A contrario block matching

Neus Sabater [1]
Supervisors : A. Almansa [2] J-M. Morel [1]

[1]CMLA, ENS Cachan

[2]TELECOM ParisTech

December 2008

What is Stereovision?

Stereo Principle

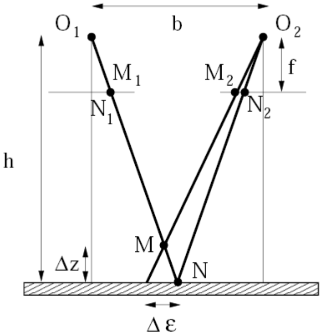
f focal lenght.

b distance between centers.

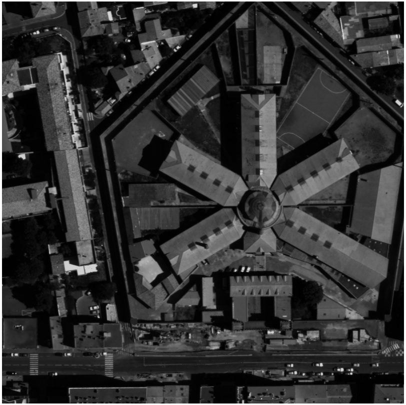
h camera-scene distance.

$$\triangle\varepsilon = \frac{M_2 N_2}{f/h}$$

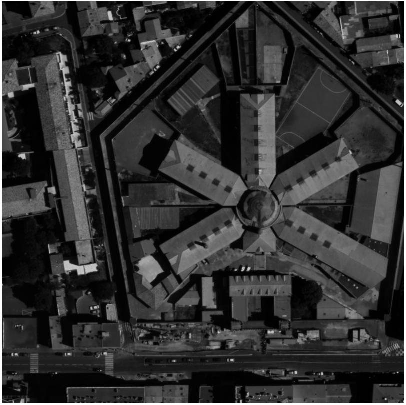
$$\triangle z \simeq \frac{\triangle\varepsilon}{b/h}$$



Real stereo pair

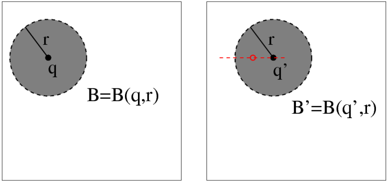


Real stereo pair



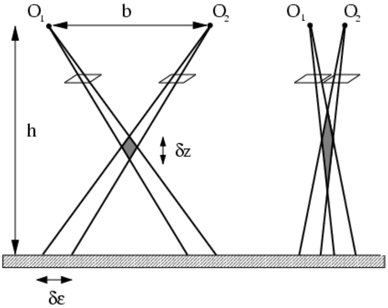
Block-matching

Given two images u and u', if q is a pixel of the reference image u and B a neighborhood (or block or patch) around that pixel,
is there a block B' in the secondary image u' that is perceptually equal to B ?



- Which method for block-matching?
  A similarity mesure has to be defined.

The small baseline case



The bigger b/h, the more precise the depth measurement.

Advantages :

- less **occlusions**
- views **quasi-simultaneous**
- less **radiometric changes**

Challenges :

- More accuracy of $\Delta\varepsilon$ for the same accuracy of Δz

$$\Delta z \simeq \frac{\Delta\varepsilon[pix]}{b/h} R[m/pix]$$

CNES' MARC software

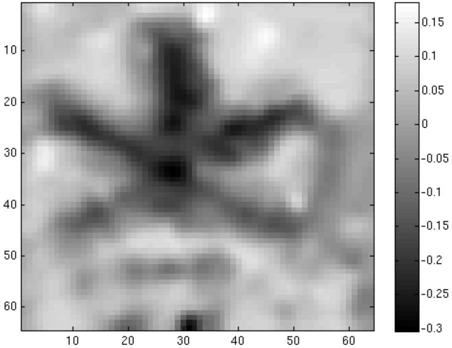
Strengths and weaknesses :

- Excellent **subpixel localization** of individual matches (small baseline case)
- Good **disparity error estimate**
  (...provided the correct local maximum is detected!)
- Only partial correction of **adhesion effects**
- Only partial elimination of **false correspondences**
  (at the price of strong regularization (smoothing) of the elevation map)

