Reflection Removal using ghosting cues

Abstract:

Photographs taken through glass windows often contain both the desired scene and undesired reflections. In this work, we use ghosting cues that arise from shifted double reflections of the reflected scene off the glass surface to exploit asymmetry between the transmission and reflections layers of the surface.

Introduction:

Ghosting provides a critical cue to separate the reflection and transmission layers, since it breaks the symmetry between the two layers. We model the ghosting as convolution of the reflection layer R with a kernel k. Then the observed image I can be modeled as an additive mixture of the ghosted reflection and transmission layers by R and T respectively:

$$I = T + R \otimes k$$

Estimation of K:

The kernel k, which is parameterized by a spatial shift vector dk and an attenuation factor ck.The 2D-autocorrelation of laplacian on the input image gives dk.The shifted copies of the reflection layer create a local maximum at dk on the autocorrelation map. To detect dk,a local maximum filter in each 5-by-5 neighborhood. For robust estimation, we discard local maxima in neighborhoods where the first and second maxima are closer than a predefined threshold.This removes incorrect maxima that are caused due to locally flat or repetitive structures.After removing local maxima within 4 pixels of the origin,the largest maxima in the remaining local maxima is taken as dk[1].The sign of dk, and we resolve this by choosing dk such that ck < 1, using the property that the second reflection has lower energy than the first.

For finding ck,We first detect a set of interest points from the input using a Harris corner detector. For most corner features within the image, the gradients of a local patch are dominated by the gradients of either R1, R2 or T.Around each corner feature, a 5 × 5 contrast normalized patch is extracted.We then estimate the attenuation between a pair of matching patches pi , pj as the ratio $a_{ij} = \sqrt{\frac{\mathrm{var}[p_i]}{\mathrm{var}[p_j]}}$, where var[pi] is the variance of the pixels in patch pi , and we choose the order of (i, j) such that aij < 1. Finally, we sum over all such pairs to give an estimate of ck:

$$c_k = \frac{1}{Z} \sum_{ij} w_{ij} a_{ij}$$

where $Z = \sum_{ij} w_{ij}$ wij is the normalization factor, wij = $e^{-\frac{\|p_i - p_j\|^2}{2\theta^2}}$, θ = 0.2.

Layer separation algorithm:

Our formation model for the observed image I, given the transmission T, reflection R and ghosting kernel k, is

$$I = T + R \otimes k + n$$

where n is additive i.i.d. Gaussian noise with variance σ 2.Given k, the above formation model leads to a data (log-likelihood) term for reconstruction of T and R:

$$L(T, R) = \frac{1}{\sigma^2} \|I - T - R \otimes k\|_2^2$$

However, minimizing L(T, R) for the unknowns T and R is ill-posed. Additional priors are needed to regularize the inference.The best performing priors are found using Gaussian mixture models(GMM).The GMM prior captures covariance structure and pixel dependencies over patches of size 8×8, thereby giving superior reconstructions to simple gradient-based filters, which assume independence between filter responses of individual pixels.Therefore the following cost has to be minimised.

$$-\sum_i \log(\text{GMM}(P_i T)) - \sum_i \log(\text{GMM}(P_i R))$$

where GMM(PiX) = PK j=1 π jN (PiX; 0, Σ j). The cost sums over all overlapping patches PiT in T, and PiR in R; where Pi is the linear operator that extracts the i th patch from T or R.A pre-trained zero-mean GMM model with 200 mixture components, and patch size 8 × 8. The mixture weights are given by { π j}, and the covariance matrices by { Σ j}.

So our final cost function with non-negativity constraint on T and R is:

$$\begin{aligned} \min_{T,R} \quad & \frac{1}{\sigma^2}\|I - T - R \otimes k\|_2^2 - \textstyle\sum_i \log(\text{GMM}(P_i T)) \\ & - \textstyle\sum_i \log(\text{GMM}(P_i R)), \text{ s.t. } 0 \le T, R \le 1 \end{aligned}$$

The constraints for the above equation are per-pixel and per channel.

Separation of T and R and optimisation:

An optimisation scheme based on half quadratic regularisation[2] method is used here.A set of patches ziN 1 , one for each overlapping patch Pix in the image, yielding the following cost function:

$$\begin{aligned} c_{p,\beta}(\mathbf{x}, \{\mathbf{z^i}\} \,|\mathbf{y}) = & \frac{\lambda}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 \qquad (\\ & + \sum_i \frac{\beta}{2}\left(\|\mathbf{P}_i\mathbf{x} - \mathbf{z}^i\|^2\right) - \log p(\mathbf{z}^i) \end{aligned}$$

As $\beta \rightarrow \infty$ the patches Pix are restricted to be equal to the auxiliary variables z i and the solution of the following equation will converge.

$$f_p(\mathbf{x}|\mathbf{y}) = \frac{\lambda}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 - EPLL_p(\mathbf{x})$$

For a fixed value of β , optimizing can be done in an iterative manner, first solving for x while keeping z_i constant, then solving for z_i given the newly found x and keeping it constant.The following function will be the cost function for this problem.

$$\min_{T,R,z_T,z_R} \frac{1}{\sigma^2}\|I - T - R \otimes k\|_2^2 \qquad (7a)$$

$$+ \frac{\beta}{2}\sum_i \left(\|P_iT - z_T^i\|^2 + \|P_iR - z_R^i\|^2\right) \qquad (7b)$$

$$- \sum_i \log(\mathrm{GMM}(z_T^i)) - \sum_i \log(\mathrm{GMM}(z_R^i)) \qquad (7c)$$

$$\text{s.t. } 0 \leq T, R \leq 1 \qquad (7d)$$

Starting with β = 200, and increase the value by a multiple of 2 after each iteration for 25 iterations in all.As β is increased, the values of PiT and ziT are forced to agree; similarly for the values of PiR and ziR.Value of σ =5 × 10−3.We solve for T and R simultaneously by transforming the term || I− T − R ⊗ k|| 2 to ||I − AX|| 2 .Here X vertically concatenates vectors T and R, i.e., X = [T; R], and A horizontally concatenates the identity matrix I and convolution matrix k, i.e. A = [I|k].An extended **L-BFGS** is used to handle box constraints. Since Pi contains only diagonal elements, and k contains only two non-zero entries for each pixel, the pixel domain LBFGS solver is very efficient.

Fixing T and R we update zi .The component with the largest likelihood in the GMM model, and then perform Wiener filtering using only that component; this is a simple least squares update.This is EPLL framework which is used.

$$\hat{\mathbf{x}} = \left(\mathbf{\Sigma}_{k_{max}} + \sigma^2\mathbf{I}\right)^{-1}\left(\mathbf{\Sigma}_{k_{max}}\mathbf{y} + \sigma^2\mathbf{I}\mu_{k_{max}}\right)$$

A good initialization is crucial in achieving better local minima. The initialisation for the GMM-based model is done with a sparsity inducing based model, with a convex L1 prior penalty:

$$\min_{T,R} \quad \frac{1}{\sigma^2}\|I - T - R \otimes k\|_2^2 + \sum_j \|f_j \otimes T\|_1 + \sum_j \|f_j \otimes R\|_1$$

The L1 optimization can be efficiently performed using ADMM and the sparsity inducing filters {fi} are set that include gradients and Laplacians.

References:

[1] Y. Diamant and Y. Y. Schechner. Overcoming visual reverberations. In IEEE Conference on Computer Vision and Pattern Recognition

[2] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration