DOI: 10.1111/j.1467-8659.2012.03225.x

COMPUTER GRAPHICS *forum*
Volume 31 (2012), number 8 pp. 2528–2540

Content-Aware Automatic Photo Enhancement

Liad Kaufman, Dani Lischinski and Michael Werman

The Hebrew University of Jerusalem, Israel
liadkaufman@gmail.com, danix@mail.huji.ac.il, werman@cs.huji.ac.il

Abstract

Automatic photo enhancement is one of the long-standing goals in image processing and computational photography. While a variety of methods have been proposed for manipulating tone and colour, most automatic methods used in practice, operate on the entire image without attempting to take the content of the image into account. In this paper, we present a new framework for automatic photo enhancement that attempts to take local and global image semantics into account. Specifically, our content-aware scheme attempts to detect and enhance the appearance of human faces, blue skies with or without clouds and underexposed salient regions. A user study was conducted that demonstrates the effectiveness of the proposed approach compared to existing auto-enhancement tools.

Keywords: photo enhancement, tone and color manipulation, face enhancement, sky enhancement

ACM CCS: Image Processing and Computer Vision—1.4.3 Enhancement; 1.4.8 Scene Analysis; 1.4.9 Applications.

1. Introduction

The last decade has witnessed a dramatic growth in the amount of digital imagery captured and stored. This growth has been fueled by the advent of inexpensive digital cameras and camera-embedded mobile devices, as well as the abundance and increasing popularity of various channels for sharing images on the World Wide Web. The vast amounts of captured digital imagery and the fact that most of it comes from non-professional photographers underscores the need for effective automatic photo enhancement tools.

Indeed, virtually all existing photo management packages offer various automatic enhancement tools. However, most of these automatic approaches operate obliviously to higher level image content and their results leave considerable room for improvement. A number of content-aware enhancement methods have been recently proposed, however the automatic methods among these mostly focus on global tone and colour corrections.

In this paper, we describe a new automatic photo enhancement framework that intelligently combines a variety of global and local tone mapping, colour correction and detail enhancement operations, while attempting to take into account image content. This is a lofty goal since general image understanding is one of the long-standing grand challenges of computer vision. However, a large portion of images in a typical personal photo collection feature *people*, and virtually all outdoor photographs feature *sky*. Robust detectors for these already exist and are already in use in some consumer cameras. Thus, in addition to global manipulation of contrast and saturation, we detect image areas containing human faces, skin, blue sky and clouds, as well as salient underexposed parts of the scene, and apply customized enhancement operators in these areas. Two examples of our results are shown in Figure 1.

User studies we conducted show that the proposed approach consistently and significantly improves a large majority of a set of 100 images (randomly selected from a database of unprocessed images [BPCD11]), outperforming the automatic enhancement operators available in several popular products, and compares favourably with some recently published automatic global adjustment approaches.

Figure 1: *Two results produced by our method. The input image is on the left and our result is on the right. Note the improved illumination on the face, the different appearance of sky and clouds and the enhanced detail in textured regions.*

2. Related Work

There are many tools for automatically enhancing photographs. Basic contrast enhancement may be achieved through histogram equalization [GW06], or by stretching the tonal range (also known as *Auto Levels* in Adobe Photoshop). Despite the simplicity of the latter tool, it is often quite effective and appears to be in use in several popular commercial software packages. Examples include the *Auto Correct* feature in Microsoft's Office Picture Manager [Mic], the *I'm Feeling Lucky* button in Google's Picasa [Goo] and the *Auto Smart Fix* operation in Adobe's Photoshop Elements [Ado].

In addition to global contrast enhancement, photographs typically benefit from a modest amount of sharpening or detail enhancement, which may be achieved via unsharp masking [GW06]. Recent works advocated the use of non-linear filters for performing detail enhancement with fewer artefacts, such as the *weighted least-squares (WLS) filter* [FFLS08], edge-avoiding wavelets [Fat09] and several other filters [HST10, GO11].

Photographs in which interesting or important details are hidden in shadows may be improved using a variety of tone mapping operators originally developed for contrast reduction in HDR images [RWPD05].

All of the above methods are typically applied to the entire image and operate obliviously to higher level image content. In this work, we also use some of these tools, but we apply them in a highly selective and local fashion, driven by detection and analysis of image content.

Other content-aware image enhancement approaches include the work of Joshi *et al.* [JMAK10], which improves the quality of faces in a personal photo collection by leveraging better photos of the same person. In [vdWSV07], high-level visual information is used in the context of colour constancy. In the context of detail enhancement, the operator is sometimes applied only in high detail areas and not in smooth regions (e.g. [PRM00]). Wang *et al.* [WYW*10] describe a content-aware method for changing the colour theme of an image.

Other examples of content-centred manipulation include automatic touch-up of facial images by making inferences about undesirable textures in the image [BP08], automatic sky replacement [TYS09] for changing the mood of a photograph or automatic swapping of faces that are similar in pose and appearance [BKD*08]. Hasinoff *et al.* [HJDF10] describe a system that automatically propagates user-specified local edits in an image to other images in the same photo collection. Such methods introduce or remove content in the image, whereas our approach restricts itself to manipulating the tone and colour of existing content.

Dale *et al.* [DJS*09] describe a content-aware image restoration method that leverages a large database of images gathered from the web. Given an input image, they use the closest images found in this database for contrast enhancement, exposure correction and white balancing. Although their process of determining the correction parameters involves co-segmentation of a pair of images and local colour transfer between matching segments, the corrections are eventually applied globally to the entire image. In contrast, our approach involves both global and local corrections and does not require a large database of images.

Kang *et al.* [KKL10] and Bychkovsky *et al.* [BPCD11] describe methods that learn how different individuals prefer to adjust photographs, resulting in image enhancement methods that attempt to automatically correct images based on the learnt individual preferences. Caicedo *et al.* [CKK11] take this approach a step further and show that these individual preferences tend to cluster, and can be used to construct statistical preference models from a user group. These works use only global tone adjustment operators and do not attempt to apply them locally based on the image content as we do. We believe that the approach we present here may also benefit from personalization, but leave this to future work.

