

Data-Optimal Rectification for Fast and Accurate Stereovision*

Martin Matoušek, Radim Šára and Václav Hlaváč
Center for Machine Perception, Czech Technical University,
Prague, Czech Republic
<http://cmp.felk.cvut.cz>

Abstract

In this paper we propose rectification procedure for binocular stereoscopic vision that minimizes the loss of local image neighbourhood discriminability in rectified images. The optimality of the rectification is thus influenced by image contents. Such rectification helps seek for precise dense correspondences.

1. Introduction

In computational stereopsis, rectification has been introduced as image pre-processing the goal of which is to make the subsequent task of dense matching easier. Given a set of images with general epipolar geometry, the task is to find a set of homographies, one per image, which transform the images so that the pairwise corresponding epipolar lines become parallel to image rows. In the case of two cameras, the two homographies have 6 degrees of freedom (DOF): common *vertical scale* and *slant*, individual *horizontal scale*, and *skew* respectively, and three translations that are of little interest in rectification.

Any homography from this 6DOF class performs the rectification task. It has been noticed early [1] that not all of them are equally suited because some subject the image to unacceptable distortion. The problem arises to choose the free parameters that minimize the unwanted distortion. The goal of previously reported methods [3, 4, 5, 6, 7] was to minimize image distortion or to maximize the computational efficiency of the rectification procedure. The best results are based on a functional minimisation of local image grid distortion [3].

Since the purpose of image rectification is efficient matching, the problem cannot be limited to just minimiz-

ing image distortion: *In our view, image rectification serves as the optimal image pre-processing for dense matching.* The rectification therefore must take into account the properties of the matching task. Matching uses images to calculate image similarity statistics. Local features computed from pixel neighborhood are used: they may be as simple as pixel values or as complex as a set of affine invariants constructed from the response set of a filter bank. There are two essential properties associated with the image similarity statistics:

Discriminability. It is the probability that a correct match is recognized as such based on the local image features alone. The more discriminable the features are the more likely it is they provide a unique label for a particular image location. Discriminability of image pixel neighbourhood increases with de-correlation of the image values over the neighbourhood.

Localization. Image similarity, when expressed as a function in disparity space, should have sharp maxima around the correct matches. The curvature of this function at the maximum is directly proportional to accuracy ε of the 3D reconstruction.

Consider an image in which all spatial frequencies are low. Dense matching at full resolution is a waste of computational resources: if image size was reduced all the way to the Nyquist limit then the localization would not change and the discriminability of local image features would increase because of partial de-correlation of image values in the neighbourhood. As a result, dense matching will run faster (because images are smaller, and the features are more discriminable) while preserving the accuracy of the 3D structure reconstructed from the images.

The goal of this paper is to propose a functional which *uses the image content* to determine optimal values of the free rectification parameters, *including the overall image scale*. In Sec. 2, we define local frequency loss functional which takes into account image content and describe the corresponding optimization task. We discuss the behaviour of the proposed rectification procedure and show experimental results on both simulated and real data.

* The 1st author was supported by the Czech Ministry of Education under project LN00B096, the 2rd author was supported by the Austrian project CONEX GZ 45.535 and by the Czech Academy of Sciences under project IET101210406, the 3rd author was supported by the project GACR 102/03/0440 and by CONEX GZ 45.535 and by the European Commission under project IST-004176.

2. Measuring Rectification Quality

Both localization and discriminability favor high image frequencies. We therefore pose the rectification problem as selecting the free rectification parameters \mathbf{p} such that a frequency loss functional $C(\mathbf{H}(\mathbf{p}) | f)$ is minimized. Given image instance f , this functional essentially integrates local frequency content over the image domain assuming $\mathbf{H}(\mathbf{p})$ represents the rectification homographies parameterized by the vector of free parameters \mathbf{p} . The elements of \mathbf{p} represent the individual geometric modes of the homography, i.e., the scales, skew and slant. We are looking for

$$\mathbf{p}^* = \underset{\mathbf{p}}{\operatorname{argmin}} C(\mathbf{H}(\mathbf{p}) | f). \quad (1)$$

In Subsec. 2.3 we show that, in addition to frequency loss (denoted C_F later) this functional C must include an image size penalty term (denoted A).

