# Robust Image-to-Image Color Transfer Using Optimal Inlier Maximization Supplementary material

Magnus Oskarsson
Centre for Mathematical Sciences, Lund University, Sweden
magnus.oskarsson@math.lth.se

## 1. Derivation of polynomial constraints

Given three corresponding intensity measurements $\{(v_1, v_1'), (v_2, v_2'), (v_3, v_3')\}$ we would like to find a third-degree polynomial $p_{abc}$ going through the origin so that

$$D(p_{abc}, (v_i, v_i')) = \epsilon_v, \quad i = 1, 2, 3. \tag{1}$$

The distance from a point $(v_1, v_1')$ is given by

$$D(p_{abc}, (v_i, v_i')) = \frac{|p_{abc}(v_i) - v_i'|}{\sqrt{1 + (\frac{dp_{abc}}{dx}(v_i))^2}}. \tag{2}$$

Combining these equations gives for each $i$

$$D(p_{abc}, (v_i, v_i')) = \epsilon_v \Leftrightarrow \tag{3}$$

$$\frac{|p_{abc}(v_i) - v_i'|}{\sqrt{1 + (\frac{dp_{abc}}{dx}(v_i))^2}} = \epsilon_v \Leftrightarrow \tag{4}$$

$$|p_{abc}(v_i) - v_i'| = \epsilon_v \sqrt{1 + (\frac{dp_{abc}}{dx}(v_i))^2} \Leftrightarrow \tag{5}$$

$$(p_{abc}(v_i) - v_i')^2 - \epsilon_v^2 (1 + (\frac{dp_{abc}}{dx}(v_i))^2) = 0. \tag{6}$$

Inserting the expression of $p_{abc}$ gives us our second degree polynomial in the unknown parameters $(a, b, c)$

$$g_i = (av_i^3 + bv_i^2 + cv_i - v_i')^2 - \epsilon_v^2 (1 + (3av_i^2 + 2bv_i + c)^2). \tag{7}$$

## 2. Transformation modelling

We have compared using different numbers of feature points, and different transformation models on the synthetic dataset [12]. We used a fixed number of 50 features and compared it to using all the extracted features as well as using all pixel values. We varied the transformation models as well. In this experiment we assume that there are no outliers and only want to test the fidelity of different models. We have tested four different transformations, the Monge-Kantorovitch linear model, a 3D affine model, a 2D projective model, and a second degree polynomial model in 3D. The least squares estimates of these transformations can all be estimated directly from the data. The results for various cases are shown in Table 1. We have run hypothesis tests on that the differences in mean are significant or not. On a 5% level, there is no significant difference on using 50, all features or all pixels for any of the tested transformations, for both PSNR and SSIM means. On the transformation level we get for the PSNR means that the projective model is significantly worse than the others, and that the polynomial model is significantly better. For the SSIM means, the only significant is that the projective is worse that the others. Significantly better results on a 5% level are marked with bold, and significantly worse results are marked with italics, in the table. We did not test the projective model for all pixels since the estimation then involves taking the SVD of a very large matrix. In our method we model and handle non-linearities in the intensity transformation, and hence we have chosen to use the affine model as the post-processing transformation on the inlier set, in order to combine model fidelity with robustness in a balanced way. The polynomial model is prone to give artifacts in the presence of also very little amounts of incorrect matches.

## 3. Benchmark tests

Distributions of PSNR and SSIM values for different inlier levels (and different methods) on the INTEL-TUT dataset [2] are shown in Fig. 1. More output from the benchmark tests can be seen in Fig. 2 and 3.