Finally, Berthouzoz *et al.* [BLDA11] describe a framework for creating content-adaptive macros for transferring photo manipulations to new target images. The key idea behind their method is to learn (from a set of training demonstrations) dependencies between certain image features and the location of selection regions, the paths of brush strokes and the parameters of image processing operations. Similarly to our approach, they rely on the ability to automatically detect face, skin and sky regions in images. There are, however, a number of important differences between the two approaches. Berthouzoz *et al.* treat each image pixel as either selected or not selected for the operation based on a low-dimensional pixel-level feature vector, and the adjustment parameters are learned *from the differences* between the features inside the selected region and its complement. Also, the inferred adjustment is applied to *all* the pixels inside the transferred selection. In contrast, our approach performs higher level image analysis (e.g. examining the entire histogram inside a face region or separating a sky region into its sky and clouds components). Based on this analysis, a piecewise smooth adjustment map is generated, which results in different amounts and types of adjustment for different pixels inside the same semantic region. For example, as described in Section 5., we correct the blue sky and the white cloud components of each pixel inside the sky region differently, and later combine the results together. Thus, we see their work as largely orthogonal to ours.

3. Overview

Our automatic *Content-Aware Photo Enhancement* approach operates by applying a sequence of global and local operators to the image, following the pipeline:

(1) Detection,
(2) Global contrast and saturation correction,
(3) Face enhancement,
(4) Sky enhancement,
(5) Shadowed-saliency enhancement and
(6) Detail and texture enhancement.

This section provides a high-level overview of our pipeline, whose main steps are described in more detail in subsequent sections. Steps 1 and 2 adapt existing techniques, whereas the remaining steps (3–6) are novel contributions of this work.

The first step consists of detecting faces, skin and sky in the input image. The results of this step are then used at various subsequent steps of our approach. We adapted standard detection algorithms, which can easily be replaced by better ones, as they become available. These detection tasks are not the focus of our work, but the specific detectors used in our implementation are described in the appendices, for completeness.

Figure 2: *Top: A sidelit face before and after our sidelight correction. Bottom: An underexposed face before and after our exposure correction.*

In the second step, we perform a global enhancement of the image by stretching the contrast and increasing the saturation. This is similar, to the best of our knowledge, to the automatic enhancement in most commercial tools (Picasa, Photoshop Elements, etc.) Specifically, we stretch the image contrast to full range by clipping 0.5% of the darkest and the brightest pixels and increase the saturation of each pixel by 20%[1]. These operations are applied to all pixels except those that were classified as skin or sky, as they are treated separately in subsequent steps.

In the *face enhancement* step (Section 4.), we analyze the illumination of detected faces and correct some common problems, as demonstrated in Figure 2. Specifically, we reduce the contrast of sidelit (partially shadowed) faces and increase the exposure of underexposed ones.

Next, we perform *sky enhancement* (Section 5.). This enhancement applies to blue skies with or without clouds. We assume that each sky pixel is a linear combination of a blue sky colour with a grey cloud colour. Using a process inspired by image matting, we separate the two colours at each

[1] It is common practice to make the colours more vivid by boosting the saturation. Professional products such as Adobe Lightroom or Apple's Aperture provide a *vibrance* slider, which avoids saturating skin tones. Similarly, we exclude skin pixels from this step.

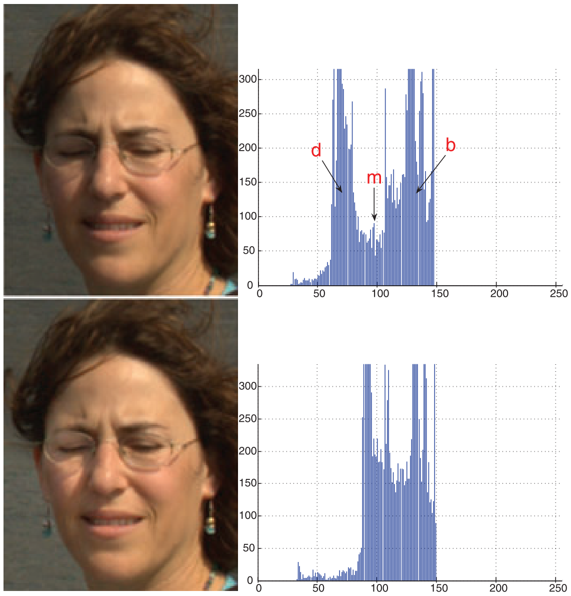

Figure 3: *The histograms of the skin pixels in a sidelit face before (top) and after (bottom) our correction. d, b denote the detected dark and bright modes and m the minimum between them.*

pixel and process each component separately. The sky component is adjusted towards a more attractive shade of blue, whereas the cloud component is corrected towards white (see Figure 6).

Many images contain dark, underexposed regions, which might be visually important. Indiscriminately brightening such regions results in an undesirable global contrast reduction, and therefore our next step attempts to only improve the visibility of the salient details, if present in underexposed image regions (*shadowed-saliency enhancement*, Section 6.), while preserving the brightness of other regions. A salience map is computed, and the exposure of each underexposed region is increased proportionally to its estimated salience (see Figure 7).

The perceived visual quality of an image is typically improved by slightly boosting fine scale detail and texture. Thus, the final step of our pipeline (*detail and texture enhancement*, Section 7.), performs such boosting using an edge-preserving decomposition of the image. Details are boosted in a selective, spatially varying manner, excluding image regions that could be visually harmed by this operation. The following sections explain steps 3–6 in greater depth.

4. Face Enhancement

We detect faces during the pre-processing stage. To cope with differently exposed faces, we first normalize the lighting in the image using the method proposed by Tao and Asari [TA05], and then apply the well-known Viola-Jones detector [VJ04] that produces a bounding rectangle for each detected face. Note that the normalization is only carried out to assist the face detection and not used afterwards.