MARC's approach to avoid false correspondences

1. Multiscale structure to avoid false local minima

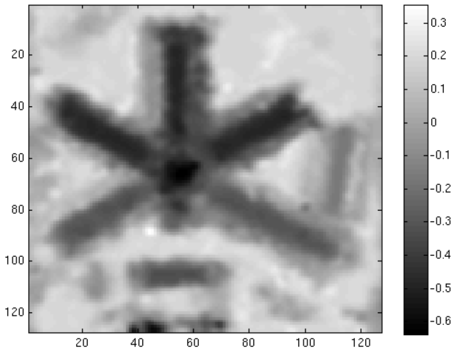
scale 3 (1 graylevel = 4 pixel disparity)



MARC's approach to avoid false correspondences

1. Multiscale structure to avoid false local minima

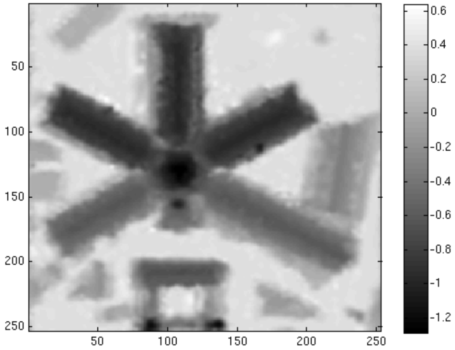
scale 2 (1 graylevel = 2 pixel disparity)



MARC's approach to avoid false correspondences

1. Multiscale structure to avoid false local minima

scale 1 (1 graylevel = 1 pixel disparity)

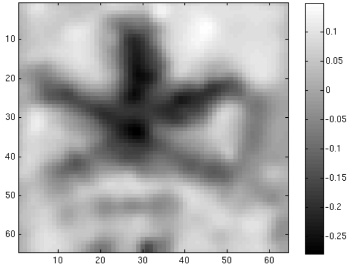


MARC's approach to avoid false correspondences

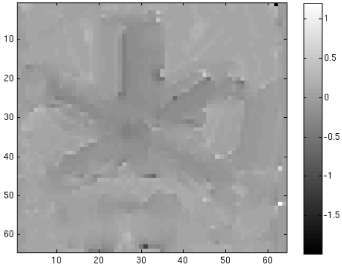
2. At each scale smoothing interpolation is used

scale 3 (1 graylevel = 4 pixel disparity)

after smoothing interpolation



before smoothing interpolation



Approach & Methodology

Our approach

- **Eliminate false correspondences** before regularization/interpolation
- **Detect geometrical 3D structure** from this point cloud + image information (planes, edges, curved surfaces)
- regularize/interpolate based on detected structure

Methodology

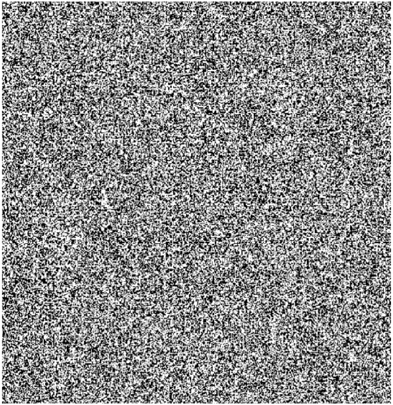
A contrario detection to elliminate false detections

A contrario detection methods

- Inspired on **Gestalt theory** of visual perception
  ⇒ *"qualitative" laws of perceptual grouping*
- Augmented with **sampling** theory and **probabilistic** approach
  ⇒ *mathematical and computational framework*
- Compute **automatic detection thresholds** using common sense principle *"no detection in white noise "*
- **Helmholtz principle**
  A grouping G of objects having a common characteristic is perceptually significant, if the the expected number of occurrences of such an event by chance is very small
- What do we mean by "chance" ? *This is the "a contrario" model. The least structured statistical model that can be reasonably conceived. Usually white noise.*

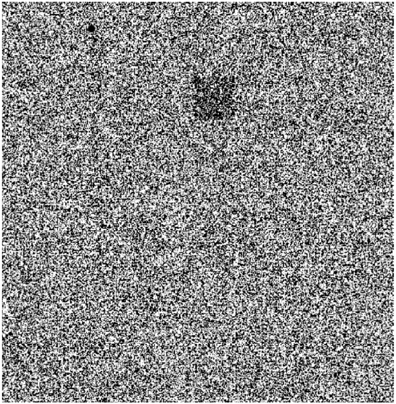
A toy example

Do you see a dark square here ?



A toy example

And here ?



Which a contrario model for block-matching ?

- **Point matching is hopeless** (few detections)

$$NFA(q,q') = N_{test} * P(|u(q) - u'(q')| < \theta) = \frac{N_{tests}}{SNR} < 1(\text{for } \theta = \sigma)$$

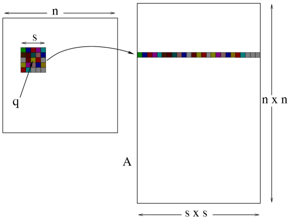
- Some **neighborhood information** is needed

$$NFA(q,q') = N_{test} * P(\frac{1}{B(0)} \sum_{d \in B(0)} |u(q+d) - u'(q'+d)|^2 < \theta^2)$$

Which (a contrario) statistical model for image patches $u'|_{B(q')}$?

- **learning empirical distribution?** hopeless! (very high dimensional space)
- learn some **marginals** and assume **independence**
  - **independent pixels**: too naïve! leads to false matches
  - **independent PCA coefficients** : OK!

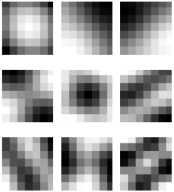
Extracting neighborhood features : PCA



- An $n \times n$ image contains $\approx n^2$ (overlapping) $s \times s$ blocks B
- The grayvalues $u(B)$ are considered as instances of a random s^2-vector $X_1, \ldots, X_{s^2}$
- Matrix A stores the n^2 instances of $X_1, \ldots, X_{s^2}$
- Build the covariance matrix $C = Cov(A)$. The eigenvectors of C provide the Principal Component (PC) basis.
- The projections of a block $u(B)$ in the direction of each eigenvector provides the PC coefficients.

PCA analysis of 7x7 blocks

(a)



(b)

Figure 2: (a) Reference Image I. (b) The nine first principal components with larger eigenvalues of I, for 7×7 neighborhoods.

PCA analysis of 7x7 blocks

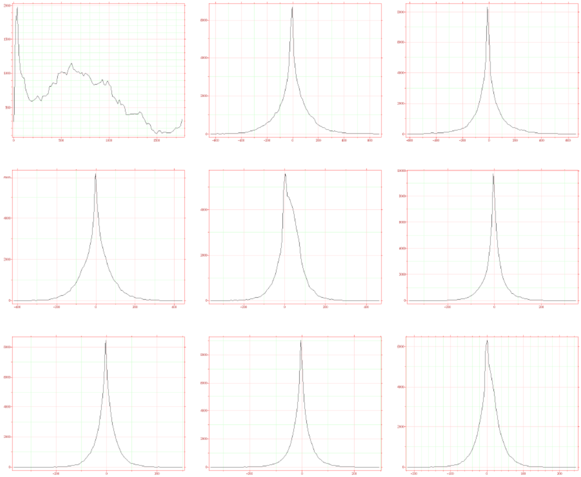
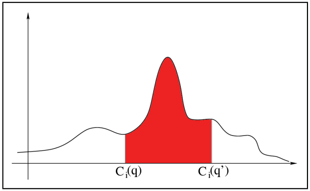


Figure 3: Histograms of the nine first principal components.

Similarity measure

- Choose the N PC's having the largest eigenvalues among the s^2.
- Let $c_i(q)$ be the value of the i-th PC coefficient at point q, $i = 1, \ldots, N$.
- Let $H_i(q) := H_i(c_i(q))$ be the cumulative histogram of $c_i(q)$ for the secondary image.



