REFLECTION SEPARATION USING GUIDED ANNOTATION

Ofer Springer [*†], *Yair Weiss* [†]

School of Computer Science and Engineering
The Hebrew University of Jerusalem, Israel

ABSTRACT

Photographs taken through a glass surface often contain an approximately linear superposition of reflected and transmitted layers. Decomposing an image into these layers is generally an ill-posed task and the use of an additional image prior and user provided cues is presently necessary in order to obtain good results. Current annotation approaches rely on a strong sparsity assumption. For images with significant texture this assumption does not typically hold, thus rendering the annotation process unviable.

In this paper we show that using a Gaussian Mixture Model patch prior, the correct local decomposition can almost always be found as one of 100 likely modes of the posterior. Thus, the user need only choose one of these modes in a sparse set of patches and the decomposition may then be completed automatically. We demonstrate the performance of our method using synthesized and real reflection images.

Index Terms— Natural image statistics, reflection separation

1. INTRODUCTION

Many real-world photographs contain reflections and photographers often expend significant effort to avoid their effect. Ideally, we would like a fully automatic method that can remove reflections in post-processing of a single image. In the general single image reflection separation problem, assuming a linear response of the camera sensor then the input image $y = x_1 + x_2$ is a sum of two unknown reflection and transmission images. Automatically recovering x_1 and x_2 given y is a highly ill-posed task. This is due to the fact that the number of equations is generally half the number of unknowns.

The problem becomes less difficult when additional cues exist. Some existing approaches use the presence of a double image [1] or polarization [2, 3] in the reflection layer. Others use the availability of more than a single composite image due to camera motion [4, 5, 6, 7] or additional a priori known differences between the reflected and transmitted layers, e.g. a smoothness disparity [8] or a disparity in the relative layer attenuation factor [9]. However, in many situations, these extra cues are not available.

While fully automatic separation of reflection images from a single image is extremely difficult, photographers would also be happy with a *semi-automatic* method requiring user annotation. If the amount of annotation required is manageable then a method that completes the separation given this annotation may be of great utility. This was the motivation behind the work of Levin and Weiss [10] (LW) who presented a user-assisted method for separating reflections. Their method was based on the well-known property that derivative filters of natural images tend to have a sparse distribution. This property both served as an image prior and was the basis of their proposed annotation mechanism: the user was tasked with selecting pixels in which the (first and second order image derivative) filter responses fully originated from only one of the layers. They then optimized for a likely reflection separation that is also consistent with these annotations.

The LW method is the only currently available approach to solving the single image user-assisted reflection separation task that we know of. As shown in [10] it is often possible to get very good separations on real images with a modest amount of user input. However, for many images, we find that the annotation mechanism of the LW method is not applicable. Often a large fraction of the pixels in one layer contain non-sparse texture overlapping edges in the other layer. In such cases it is often impossible to locate a sufficient number of pixels having filter responses originating from only a single layer. Figure 1 shows one such example. The input image is a sum of two images, the first of which contains significant texture. Due to this texture, filter responses in both images are nonzero at edges of the second image, and thus the LW labeling mechanism fails. We show in figure 1 the results of running the LW method and our method on this input.

In section 2 We study the limitations of the sparsity assumption as an annotation mechanism. In section 3 we introduce an alternative annotation mechanism based on the GMM patch prior which overcomes the shortcomings of the LW annotation mechanism. The full algorithm utilizing both the revised prior and revised annotation mechanism (GMM-C) is described in section 4. In section 5 we asses the accuracy of our proposed method on reflection images synthesized from

[*]springer@cs.huji.ac.il
[†]Supported by Intel ICRI-CI and the ISF.

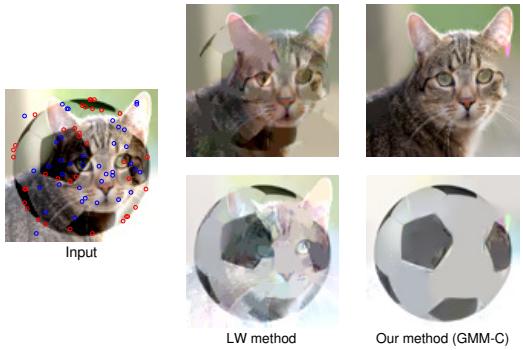

Fig. 1: Visual comparison of the results of running the sparsity based LW method and our method on a synthetic reflection image. Circles denote the locations at which user annotation was provided. For the LW method, circle colors (blue or red) indicate where the local filter response has originated from (layer 1 or 2) according to the user provided annotation.

the BSDS300 dataset [11] and show separation results on real reflection images.