## 4. Background/foreground estimation

We will start to look at background/foreground segmentation. We assume that we have an estimate of a background image, and using this we want to detect parts in an input image, that differ from the background. In Fig. 4 an example is shown. Here we have a fixed camera viewing a traffic intersection, so the images are geometrically aligned. The objective is to detect the vehicles, and we have estimated a background image, by median filtering a number of frames. Due to lighting changes over time, the input image

Table 1. Synthetic model fidelity results using different number of correspondences and different transformation models.

| Transformation | *MKL* | | | *Affine* | | | *Projective* | | *Polynomial* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | 9 | | | 12 | | | 8 | | 30 | | |
| Points | 50 features | all features | all pixels | 50 features | all features | all pixels | 50 features | all features | 50 features | all features | all pixels |
| PSNR | 36.71 | 36.98 | 37.05 | 37.15 | 37.61 | 37.77 | *26.05* | *28.33* | **39.55** | **41.65** | **41.98** |
| SSIM | 0.968 | 0.969 | 0.969 | 0.971 | 0.971 | 0.972 | *0.875* | *0.926* | 0.971 | 0.977 | 0.978 |
| Time (ms) | 68 | 71 | 14 | 68 | 69 | 19 | 81 | 93 | 71 | 74 | 45 |

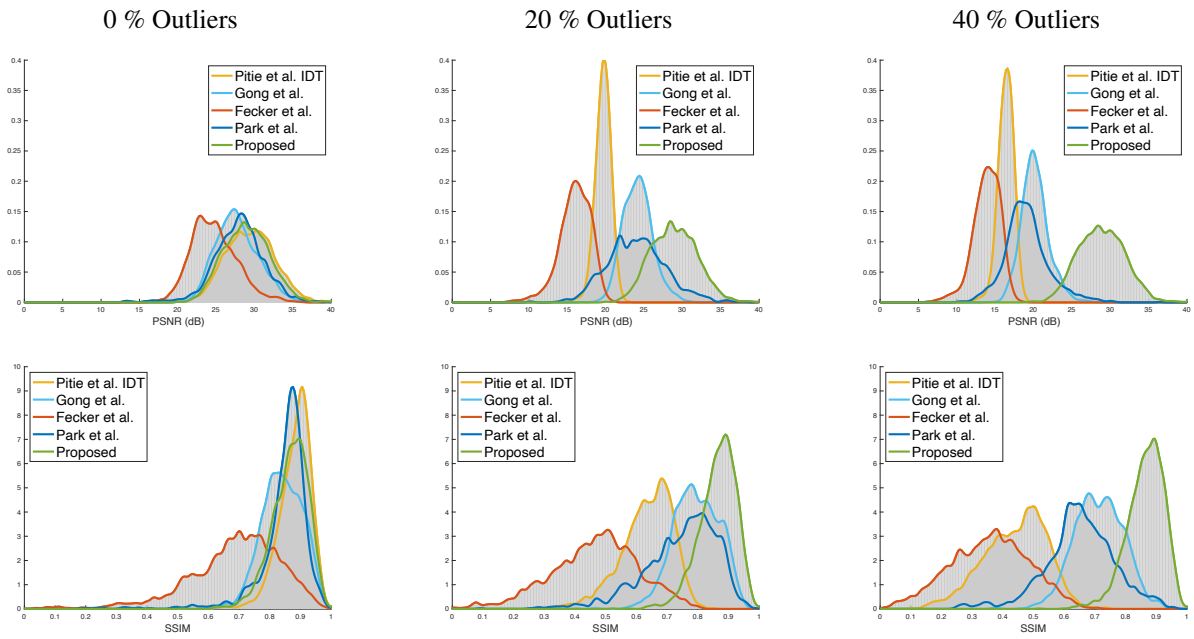


Figure 1. Distribution of PSNR (top) SSIM (bottom) values on the INTEL-TUT dataset [2], for varying outlier levels.

will not have the same color distribution as the estimated background image. We would like to transform the background image to the correct color space, but due to large extent of foreground objects we need a robust approach. We start with extracting feature points in the background image, and calculate the corresponding descriptors in the background and input image respectively. We then use our robust color matching to transform the background image. In Fig. 4 a number of input images are shown to the left. The next two columns show the transformed background image using [9] and the proposed method respectively. Fourth and fifth columns show the segmented output, based on the two difference images. The mask is constructed by thresholding the difference images, with an additional morphological opening with a structuring disk with a radius of three pixels.