Definition (Empirical Probability of coefficient similarity)

Given a stereo pair of images u and u' and $q \in u$, $q' \in u'$. For every $i = 1, \ldots, N$ the empirical probability that an observed block $B_{q'}$ in u' be similar to B_q for the feature i ,

$$\widehat{p^i}(B_q, B_{q'}) = 2 \cdot |H_i(q) - H_i(q')|.$$

Similarity measure

- Consider a family of Q probability thresholds:
  $p_j = \left(\frac{1}{2}\right)^{j-1}; \quad j = 1, \ldots, Q.$

Definition (Increasing probability distribution upper bound)

For a pair of points q and q', and for all $i = 1, \ldots, N$,

$$p^i_{q,q'} = inf_j\{p_j \,|\, p_j \geq \widehat{p^i_{q,q'}},\ p_j \geq p^{i-1}_{q,q'}\ si\ i \geq 2\}.$$

Definition (ε-meaningful block-match)

A pair of points q and q' is an ε-meaningful block-match if $NFA_{q,q'} \leq \varepsilon$, where

$$NFA_{q,q'} = N_{test} \prod_{i=1}^{N} p^i_{q,q'}.$$

Similarity measure

- Consider a family of Q probability thresholds: $p_j = \left(\frac{1}{2}\right)^{j-1}; \quad j = 1, \ldots, Q.$

Definition (Increasing probability distribution upper bound)

For a pair of points q and q', and for all $i = 1, \ldots, N$,

$$p^i_{q,q'} = inf_j \{ p_j \,|\, p_j \geq \widehat{p^i_{q,q'}},\ p_j \geq p^{i-1}_{q,q'}\ si\ i \geq 2 \}.$$

Definition (ε-meaningful block-match)

A pair of points q and q' is an ε-meaningful block-match if $NFA_{q,q'} \leq \varepsilon$, where

$$NFA_{q,q'} = N_{test} \prod_{i=1}^{N} p^i_{q,q'}.$$

Computation of the number of tests

- The image size: n^2.
- The size of the search domain for the second point q': n_{epipol}.
- The number of possible increasing quantized probability distributions $p^i_{q,q'}$ with N PC's and Q probability thresholds: $FC_{N,Q}$.

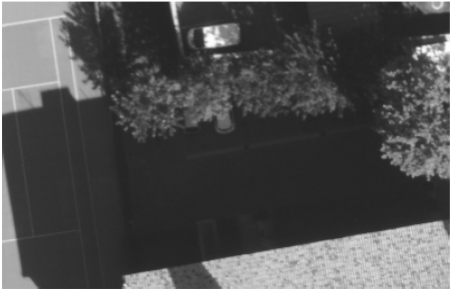
$$\boxed{N_{test} = n^2 n_{epipol} FC_{N,Q}}$$

$$FC_{N,Q} = \sharp\{f : [\![1, N]\!] \to [\![1, Q]\!] \mid f(x) \leq f(y), \forall x \leq y\}$$

$$FC_{N,Q} = \sum_{t=0}^{Q} (t+1) \binom{N+Q-t-3}{Q-t-1}$$

Shadows

A shadow seems homogeneous from the global contrast perspective but expanding contrast reveals some local structure.



Conclusion : Local PCA to avoid false detections in shadows.

Local PCA analysis of 7*x*7 blocks

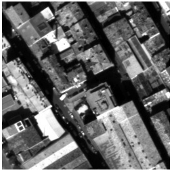
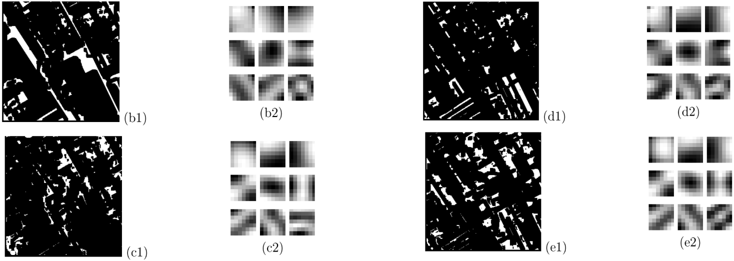


Image partition into 4 classes based on mean and variance of the patches:



Self-similarity thresold

Given a meaningful block match (q,q') between images u and u'.
The match is **rejected** if there is a point $r \in u$ near q such that

$$d_{L^2}(B_q, B_r) < d_{L^2}(B_q, B_{q'})$$

This sanity check avoids **"stroboscopic effects"** (very large disparity errors due to self-similar repeated structures) at the cost of a slightly sparser disparity map.