2.1. Local Frequency Loss Functional

The functional C captures the influence of homography \mathbf{H} on the image frequency loss. We will express it in Fourier domain. We propose to do it locally, because homography can be approximated by a linear mapping. We can then integrate the local contributions over the image domain. Let $\mathbf{A} = J(\mathbf{H})$ be the Jacobian matrix approximating homography \mathbf{H} at a given pixel (u_0, v_0) :

$$\mathbf{A}(u_0, v_0) = \begin{bmatrix} \frac{\partial u'}{\partial u}(u_0, v_0) & \frac{\partial u'}{\partial v}(u_0, v_0) \\ \frac{\partial v'}{\partial u}(u_0, v_0) & \frac{\partial v'}{\partial v}(u_0, v_0) \end{bmatrix}, \quad (2)$$

where $[u' \ v' \ 1]^T \simeq \mathbf{H} [u \ v \ 1]^T$. We will integrate amplitudes of Fourier transform of fixed-size local image windows centered at (u_0, v_0) as the elementary contributions to the loss functional C . For ease of derivation we will assume that the image function is continuous.

First we need to express the influence of geometric transformation \mathbf{A} of the image domain on the spectrum of the image in the Fourier domain. It is easy to show that the spectrum domain will transform by \mathbf{A}^\top [2]:

$$\mathcal{F}\{f(\mathbf{A}^{-1}\mathbf{x})\}(\mathbf{u}) = |\det(\mathbf{A})| \mathcal{F}\{f(\mathbf{x})\}(\mathbf{A}^\top \mathbf{u}), \quad (3)$$

where \mathcal{F} is the Fourier transform operator and \mathbf{u} is coordinate pair in the Fourier domain (Fig. 1a).

The application of the mapping \mathbf{A} may reduce the extent of the original image spectrum, as follows. The image re-sampling associated with the forward transformation by \mathbf{A} of image domain (cyan solid arrow in Fig. 1a) involves (1) original image reconstruction to continuous domain, (2) geometric transformation of the domain, followed by (3) discrete sampling. The sampling restricts the transformed image spectrum to the Nyquist interval (red

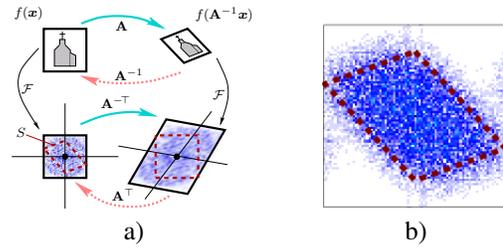


Figure 1. (a) The effect of linear mapping \mathbf{A} in the image domain (top) and in the frequency domain (bottom). (b) The restriction of the spectrum of image f (i.i.d. random noise).

dashed square, bottom right in Fig. 1a). To compute the frequency loss functional, this restriction needs to be penalized in the Fourier domain of the *original* image $f(\mathbf{x})$. Hence, we use the backward mapping \mathbf{A}^{-1} to transform the image $f(\mathbf{A}^{-1}\mathbf{x})$ back (this again consists of interpolation, transformation, sampling; pink dashed arrow in Fig. 1a). This time there is no loss. Clearly, the region S containing the frequencies after the backward mapping (bottom left in Fig. 1a) is given by the mapping by \mathbf{A}^\top of the Nyquist interval of the transformed image $f(\mathbf{A}^{-1}\mathbf{x})$. The region S induced by the local linear approximation of homography \mathbf{H} at position (u, v) will be denoted as $S_{(\mathbf{H}, u, v)}$. We define the loss of frequencies as the integral of the Fourier amplitude spectrum over the complement of the intersection of the region S and the Nyquist interval of the original image. One therefore does not need to compute the Fourier transform of the transformed image.