2. LIMITATIONS OF THE SPARSITY ASSUMPTION

Although the outputs of derivative filters applied to natural images tend to have a sparse distribution overall (c.f. [12]), this property varies greatly between images and within images. An example of this is shown in figure 2, where three natural images taken from the BSDS300 dataset show very different filter response statistics. We can also quantify this disparity between natural images by noting that although overall in BSDS300 $\sim 10\%$ of pixels have gradient magnitude > 0.1, it is also the case that $\sim 30\%$ of the images in BSDS300 have $\sim 30\%$ of pixel gradient magnitudes > 0.1. This indicates that this sparsity property is not spread out evenly. Some images tend to contain much more textured regions than others and filter responses in these textured regions tend to have a non-sparse gaussian distribution.

When one layer contains texture, it becomes difficult or oftentimes impossible to correctly annotate edges in the overlapping regions of the second layer using the LW annotation mechanism. In figure 3, we show how these unavoidable inaccuracies in annotation affect the quality of the resulting LW separation. In all cases, annotation is performed automatically using the ground truth layers and the automatic annotation protocol of [10]. In general we see in the LW results that the inaccurate annotation leads to texture transferring into the edge layer and vice versa. In section 5 we quantify this visual observation.

3. THE GMM PATCH PRIOR

While the sparsity of filter responses may often serve as a reasonable prior, recent work has shown that more powerful models of natural images can be learnt. One very successful prior is the Gaussian Mixture Model (GMM) that models the

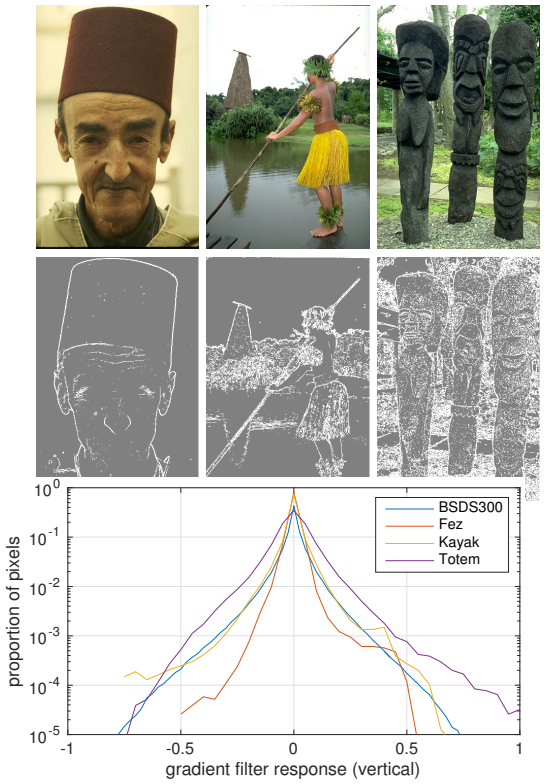

Fig. 2: Example images having a large variation in filter response statistics shown in the top row with corresponding pixels having gradient magnitude > 0.1 marked white in the middle row. We see that the fraction of high gradient magnitude pixels varies greatly and that vertical gradient histograms also display this over and under sparseness of the "Fez" and "Totem" images compared to the "Kayak" image and the full BSDS300 dataset.

statistics of 8×8 image patches (see [13]). Denoting the 64 dimensional vector representation of the patch by x, then under the GMM prior we have

$$\Pr(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Sigma_k),$$

where $\mathcal{N}(x; \mu_k, \Sigma_k)$ denotes a Gaussian density with mean μ_k and full covariance matrix Σ_k. In a similar fashion to [13] we learn two models having $K = 50$ and $K = 200$ components using the BSDS300 train dataset. Suppose we observe an image patch y that is the sum of two image patches x_1, x_2, then it can be shown[1] that the posterior probability of x_1 given y is also a Gaussian Mixture of the form

$$\Pr(x_1|y) = \frac{1}{Z} \Pr(x_1) \Pr(y - x_1)$$

having K^2 components, one for each Gaussian cross term. The mixture weights and means of the posterior GMM depend on the prior GMM as well as the input patch y, while the

[1] See derivation in online version: https://arxiv.org/abs/1702.05958v2

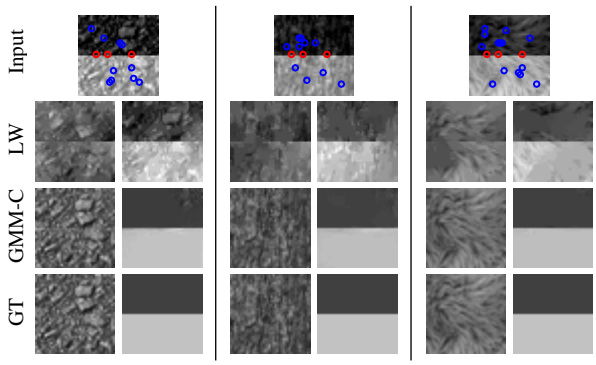

Fig. 3: Visual comparison of decomposing non sparse textures (asphalt, tree bark, fur) from a horizontal edge using the LW and GMM-C methods. Bottom row shows ground truth.

posterior covariances depend only on the prior covariances and may be precomputed. We also note that the resulting posterior means will generally be nonzero. Given an input image patch y a naive approach to finding the original patches is to find x_1 that maximizes $\Pr(x_1|y)$. Unfortunately due to the posterior probability being highly multimodal such an approach often fails. A more potent approach is to rely on the user at this point and ask that he picks x_1 from a set of candidate decompositions that would likely contain a close match to the true decomposition. We therefore seek for a set of preferably diverse decompositions that maximize $\Pr(x_1|y)$.

3.1. A new annotation method

We choose to approximate such a set of candidate decompositions by taking the means of the posterior GMM that have highest mixture weights. Figure 4 shows some examples of

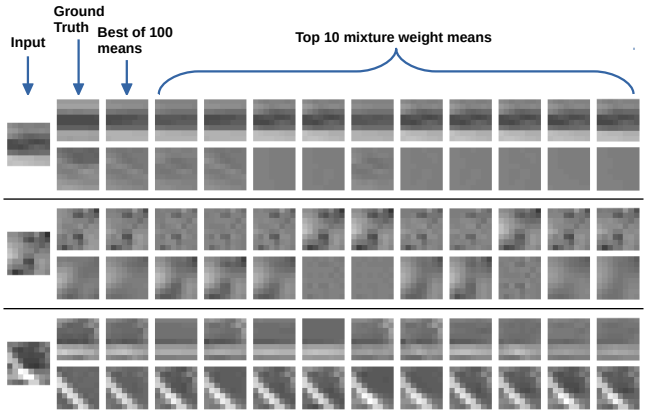

Fig. 4: Some examples of 8×8 patch decomposition candidates provided by the posterior means having top mixture weights. Here the $K = 50$ prior was used.

these local decompositions. Even when it is hard for an unguided human to decompose y into the original patch layers, the posterior GMM does a remarkable job of suggesting good

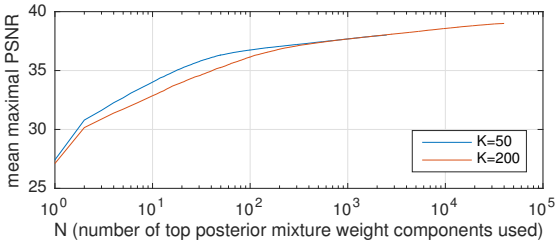

Fig. 5: Accuracy of best patch decomposition among N candidates: the N posterior means with highest mixture weights.

candidates among the top 100 posterior means and often even among the top 10.

To quantify the performance of this candidate proposal mechanism we repeatedly sampled a random pair of image patches x_1, x_2 from the BSDS300 test set, and summed them to create $y = x_1 + x_2$. Given y we created a set of N possible decompositions by taking the N posterior means with highest mixture weights, and measured the distance between the true x_1 and the closest posterior mean among these N candidates. Figure 5 shows that with as few as $N = 100$ candidates, it is possible to find a decomposition that is a very close match to the true x_1 (with accuracies of 36.7 dB for $K = 50$ and 36.2 dB for $K = 200$).

Our proposed annotation method is then to require the user to both pick an annotation point and to pick the most appropriate patch decomposition among the $N = 100$ posterior means with highest mixture weights at this point.

