Shift-map Image Registration

Svärm, Linus; Strandmark, Petter

2010

Shift-map Image Registration

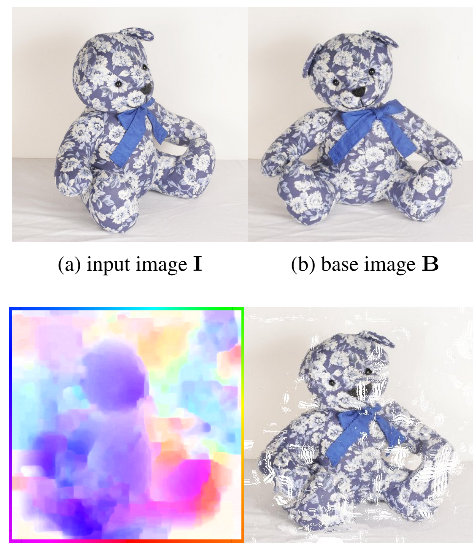
Linus Svärm Petter Strandmark
Centre for Mathematical Sciences, Lund University
{*linus,petter*}*@maths.lth.se*

Abstract

Shift-map image processing is a new framework based on energy minimization over a large space of labels. The optimization utilizes α-expansion moves and iterative refinement over a Gaussian pyramid. In this paper we extend the range of applications to image registration. To do this, new data and smoothness terms have to be constructed. We note a great improvement when we measure pixel similarities with the dense DAISY *descriptor. The main contributions of this paper are:*

- *The extension of the shift-map framework to include image registration. We register images for which* SIFT *only provides 3 correct matches.*
- *A publicly available implementation of shift-map image processing (e.g. inpainting, registration).*

We conclude by comparing shift-map registration to a recent method for optical flow with favorable results.



(a) input image $\mathbf{I}$ (b) base image $\mathbf{B}$

(c) shift-map (d) final location of pixels

Figure 1. Registration of two images using a shift-map. Each pixel in the input image is placed on the base image as described by the shift-map.

1 Introduction to shift-maps

Shift-map image processing has recently [8] been introduced and applied to image inpainting, content aware resizing, texture synthesis and image rearrangement. This paper will extend the range of applications to image registration. The motivation behind this work is the fact that we already were in possession of images we found to be impossible to register using standard methods (e.g. SIFT correspondences). The results we obtained with shift-map registration seem promising.

1.1 Problem formulation

Registration can be performed using a parametric model, e.g. an affine or a projective transformation estimated from point correspondences between the two images. In this paper, we consider a non-parametric model. We have a base image $\mathbf{B}(i,j)$ and an input image $\mathbf{I}(i,j)$. These two imaged need not have the same size. The goal is to register the pixels of the input image onto the base image using a shift-map $\mathbf{T}(i,j) = \Big(t_i(i,j)\ ,\ t_j(i,j)\Big)$. The pixel $\mathbf{I}(i,j)$ is registered onto $\mathbf{B}\Big(i + t_i(i,j)\ ,\ j + t_j(i,j)\Big)$. Figure 1 shows the input and base images and the resulting image obtained by moving all pixels in the input image as specified by the computed shift-map.

Each possible shift-map is assigned an energy, based on *a priori* assumptions on what a good shift-map typically looks like and how well the two images match each other. The goal is then to find the optimal shift-

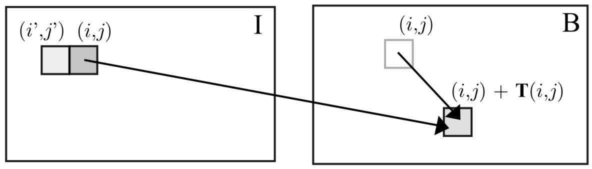


Figure 2. Shift-map between two images

map, that is, the shift-map with the lowest energy:

$$E(\mathbf{T}) = \alpha \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} E_d^{ij}(\mathbf{T}(i,j)) + \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} \sum_{(i',j')} E_s^{ij}(\mathbf{T}(i,j), \mathbf{T}(i',j')), \tag{1}$$

where the last summation refers to summations over all (i', j') in a neighborhood $\mathcal{N}(i,j)$ of (i,j). Figure 2 one such neighbor. We have used 4-connectivity of adjacent pixels throughout this paper. E_d^{ij} and E_s^{ij} are the data terms and smoothness terms, respectively. They will be described in separate sections below.

2 Registration energy terms

The methods in [8] deal with constructing a new image from an old one and the registration problem is about finding a map between two existing images. Hence the energy previously used for finding shift-maps is not suitable for registration and new energy terms must be constructed.

