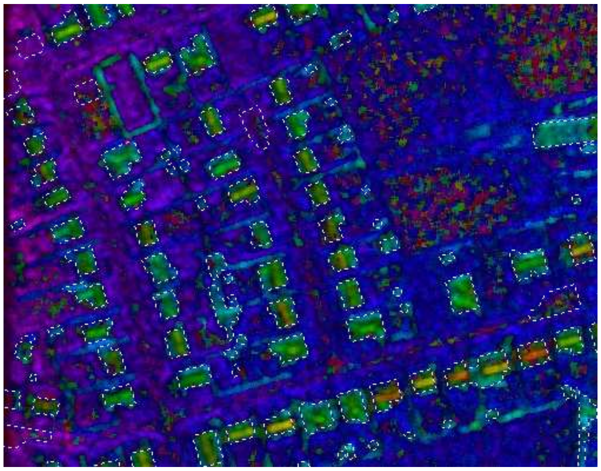

Figure 5. Polygons of the vector database superposed on the dense disparity map

6. Conclusions and perspectives

This paper has presented an original correspondence algorithm for stereo couples based on the association of left and right pixels with same horizontal gradient. The candidate disparity values are filtered by gradient orientation and intensity constraints. The estimated disparity value of each left image pixel is obtained from the best occurrence of candidate disparity values associated to pixels in a square neighborhood of the left pixel. Although isolated feature points lead to disparity values of little confidence, aggregation with the square window offers a valid way to obtain a spatially coherent disparity map as demonstrated by several examples with satellite or airborne imagery.

The implementation results in a short and quite simple source code, processing line after line and delivering a throughput of about 0.75Mpixel per second on a 2.33 GHz Intel Core 2 Duo CPU.

As suggested in the text, we intend to integrate a multi-resolution scheme to this approach in order to solve the common problem of disparity estimation in large poorly textured areas. To a smaller extent, we also plan to replace the intensity constraint by a multi-spectral condition to benefit from the different image bands of nowadays imagery.

7. Acknowledgements

This study was funded by the Belgian MoD. The author would like to thank the different data providers: the Belgian National Geographic Institute (IGN), the Brussels Region Informatics Centre (CIRB) and partners of the European GMOSS Network of Excellence.

8. References

[1] M. Brown, D. Burschka and G. Hager, "Advances in Computational Stereo", *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (PAMI)*, Vol. 25, NO 8, Aug 2003, pp. 993-1008.

[2] G. Medioni, R. Nevatia, "Segment-Based Stereo Matching", *COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING*, Vol. 31, 1985, pp. 2-18.

[3] J.L. Lotti, G. Giraudon, "Adaptive window algorithm for aerial image stereo", in *Proceedings of International Conference on Pattern Recognition*, Vol. A, IEEE Computer Society Press, Jerusalem, Israel, 1994, pp. 701-703.

[4] M. Idrissa, V. Lacroix, "A Multiresolution-MRF Approach for Stereo Dense Disparity Estimation", In *IEEE-GRSS/ISPRS Joint Urban Remote Sensing Event*, Shanghai, China, 20-22 May 2009.

[5] C. Beumier, "Building Verification from Disparity of Contour Points", In *Image Processing Theory, Tools & Applications (IPTA'08)*, Sousse, Tunisia, 23-26 Nov 2008, pp. 408-413.

[6] T. Twardowski, B. Cyganek, and J. Borgosz, "Gradient Based Dense Stereo Matching", In *Lecture Notes in Computer Science*, Vol. 3211, pp. 721-728.

[7] D. Scharstein, R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", *International Journal of Computer Vision*, Vol. 47, Apr-Jun2002, pp. 7-42.