4. FULL ALGORITHM

Given input image y which is the sum of two unknown layer images x_1, x_2 the user is first prompted to provide annotation at a set of image points as follows:

- The user picks an annotation point i^* and the surrounding input patch $P_{i^*}y$ is extracted. Here P_{i^*} is the linear operator extracting an 8×8 pixel patch from an image encoded as a column vector.
- Given $P_{i^*}y$, the corresponding set of $N = 100$ posterior means with highest mixture weights are computed and presented to the user as candidate decompositions.
- The user then picks the decomposition most appropriate at point i^*. We denote by μ_{i^*} the posterior mean selected by the user and the corresponding posterior covariance by Σ_{i^*}.

The annotation is then automatically propagated by finding image x_1 maximizing the expected patch log likelihood (EPLL) function

$$EPLL(x_1|y) = \sum_i \log \Pr(P_i x_1 | P_i y),$$

under the set of annotation constraints $P_{i^*}x_1 = \mu_{i^*}$. Here i indexes all overlapping image patches and i^* indexes patches

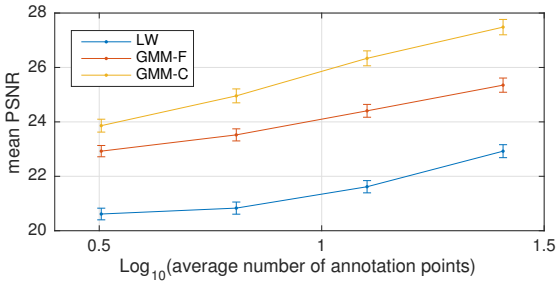

Fig. 6: Decomposition accuracy of LW, GMM-F and GMM-C measured on 8,000 40×40 pixel patches generated from random BSDS300 test patch pairs.

surrounding annotation points. In practice we replace these hard constraints by an additional cost and minimize

$$J_C(x_1) = -EPLL(x_1|y) + \frac{\lambda_C}{2} \sum_{i^*} (P_{i^*} x_1 - \mu_{i^*})^T \Sigma_{i^*}^{-1} (P_{i^*} x_1 - \mu_{i^*}).$$

This is the cost used in our proposed GMM-C method (GMM with component annotations). To optimize this cost we use the iterated least squares approach described in [14]. Once x_1 is estimated, the second image is then set to $x_2 = y - x_1$ per the problem definition.

5. RESULTS AND DISCUSSION

Accuracy was assessed by decomposing 8,000 40×40 pixel patches generated from the addition of random BSDS300 test patches. To separately measure the effects the revised prior and revised annotation each have on the relative accuracies of the GMM-C and LW methods, we also included a third control method denoted GMM-F (GMM with filter annotations). GMM-F uses the annotation cost term of the LW method (see [10]) but replaces the prior cost term with the negative EPLL cost defined in the previous section.

Filter annotations for the LW and GMM-F methods were generated automatically using the protocol used by [10]. Essentially a small random subset of the canny edges in the input image were classified as originating from either layer, depending on which ground truth layer had greater gradient magnitude there. Automatic annotation for GMM-C was computed at this same set of locations. At these locations one of the top two posterior means closest to the ground truth among the set of $N = 100$ posterior means with highest mixture weights was randomly picked. All results presented in this section were generated using the $K = 200$ GMM prior (both in the auto-annotation protocol and in the EPLL cost).

Accuracy statistics are presented in figure 6. The overall PSNR gain between the LW and GMM-C methods ranges between 3.3 dB at an annotation density of $1/22\,\mathrm{px}^{-2}$ and 4.6 dB at $1/8\,\mathrm{px}^{-2}$. Of this gain, approximately 2.5 dB is due to the modified prior (GMM-F) and approximately 1.6 dB is due to the modified annotation.

In figure 7 we show decompositions of real reflection images generated using the GMM-C and LW methods. In the GMM-C results we see successful separation of textured regions in one layer from overlapping edges and low frequency texture in the second layer (see e.g. the mannequin shirt appearing in the left-most column). These textured regions are not successfully separated by the LW method.

Using non optimized code in Matlab on a standard PC the run time for optimizing the costs of the previous section on an image from BSDS300 is approximately two minutes.