## 5. Image stitching

The proposed method can be used for image stitching. Misalignments and moving objects make robust methods desirable. We will describe the approach using two images, but it readily generalizes to larger image sets. We use standard methods for the geometric alignment, *i.e.* extracting SIFT features, matching using RANSAC and fitting a homography. Mapping the image using the found homography we have the two images in the same geometric coordinate system. We then run our feature detector on the overlapping region, and extract our color descriptor from the two images. We run our robust color matching, and use the estimated transform on the whole image. This gives us our two images now also hopefully in the same color coordinate system. Often some form of blending function is used to avoid borders between images in the stitching, but here we don't use any blending. For overlapping pixels, we simply choose the pixel-wise max. The result on examples can be seen in Fig. 7. To the left is the stitched images without the estimated color transform, and to the right using the estimated color transform. One can clearly see that we in this case get a very seamless stitching result.

Figure 2. Some examples from the synthetic test. From left to right: Altered image, original image, proposed approach with 20% outliers, proposed approach with 60% outliers.

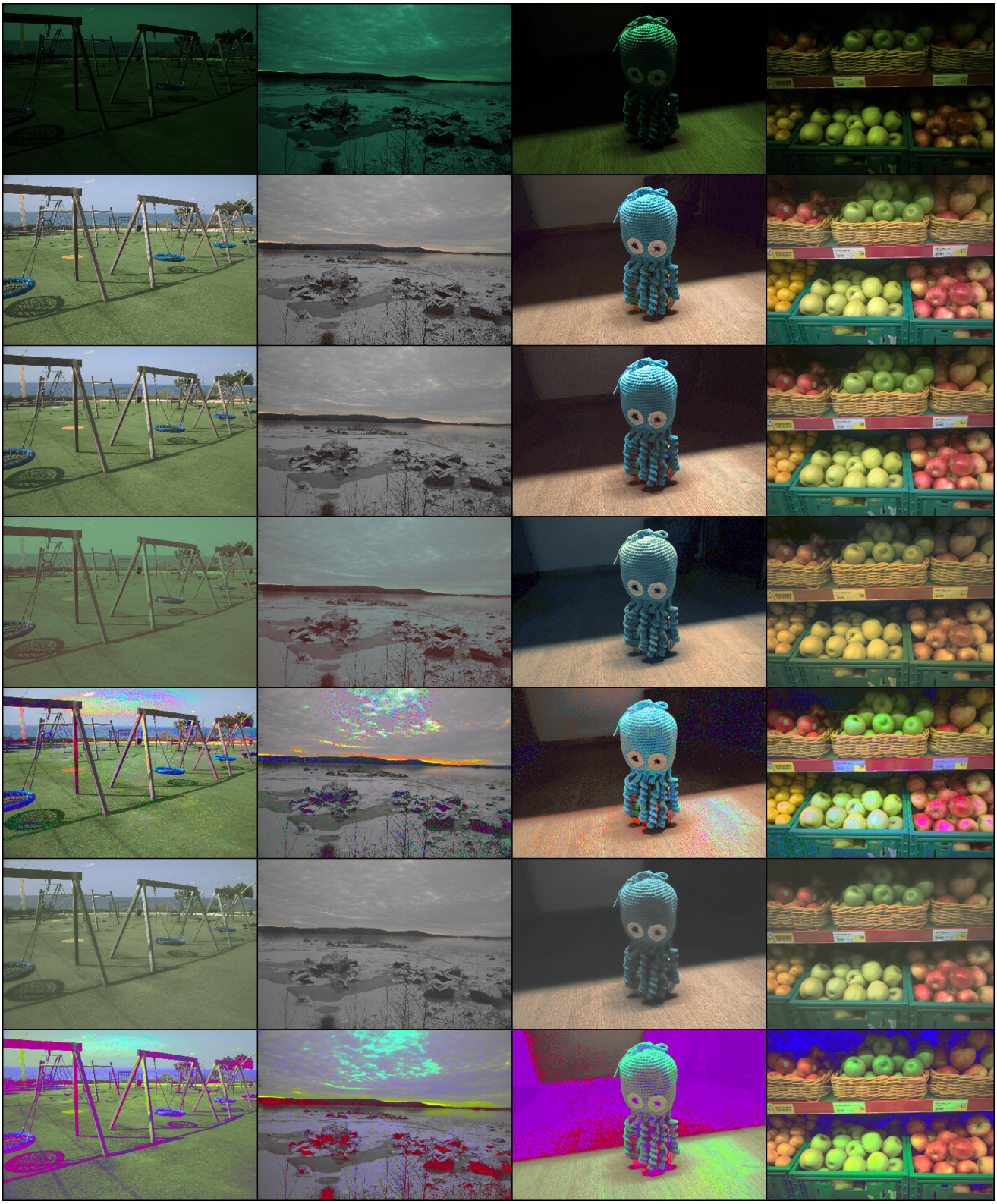

Figure 3. Example results on the INTEL-TUT dataset [2], with 20% added outliers. Top row shows the RAW input images, second row shows the target JPG images. Third row shows the output of our proposed method. Then follows the transformed images using the compared methods [8, 9, 5, 4].

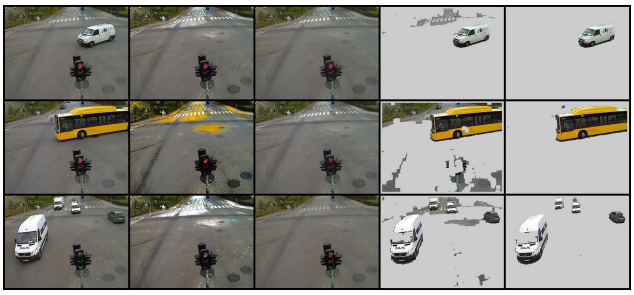

Figure 4. Example images for background foreground segmentation. Left: Input images. Second column: Transformed background image using [9]. Third column: The same using proposed method. Fourth and fifth columns: Segmented output based on the two methods. See also supplemental material.

## 6. HDR estimation from multiple exposure LDR brackets

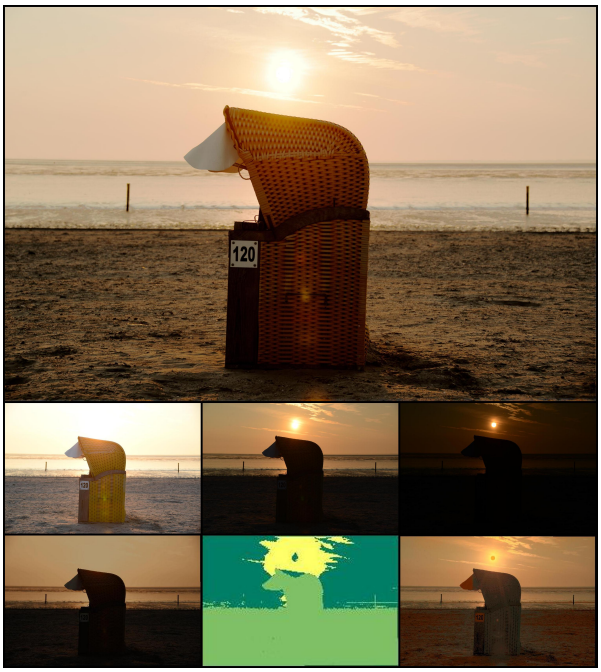


Figure 5. Middle shows the three input LDR input images [1]. Bottom shows to the left and right the the corresponding input images transformed to the color space of the middle image. Bottom middle shows the confidence measure choice for each pixel. Top shows the final tone mapped output after using the proposed method.