Hence, the integral over the “lost frequencies” is

$$L(u, v | \mathbf{H}, f) = \iint_{S_{(\mathbf{H}, u, v)}} h(\mathcal{F}\{w_{(u, v)}(f)\}(i, j)) di dj, \quad (4)$$

where w is the local window of image function f around position (u, v) , \mathcal{F} is Fourier transform and h is a kernel that weights the different elementary contributions. In this paper, we have chosen $h(x) = |x|$. The local penalty behaviour for two chosen homographies is illustrated on Fig. 2.

The global influence of the homography \mathbf{H} on the quality of the rectified image is the local penalty integrated over the input image domain \mathcal{I}

$$C_F(\mathbf{H} | f) = \iint_{\mathcal{I}} L(u, v | \mathbf{H}, f) du dv. \quad (5)$$

2.2. Computational Complexity Reduction

The most demanding computational operation in solving the problem (1) is calculating the integral (4), since it must be evaluated (in every pixel) for every homography \mathbf{H} . The local Fourier transforms can be pre-computed.

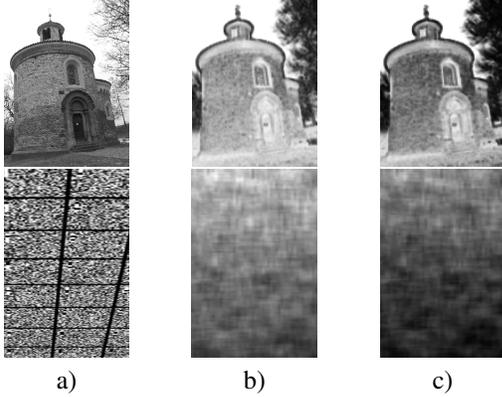


Figure 2. Local penalty values (darker is higher). (a) The original images. (b) Local penalty for horizontal scale change of 0.8. (c) The same plus horizontal skew of 0.5 (c).

Suppose the Jacobian matrix is constant over an extended image region \mathcal{R} . Then the $\overline{\mathcal{S}_{(H,u,v)}}$ is constant as well, and we can denote it as $\overline{\mathcal{S}_{(H,\mathcal{R})}}$. The contribution of \mathcal{R} to the global penalty $C_F(\mathbf{H} | \mathcal{I})$ will be

$$\begin{aligned}
 C_F(\mathbf{H} | \mathcal{R}) &= \iint_{\mathcal{R}} \iint_{\overline{\mathcal{S}_{(H,\mathcal{R})}}} h(\dots) dudvdidj = \\
 &= \iint_{\overline{\mathcal{S}_{(H,\mathcal{R})}}} \underbrace{\iint_{\mathcal{R}} h(\mathcal{F}\{w_{(u,v)}(f)\}(i,j))}_{F(i,j|\mathcal{R},\mathcal{I})} dudv didj.
 \end{aligned} \tag{6}$$

The inner integral $F(i,j | \mathcal{R}, \mathcal{I})$ can now be pre-computed for every region, since it is independent of \mathbf{H} . Hence, the global penalty can be integrated over regions whose number is significantly smaller than the number of pixels,

$$C_F(\mathbf{H} | f) = \sum_{\mathcal{R}_k} C_F(\mathbf{H} | \mathcal{R}_k), \tag{7}$$

where $\bigcup_k \mathcal{R}_k = \mathcal{I}$. In our experiments, we divided the image to 30×40 regions assuming that mild violation of constancy over the regions will lead to small error. The number of regions in the image is a parameter of the rectification procedure. The analysis of the influence of this simplification on the accuracy of the result is left for future work.

2.3. Frequency Loss Functional Versus Image Size

As illustrated in Fig. 3a the penalty C_F is monotonic with respect to scale change. This is already obvious from definition (5). Image shrinking results in equal shrinking of the region \mathcal{S} in the spectrum (Fig. 1), since $\mathbf{A}^\top = \mathbf{A}$ in this case. Hence, we always *lose* some frequencies (more precisely, we never get back any lost frequencies). Under changing the other parameters, the penalty need not be

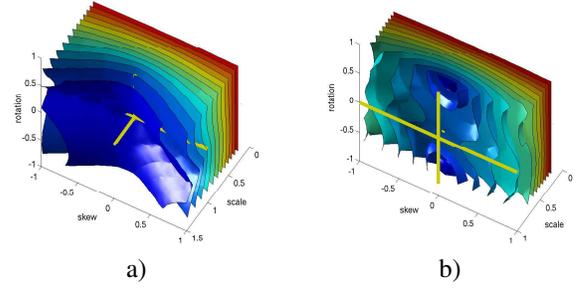


Figure 3. Penalty function isosurfaces for a synthetic image of a plane. The homography is parameterized by isotropic scale, skew and rotation. The coordinate origin is at identity ($\mathbf{H} = \mathbf{E}$), as marked by the thick cross. (a) The frequency loss penalty only, (b) the frequency loss penalty combined with the image size counter balance. The isosurfaces have quadratic spacing.

monotonic any more. In the practical case, when image frequencies have no dominant direction, the penalty is often monotonic in other parameters as well.

Clearly, the inherent monotonicity in scale must be counter-balanced to prevent unlimited growth of rectified image due to the tendency to achieve zero loss in all parts of the image. We have therefore chosen a counter-balance based on image size (the portion that contains rectified image data), Fig. 3b. This is motivated by a tradeoff between dense matching accuracy and speed.

2.4. The Optimization Problem

The full optimization problem is non-convex. Simple minimization can get trapped in a local minimum. We use a simple gradient descent optimization initialized by homography \mathbf{H}_0 obtained by Gluckman’s method [3]. Obviously, this method does not guarantee that the global optimal solution will be found. The problem then is to find the best update $\mathbf{H}(\mathbf{p})$ with \mathbf{H}_0 with $\mathbf{H}(\mathbf{p})$ such that the resulting homography stays in the class of valid rectification homographies.

The optimization problem associated with image pair rectification is as follows. Let $\mathbf{H}_l(\mathbf{p})$, $\mathbf{H}_r(\mathbf{p})$ be the homographies applied on the left image f_l and right image f_r , where \mathbf{p} are the common parameters of the rectification homography class. Let $A(\mathbf{H} | \mathcal{I})$ be the rectified image area. We are looking for

$$\mathbf{p}^* = \underset{\mathbf{p}}{\operatorname{argmin}} \left\{ C_F(\mathbf{H}_r(\mathbf{p}) | f_r) + C_F(\mathbf{H}_l(\mathbf{p}) | f_l) + \lambda \left(A(\mathbf{H}_r(\mathbf{p}) | f_r) + A(\mathbf{H}_l(\mathbf{p}) | f_l) \right) \right\}, \tag{8}$$

where λ is a parameter balancing both components.

3. Experimental Results

We performed experiments on synthetic data. A slanted plane was viewed by a pair of cameras in canonical configuration (i.e., the images were already rectified, although not necessarily in the optimal way with respect to the image). The plane was textured by random texture (i.i.d. noise scaled into 0–255). The images were generated from the known plane-camera homographies. Gaussian noise of standard deviation $\sigma = 2$ was added independently in each image, see Fig. 4a,b. Camera parameters were chosen so that the texture in the resulting images was super-sampled by approximately the factor of two. Since the images were already rectified, we started the search at identity $\mathbf{H}_0 = \mathbf{E}$. Note that any method based on image grid optimization [3, 4, 5, 6] would result in no improvement over \mathbf{H}_0 .

The results shown in Fig. 4c are insensitive to λ in a wide range. Note that the slant is mostly eliminated and the size is reduced approximately to 1/2, essentially removing the oversampling.

