ERROR-FREE STEREO REACHES OPTIMAL THEORETICAL ACCURACY

Neus Sabater
Sabater@cmla.ens-cachan.fr

Summary

A new parameterless method for detecting matches in pairs of stereo images with virtually no false matches [1] is proposed. A very precise disparity map is provided avoiding the fattening effect. Matches are found *a contrario* and Principal Component Analysis (PCA) is used to extract a feature vector in each point. The subpixel accuracy is obtained by interpolating and minimizing the quadratic distance under Shannon conditions.

Full Control of False Matches

- Feature Extraction: Local PCA. The patch features are the PCA coefficients trained separately for patches with similar mean and variance. ADVANTAGE: More precise results in regions with lack of information (shadows and saturated regions).
- The *a contrario* model is a statistical model for image patches deriving from the Helmholtz Principle: A geometric structure is meaningful if the expected number of occurrences is very small in white noise. In our case: An empirical background model is fixed from the images. The relevant matches are detected as rare events for this model.
  - Let B_q be a block in the reference image. The empirical probability that an observed block $B_{q'}$ in the secondary image be similar to B_q for the feature i is

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

  where H_i is the cumulative histogram of the principal components coefficients for the secondary image.
  - The Number of False Alarms (NFA) is the expected number of appearances of the match (q, q') :

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

  where N_{test} is the number of matches to be tested, N is the number of principal components considered and $p^i_{q,q'}$ is a quantification of $\widehat{p^i_{q,q'}}$ and is imposed to be non-decreasing in i.
  - A match (q, q') is ϵ-meanigful when $NFA_{q,q'} < \epsilon$.

Theorem 1. *Under the a contrario assumption that all principal components are independent,* $\mathbb{E}\#\{\varepsilon - meaningful\ matches\} \leq \varepsilon$.

CAUSE OF FALSE MATCHES	PROPOSED SOLUTIONS
Occlusions , moving objects and noise	*a contrario* model
Poor textured regions	Local PCA
Stroboscopic effects	Self-similarity threshold

Subpixel Accuracy:

- Once a meaningful match have been found in x_0 its disparity is refined.

$$\mu^d(x_0) := \arg\min_{\mu \in \mathbb{R}} e^d_{x_0}(\mu), \quad e^d_{x_0}(\mu) := \|\tau_\mu u - \tilde{u}\|^2_{\varphi_{x_0}} \, ,$$

where $\langle u, v \rangle_{\varphi_{x_0}}$ is the weighted discrete scalar product and $\| \cdot \|_{\varphi_{x_0}}$ the corresponding weighted norm. We write $\tau_\mu u(x) = u(x+\mu)$, and $\varphi_{x_0} = \varphi(x - x_0)$ is a symmetric and normalized window in a compact support (e.g. prolate).

Theorem 2. *Assuming that* $\tilde{u}(x) = u(x + \varepsilon(x)) + n(x)$ *where the noise* $n \sim N(0, \sigma)$ *. Then*

$$\mu^d(x_0) - \varepsilon(x_0) = \underbrace{\frac{\langle u'^2, \varepsilon \rangle_{\varphi_{x_0}}}{\|u'\|^2_{\varphi_{x_0}}} - \varepsilon(x_0)}_{\text{fattening error}} + \underbrace{\frac{\langle u', n \rangle_{\varphi_{x_0}}}{\|u'\|^2_{\varphi_{x_0}}}}_{\text{noise error} \sim N\left(0, \frac{\sigma \|u'\|_{\varphi^2_{x_0}}}{\|u'\|^2_{\varphi_{x_0}}}\right)} + o(|\mu^d - \varepsilon|^2)$$

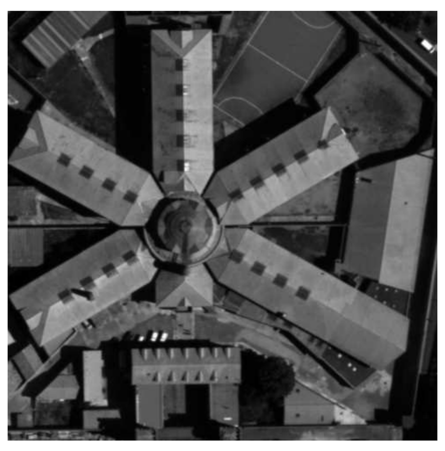
Remark By Thm 2 the noise error is estimated at each point x_0.

Correcting the Fattening Effect

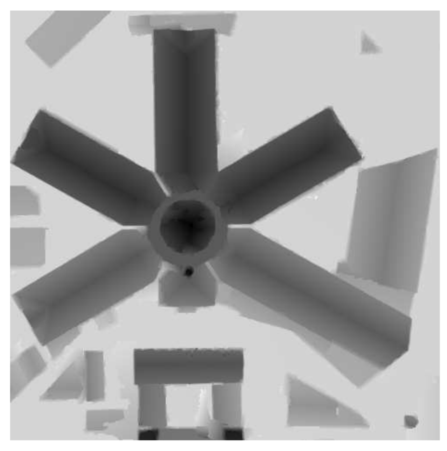
- Dilation of the size of the patch of structures with boundaries coinciding with depth discontinuities.
- PROPOSED APPROACH: Compute a new disparity map which considers the disparity computed at pixel x as a feasible disparity for a whole set of pixels: namely, all pixels in the patch with best matching gradient angle. Reject the match when the two solutions differ by more than the allowed precision.