Comparison of pixels A related problem to image registration is dense depth estimation from two images of the same object with known camera positions. This problem has been studied extensively, see for example [5]. Recently a new descriptor, DAISY, was proposed by Tola et al. [9], tailored to dense stereo estimation where the position of the two cameras differ by a large amount. This descriptor is shown to outperform other approaches (e.g. SIFT, SURF and pixel differences) in extensive experiments. Therefore, it seems relevant to try and apply this descriptor to the related problem of estimating a dense image registration.

Not unlike SIFT [7], a DAISY descriptor samples the image derivative in different directions. Eight different directions and three different scales are used. By sampling these fields at different points around the feature location, a descriptor of dimensionality 200 is obtained. Since the same fields are used for all image locations, a dense field of descriptors can be computed in a couple of seconds. The main goal of the DAISY descriptor was efficient dense computation. In order to choose relevant parameters, we found [10] helpful.

Data terms The data terms E_d^{ij} were previously used in [8] to enforce hard constraints on the shift-map. When inpainting an image, the data term makes sure no pixels in the “hole” are used in the output image by assigning such shifts a cost of ∞.

In this paper, where image registration is considered, we need to develop more complex data terms to incorporate the fact that we want to find a mapping between two images such that similar pixels are mapped to similar pixels The data terms dictates that similar parts of the images should end up on top of each other. To measure similarity, dense DAISY is used.

It might only be possible to register parts of the input image, so shifting pixels outside the base image is permitted, at a constant cost P per pixel. The data terms are then given by

$$E_d^{ij}(\mathbf{T}) = \begin{cases} \left|\left|\hat{\mathbf{I}}(i,j) - \hat{\mathbf{B}}\Big((i,j) + \mathbf{T}(i,j)\Big)\right|\right|_2 \\ P \text{ when } (i,j) + \mathbf{T}(i,j) \text{ is outside } \mathbf{B}, \end{cases} \tag{2}$$

where $\hat{\mathbf{I}}(i,j)$ is the DAISY descriptor describing the image $\mathbf{I}$ at pixel location (i,j). If the shift takes pixel (i,j) outside the bounds of the base image, a constant cost is issued. Otherwise, dissimilarity of the pixels determines the cost of the assignment. Figure 6f shows a heat map of the distance from the circled feature in the first row to all locations in the image in row 2.

Smoothness terms The smoothness terms are used to enforce global consistency to the shift-map, while allowing discontinuities at a limited number of places. In [8], the smoothness terms compared the color and gradient pixel-wise. Where a discontinuity in the shift-map occurs, the penalty is computed as the difference in color and gradients.

Our smoothness function takes the form of the Euclidean distance between the endpoints of the two shifts:

$$E_s^{ij}(\mathbf{T}(i,j), \mathbf{T}(i',j')) = ||(i',j') + \mathbf{T}(i',j') - (i,j) - \mathbf{T}(i,j)||_2. \tag{3}$$

Here, (i,j) and (i',j') are neighboring pixels, see (1). Using the shift difference $||\mathbf{T}(i',j') - \mathbf{T}(i,j)||_2$ will penalize smoothly varying shift-maps too much, and hence it is important to compare the end points (as in (3)).

Color information The DAISY descriptor does not use color information, yet intuitively it makes little sense to match pixels of very different colors. Because of this, we have also made experiments where the color information of the images are incorporated into the above data terms. The color model used assigned a cost of P to pixels with large difference in hue, given that the intensity and saturation allowed a reliable value of the hue. This model improved the result of the registration in Fig. 6. We did not use color information in the experiment shown in Fig. 1.

3 Experiments

To minimize the energy (1), we used α-expansion as described by Boykov and Veksler [3] with the graph algorithms described in [2, 4]. Each possible shift value $\mathbf{T}(i,j) \in \{-m \ldots m\} \times \{-n \ldots n\}$ is mapped to a 1D label space. The number of labels for even moderately sized images then becomes very large. In order to make it tractable, a Gaussian pyramid was used. For the images in Fig. 6, an initial size of 128×23 was used. The size was doubled 3 times until the final resolution of 1024×179 was reached. Each doubling of the image size is followed by a linear interpolation of the shift-map. This shift-map was used as a starting guess for the optimization at the larger level. At each level after the first, only 9 possible shifts then need to be considered: $\{-1, 0, 1\}$ in each direction.

To verify our implementation, we inpainted an example image used in [8], see Fig. 5. We tried to follow their implementation as closely as possible and got different, but qualitatively similar results. We did not allow the pixels outside the area to be removed to move at all, which is in contrast to [8], where all pixels except the border of the image were allowed to be shifted.