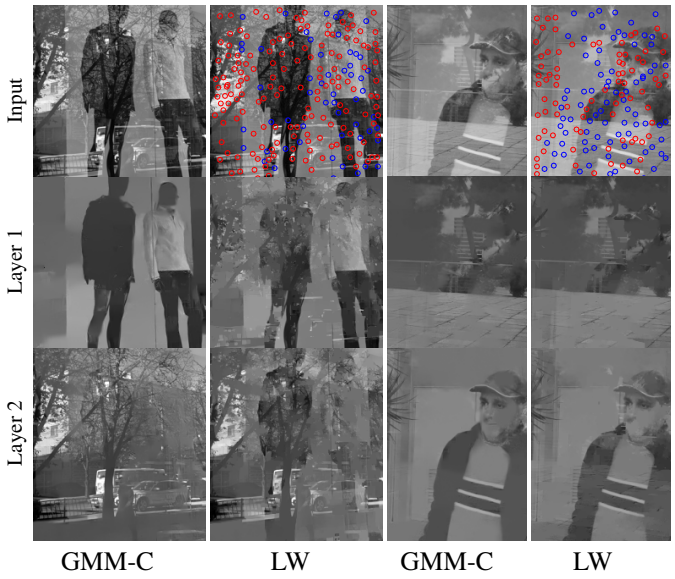

Fig. 7: Decomposition results for real reflection images using the LW and GMM-C methods. Note that GMM-C annotations are not shown but that they were provided at the same locations as the LW annotations. Best viewed on screen.

6. DISCUSSION

The single image reflection separation problem is inherently ill-posed and requires additional constraints provided by user annotations and a natural image prior. Previous work was based on a sparsity based prior and annotation mechanism which enabled semi-automatic separation of images in which both layers had little texture and the sparsity assumption was a good fit. However, for many real images, sparsity based methods cannot yield good separations even when a large fraction of the edges in the input image are correctly labeled. In this paper we have proposed a new user-assisted algorithm that is based on a much stronger prior of natural images - a GMM prior learnt from training examples. We have shown that this GMM, which has already been used successfully in image restoration tasks, can also be used to define a new annotation mechanism for user-assisted reflection separation. Our results show that high quality decompositions can be obtained with a relatively small amount of user interaction, even in the presence of significant texture. We thank the anonymous reviewers for their helpful suggestions.

7. REFERENCES

[1] YiChang Shih, Dilip Krishnan, Fredo Durand, and William T Freeman, "Reflection removal using ghosting cues," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3193–3201.

[2] Hany Farid and Edward H Adelson, "Separating reflections and lighting using independent components analysis," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.* IEEE, 1999, vol. 1, pp. 262–267.

[3] Yoav Y Schechner, Joseph Shamir, and Nahum Kiryati, "Polarization-based decorrelation of transparent layers: The inclination angle of an invisible surface," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. IEEE, 1999, vol. 2, pp. 814–819.

[4] Richard Szeliski, Shai Avidan, and P Anandan, "Layer extraction from multiple images containing reflections and transparency," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*. IEEE, 2000, vol. 1, pp. 246–253.

[5] Kun Gai, Zhenwei Shi, and Changshui Zhang, "Blind separation of superimposed moving images using image statistics," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 1, pp. 19–32, 2012.

[6] Tianfan Xue, Michael Rubinstein, Ce Liu, and William T Freeman, "A computational approach for obstruction-free photography," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, pp. 79, 2015.

[7] Xiaojie Guo, Xiaochun Cao, and Yi Ma, "Robust separation of reflection from multiple images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2187–2194.

[8] Yu Li and Michael S Brown, "Single image layer separation using relative smoothness," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2752–2759.

[9] Qing Yan, Yi Xu, and Xiaokang Yang, "Separation of weak reflection from a single superimposed image using gradient profile sharpness," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 937–940.

[10] Anat Levin and Yair Weiss, "User assisted separation of reflections from a single image using a sparsity prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, 2007.

[11] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, July 2001, vol. 2, pp. 416–423.

[12] Aapo Hyvärinen, Jarmo Hurri, and Patrick O Hoyer, *Natural Image Statistics: A probabilistic approach to early computational vision*, vol. 39, Springer-Verlag New York Inc, 2009.

[13] Daniel Zoran and Yair Weiss, "From learning models of natural image patches to whole image restoration," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 479–486.

[14] Effi Levi, *Using natural image priors-maximizing or sampling?*, Ph.D. thesis, The Hebrew University of Jerusalem, 2009.