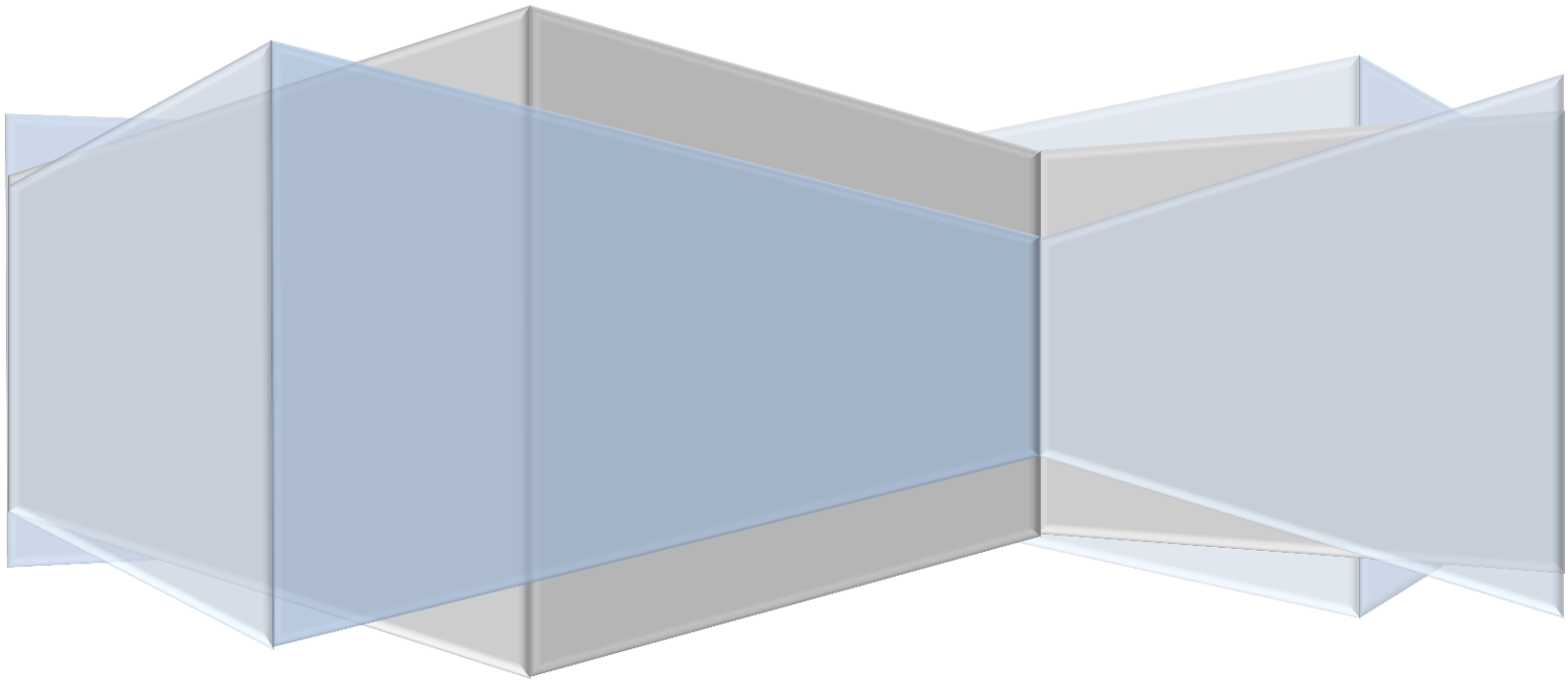
Technical University of Denmark

Two-View Stereovision

02503 – Advanced Image Analysis

s090763 - Olivier Jais-Nielsen



1. Introduction

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. As an illustration, some steps of a particular reconstruction are shown; however, for the sake of readability, all the figures might not come from the same instance of the reconstruction although none were artificially enhanced.

2. Calibration Matrices Estimation

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 ($\frac{W}{2}$, $\frac{H}{2}$).

Finally, we evaluate the calibration matrix of each camera according to the following formula:

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

The images can then be cropped to an interest area. The optical center coordinates are updated by subtracting the coordinates of the new top-left hand corner.

3. Interest points extractions and matches

Three options are implemented for the extraction of interest points.

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

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

Finally, it is also possible to point and click manually pairs of matches.