Précision subpixellique : Raffinement par corrélation

On détermine la disparité μ^d sur chaque point x_0:

$$\mu^d(x_0) := \arg\min_{\mu} e^d(\mu, x_0) \ , \tag{1}$$

$$e^d(\mu, x_0) := \|\tau_\mu u - \tilde{u}\|^2_{\varphi_{x_0}} \ . \tag{2}$$

où

- $\tau_\mu u(x) = u(x + \mu)$,
- $\varphi_{x_0} = \varphi(x - x_0)$ fenêtre symmetrique à support compact.
- $\|u\|^2_{\varphi_{x_0}} = \langle u, u \rangle_{\varphi_{x_0}}$,
- $\langle u, v \rangle_{\varphi_{x_0}} = (\frac{1}{2})^2 \sum_{x \in \frac{1}{2}\mathbb{Z}^2} u(x)v(x)\varphi_{x_0}(x)$.

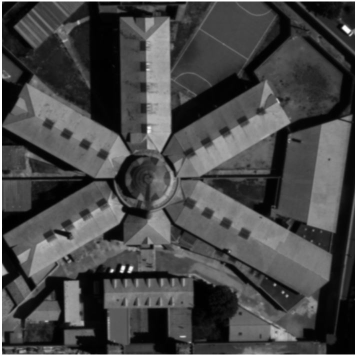
Subpixel discrete correlation requires x2 zoom.

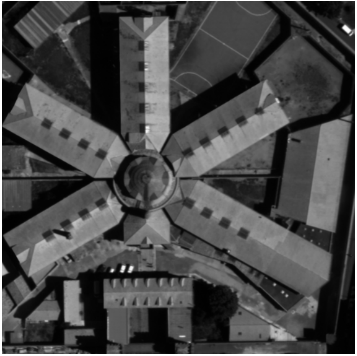
Causes de l'Erreur.

$$\mu^d(x_0) = \varepsilon(x_0) + \underbrace{A(x_0)}_{\text{adhesion error}} + \underbrace{r(x)}_{\text{noise error}} + \underbrace{\mu^d(x_0) - \mu^c(x_0)}_{\text{discretization error}}, \quad (3)$$

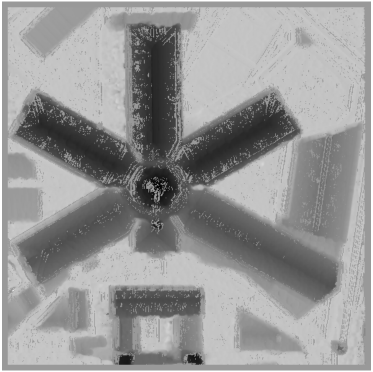
- Gross errors ($\sim$ 10 pixels).
- Adhesion errors ($\sim$ 1 pixels).
- Noise errors ($\sim$ 0.1 pixels).
- Discretization errors ($\sim$ 0.01 pixels).