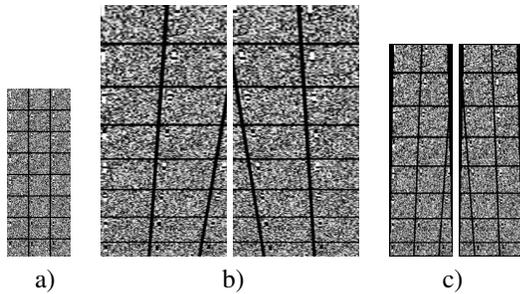


Figure 4. Optimal rectification. (a) Synthetic texture 100×266 pixels, (b) synthetic images of slanted plane 200×400 pixels, (c) rectification results 104×350 pixels. Grid superimposed for clarity; it was not present in data.

Next, we run dense matching algorithm [9] followed by sub-pixel disparity estimation [8] on the optimally rectified images, Fig. 5. This experiment illustrates the influence of the optimal rectification on image neighbourhood discriminability. More contents is squeezed in the local image window by the size reduction. This is especially noteworthy in the upper part of the image where the original image frequencies are relatively slow, hence image window discriminability is insufficient to find confidently stable matching. Many holes result in that region in the non-optimal images.

4. Conclusion

We have introduced optimal image rectification that takes into account the image contents. We defined the

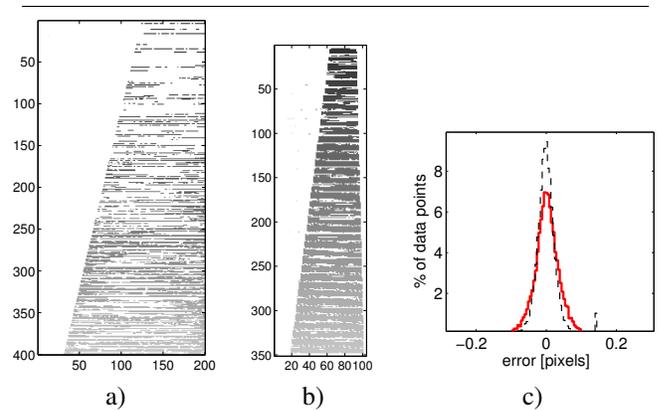


Figure 5. Dense matching experiment. (a) Disparity map computed from the original images, density 34%. (b) Disparity map from the optimally rectified images, density 80%, rectified area reduction 38%. (c) Ground-truth error in disparity, dashed line corresponds to (a), solid red corresponds to (b).

frequency-loss functional which penalize the loss of power spectrum components after geometric transformation of image. Minimization of this functional together with image size penalty term is then used for choosing the free degrees of freedom of rectifying homographies optimally.

Experiments show that the method is able to discover the correct rectified image size and compensate other components of the distortion well. Simulated experiment results show a significant increase in disparity map quality.

References

- [1] N.Ayache, C.Hansen. Rectification of images for binocular and trinocular stereovision. In *Proc. ICPR*, 1:11–16, 1988.
- [2] R.N.Bracewell et al. Affine theorem for two-dimensional Fourier transform. *Electronics Letters*, 29:304, Feb. 1993.
- [3] J.Gluckman, S.K.Nayar. Rectifying transformations that minimize resampling effects. In *Proc. CVPR*, 1:111–117, 2001.
- [4] R.I.Hartley. Theory and practice of projective rectification. *Intl. J. of Computer Vision*, 35(2):115–127, Nov. 1999.
- [5] C.Loop, Z.Zhang. Computing rectifying homographies for stereo vision. In *Proc. CVPR*, 1:125–131, 1999.
- [6] M.Pollefeys et al. A simple and efficient rectification method for general motion. In *Proc. ICCV*, 1:496–501, 1999.
- [7] S.Roy et al. Cylindrical rectification to minimize epipolar distortion. In *Proc. CVPR*, 393–399, 1997.
- [8] R.Šára. Accurate natural surface reconstruction from polynocular stereo. In *Proc. NATO Advanced Res. Wkshp Confluence of Comp. Vision and Comp. Graphics*, NATO Science Series 3, no.84:69–86, 2000.
- [9] R.Šára. Finding the largest unambiguous component of stereo matching. In *Proc. ECCV*, 3:900–914, 2002.