The reason for implementing three method is because a very large amount of accurate matches, well spread over the photos is required to achieve a good reconstruction and that some photos will give numerous SIFT features (natural complex scenes with complicated textures) while more simple scenes with uniform textures might not give enough SIFT features. In those case Harris corners might give better results. However, it should be noted that the matching of SIFT features, thanks to their descriptors is more reliable than matching points based on normalized cross correlation.

4. Robust matches and fundamental matrix estimation

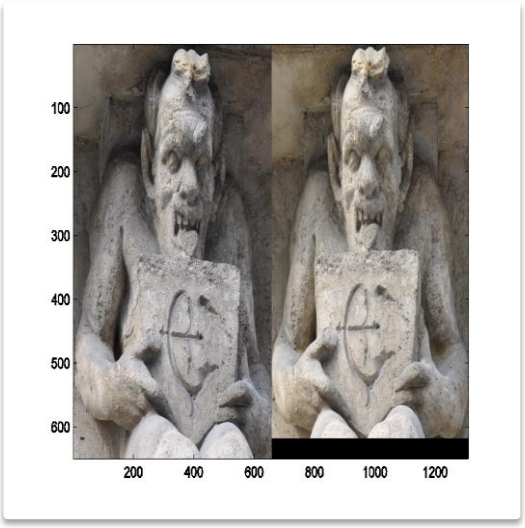
From the previous pair candidates, a selection is performed using a RANSAC algorithm to determine the fundamental matrix.

At each iteration, a fundamental matrix is evaluated using the eight-point algorithm one eight random candidate matches. The distance determining inliers is the Sampson distance: $d\big((P,Q),F\big) = \frac{(Q^T FP)^2}{\|FP\|^2+\|F^T Q\|^2}$ where (P,Q) are a pair of normalized points belonging to a candidate match.

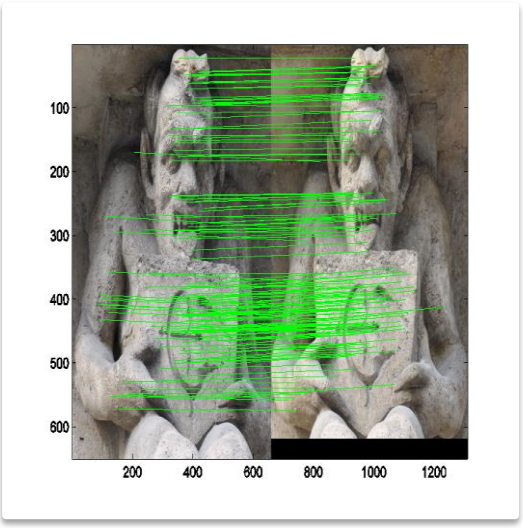
Another option is to constrain the RANSAC algorithm regarding an homography. In that case four matches are used at each iteration. The distance used is the following: $d\big((P,Q),H\big) = \|Q - HP\|^2 + \|P - H^{-1}Q\|^2$ where where (P,Q) are a pair of non-normalized points belonging to a candidate match.

In any case, the fundamental matrix is evaluated using the eight-point algorithm on the normalized matches considered as inliers.

Original cropped images



Some robust matches

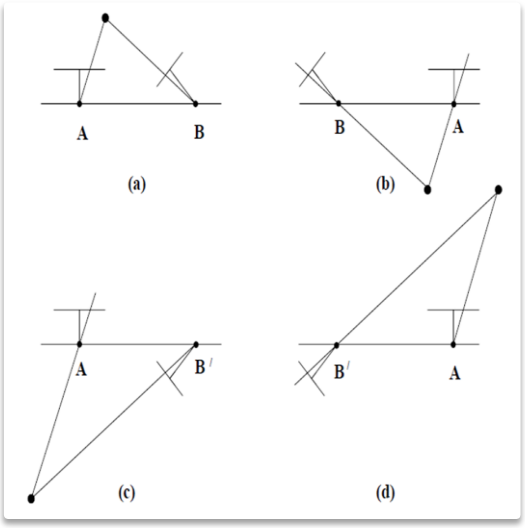


5. Camera matrices estimation

The camera projection matrices are evaluated as $P_1 = K_1[I|0]$ and $P_2 = K_2[R|t]$, where R and t are the rotation and translation defining both camera relative positions. We thus consider that the first camera's coordinates are the world coordinates.

R and t are deduced from the singular value decomposition of the essential matrix: $E = K_2^T F K_1$ However this gives four possible solutions.

Four solutions ambiguity



Only one solution has its object in front of both cameras; this solution is the correct one. To find it, the robust pairs are triangulated in both camera systems; the sign of the depths of the resulting three-dimensional points is checked, for each possibility. Only for the actual solution, the depths are positive in both systems.

