AN HYBRID MATCHING METHOD

1YVES HECKEL
2RAUL QUEIROZ FEITOSA

Pontificia Universidade Católica do Rio de Janeiro
Centro Técnico-Científico
Departamento de Engenharia Elétrica - PUC-RJ
1yveshk@gmail.com; 2raul@puc-rio.br

ABSTRACT – The increasing resolution of satellite images made the construction of 3D models of urban areas from satellite stereo image pairs a viable approach. The accuracy in the construction of such models essentially relies on the matching of corresponding points in the image pair. Whereas the area-based matching algorithms, such as Least-Square Correlation, tend to be accurate in most cases, they fail in regions with significant height changes, for example over high-inclination zones or along the façades of high-buildings. Moreover Least-Square Correlation requires a good initial solution in order to attain good performances. The alternative to area based methods is the feature-based approach, such as the Scale-Adapted Harris Corners or the Scale-Invariant Feature Transform. They are usually less accurate but more robust to abrupt high changes and do not depend on good starting solutions. This work proposes a matching method that combines both approaches in order to obtain automatically a dense matching map independent on terrain inclination or on the height variation due to buildings. The basic idea consists of using feature-based matching as an initialization step for the area-based matching, which will refine and complete the correspondence map.

1 INTRODUCTION

After high spatial resolution satellite images became commercially available, 3D surface models built from satellite stereo pairs became a viable alternative for applications like telecommunication planning, disaster management and urban planning in general.

The quality of 3D models built from a pair of stereo images depends essentially on how accurately matching points are located. Automatic matching algorithms can be grouped into two main categories: area based and feature based methods.

Generally, area based methods perform better than feature based approaches, since they consider image patches and therefore a comparatively larger amount of surrounding pixels to establish the match. However, their performance degrades considerably for inclined scene’s regions where the image is distorted. The resulting region deformation can be partially compensated by geometric transformation whose parameters may be estimated using least square techniques. Yet such methods can only perform satisfactorily if a good starting solution can be provided.

Dense matching can be achieved with area based methods working in conjunction with the region growing technique. The approximated location of an initial set of matching points (seeds) is provided somehow. These locations are refined by an area based method. The position of new matching points is predicted in the neighbourhood of the initial set, and the search for new matching points continues from them on spreading over the image. At the end, a dense distribution of matching points over the image can be produced.

Figure 1 shows the outcome of the region-growing algorithm that started from 4 seed points (large numbered crosses). The matching points found this way appear as white points on the figure. On the lower right corner of Figure 1 a limitation of method becomes apparent. It shows a building and no matching points found around it. In fact the region growing process does not advance over areas where an abrupt height change occur, i.e., where area based matching fails.

Figure 1: Application of region growing combined with LS-Correlation.
The usual way to circumvent this limitation consists of plating new seeds manually somewhere within each region not reached by the region growing procedure and restart the search from them on. This manual intervention may be necessary on hundreds or even thousands of regions depending on the image size and on the geographical area.

In general, feature based methods are less accurate than area based methods and do not produce a so dense matching map. However they are not sensitive to height variations on the imaged scene.

The present work aims at the development of an automatic matching method that works well for flat as well as for inclined scene’s regions. An hybrid method is proposed that provides an accurate and dense set of matching points even across regions with strong height variation.

This work is organized as follows. Section 2 shortly describes the matching techniques underlying the proposed approach. Section 3 presents an overview of the software prototype built to evaluate the method, as well as the research being performed upon it. Some concluding remarks are presented in the final section.

2 UNDERLYING TECHNIQUES

The method proposed in this paper aims at combining the advantages of both, area based and feature based methods. It consists of two main steps. In the first step a feature-based algorithm, called Shift Invariant Feature Transform (SIFT) [2], generates a fairly dense cloud of matching points. These act as seeds for the second step that performs region growing combined with Least Square Correlation (LS-Correlation) [1], producing a denser and more accurate set of matching points. An overview about SIFT and LS-Correlation is presented in the next subsections.

2.1 Least Square Correlation

In traditional correlation the grey values ($g_1$ and $g_2$) of two matching patches centred respectively at $(x_1,y_1)$ and $(x_2,y_2)$ in each images of the stereo pair are assumed to meet the model shown on the left hand side of Table 1. The model is invariant to brightness and contrast changes, as well as to 2D spatial translation, but it does not care for changes in spatial scale, in 3D viewpoint or for 3D rotation.

$$\begin{align*}
x_2 &= x_1 + a \\
y_2 &= y_1 + b \\
g_2 &= \alpha + \beta g_1
\end{align*}$$

An improvement of traditional correlation, called Least Square Correlation (LS-Correlation), accounts for geometric image distortions, by means of an affine transformation between the spatial coordinates $(x_1,y_1)$ and $(x_2,y_2)$, as shown on the right hand side of Table 1. Conventional least-square methods are applied to estimate the 8 parameters of LS-Correlation model for each pair of putative matching patches. The absolute difference between the grey values measured on one image and the grey values predicted by the model based on the second image is the residual that is related to the level of matching.

$$\begin{align*}
x_2 &= a + bx_1 + by_1 \\
y_2 &= d + ex_1 + fy_1 \\
g_2 &= \alpha + \beta g_1
\end{align*}$$

Table 1: Comparison between traditional and LS-Correlation

2.2 Shift Invariant Feature Transform

SIFT has found many successfully applications in robotics [2]. It provides a set of features that are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. The method involves three steps described below.

Scale-space extreme point detection:

Scale invariance is obtained by building a scale-space pyramid, where each level is obtained by smoothing the previous one, usually with a Gaussian kernel, and eventually downsampling the result (see figure 2).

Key points are then selected in the scale-space pyramid. Several approaches were proposed in the literature to find the best key points. An extensive study on it was made by K. Mikolajczyk [3] [5]. Two kinds of approaches are investigated in this work (see figure 3):

- DoG (Difference of Gaussians): each pyramid level is subtracted from the preceding one. The extremis of so computed DoG pyramid across scale and space are selected as key points. This is a 3 dimensional search.
- Scale-Adapted Harris: the key points are the corners on each level of the scale-space pyramid. This is a 2 dimensional search.

In both cases the key points on low contrast neighbourhoods or over edges are discarded. Figure 3 shows clouds of key points obtained by both methods.

Descriptor Computation

Once the key points have been found, their descriptors are obtained in the following way. The gradient on each key point is computed, which will define
the global orientation of that feature. The neighbourhood with predetermined size around the key point is decomposed in four quadrants, as illustrated in figure 4.

The gradient on each pixel belonging to the neighbourhood is computed and its magnitude is weighted by a factor that decays as we move away from the key point according to a Gaussian 3D function. For each quadrant of the neighbourhood around the key point a histogram of the gradient amplitudes is computed as a function of their relative directions. In the example shown in figure 4 there are $2 \times 2 = 4$ histograms with 8 directions bins, i.e., a total of 36 values. The use of relative directions makes the descriptor invariant to rotation.

Matching
The grade of matching is given by the Euclidian distance between the corresponding descriptors. A pair of putative matching points is kept if they meet the following constraints:

- the distance between their descriptors is lower than a given low threshold,
- the Euclidian distance to the second closest key point descriptor is higher than a second high threshold, and
- the key points are close to the corresponding epipolar lines.

Figure 3: example of key points obtained by the DoG (left) and by the Scaled-Adapted Harris methods (right)

Figure 4: The SIFT descriptor

Figure 5: the main window of prototype’s GUI
3 SOFTWARE PROTOTYPE AND PRELIMINARY RESULTS

A software prototype implementing the proposed method has been built to assess its performance. Figure 5 shows its graphical user interface (GUI). Beyond basic functionalities, the GUI offers different tools for image visualisation, image enhancement, parameter set up, calibration, as well as RPC 3D reconstruction [6].

Figure 3 shows examples of results produced by our prototype for both SIFT variants mentioned in section 2.1.

Most algorithms, including SIFT, were written in MATLAB specifically for this prototype. A routine called DPCOR, developed by Karsten Jacobsen [4] at the Leibniz University of Hannover (Institut für Photogrammetrie und GeoInformation), that implements LS-Correlation, has been integrated to the prototype.

All these functionalities have been extensively tested and are operational.

4 FURTHER WORKS

The next step in this research is the validation of the proposed method. The analysis aims at determining the ability of the proposed method to automatically locate matching points where the region growing together with LS-Correlation fail due to abrupt height changes.

The prototype described in the previous section contains all tools required for a thorough experimental evaluation. Experiments are currently being performed upon a stereo pair of IKONOS images taken in January 2007. The images cover approximately 100 Km2 of the West Side of Rio de Janeiro, encompassing different types of settlements and forest areas with heights in the range of 0 to approximately 500 m.

ACKNOWLEDGEMENTS

The authors acknowledge CNPq and DLR for the financial support, and Dr. Karsten Jacobsen from the Leibniz University of Hannover (Institut für Photogrammetrie und GeoInformation), for sharing his software, data, experience and time.

REFERENCES


