21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Learning Human Preferences to Sharpen Images

Myra Nam and Narendra Ahuja
University of Illinois at Urbana-Champaign
myranam2, n-ahuja@illinois.edu

Abstract

We propose an image sharpening method that automatically optimizes the perceived sharpness of an image. Image sharpness is defined in terms of the one-dimensional contrast across region boundaries. Regions are automatically extracted for all natural scales present that are themselves identified automatically. Human judgments are collected and used to learn a function that determines the best sharpening parameter values at an image location as a function of certain local image properties. We use the Gaussian mixture model (GMM) to estimate the joint probability density of the preferred sharpening parameters and local image properties. The latter are then adaptively estimated by parametric regression from GMM. Experimental results demonstrate the adaptive nature and superior performance of our approach over the traditional Unsharp Masking method.

1 Introduction

Image sharpening is an important class of enhancement techniques. Spatial sharpening methods modify pixel values near edges [7, 5]. GradientShop method [4] sharpens salient edges and smoothens homogeneous area in gradient domain. The control parameters in any sharpening method play an important role in determining the perceptual quality of an output image. Oversharpening can result in noise amplification or loss in naturalness of photographs. In addition, different parts of an image may require different degrees of sharpening, as demonstrated by Figure 1. Adaptive parameter selection is required in order to obtain the most perceptually pleasing appearance.

It is not clear which local image properties the human eyes use to characterize the context so as to select the best parameters. It is also unknown what computational criteria may be used to automatically identify which image among a set of enhanced alternatives has been sharpened the best.

This paper is aimed at finding an automated method that maximizes the perceptual sharpness while preserving naturalness and the color of a given image. We hypothesize a set of image properties to model the context for selection of sharpening parameters. We assume these properties also imply some feature (sub)space which, once identified, could be used to uniquely specify and estimate the best sharpening parameters. We learn this (sub)space through a set of training examples for which human judgments are obtained by best mapping the space defined by the local properties of a judged sub-image to a (sub)space defined and learning the function that maps the (sub)space to the best sharpening parameter values. This function thus facilitates adaptive enhancement across an image since only the local image properties determine the value the function takes.

2 Motivation

A region is modeled as being homogeneous in intensity surrounded by a ramp. A ramp is a non-step transition between two adjacent regions in a real world image, resulting from factors such as circular lens aperture, sensor noise, round objects, out-of-focus blur and atmospheric turbulence. Edges (curves) exist inside the ramps (thick transition areas) where step transitions would occur if the contrast was increased indefinitely. Our approach explicitly identifies the ramps and the remaining, homogeneous parts of the regions using the image segmentation algorithm described in [1, 8, 2] (Figure 2). To increase the region contrast, we increase the slope of the ramp by adding an overshoot at its high end, and an equal undershoot at the other end, as shown in Figure 3. Our sharpening parameter is the magnitude of the overshoot/undershoot.

Underlying our approach is the presumption that human preferences for the best sharpening parameter values are consistent over time and across people, described and verified in [6]. We collected a dataset of the most preferred sharpening parameter values as a function of local context parameters.

Section 3 presents our algorithm to automatically estimate the sharpening parameter values from the col-

978-4-9906441-0-9

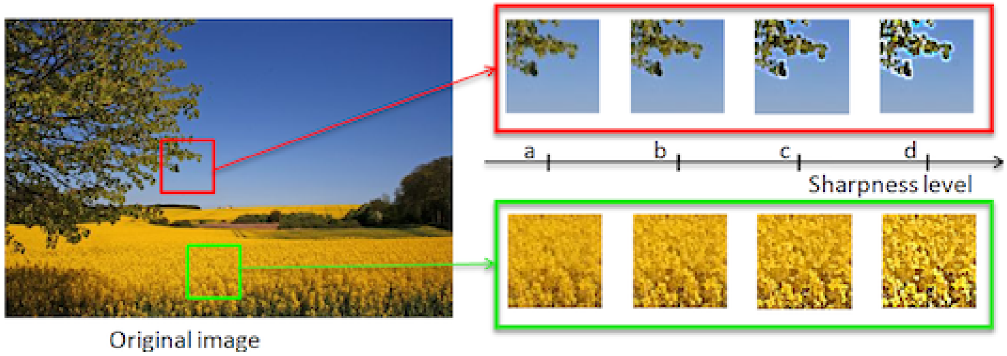
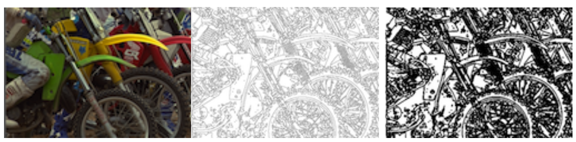


Figure 1: Two image patches are enhanced using four different sharpening levels: a,b,c,d. The red and green patches appear to be the most pleasing at level b and c, respectively. Thus, the degree of desired sharpness, and therefore the sharpening operation, must adapt to the image being sharpened. Using a single, sharpening parameter globally, across the entire image, can produce parts of the image to appear unnatural.