Figures 1 and 3 show shift-map registration results. The bear image in Fig. 1 shows the same object from two different views and is from [6]. The building images in 3 register correctly, except for the light pole, which is very thin and does not have a large enough data term.

We have also conducted an experiment where we used shift-maps to recover a known image deformation. The results are displayed in Fig. 4.

During large-scale reconstruction of a city using images taken with a cylindrical camera [1], we have encountered many difficult image pairs where SIFT is unable to provide useful correspondences. The top two rows in Fig. 6 show one of the hardest. Computed SIFT features for the two images (794 and 1019 feature points, respectively) only yielded *3 correct matches*. The main reason for this was the image geometry and large, repetitive patterns. Using shift-map we obtained a dense, mostly correct map between the images. This was then used as an aid to compute SIFT correspondences. We then obtained 28 matches, *of which 12 were correct*. The runtime for this image was about 2 minutes.

We compared shift-map registration to the optical flow algorithm described in [11]. This algorithm did not produce useful results for the street images, see bottom of Fig. 6. Optical flow techniques are arguably not suitable for these kinds of registration tasks.

4 Conclusion and further work

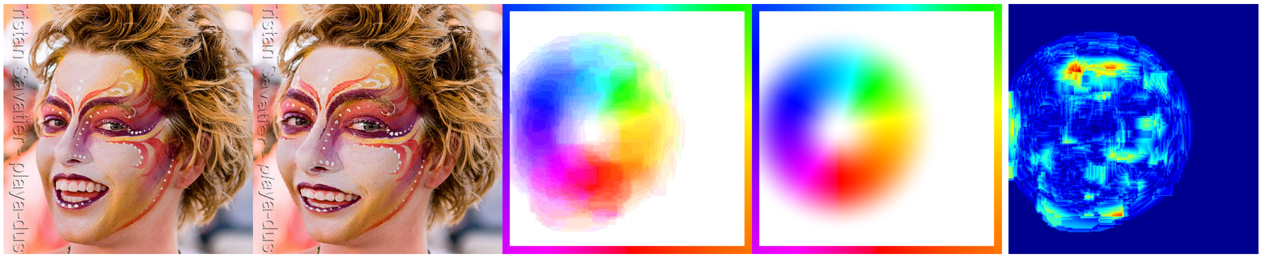
We have studied the application of shift-maps to image registration. Computing the smoothness term with color and gradient differences as in [8] did not give satisfactory results when extended to image registra-

(a) input image $\mathbf{I}$

(b) base image $\mathbf{B}$

(c) final location of pixels

Figure 3. Registration of two images of a building.



(a) base image (b) input image (c) Shift-map result (d) Ground truth (e) error

Figure 4. Recovering a known image distorsion. The maximum and mean error was 7.3 and 0.7 pixels, respectively. Photo by Tristan Savatier obtained through Flickr.

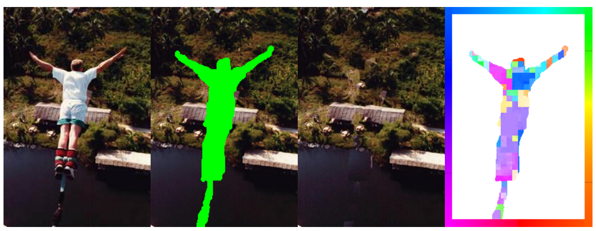


Figure 5. Our reimplementation of the algorithm in [8]. The complete running time for this example was 3.1415 seconds[2].

tion, but we found a great improvement with the dense DAISY descriptor. For relatively easy cases (Figs. 1 and 3), we obtained very good results. For very hard cases (Fig. 6) we obtained results which proved very useful for obtaining correspondences between the images. We compared shift-map registration to the optical flow algorithm described in [11] (Fig. 6e), which was significantly less accurate. One interesting future line of work would be to investigate whether shift-map inpainting can be improved by the DAISY descriptor as well. We have also not investigated large rotations in this paper, which would require additional considerations.

The source code for our shift-map implementation and for all experiments in this paper has been released to the public[1].

Acknowledgements

This work has been funded by the Swedish Foundation for Strategic Research (SSF) through the programme Wearable Visual Information Systems and by the European Research Council (GlobalVision grant no. 209480). The authors had several useful discussions with Klas Josephson and Olof Enqvist.

[1] http://www.maths.lth.se/~linus/shiftmap

[2] The fact that the running time in seconds closely approximates π is purely coincidental.

References

[1] Hitta.se streetview, C3 Technologies, http://www.hitta.se.