Our final example application is how to estimate a high dynamic range (HDR) image from several low dynamic range (LDR) images. A standard way of constructing HDR images in this way is to take several LDR images with varying exposure settings. One can then transform the LDR images into the same color coordinate system by mapping the different images to the same intensity range. After this step the different mapped LDR images are combined in some way to produce an HDR result, [11, 3, 6]. To view the HDR image on an LDR display or print it, some form of tone mapping algorithm is then used to transform the HDR image to an LDR image, see *e.g.* [10].

There are several substeps that can be, and that have been solved in a number of different ways, and each of the steps have their own difficulties that have to be addressed. We will now show how our robust color transformation method can be applied to the step of mapping the differently exposed LDR images to the same color coordinate system. In some cases, the exposure settings for the different images are known, and then this information can be used to transform the images. In many cases however the exposure settings are unknown, and even if it is known, for many cameras the exposure process is not linear. This motivates the use of an automatic process for finding the mappings. Typically, there is some motion between the images taken, and the process starts with image alignment. We will have the same problems with misalignment and moving objects as we have in image stitching applications, so this motivates the need for robust methods. In addition to this, we will in this case also need to handle severely over and under exposed parts of the images, that will give rise to outliers in our color matching. We choose one image as reference image (typically the middle-exposed image). We then map all other LDR images to the reference image using our proposed method. This gives all the images in the same color reference system. An example can be seen in Fig. 5 where the reference and mapped images are shown. In each input image we put a confidence value at each pixel depending how over- or undersaturated the pixel is. We then combine the transformed images, by choosing for each pixel the value from the most trusted transformed images based on the confidence measure. This gives an HDR image, which we then tone map in order to display it. We have used the global (non-parametric) method of [7]. Results can be seen in Fig. 6.

## References


[1] https://www.projects-software.com/hdr/sample-bracket-images. 5

[2] Caglar Aytekin, Jarno Nikkanen, and Moncef Gabbouj. Intel-tut dataset for camera invariant color constancy research. *arXiv preprint arXiv:1703.09778*, 2017. 1, 2, 4

[3] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, page 31. ACM, 2008. 5

[4] Ulrich Fecker, Marcus Barkowsky, and André Kaup. Histogram-based prefiltering for luminance and chrominance compensation of multiview video. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(9):1258–1267, 2008. 4

[5] Han Gong, Graham D Finlayson, Robert B Fisher, and Fufu Fang. 3d color homography model for photo-realistic color

Figure 6. Examples of tone mapped HDR output after using the proposed method.

transfer re-coding. *The Visual Computer*, 35(3):323–333, 2019. 4

[6] Michael D Grossberg and Shree K Nayar. High dynamic range from multiple images: Which exposures to combine.

Figure 7. The result of stitching two images, with or without the proposed color transformation. Left shows the result after taking the pointwise max at each pixel after homography registration of the two images. Right shows the same result but with the proposed color transformation on the second image. No additional blending was used.

In *Proc. ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, 2003. 5

[7] Magnus Oskarsson. Temporally consistent tone mapping of images and video using optimal k-means clustering. *Journal of mathematical imaging and vision*, 57(2):225–238, 2017. 5

[8] Jaesik Park, Yu-Wing Tai, Sudipta N Sinha, and In So Kweon. Efficient and robust color consistency for community photo collections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 430–438, 2016. 4

[9] Francois Pitié, Anil C Kokaram, and Rozenn Dahyot. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1-2):123–137, 2007. 2, 4, 5

[10] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010. 5

[11] Mark A Robertson, Sean Borman, and Robert L Stevenson. Dynamic range improvement through multiple exposures. In *IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 159–163, 1999. 5

[12] Wei Xu and Jane Mulligan. Performance evaluation of color correction approaches for automatic multi-view image and video stitching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 1