(a) (b) (c)

Figure 2: (a) Input image. (b) Edge map from segmentation. (c) Ramp map.

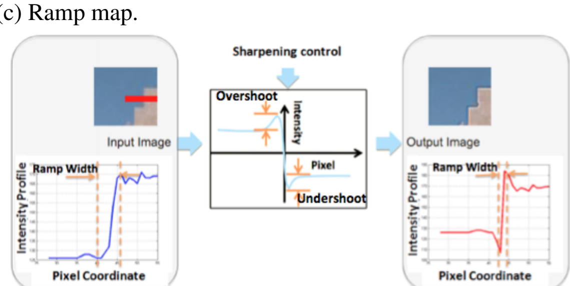


Figure 3: The region-based sharpening model reduces the perceived ramp width by adding overshoot and undershoot. The intensify profiles are taken along the red bar in the input and output images. The best overshoot/undershoot is determined by a supervised learning process from a dataset of training images and the associated human preferences for sharpness.

lected data on human preferences (Section 4). Section 4 presents the results followed by the conclusion.

3 Algorithms

We now describe our sharpening method, feature (sub)space selection, and the estimation of sharpening parameter.

3.1 Region-based sharpening

The sharpening method has a single parameter that locally controls the contrast increase in the detected edges. Edge pixels are detected by the segmentation algorithm [3]. Ramp pixels are estimated from the edge pixel location by the ramp definition [3]. At each end of the one-dimensional profile forming the ramp pixels, we add/subtract the first-order derivative of a one-dimensional Gaussian to increase the ramp value at higher end and reduce it at the lower end, thus increasing the contrast represented by the ramp. The output intensity $I_o(\mathbf{p})$ at a ramp pixel $\mathbf{p}$ is then obtained by

$$I_o(\mathbf{p}) = I_i(\mathbf{p}) + \alpha(\mathbf{p}) \cdot c(\mathbf{p}) \left\{ \frac{-d_{\mathbf{p}}}{\sigma_e^3} exp\left(-\frac{d_{\mathbf{p}}^2}{2\sigma_e^2}\right) \right\}, \quad (1)$$

where $d_{\mathbf{p}}$ is the distance from $\mathbf{p}$ to the nearest edge pixel. The local contrast $c(\mathbf{p})$ is estimated as a difference between the average intensity of the neighboring regions along the 1-D intensity profile. In Equation 1, we fixed $\sigma_e = 0.7$ so that the overshoot and undershoot are located inside the spatial ramp area. The control parameter is the scaling factor $\alpha \geq 0$, which is determined via supervised learning from our collected data (Section 4).

3.2 Feature selection

As illustrated in Figure 4, different sharpening parameter values are chosen to yield images that appear the most pleasing to human eyes. Now the eye appeal depends on more than just the ramp profile. Several other image features affect our judgment of preferred sharpness. Key features include interior color of a region, local contrast, edge orientations and ramp widths. These low-level visual features are extracted from a 20×20 neighborhood of a ramp pixel. In CIE-L*a*b color space, the mean, variance and maximum pixel values (3×3) are collected from non-ramp pixels. The mean, variance and maximum contrast values (3×3) are also collected from ramp pixels. The mean and variance of ramp widths (2) and edge orientations in horizontal/vertical and diagonal directions (2) are extracted as features. This yields a 22-dimensional feature vector $(3 \times 3 + 3 \times 3 + 2 + 2)$ is formed, whose dimensionality is reduced to 14 by Principle Component Analysis, and the feature space is transformed so that feature

