Two-view Stereovision

Olivier Jais-Nielsen
Technical University of Denmark (DTU)

Abstract

The stereo 3D reconstruction denotes a family of problems: reconstruction of interest points, reconstruction of surfaces, determination of camera motion, use of calibrated / uncalibrated cameras, …

We will investigate the problem of textured surface reconstruction from a pair of photos knowing only standard EXIF information.

The program uses Matlab with the Image Processing Toolbox and three third-party modules.

Calibration matrices estimation

$$K = \begin{pmatrix} f.r_x & 0 & \frac{W}{2} \\ 0 & f.r_y & \frac{H}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

We don't consider any skew. The focal length (f) and the horizontal and vertical resolutions (r_x and r_y) are read in the EXIF tags.

The optical center is evaluated as the image center $(\frac{W}{2}, \frac{H}{2})$.

Interest points extractions and matches

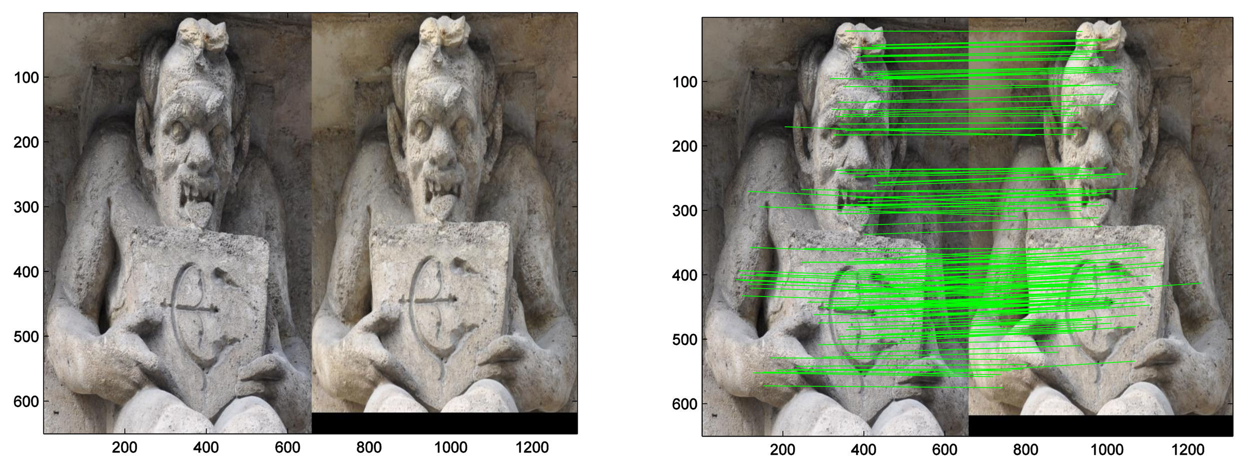
SIFT features are extracted and matched using the VLFeat library (1). Matching is performed using Euclidian distances on the descriptors.

Harris corners are extracted using a function from Peter Kovesi (2). The Matching is performed using normalized cross correlation on square patches centered on the interest points.

Robust matches and fundamental matrix estimation

From the previous pair candidates, a selection is performed using a RANSAC algorithm to determine the fundamental matrix.

The fundamental matrix is evaluated using the 8-point algorithm. The distance determining inliers is the Sampson distance: $d\big((P,Q),F\big) = \frac{(Q^T F P)^2}{\|FP\|^2 + \|F^T Q\|^2}$

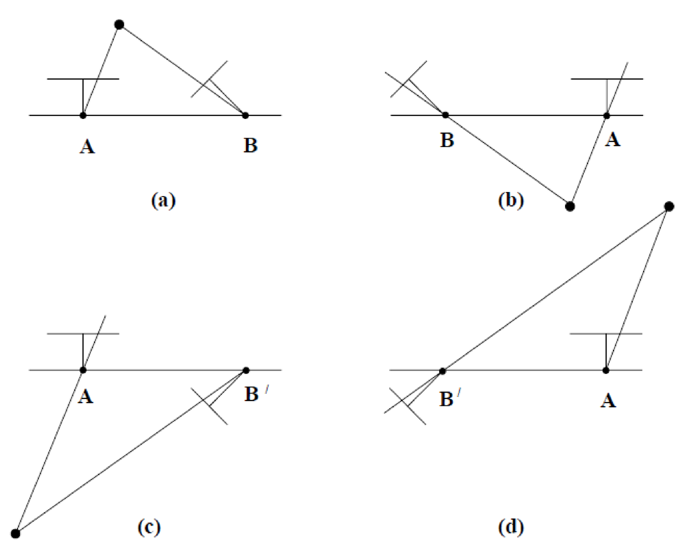


Camera matrices estimation

$P_1 = K_1[I|0]$ and $P_2 = K_2[R|t]$

R and t are deduced from the singular value decomposition of the essential matrix: $E = K_2^T F K_1$

Four solutions ambiguity:

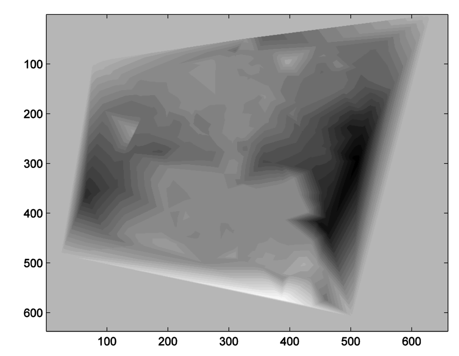


Only one solution has its object in front of both cameras → the robust pairs are triangulated in both camera systems and the sign of the depths is checked, for each possibility. In one case the depths are positive in both systems.