Simulated stereo pair

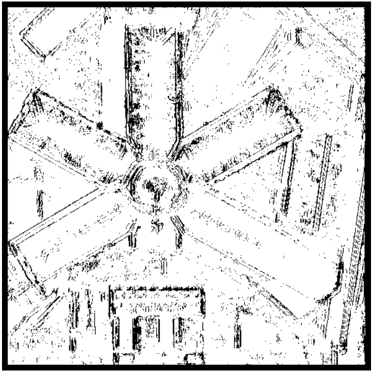




Simulated stereo pair

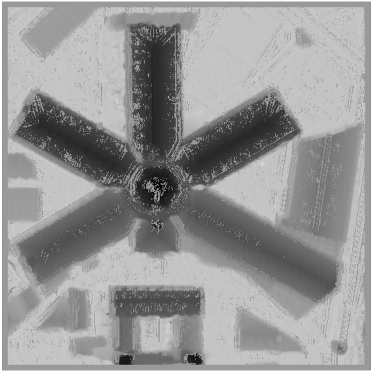


Disparity Map

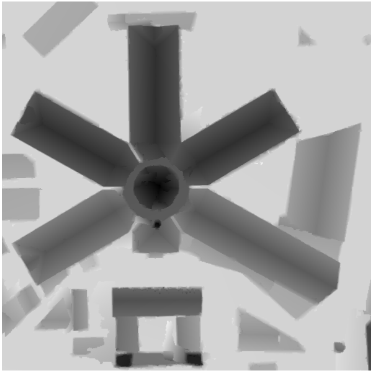


Associated mask of points

Simulated stereo pair



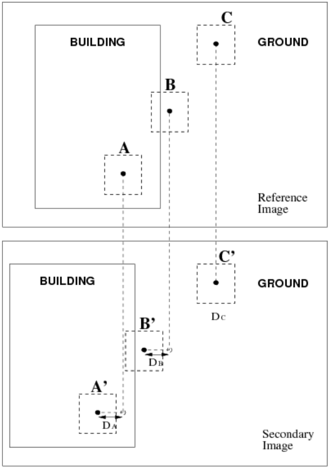
Disparity Map



Groundtruth

Solving the adhesion problem

- Block-matching methods suffer distortions of the DEM: **Adhesion Artifact**.
- Dilation of structures due to the prevalence of borders.
- Depends on the size of the block/window.



Solving the adhesion problem

- Let ε be the estimated disparity map between u and v.
- $\forall c \in u$, we consider pixels in B_c minimizing

$$\alpha^c(x) = Angle\Big(\frac{\nabla u}{|\nabla u|}(x)\,,\ \frac{\nabla v}{|\nabla v|}(x+\varepsilon(c))\Big).$$

- $\forall q \in u$ such that $|\nabla u|(q) > 3\sigma$, we define the corrected disparity as:

$$\tilde{\varepsilon}(q) = \underset{y \in B_q}{Med}\Big\{\varepsilon(y)\ \Big|\ \alpha^y(q) < \underset{z \in B_y}{Med}\ \alpha^y(z)\Big\},$$

- Points with non tolerated error:

$$M = \Big\{x \in u\ \Big|\ |\varepsilon(x) - \tilde{\varepsilon}(x)| > \theta\Big\}$$

Solving the adhesion problem

- Risk points: $\mathrm{RP} = \bigcup_{c \in M} E_c$

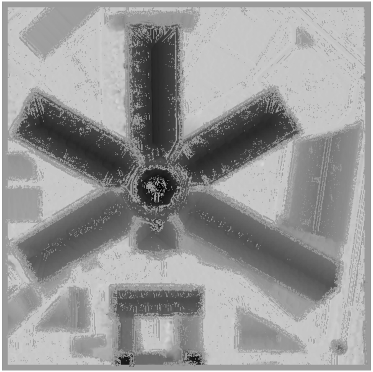
$$\forall c \in M, \qquad E_c = \left\{ x \in B_c \;\middle|\; |\nabla u|(x) \simeq \underset{z \in B_c}{Max} |\nabla u|(z) \right\}$$

- The final merged disparity:

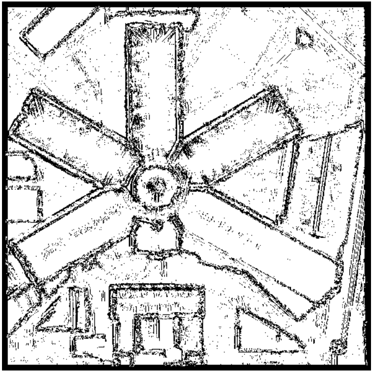
$$\varepsilon_F(x) = \begin{cases} \tilde{\varepsilon}(x) & \text{if} \quad B_x \cap \mathrm{RP} \neq \emptyset \\ \varepsilon(x) & \text{otherwise} \end{cases} \tag{4}$$