We also compute a *skin probability map*, which estimates the probability of a pixel belonging in a skin region based on its colour. More details on this process are provided in Appendix A.

In the face enhancement step, we locally adjust the faces detected earlier, performing two possible corrections:

(1) *Sidelight correction*: reduce the contrast across faces that are partially brightly lit and partially shadowed.
(2) *Exposure correction*: brighten up underexposed faces.

Since a face may be both sidelit and underexposed, it may be necessary to apply both corrections. In this case, they are applied in the order listed earlier.

4.1. Sidelight and exposure correction

Sidelight and exposure corrections are essentially local tone mapping operators that manipulate the luminance of faces. We use the WLS filter of Farbman *et al.* [FFLS08] to decompose the monochromatic luminance channel into a *base layer* and a *detail layer*. The base layer is assumed to capture the illumination of the face. Thus, both the sidelight and the exposure correction operate on the base layer, as described later. Finally, the detail layer is added back in and the colour is restored, as typically done in tone mapping algorithms (e.g. [DD02]).

A sidelit face is characterized by a bimodal luminance histogram where one mode comes from the shadowed pixels in the face and the other from the brightly lit ones (Figure 3). Once we detect the two modes, we move the *dark mode* towards the *bright mode*. Although this reduces the overall contrast across the face, it increases the illumination in the shadowed region, leaving the brighter region unchanged. In our experience, this typically produces better results than moving the brighter region as well.

Thus, for each face, we first compute the histogram of the skin pixels inside the face's bounding rectangle. We then smooth the histogram and look for local maxima that exceed a certain threshold (at least 5% of the face pixels). If two such peaks are found, such that the minimum between them is at least 20% lower than each of the peaks, we choose these peaks as our dark and bright modes. If a bimodal structure is not detected, the face is not modified by this operator.

Having found the two modes, we pull the dark mode towards the bright mode in proportion to the distance between them. Specifically, let d and b denote the

Figure 4: *The result of the complete face enhancement step (with no other enhancements applied).*

intensities of the dark and bright modes, and m denotes the intensity at the minimum between them (see Figure 3). We construct a multiplicative adjustment map A where every skin pixel inside the face with intensity below m is multiplied by $f = (b - d)/(m - d)$, and use edge-aware constraint propagation [LFUS06] to smooth A. The resulting adjustment map is then applied to the base layer. Note that smoothing the map with edge-aware constraint propagation is instrumental for applying the changes in a piecewise smooth fashion to the entire face, while avoiding visible boundaries between modified and unmodified regions.

The sidelight correction demonstrates some important differences between our approach and that of Berthouzoz *et al.*: we analyze the structure of the luminance histogram inside the detected facial region rather than just considering its range, peak and median. We then modify only certain pixels in the region of interest, whereas others remain unchanged. A macro of this sort has not been demonstrated by Berthouzoz *et al.*

After sidelight correction, it still may be necessary to correct the overall exposure of the face to ensure satisfactory brightness and visibility. Having gathered statistics from 400 well-exposed faces, we found that the 75th percentile corresponds to a luminance of roughly 120 (of 255). Therefore, we correct the exposure by multiplying all the face pixels by a factor so that the 75th percentile shifts halfway towards 120. We empirically found this to produce better results than shifting all the way to 120, which may change the picture too much.

In order to prevent correction of well-exposed faces and avoid visual artefacts due to an overly aggressive exposure correction, the correction factor is bounded to be between 1 and 2. The entire face correction process is summarized in Algorithm 1. See Figure 2 for examples of the result of the sidelight and of the exposure correction, and Figure 4 for the combined result of applying both corrections.

Algorithm 1. *CorrectFace*(I, F, M_{skin})

Require: I – input image (luminance channel).
Require: F – detected face rectangle.
Require: M^{skin} – skin mask.
Ensure: I^{out} – adjusted image (luminance channel).

// Perform edge-preserving base/detail decomposition
$(Base, Detail) = WLSFilter(I)$
$I^{out} = Base$

// **Sidelight correction**
$S = F \bigcap M^{skin}$ // skin pixels inside F
$H =$ Smoothed histogram of intensities in $I^{out}[S]$
if H is bimodal (sidelit face) **then**
  $d =$ intensity of dark mode in H
  $b =$ intensity of bright mode in H
  $m =$ intensity at local minimum between d and b
  $f = \frac{b-d}{m-d}$
  $A =$ adjustment map scaling by f every pixel $\in S$ with intensity $\leq m$
  Apply edge-aware constraint propagation to A
  $I^{out} = I^{out} \cdot A$ // pixelwise multiplication
end if

// **Exposure Correction**
$p = 75^{th}$ percentile of face skin pixels S
if $p < 120$ (underexposed face) **then**
  $f = \frac{120+p}{2p}$; ensure $1 \leq f \leq 2$
  $A =$ adjustment map scaling by f every pixel $\in S$
  Apply edge-aware constraint propagation to A
  $I^{out} = I^{out} \cdot A$ // pixelwise multiplication
end if

$I^{out} += Detail$ // restore detail

5. Sky Enhancement

Recall that in the pre-processing stage we compute a sky probability map: the probability of each pixel to belong to the regions containing sky and clouds. This process takes into account pixels' smoothness, colour and positions, as described in more detail in Appendix B. In addition, we detect the largest connected sky region in the image, referred to below as the *sky reference patch*.