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, Sept. 2004.

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.

[4] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.

[5] V. Kolmogorov and R. Zabih. Graph cut algorithms for binocular stereo with occlusions. In *Handbook of Mathematical Models in Computer Vision*. Springer, 2006.

[6] A. Kushal and J. Ponce. Modeling 3D objects from stereo views and recognizing them in photographs. In *European Conf. Computer Vision*, 2006.

[7] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal Computer Vision*, 20:91–110, 2004.

[8] Y. Pritch, E. Kav-Venaki, and S. Peleg. Shift-map image editing. In *Int. Conf. Computer Vision*, 2009.

[9] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide baseline stereo. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2009.

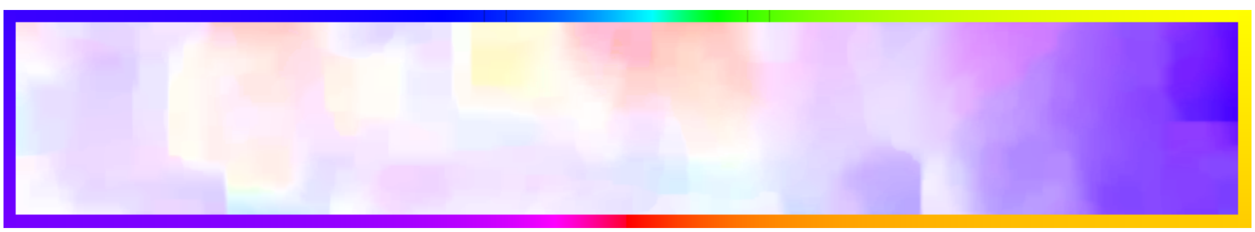
[10] S. Winder, G. Hua, and M. Brown. Picking the best DAISY. In *Conf. Computer Vision and Pattern Recognition*, pages 178–185. IEEE, 2009.

[11] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L1 optical flow. In *Pattern Recognition (Proc. DAGM)*, pages 214–223, 2007.

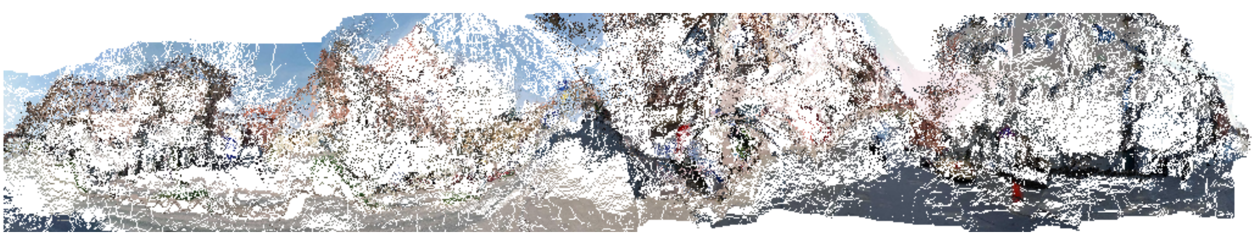
(a) input image $\mathbf{I}$

(b) base image $\mathbf{B}$

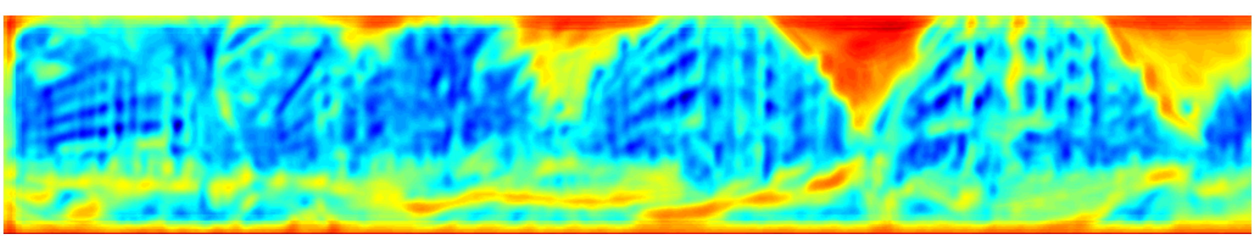
(c) final locations of the pixels in $\mathbf{I}$



(d) resulting shift-map



(e) result using the method in [11]



(f) DAISY distance between the circled feature in $\mathbf{I}$ to all pixel locations in $\mathbf{B}$

Figure 6. Registration of 1024×179 Hitta images. We note that we achieved a dense, highly non-linear registration. This shift-map allowed us to obtain useful point-correspondences between the images, which was not possible using SIFT alone.