- Post-processing

Final result of the simulated stereo pair

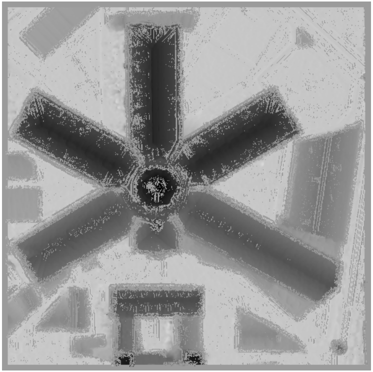


Merged Disparity Map

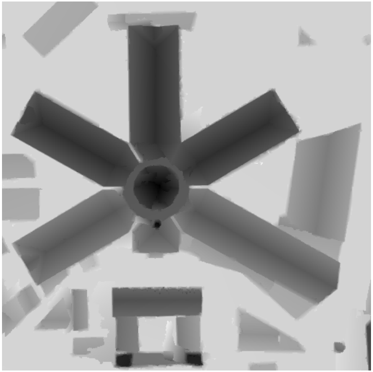


Associated mask of points

Final result of the simulated stereo pair

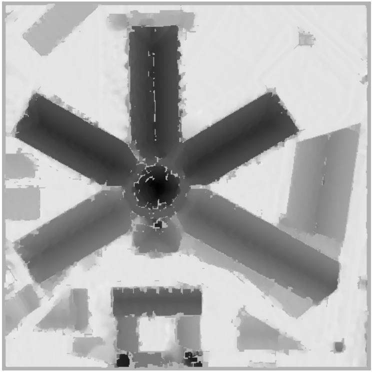


Merged Disparity Map

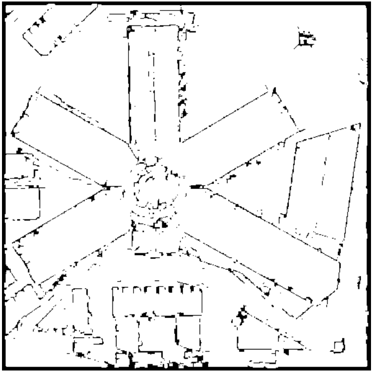


Groundtruth

Final result of the simulated stereo pair



Post-processing



Associated mask of points

Paire simulée

	RMSE	Matched points	Bad matches
ε	0.163	89.41%	0.46%
ε_{lsd}	0.075	45.39%	0.01%
ε_{final}	0.093	95.49%	0.12%

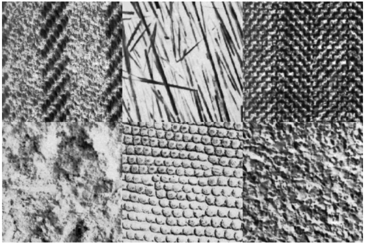
- $RMSE = \left(\frac{\sum_{i \in I}(u_i - gt_i)^2}{|I|}\right)^{\frac{1}{2}}$,
- Bad matches: $|u_i - gt_i| > 1$.

Noise sensitivity

SNR	RMSE	Matched points	Bad matches
∞	0.063	63.9	0.01
357.32	0.075	45.39	0.01
178.66	0.081	33.15	0.01
125.06	0.096	24.73	0.02

- $RMSE = \left(\frac{\sum_{i \in I}(d_i - gt_i)^2}{|I|}\right)^{\frac{1}{2}}$,
- $SNR = \frac{\| u \|_2}{\sigma_n}$,
- Bad matches: $|d_i - gt_i| > 1$.

Textured images



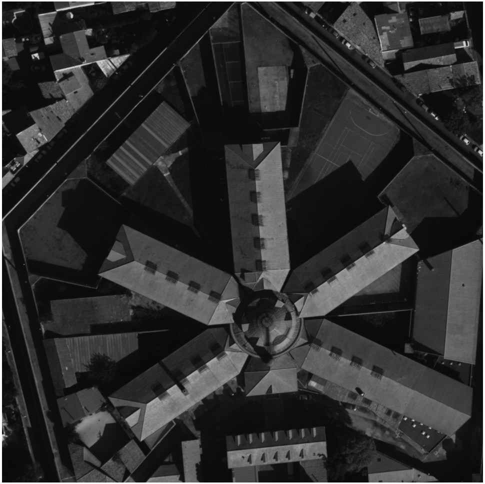
SNR	RMSE	Matches	Bad matches
∞	0.0053	100	0.0
96.38	0.0073	99.8	0.0
48.19	0.0109	99.8	0.0
32.12	0.0160	98.7	0.0
24.09	0.0203	87.1	0.0

Paire réelle

50194-60194. B/H = 0.04. Résolution = 20cm.

