Error-Free Stereo Reaches Optimal Theoretical Accuracy. Application to High Resolution Topography.

Neus Sabater

CMLA - ENS Cachan
Advisors: A. Almansa and J.-M. Morel

March 30th, 2010

The MISS Project
(Mathematics for Space Stereo Images)

- Collaboration agreement with CNES (French Space Agency)
- Main goal: Automatically computating Digital Elevation Models (DEM) from low-baseline stereo-pairs in urban areas
- Application to Pleiades images (Along track stereo capability, 0.7m resolution, $b/h \sim 0.12$)
- Control every step from image acquisition to final 3D model

Calibration → Rectification → Stereo → 3D Reconstruction

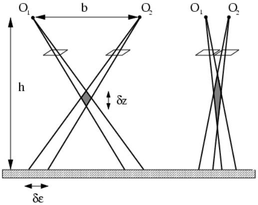
Image Matching Problems

1. Occlusion phenomenon
2. Radiometric changes
3. textureless region
4. Repetitive structures
5. Moving or disappearing objects
6. Fattening phenomenon (especially in urban areas)

Getting Sparse Reliable Matches without Parameters

- Need for rejecting false matches due to ambiguities and non existence of corresponding points
- Required technique: *A Contrario* Detection Theory
- Results: density of more than 50%

The Low-Baseline Case (Small b/h Ratio) [Delon and Rougé 07]



b=baseline
h= satellite height

Advantages:

- Less occlusions
- Quasi-simultaneous views
- Fewer radiometric changes

$\Rightarrow$ Original images are more similar

Challenges:

Let $R = \dfrac{h}{\text{focal length}}$ the image resolution then height $\simeq \dfrac{\text{disparity}}{b/h} R$

- To obtain the same accuracy for height, a higher accuracy has to be computed for disparity $\Rightarrow$ subpixel accuracy
- Required technique: Shannon Sampling Theory

Stereo Subpixel Accuracy

- Deformation Model: Let $\mathbf{x} = (x, y)$ and u_1, u_2 the rectified stereo images defined in $[0, a]^2$. We consider the deformation model

$$u_1(\mathbf{x}) = u(x + \varepsilon(\mathbf{x}), y) + n_1(\mathbf{x}) \ ,$$
$$u_2(\mathbf{x}) = u(\mathbf{x}) + n_2(\mathbf{x}).$$

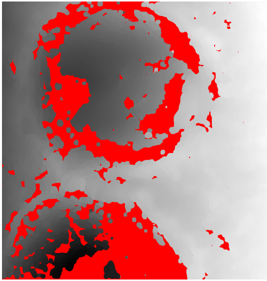
  where $n_i \sim \mathcal{N}(0, \sigma^2)$ is the noise, $u(\mathbf{x})$ the ideal image ant $\varepsilon(\mathbf{x})$ the disparity.
- This is a realistic model only in the low-baseline framework.

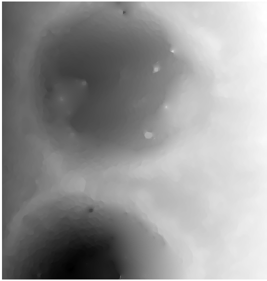
- Block-matching approach: ε is estimated at $\mathbf{x}_0$ by minimizing the quadratic distance

$$e_{\mathbf{x}_0}(\mu) := \int_{[0,a]^2} \varphi_{\mathbf{x}_0}(\mathbf{x})\big(u_1(\mathbf{x}) - u_2(\mathbf{x} + (\mu, 0))\big)^2 d\mathbf{x}.$$

  where $\varphi_{\mathbf{x}_0}(\mathbf{x}) := \varphi(\mathbf{x} - \mathbf{x}_0)$ is a window function.

MARS images





Stereo pair of images

Disparity map.
Red points were rejected

Interpolated disparity map

Stereo Subpixel Accuracy

Subpixel discrete correlation requires

1. An initial zoom x2 of the images
   $\longrightarrow$ then the quadratic distance is well-sampled
2. A band-limited window function $\varphi(\mathbf{x})$
   $\longrightarrow$ then the continuous and the discrete quadratic distance are equal
3. Fourier interpolation

- Margin of error?