Although the matches are supposed robust, to absorb a potential incorrect match, the average of the signs of the depths is considered.

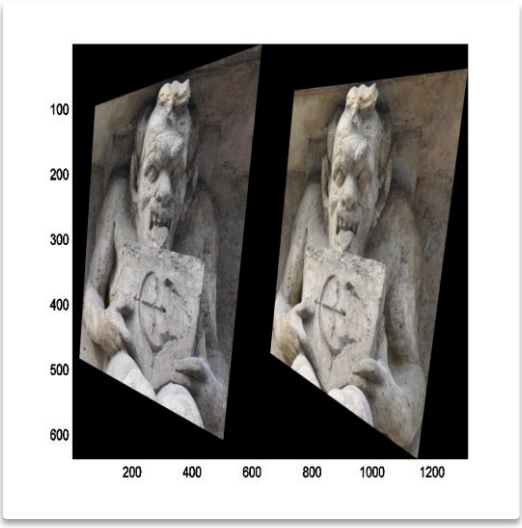
6. Rectification

Both images are rectified, that is we define two projective transformations such that the resulting images correspond to the following camera matrices: ${P_1}' = {K_1}'[R'|0]$ and ${P_2}' = {K_2}'[R'|t']$

R' is defined as the rotation matrix transforming the first camera coordinates (i.e. the world coordinates) into a direct orthonormal coordinate system which first direction is defined by both original camera centers. We deduce t' such that the new rectified cameras have the same centers as the original ones.

${K_1}'$ and ${K_2}'$ are initially defined arbitrarily as the arithmetic mean of K_1 and K_2. However, to reduce the out-of-bound area of the rectified images, the displacement for both optical centers is measured and ${K_1}'$ and ${K_2}'$ are redefined to compensate this displacement.

Rectified images



7. Rectification error estimation and bounding geometry determination

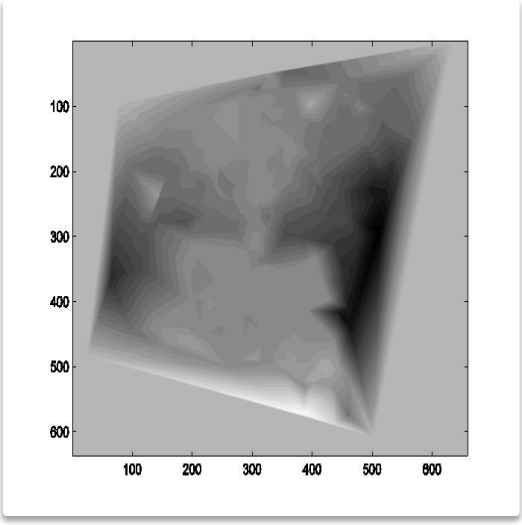
The bounding geometry is determines by applying the previously defined transformations to the original images corners coordinates. This allows skipping the out-of-bound area from any computations after this step.

It is also possible to manually reduce the bounding geometry of the first rectified image to limit the reconstruction to a specific area.

Robust matches are extracted from the rectified images. At one point of the first rectified image, the rectification error is the vertical distance between this point and its matching point. To estimate this rectification error on the entire bounding geometry, a linear interpolation is used with a zero-value error set on the bounding geometry.

Additionally, these matches are used to estimate the mean disparity, that is the mean horizontal distance between the points of each pair.

Rectification error estimation



8. Disparity estimation

Matching points are searched along the same horizontal line on both rectified image.

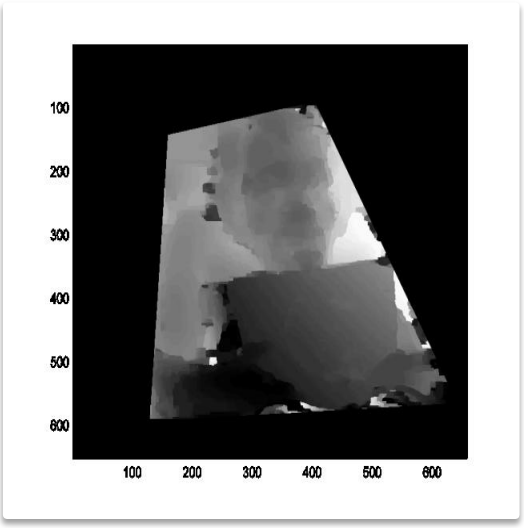
For each considered disparity, a new bounding geometry is defined for the first rectified image: it is composed of each point of the initial bounding geometry that, horizontally shifted by the considered disparity lands in the second image bounding geometry. The union of these bounding geometries is also kept. It is referred to as the interest region.