In the sky enhancement step, we decompose each high-probability sky pixel into a blue sky component and a grey cloud component. The goal is to change the sky colour to a 'nicer' shade of blue and make the clouds whiter and 'cleaner'. It should be noted that our method only enhances blue skies, not sunset ones, since we don't want to turn a

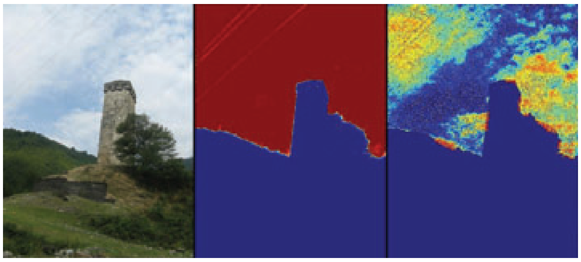

Figure 5: *An example of cloud/sky separation. Left: input image; middle: sky probability map, red indicates high sky/cloud probability; right: the α channel recovered by our method, hotter colours indicate higher cloud coverage.*

Figure 6: *An example of the results obtained by running the sky enhancement step alone. Left column: input images; middle column: the result of sky enhancement alone; right column: sky enhancement without separating the clouds produces bluish clouds.*

colourful sunset sky into a blue one. Since our sky detector is designed to only detect blue skies, the sky in sunset photos will not be detected, and the sky enhancement will not take place.

To decompose each sky pixel into its blue sky and grey cloud components, we assume the following simple image formation model:

$$p_i = \alpha_i \cdot \underbrace{c_i \cdot (1,1,1)^T}_{\text{Cloud}} + (1-\alpha_i) \cdot \underbrace{s_i \cdot (S_R, S_G, S_B)^T}_{\text{Sky}}, \quad (1)$$

where $S = (S_R, S_G, S_B)$ is the average colour of the sky reference patch, c_i accounts for the spatially varying grey level of the cloud, α_i is the 'cloud coverage' of pixel i and s_i is the spatially varying blue sky intensity.

Our goal is to recover the three maps (α, c, s), which we find by minimizing

$$J(\alpha, c, s) = \sum_i (D(\alpha_i, c_i, s_i) + \lambda R(s_i)), \quad (2)$$

where the *data term* D and the *regularization term* R are defined independently at each sky pixel $p_i = (r_i, g_i, b_i)$ as

$$\begin{aligned} R(s_i) &= (s_i - 1)^2 \\ D(\alpha_i, c_i, s_i) &= \\ &\left(\alpha_i c_i \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + (1-\alpha_i) s_i \begin{bmatrix} S_R \\ S_G \\ S_B \end{bmatrix} - \begin{bmatrix} r_i \\ g_i \\ b_i \end{bmatrix} \right)^2 . \end{aligned} \quad (3)$$

The R term ensures that the clear sky colour does not deviate much from S, whereas the D term attempts to satisfy the formation model (1). Figure 5 shows an example of the separation obtained in this manner.

Having obtained the sky/cloud decomposition, we perform the sky enhancement. After collecting and examining a set of professional manually enhanced photographs with blue skies, a bright blue target colour was chosen. We adjust the blue sky portion of each sky pixel towards this target colour. We compute a multiplicative correction factor $f^{\text{sky}} = (f_L, f_A, f_B)$ as the channel-wise ratio (in CIELAB colour space) between the target colour and the average sky reference patch colour. This correction is then carried out for each pixel i taking into account its sky probability P_i^{sky}. We also adjust the cloud component of each sky pixel towards white.

Let β_i^{old} denote the extracted blue sky colour $s_i(S_R, S_G, S_B)$ converted to the CIELAB colour space, and κ_i^{old} the extracted grey cloud colour (c_i, c_i, c_i) (also in CIELAB). We correct these two colours separately:

$$\beta_i^{\text{new}} = P_i^{\text{sky}} f^{\text{sky}} \cdot \beta_i^{\text{old}} + (1 - P_i^{\text{sky}})\, \beta_i^{\text{old}}, \quad (4)$$

$$\kappa_i^{\text{new}} = P_i^{\text{sky}} \frac{W + \kappa_i^{\text{old}}}{2} + (1 - P_i^{\text{sky}})\, \kappa_i^{\text{old}}. \quad (5)$$

Here, W is the reference white of CIELAB (100, 0, 0). We convert the new colours back to RGB where they are recombined using α_i. See Figure 6 for an example of this step's result.

6. Shadowed-Saliency Enhancement

In many photographs, parts of the scene appear darker than intended by the photographer. In this step, our goal is to increase the visibility of the details in salient regions of the image, while attempting to preserve global brightness relationships. This is done locally by increasing the exposure of underexposed regions proportionally to their estimated saliency. The exposure correction is bounded to avoid making

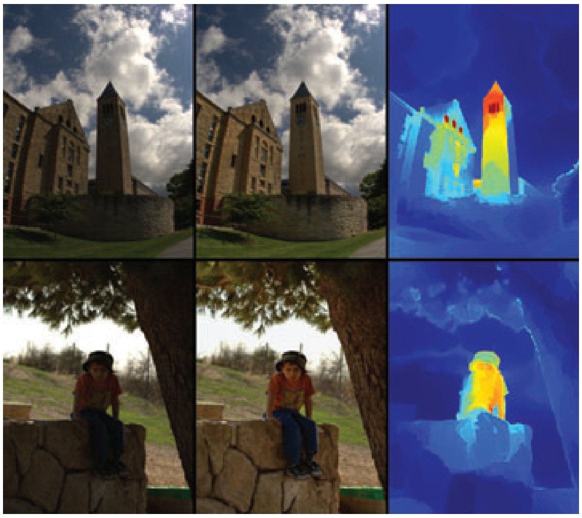

Figure 7: *Examples demonstrating the saliency enhancement step alone. Left: input image; middle: shadowed-saliency correction result; right: saliency map (after edge-preserving propagation).*

shadowed objects appear too bright. Since we avoid increasing the brightness of all shadowed regions indiscriminately, there is a smaller risk of reducing the global contrast in the image.

There are many methods for estimating the salience of image regions (e.g. [JEDT09, GZMT10]). Having experimented with several techniques, the one that proved best for our purpose is based on the energy map originally proposed by Kapadia [Kap08] for image resizing by seam carving [AS07]. This technique combines a number of cues:

(1) Greyscale gradient magnitude.
(2) Colour gradient magnitude (using the *A*,*B* components of CIELAB).
(3) Colour histogram: the rarer the colour the higher the saliency. (Again, using the *A*,*B* components.)
(4) Skin pixels are assigned higher saliency.
(5) Distance from centre.