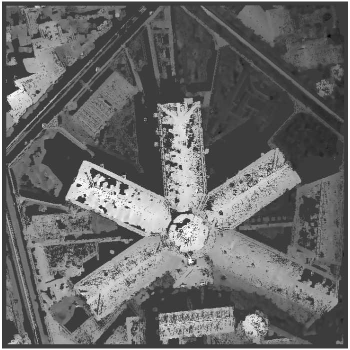
Paire réelle

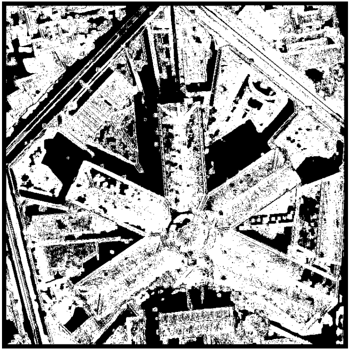
50194-60194. B/H = 0.04. Résolution = 20cm.



Paire réelle

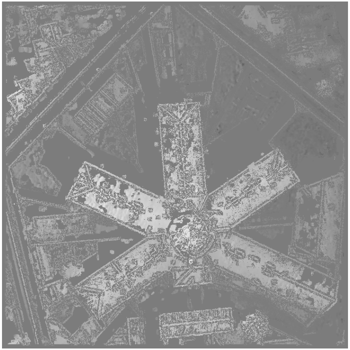
50194-60194. B/H = 0.04. Résolution = 20cm.

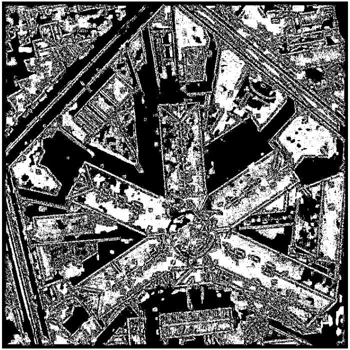




Paire réelle

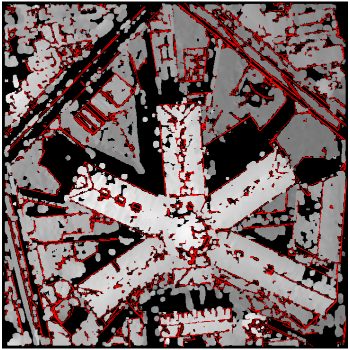
50194-60194. B/H = 0.04. Résolution = 20cm.





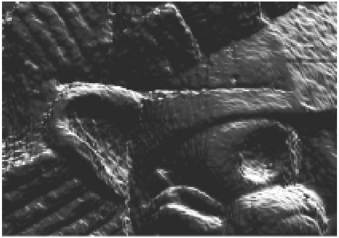
Paire réelle

50194-60194. B/H = 0.04. Résolution = 20cm.



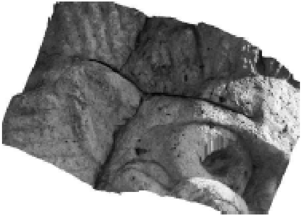
Lion

Lion



(a) upper view of a 3D rendering of the computed surface

(b) slanted view of a 3D rendering of the computed surface



(c) slanted view of the computed surface with the reference image rendered as texture on the surface

Conclusions

- Block-matching with nearly zero false matching rate is possible !!
  (at the cost of 50% to 80% sparse disparity map)
- Fine-localization error for these matches is between 0.1 and 0.01 pixels in typical SNR settings.

Improvements

- Extend to color and vector-valued images

Perspectives

- Use the resulting point cloud to automatically detect 3D structure
- Use detected 3D structure for interpolation and regularization