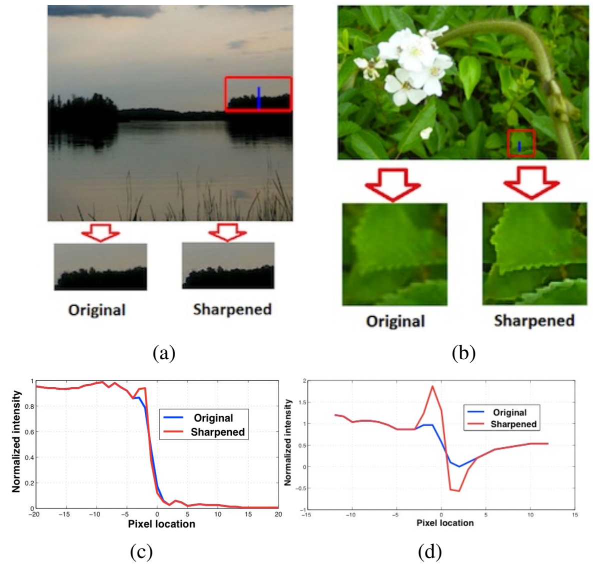


Figure 4: (a) and (b) Human subjects were asked to sharpen the parts highlighted in red, using our region-based sharpening model. They chose the sharpening parameter values to obtain the sharpest and most pleasing images. They chose different values: α=0.1 for (a) and 0.73 for (b). (c) and (d) display normalized one-dimensional intensity profiles along the blue lines in (a) and (b), showing different overshoot/undershoot. Our goal is to define a function that estimates the parameter α from the local image properties.

points and their nearest neighbors have the same sharpening parameter values, thus enabling similar images to be sharpened similarly [9].

3.3 Sharpening parameter estimation

We formulate the problem of estimating the control parameter α as a parametric regression problem to learn the data distribution by Gaussian Mixture Model (GMM). At each ramp pixel, we extract a feature vector $\mathbf{t} \in R^D$ as described in Section 3.2. GMM is used to estimate the joint probability of $\mathbf{x} = [\mathbf{t}^T \alpha] \in R^{D+1}$, where α is the optimal sharpening parameter value. In the training stage, the data set of N points is represented as an $N \times (D+1)$ matrix X in which the n^{th} row is given by $\mathbf{x}_n$. We maximize the log likelihood function

$$lnP(X|\pi,\mu,\Sigma) = \sum_{n=1}^{N} ln\{\sum_{k=1}^{K} \pi_k N(x_n|\mu_k,\Sigma_k)\}, \quad (2)$$

where $N(\cdot)$ is a Gaussian with mean μ and covariance Σ, and π is weight to the Gaussian. By the expectation-maximization algorithm, these GMM parameters are estimated, and the value K is chosen to minimize the Akaike information (AIC).

In the testing stage, given a test ramp pixel $\mathbf{t}$, we find α that maximizes $p(\mathbf{x}|\mu,\Sigma,\pi)$, $\mathbf{x} = [\mathbf{t}^T \alpha]$. For simplicity of notation, let Λ^k denote Σ_k^{-1}.

$$\begin{aligned} P(\alpha) &= p(\mathbf{x}|\mu,\Sigma,\pi) = \sum_{k=1}^{K} \pi_k N(\alpha|\mu_k,\Sigma_k) && (3) \\ &= \sum_{k=1}^{K} D_k exp\{-(A_k\alpha^2 + B_k\alpha + C_k)\} && (4) \end{aligned}$$

where for efficiency reason, $A_k = \Lambda^k_{D+1,D+1}$, $B_k = 2\sum_{j=1}^{D} \Lambda^k_{D+1,j}(t_j - \mu_j)$, $C_k = \sum_{i=1}^{D}\sum_{j=1}^{D}(t_i - \mu_i)\Lambda^k_{i,j}(t_j - \mu_j)$, and $D_k = \pi_k / \left((2\pi)^{\frac{(D+1)}{2}}|\Sigma|^{\frac{1}{2}}\right)$. Since $exp(-x)$ is monotonically non-increasing, the likelihood is maximized for the following values of α_k^*: $\alpha_k^* = arg_\alpha min(A_k\alpha^2 + B_k\alpha + C_k)$.

4 Results

We have collected 770 images while attempting to represent their natural diversity by including landscapes, urban scenes, objects, etc. From 9 human subjects at 4 different times, sharpening parameter values were obtained to yield the output images most pleasing to their eyes. In total, we collected 3440 training and 3440 testing sub-image windows of size 20 by 20. Median values of the sharpening parameters were taken as the ground truths. GMM was estimated with 25 Gaussian components.