We refer the reader to [Kap08] for more details. To complete the computation of the saliency map, we apply edge-preserving propagation [LFUS06] (using the edges of the original image). Two examples of the resulting maps are shown in Figure 7.

Next, we determine how to correct each underexposed region. We split the luminance channel into two parts: *DARK*, which includes all pixels in the image that have a value below 50 (of 255), and *BRIGHT*, which includes all the other pixels. We chose a factor of $f^{\mathrm{sal}} = min\left\{2, \frac{PT(BRIGHT,35\%)}{PT(DARK,95\%)}\right\}$ for multiplying the luminance of each pixel, where $PT(A, b)$ is the bth percentile of A's values. Bounding the factor by 2 is necessary to prevent unnatural brightening. Also, we exclude from *DARK* any pixel whose difference between its maximum and minimum RGB values exceeds 5. This limits the brightening to colourful pixels and avoids turning black-looking pixels to grey.

To avoid discontinuities resulting from the thresholding, we apply edge-aware smoothing to *DARK* using the WLS filter. The same filter is applied on the luminance channel to obtain an edge-preserving base-detail decomposition. Let M_i^{sal} denote the saliency of pixel i, and B_i its base layer luminance before the correction. For each pixel in the *DARK* set, we compute a corrected base layer luminance B_i^{new} as follows:

$$B_i^{\mathrm{new}} = f^{\mathrm{sal}}\, M_i^{\mathrm{sal}}\, B_i + \left(1 - M_i^{\mathrm{sal}}\right) B_i. \qquad (6)$$

The detail layer and colour are then restored. Note that the salient areas which are not underexposed are not affected.

7. Detail Enhancement

Photographs often benefit from a moderate amount of detail enhancement. Thus, the final step of our pipeline boosts the fine details and textures in the image. Doing this globally without accounting for image content can increase noise in the sky or emphasize skin blemishes. Also, we distinguish between images that have an identifiable focus of interest and those that do not. We assume that faces in images constitute a focus of interest proportionally to their size—the larger the face is, the less we would like to distract the viewer's attention from it. Thus, apart from avoiding detail boosting in sky and skin regions, we attenuate the degree of boosting in images that contain large faces, proportionally to the largest face size.

Let P^{ns} be a map that gives the probability of a pixel *not* being sky or skin, L the log-luminance of the image and D the detail layer extracted from L using the WLS filter. A detail-enhanced log-luminance image is obtained as

$$L^{\mathrm{new}} = L + c\, P^{ns} \cdot D, \qquad (7)$$

where $P^{ns} \cdot D$ is a pixelwise product. The factor $c \in [0, 0.25]$ determines the amount of detail boosting we give our image; c is bounded from above by 0.25 to avoid unnatural exaggerated details, and its magnitude is inversely proportional to the size of the detected faces (when faces cover more than a quarter of the image c becomes zero). Figure 8 shows two examples of the output of this step.

Figure 8: *Two examples demonstrating the effect of detail enhancement alone. The effect is subtle, and may be difficult to see: please view at 400% magnification. Note the increased definition in textured areas.*

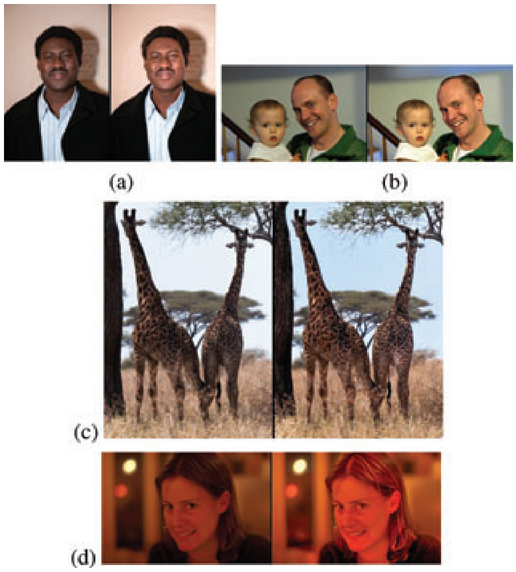

Figure 9: *A few examples demonstrating limitations of our method. The left image in each pair is the original, whereas our result is on the right. A few additional examples are included in the supplementary materials.*

8. Limitations

Although our method produces satisfactory results for most images, we experimented with, as evidenced by the results reported in Section 9.), we observed a few limitations.

Our method relies on a series of detectors to detect faces, skin, sky and salient regions. Thus, the effectiveness and the quality of our results depends on the success of these detectors. For example, sidelit or underexposed faces will only be corrected provided they are detected by the face detector, and a sufficient number of pixels inside the face is identified as skin. The same is true for the sky.

In particular, our skin and sky detectors both depend on colour. Thus, they might fail to detect skin or sky pixels in images with a sufficiently strong colour cast, or erroneously produce false positives, in which case some undesirable correction might occur. Our method is geared towards enhancing blue skies, and will not apply to sunset, or night skies.

Figure 9 demonstrates some of these limitations. Figure 9(a) shows a face with skin-coloured pixels on the wall behind the person. As a result, these pixels are brightened more than other pixels on the same surface. Figure 9(b) contains no sky, but a smooth patch in the top part of the image with sky-like colours was classified as sky, and its colour was shifted towards blue. In Figure 9(c), the sky is fragmented by the tree branches and some of the smaller fragments were not classified as sky. As a result after correction by our method, their colour differs slightly from the rest of the sky. Finally, Figure 9(d) features a strong red colour cast. The face skin pixels were classified correctly despite this colour cast, but our luminance correction produces an unappealing result in this case.

While our method avoids boosting details in sky or skin regions, in other regions of noisy images the noise will be amplified together with the fine scale details. Some noise may also become visible in salient regions brightened by our shadowed-saliency enhancement.

Finally, our method currently involves a number of empirically determined constants and thresholds. Further work is needed in order to determine these values in a more principled and robust manner, taking into account the input image.