Rectification

$P_1' = K'[R'|0]$
$P_2' = K'[R'|t']$

Rectification error estimation / bounding geometry

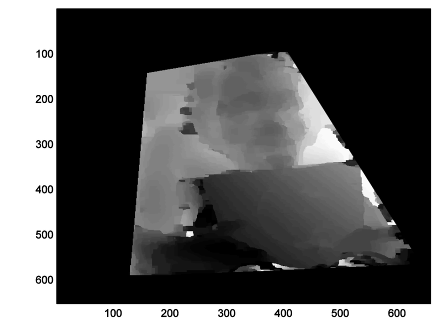


Robust matches are extracted from the rectified images. The rectification error is the vertical distance. A linear interpolation covers the whole bounding geometry.

Disparity estimation

Matching points are searched along the same horizontal line. For a point in the first image, each disparity is associated to a cost: $V(P,d) = -\log\left[c\left(S_1(P), S_2\left(P + \begin{pmatrix} d \\ 0 \end{pmatrix}\right)\right)\right]$ where $S_i(P)$ is a square patch of the image i centered on point P and $c(.,.)$ is the normalized cross correlation.

A 1-clique cost is defined as: $V(d_1, d_2) = -\beta \exp(-\frac{|d_1 - d_2|}{100})$

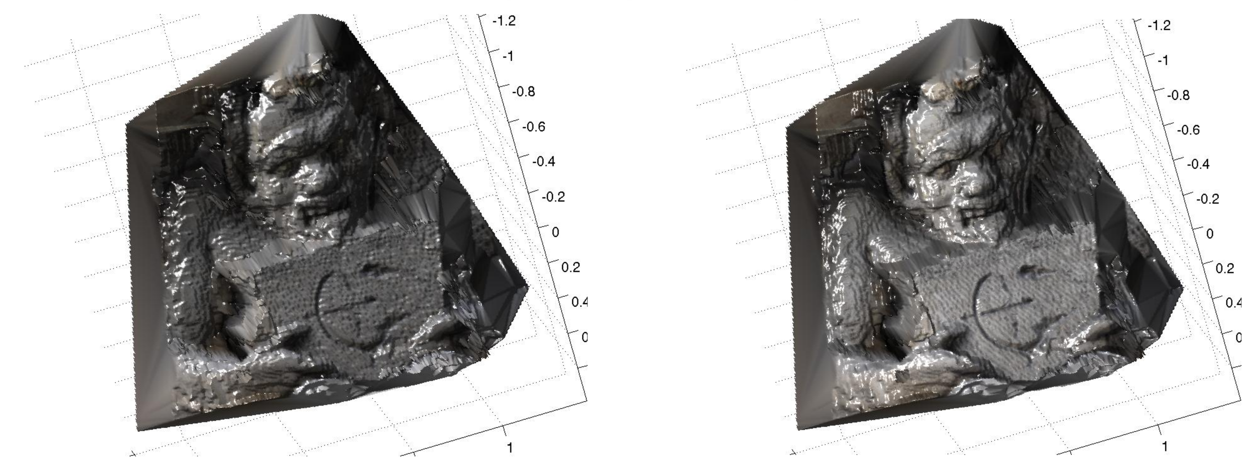


This MRF is solved with graph-cut/alpha-expansion. Each graph-cut is performed with Vladimir Kolmogorov's module (3).

Surface reconstruction

Triangulation of the pairs: $\begin{cases} p_1 = P_1 P \\ p_2 = P_2 P \end{cases} \Rightarrow \begin{cases} p_1 \wedge P_1 P = 0 \\ p_2 \wedge P_2 P = 0 \end{cases}$

A linear interpolation of the depth/color over the first two dimensions of the resulting 3D points gives a surface.



Conclusion

Surface reconstruction is well adapted to two-view geometry as it does not need occlusions to be managed. The rather raw calibration seems sufficient to obtain a usable rectification, coupled with an error estimation.

However, building the cost function for the MRF is extremely time consuming and reducing the resolution makes the number of disparities lower thus the result much worse. A better smoothing including 3-cliques might remove the quite prominent noise.

References

1. VLFeat – A. Vedaldi, B. Fulkerson
2. Harris corner detector – P. Kovesi
3. Maxflow – V.Kolmogorov
4. Multiple View Geometry in computer vision - R. Hartley, A. Zisserman
5. Epipolar Rectification – A. Fusiello