For each disparity d, for all points in the first image bounding geometry associated to d, we compute a cost: $V_1(P,d) = -log\left[c\left(S_1(P), S_2\left(P + \begin{pmatrix} d \\ e(P) \end{pmatrix}\right)\right)\right]$ where $S_i(P)$ is a square patch of the rectified image i centered on point P and $c(.,.)$ is the normalized cross correlation and $e(P)$ is the rectification error at the point P. The points outside of the bounding geometry are given the mean of the previously determined costs.

A 2-clique cost is defined as: $V_2(d_1,d_2) = -\beta exp(-\frac{|d_1-d_2|}{100})$ where d_1 and d_2 are two disparities considered for the two sites of a clique and β is a homogeneity parameter.

This MRF is solved with graph-cuts and alpha-expansion on the domain defined by the interest region. Each graph-cut is performed with Vladimir Kolmogorov's module (3).

Disparity map



9. Surface reconstruction

Each point p_1 of the first rectified image lying in the interest region is triangulated into the three-dimensional point P by solving the following this linear system:

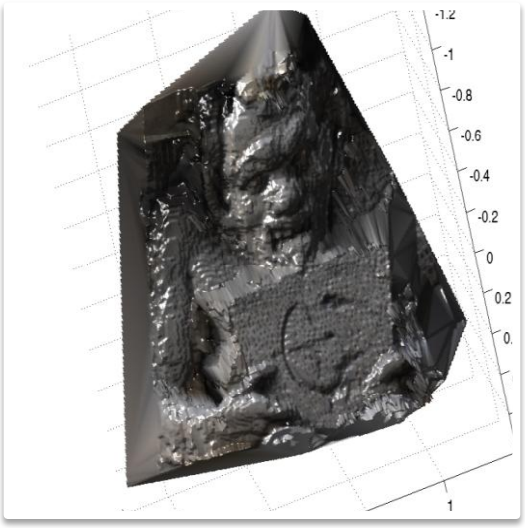
$$\begin{cases} p_1 = P_1 P \\ p_1 + \begin{pmatrix} d(p_1) \\ 0 \end{pmatrix} = P_2 P \end{cases} \Rightarrow \begin{cases} p_1 \wedge P_1 P = 0 \\ p_1 + \begin{pmatrix} d(p_1) \\ 0 \end{pmatrix} \wedge P_2 P = 0 \end{cases}$$ where $d(p_1)$ is the disparity associated to p_1.

This results in a three-dimensional point cloud where additionally, each point is associated to a color (i.e. the color of p_1 in the first rectified image). This can also be considered as a two-dimensional (the first two coordinates) point cloud where each point is associated to a color and a depth (the third coordinate).

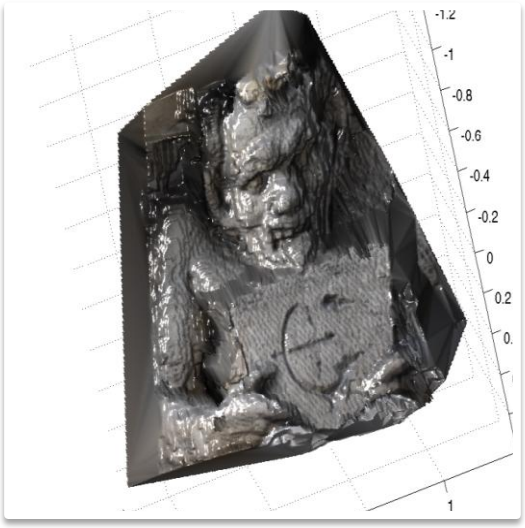
A linear interpolation of the depth and the color over the domain defined by this two-dimensional point cloud gives the surface.

The reason for not interpolating over the three-dimensional point cloud is that it requires tremendous computational time.

Reconstructed surface



Reconstructed surface with a different lighting



10. Note on null-space vector computation

The eight-point algorithm, the homography estimation and the triangulation require at some point to determine a null-space vector of a rectangle matrix. This is performed using the singular value decomposition and taking the column vector corresponding to the least singular value. Although it is mathematically equivalent to use eigen-value decomposition on the square matrix formed of the rectangle matrix multiplied by its transpose, it seems to give much bigger errors and thus is avoided here.

11. Conclusion

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