Experiments

Simulated Pair:

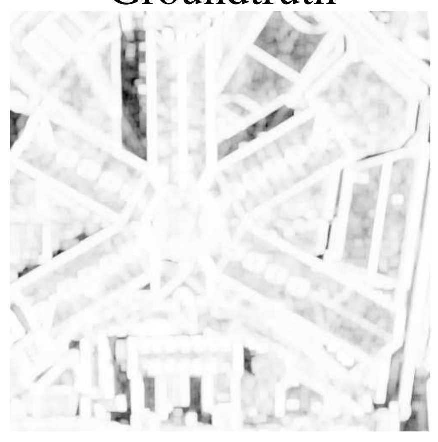


First image



Groundtruth

Disparity. Red pixels aren't matched (rejected by the *a contrario* method or the fattening correction.)



Noise error estimation at each point. The darker the pixel, the higher the error.

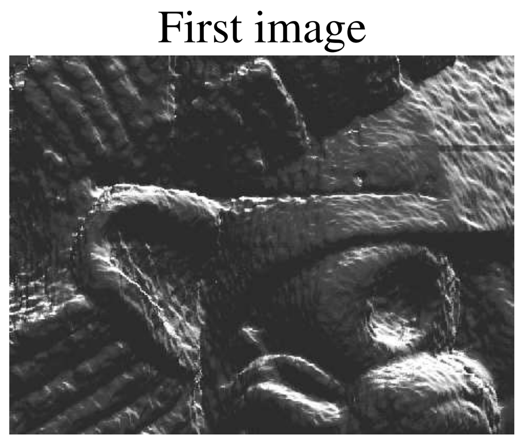
SNR	Matches	Bad matches	RMSE	Predicted noise error
∞	70.6%	0%	0.023	0
357.32	63.3%	0%	0.033	0.023
125.06	41.5%	0.02%	0.058	0.065

Root Mean Squared Error for different Signal to Noise Ratios.

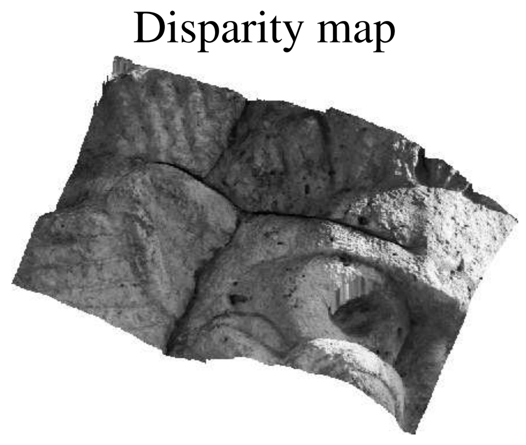
Lion Experiment:

First image

Disparity map



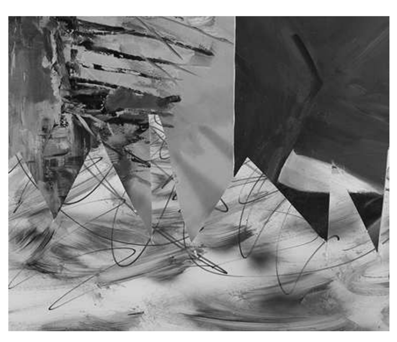
Upper view of the 3D rendering of the computer surface



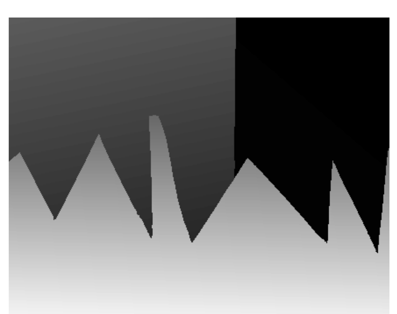
Slanted view of the computed surface with the reference image

Stereo benchmark: The Middlebury dataset [2] provides 9 orthorectified views at uniform intervals of piecewise planar scenes. A cross-validation is done taking the central view as reference:

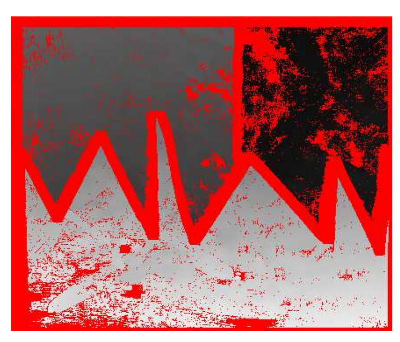
	Matches	Bad Matches	RMSE	Predicted noise error
Sawtooth	45.2 %	0.1 %	0.090	0.081
Venus	47.2 %	0.1 %	0.056	0.061



First image



Groundtruth



Disparity map

References

[1] N. Sabater, A. Almansa et J.-M. Morel, *Rejecting Wrong Matches in Stereovision.* CMLA preprint 2008-28.

[2] Middleburry dataset. *http://vision.middlebury.edu/stereo/*.