9. Results

The enhancement pipeline described in the previous sections was implemented in Matlab. An unoptimized implementation of our method takes 2–4.5 minutes to process a 640×480 image, the exact time depends on the number and the sizes of the faces in the image. These times were measured on a 2.6 GHz Pentium Dual-Core E5300 PC with 3 GB of RAM. About a one-third of the running time is spent on WLS filtering (using the Matlab code provided by the authors of [FFLS08] on their project web page). We expect that switching to an optimized C/C++ implementation would dramatically decrease the running times. Another significant speed-up might be obtained by replacing the WLS filter with the use of edge-avoiding wavelets [Fat09], or the domain transform of Gastal and Oliveira [GO11].

To objectively evaluate the effectiveness of our method, we conducted several user studies. In the first user study, we asked users to compare the results produced by our method to the original input images, as well as to enhancements produced by several popular commercial products. We randomly selected 100 images from the *MIT-Adobe FiveK Dataset* [BPCD11], and each image was automatically enhanced using four different methods: our method; Google's Picasa *'I'm Feeling Lucky'*; Microsoft's Office Picture Manager *'Auto Correct'* and Adobe's Photoshop Elements *'Auto Smart Fix'*. Figure 10 shows several test images and their enhancements by these different methods. The full set of images is available in the supplementary material.

There were 71 participants in this user study. The participants were mostly students between the ages of 20 and 30, 21 females and 50 males, without particular experience with photo enhancement, beyond the use of products such as those mentioned earlier. The experiment was carried out by each of the participants connecting to a website using his/her own computer and monitor. Each participant was shown a set of 30 randomly selected image pairs. Each pair consisted of our method's result side-by-side with either the original image or the result of one of the other methods. The left/right ordering of the images in each pair was also chosen randomly. For each pair, the participant was asked to choose among three choices: 'The left image looks better', 'The right image looks better' and 'They look the same to me'. No further information on the goals of the experiment or the origin of the images was provided to the participants.

A summary of the results is presented in Table 1. These results show that in a significant majority of the pairs the participants preferred our result. This is true both overall, as well as individually for each of the other methods. If we exclude those image pairs for which participants were not able to prefer one image over another, the tendency to prefer our method becomes even more pronounced.

Unsurprisingly, the preference of our method was the strongest when compared against the original input images (79%). Still, in 13% of the pairs the participants indicated that they prefer the original image. This may be attributed in part to the differences between the subjective preferences of the participants. In addition, after completing the experiment, some of the participants who preferred the original images pointed out that in several noisy input images the noise was amplified and in several others the global contrast manipulation was too strong. In a small number of cases, participants indicated that they chose the original image because it looked more 'genuine', whereas the enhanced image did not fit the general 'atmosphere' of the scene.

Table 1: *User study results. 'Our' means that our method was preferred, 'Other' means that the method compared to was preferred. 'Same' means the participant saw no difference between the images in the pair.*

Competitor	Statistic	No. of Pairs	Same	Our	Other
All Others	All pics	1780	11%	64%	24%
	Face or Sky	1471	9%	66%	24%
	Face and Sky	186	13%	70%	17%
	No Face or Sky	309	20%	55%	25%
Office Picture Manager	All pics	458	15%	54%	31%
	Face or Sky	390	12%	56%	32%
	Face and Sky	40	13%	75%	13%
	No Face or Sky	68	29%	41%	29%
Photoshop Elements	All pics	485	11%	64%	26%
	Face or Sky	393	8%	65%	26%
	Face and Sky	54	11%	65%	24%
	No Face or Sky	92	20%	58%	23%
Picasa	All pics	432	12%	62%	27%
	Face or Sky	355	11%	63%	26%
	Face and Sky	49	16%	67%	16%
	No Face or Sky	77	16%	56%	29%
Original	All pics	405	8%	79%	13%
	Face or Sky	333	6%	82%	11%
	Face and Sky	43	12%	77%	12%
	No Face or Sky	72	17%	65%	18%

Since our method specifically targets faces and the sky, images containing one or both of these elements were preferred more often than images that did not contain them, demonstrating the importance of content-specific processing. Notice, however, that even when images contained neither faces nor sky, our method was still preferred more often. This is attributed to our use of the shadowed-saliency enhancement, as well as the detail enhancement. Both of these enhancements are also content-aware local operators, which apply to all images regardless of presence of faces and/or skies, but their effect tends to be more subtle because they are applied rather conservatively to reduce the risk of introducing unnatural artefacts.

Next, we conducted another study to compare our method to the recent automatic method described by Bychkovsky *et al.* [BPCD11]. They hired five trained photographers to manually retouch each of the 5000 images in the MIT-Adobe dataset, using only global adjustments. Bychkovsky *et al.* then trained an automatic method to predict the global adjustments made by one of the photographers (retoucher C, whose images were ranked the highest in a user study). The resulting automatic method was used to enhance about half of the 5000 images in the dataset. Our set of 100 randomly chosen images (from the first user study) includes 49 images that were enhanced using this method. Thus, we conducted another study in a similar manner to the previous one, where 22 participants were each presented with 25 randomly

Table 2: *A comparison between the method trained by Bychkovsky et al. [BPCD11] and our method.*

Competitor	Statistic	No. of Pairs	Same	Our	Other
[BPCD11]	All pics	550	4%	70%	26%
	Face or Sky	475	3%	72%	25%
	Face and Sky	—	—	—	—
	No Face or Sky	75	9%	57%	34%

chosen image pairs, of these 49 images. One image in each pair was enhanced using our method, whereas the other was enhanced using the method trained by Bychkovsky *et al*. The results can be seen in Table 2. No statistics are reported for the category of both face and sky as the set contained only two such images.