Performance was evaluated on the test set by comparing the estimated parameter values with the ground truth. Mean of the resulting error was 0.0889, which is better than the performances of the algorithms in Two other estimation methods, Support vector machine and K-nearest neighbors ($K = 1$), were also compared with the GMM algorithm we have chosen (Table 1).

Table 1: Error performance for 3 algorithms. GMM achieved the best performance with the lowest error.

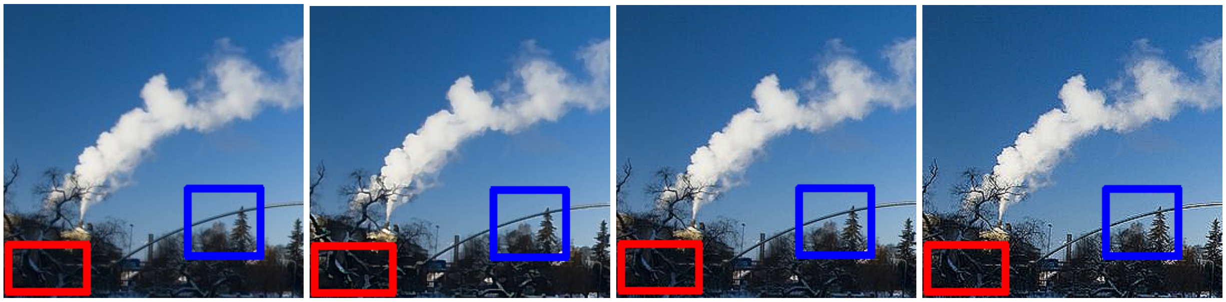
	Mean	Median	Variance
SVM	0.0914	0.1250	0.0115
KNN	0.1149	0.0920	0.0121
GMM	0.0889	0.0750	0.0071

We compare our results with corresponding results of unsharp masking (UM) which is the most commonly used method of sharpening (Figures 5, 6 and 7.) These demonstrate the superior quality of our results.

To conclude, our approach uses low-level image features to characterize human perceptual preferences and to learn to quantify subjective, psychophysical judgments to enhance images that appear both pleasing and realistic.

References

[1] N. Ahuja. A transform for multiscale image segmentation by integrated edge and region detection. *in PAMI*, 18(12):1211–1235, 1996.



(a) (b) (c) (d)

Figure 5: Results: (a) An input image. (b) Our result. (c) and (d) Photoshop UM applied once (c) and twice (d). (b) provides the maximum sharpness without distortion in both boxes. The blue box in (c) and the red box in (d) have sharpness similar to ours. However, the sharpness in the red box of (c) is below the desired level, while that in the blue box of (d) brings introduces artifacts near edges. Our adaptive approach chooses the best sharpness parameters.

(a) (b) (c) (d) (e)

Figure 6: Results: (a) Input image. (b) Our result. (c) UM applied by Photoshop twice. (d) Our estimates of sharpening parameter values in ramp pixels: brighter the pixel, larger the estimate. (e) Zoomed in displays of a window near the top left from: (top) input image, (middle) our result, and (bottom) UM. UM amplifies noise in the smooth area while our result does not.

(a) (b) (c) (d)

Figure 7: (a),(c) Input images. (b), (d) Our results.

[2] E. Akbas and N. Ahuja. From ramp discontinuities to segmentation tree. *In ACCV*, 2009.

[3] H. Arora and N. Ahuja. Analysis of ramp discontinuity model for multiscale image segmentation. *in ICPR*, 2006.

[4] P. Bhat, L. Zitnick, M. Cohen, and B. Curless. Gradientshop: A gradient-domain optimization framework for image and video filtering. *ACM SIGGRAPHS*, 2009.

[5] R. C. Bilcu and M. Vehvilainen. Constrained unsharp masking for image enhancement. In *Proceedings of the International Conference on Image and Signal Processing*, pages 10–19, 2008.

[6] A. J. Calabria and M. D. Fairchild. Perceived image contrast and observer preference: The effects of lightness, chroma, and sharpness manipulations on contrast perception. *Journal of Imaging Science and Technology*, 47:479–508, 2003.

[7] N. Joshi, R. Szeliski, and D. J. Kriegman. Psf estimation using sharp edge prediction. *In CVPR*, 2008.

[8] M. Tabb and N. Ahuja. Multiscale image segmentation by integrated edge and region detection. *IEEE Trans. on Image Processing*, 6(5), 1997.

[9] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10:207–244, 2009.