Theorem (noise estimation)

Let $\mu(\mathbf{x}_0)$ be the estimation of $\varepsilon(\mathbf{x}_0)$ obtained minimizing $e^d_{\mathbf{x}_0}(\mu)$. Then

$$\mu(\mathbf{x}_0) = \frac{\int_{\varphi_{\mathbf{x}_0}} [u(\mathbf{x}+\varepsilon(\mathbf{x}))]_x^2 \varepsilon(\mathbf{x}) d\mathbf{x}}{\int_{\varphi_{\mathbf{x}_0}} [u(\mathbf{x}+\varepsilon(\mathbf{x}))]_x^2 d\mathbf{x}} + \mathcal{E}_{\mathbf{x}_0} + \mathcal{F}_{\mathbf{x}_0} + \mathcal{O}_{\mathbf{x}_0} \,,$$

where

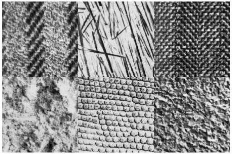
$$\mathrm{Var}(\mathcal{F}_{\mathbf{x}_0}) \ll \mathrm{Var}(\mathcal{E}_{\mathbf{x}_0})$$

$$\mathbb{E}\mathcal{O}_{\mathbf{x}_0} = O(\max_{\mathbf{x}\in B_{\mathbf{x}_0}} |\varepsilon(\mathbf{x}) - \mu|) \,, \quad \mathrm{Var}(\mathcal{O}_{\mathbf{x}_0}) = O(\max_{\mathbf{x}\in B_{\mathbf{x}_0}} |\varepsilon(\mathbf{x}) - \mu|^2).$$

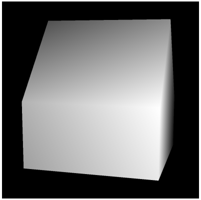
$$\boxed{\mathrm{Var}(\mathcal{E}_{\mathbf{x}_0}) = 2\sigma^2 \frac{\int [\varphi_{\mathbf{x}_0}(\mathbf{x}) u_x(\mathbf{x})]^2 \, d\mathbf{x}}{\left(\int \varphi_{\mathbf{x}_0}(\mathbf{x}) u_x(\mathbf{x})^2 d\mathbf{x}\right)^2}} \,,$$

Experimental Results (Datasets with Groundtruth)

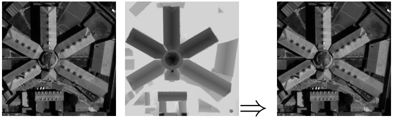
- Non-integer translation of textured images (Brodatz)



- Simulation of both images (CNES courtesy)



- 2nd image simulation from the reference and the groundtruth (CNES courtesy)



- Cross-correlation on Middlebury benchmark (7 to 9 images)

Experimental Results

- Experiments:
  - Study of noise sensibility by adding independent white noises (SNR=∞,...,20)
  - Study of the algorithm behavior depending on the baseline (b/h=0.05,...,0.5)

Results

- **Our algorithm reaches theoretical accuracy bounds**
- **Accuracy of about $\frac{5}{100}$ pixels**
  $\Longrightarrow$ **Height accuracy of about 0.3m (b/h$\sim$0.12, res.=0.7m)**

- Great improvement over state of the art algorithms

Improvements - Work in Progress

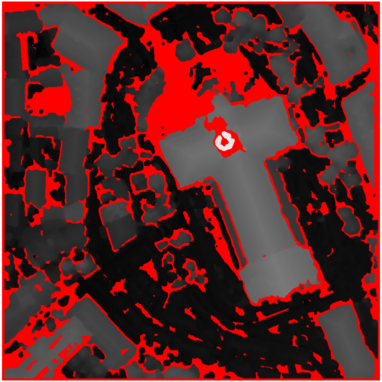
- Application to high resolution topography: A higher density of matched points is requied
- How?
  - Computation of denser results in objects with a strong projective transformation (building walls)
  - Generalization to multi-images
  - Contrast invariant
  - Disparity map interpolation with a global method

Future Work at Caltech

- Detection of geometrical surface Earth variations in space/time (Earthquakes,...)

- Further generalization of stereovision:
  1. The scene is not rigid
  2. The epipolar direction is not the only one to be studied
  3. There is no restriction on the number of images
  4. The images to be studied do not necessarily come from the same acquisition system (different resolutions)

- St. Sernin church, Toulouse (CNES courtesy).



Resolution: 20cm. b/h=0.08

Density 82%

Thank you for your attention!