This comparison shows that the result of our method was preferred in a significant majority of the cases. Again, the preference was stronger for images containing faces or the sky. The results are not surprising, since although the method of Bychkovsky *et al*. attempts to predict the adjustments made by a human retoucher, these adjustments are restricted to global operators. Also, their method only adjusts the luminance, while preserving the chrominance of the image, whereas our method modifies the colours as well. Figure 11 shows two image pairs from this experiment. The top pair demonstrates a case where most participants preferred our method, whereas the bottom pair shows a case where the other method was preferred by most, despite the presence of the sky. This image contains almost only sky, and in this case the participants preferred the original deeper blue shade of the sky over our method's more standard sky colour. All image pairs from this experiment are provided in the supplementary materials.

Finally, we performed an experiment to compare our results to those produced by Caicedo *et al*. [CKK11], corresponding to one of the three main clusters of personalized enhancement preferences identified in that work. Unfortunately, the authors were only able to provide us with

Figure 10: *Several images from the MIT-Adobe FiveK dataset. In the first (left) column are the input images; in the second are our results; in the third—Photoshop Elements; in the fourth—Microsoft Office Picture Manager; in the fifth—Google's Picasa.*

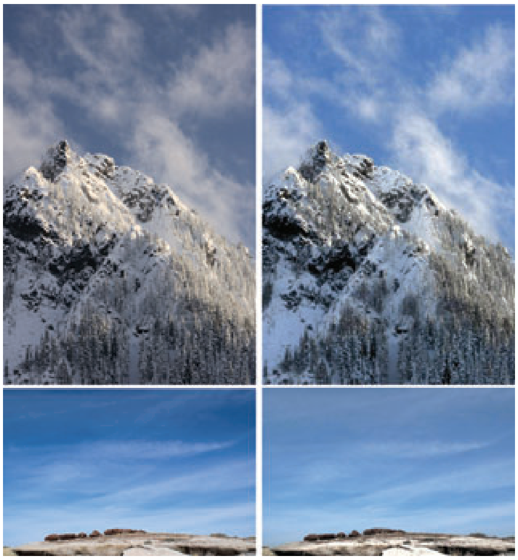

Figure 11: *Comparison between the automatic method in Bychkovsky et al.*[BPCD11]*(left column) and our method (right column). Top row: an image pair where our method was strongly preferred; Bottom row: an image pair where Bychkovsky's method was strongly preferred.*

eight low resolution results of their method, a dataset which is too small to draw any strong conclusions. Nevertheless, we conducted a third user study, where 19 participants were each shown eight image pairs (our result vs. [CKK11]). Our method was preferred in 68.6% of the cases, the collaboratively enhanced image was preferred in 25.5% of the cases and 5.9% of the cases were ranked the same. The images are provided in the supplementary materials.

10. Conclusion and Future Work

We described a new automatic photo enhancement framework that combines several tone mapping and colour correction operations in a single pipeline, applying them in a selective manner driven by the specific content detected in the image. We believe that our framework is modular, and will benefit not only from improvements in existing face, sky and saliency detectors, but also from incorporation of detectors for other kinds of content, such as vegetation or water.

We believe that our approach could benefit from personalization [KKL10, BPCD11]. For example, the system could learn the preferences of individual users with respect to global saturation, exposure of faces, skin tones and sky colour, and use these learned preferences instead of the generic parameters that we currently use.

Appendix A: Skin Detection

To determine whether a pixel is skin or not, we extracted about 300,000 skin patches of size 5×5 from a variety of images and calculated the mean of their A and B channels (in CIELAB).

As may be seen in Figure A1, most skin patch colours fall within an ellipse-shaped region of the AB plane. The ellipse fitted to this region is given by

$$\left(\frac{A-143}{6.5}\right)^2+\left(\frac{B-148}{12}\right)^2<1. \qquad \text{(A.1)}$$

Colours outside this ellipse are assigned skin probability that decreases proportionally with distance from the ellipse. This 'ellipse test' is further refined by thresholding again in the HSV domain. The threshold found in this colour space isn't tight (so it isn't good by itself), but it discards some of the non-skin pixels that passed our 'ellipse test'. We discard pixels that have a saturation value outside the range $0.25 \leq s \leq 0.75$, or a hue value larger than 0.095 (see Figure A1).

To avoid too many holes in the detected skin regions (due to noise and other artefacts in the colour), we also examine the colours that fall within the same ellipse after expanding it by a factor of 1.25. Patches in the image whose colour falls between the two ellipses are also classified as skin, but only if they are adjacent to skin pixels that passed the previous, more conservative test. Two examples of skin detection results are shown in Figure A2.

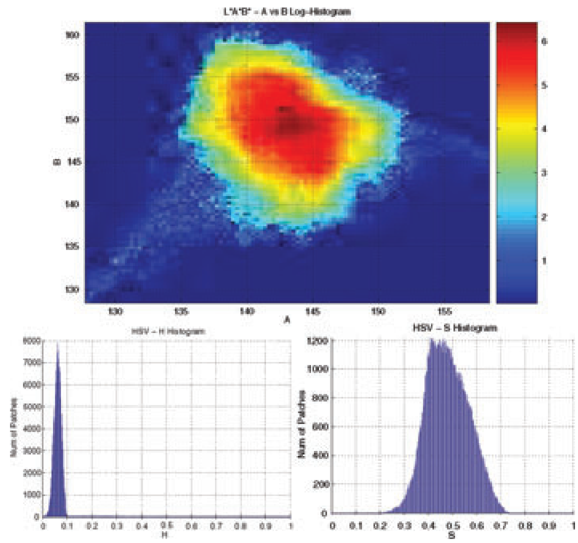

Figure A1. *Distributions of skin patch mean colours. Top: Log-histogram of the A and B coordinates in CIELAB colour space. Bottom: Histograms of the H and S channels in HSV colour space.*

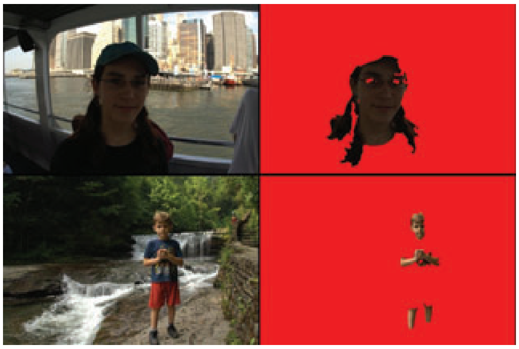

Figure A2. *Example skin-detection results. The pixels with skin probability less than 90% are coloured red on the right.*

Appendix B: Sky Detection

In this work, we restricted ourselves to non-sunset skies, so we assume that the sky has a blue hue. We take a similar approach to [SP09] and analyze the colour, position and shape of the different regions in the image in order to determine where the sky is (and if it is present at all), and create a *sky probability map* for each pixel.

By examining professional manually enhanced photos containing sky, we established the mean colour of a 'nice-looking' blue sky, as well as its variance in each colour channel. This gives us an 'ideal blue range'. We use this range to initially detect a large *sky reference patch* in the input image. Next, we use the more specific colour statistics of this reference patch to find the rest of the sky pixels in this image by requiring them to be in a colour range that is set according to the mean and variance of the reference patch.

Thus, we begin by assigning any pixel outside the ideal blue range a sky probability of zero, whereas pixels inside that range are assigned probability of one. Next, we refine the resulting binary map by keeping a probability of one for pixels with small gradients (under 5% change), and assigning an exponentially decreasing probability for blue pixels with larger gradients.

In many landscape images, the sky patches detected as described earlier also capture distant objects, due to attenuation and colour cast caused by the atmosphere. We handle this case by detecting a bimodal structure in the luminance histogram of detected sky patches, and exclude pixels that correspond to the darker mode, if present.

We assume that if the sky is present in an image, at least some part of it is located in the top third of the image, and classify an image as skyless if no such patch is found. Otherwise we designate the largest connected component in the top third of the image that passed the tests so far as the *sky reference patch*. We extract the mean and the variance inside this patch, and use them to assign each of the other sky candidate pixels a probability (assuming normal distribution).

At this stage, we have a sky probability map that contains blue sky pixels, as well as some of the clouds. We expand this map by adding to it all of the grey-coloured patches that are adjacent to high-probability sky pixels. This is the final map that is used in the sky enhancement step (Section 5.).

Acknowledgments

This work was supported in part by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities.

References

[Ado] Adobe Systems Inc.: Adobe Photoshop Elements 10. http://www.adobe.com/products/photoshopel.

[AS07] Avidan S., Shamir A.: Seam carving for content-aware image resizing. *ACM Transactions on Graphics 26*, 3 (July 2007).

[KD*08] Bitouk D., Kumar N., Dhillon S., Belhumeur P., Nayar S. K.: Face swapping: Automatically replacing faces in photographs. *ACM Transactions on Graphics 27* (August 2008), 39:1–39:8.

[BLDA11] Berthouzoz F., Li W., Dontcheva M., Agrawala M.: A framework for content-adaptive photo manipulation macros: Application to face, landscape, and global manipulations. *ACM Transactions on Graphics 30* (October 2011), 120:1–120:14.

[BP08] Brand M., Pletscher P.: A conditional random field for automatic photo editing. In *Proc. IEEE CVPR* (2008).

[BPCD11] Bychkovsky V., Paris S., Chan E., Durand F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In *Proc. IEEE CVPR* (2011).

[CKK11] Caicedo J. C., Kapoor A., Kang S. B.: Collaborative personalization of image enhancement. In *Proc. IEEE CVPR* (2011), pp. 249–256.

[DD02] Durand F., Dorsey J.: Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics 21*, 3 (July 2002), 257–266.

[DJS*09] Dale K., Johnson M. K., Sunkavalli K., Matusik W., Pfister H.: Image restoration using online photo collections. In *Proc. ICCV* (2009), pp. 2217–2224.

[Fat09] Fattal R.: Edge-avoiding wavelets and their applications. *ACM Transactions on Graphics 28*, 3 (July 2009), 22:1–22:10.

[FFLS08] FARBMAN Z., FATTAL R., LISCHINSKI D., SZELISKI R.: Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics 27* (August 2008), 67:1–67:10.

[GO11] GASTAL E. S. L., OLIVEIRA M. M.: Domain transform for edge-aware image and video processing. *ACM Transactions on Graphics 30*, 4 (August 2011), 69:1–69:12.

[Goo] Google: Picasa. http://picasa.google.com/.

[GW06] GONZALEZ R. C., WOODS R. E.: *Digital Image Processing* (3rd edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[GZMT10] GOFERMAN S., Zelnik-Manor L., TAL A.: Context-aware saliency detection. In *Proc. IEEE CVPR* (2010).

[HJDF10] HASINOFF S. W., Józ'wiak M., DURAND F., FREEMAN W. T.: Search-and-replace editing for personal photo collections. In *Proc. ICCP* (2010).

[HST10] HE K., SUN J., TANG X.: Guided image filtering. In *Proc. ECCV, Part I* (2010), Springer-Verlag, pp. 1–14.

[JEDT09] JUDD T., EHINGER K., DURAND F., TORRALBA A.: Learning to predict where humans look. In *Proc. ICCV* (2009).

[JMAK10] JOSHI N., MATUSIK W., ADELSON E. H., KRIEGMAN D. J.: Personal photo enhancement using example images. *ACM Transactions on Graphics 29*, 2 (April 2010), 12:1–12:15.

[Kap08] KAPADIA D.: *Improvements to Seam Carving: A Content Aware Image Resizing Technique*. Master of Science Graduation Report, Technische Universiteit Eindhoven, July 2008.

[KKL10] KANG S. B., KAPOOR A., LISCHINSKI D.: Personalization of image enhancement. In *Proc. IEEE CVPR* (2010).

[LFUS06] LISCHINSKI D., FARBMAN Z., UYTTENDAELE M., SZELISKI R.: Interactive local adjustment of tonal values. *ACM Transactions on Graphics 25*, 3 (July 2006), 646–653.

[Mic] Microsoft Corporation: Microsoft office picture manager. http://office.microsoft.com/en-us/help/about-picture-manager-HP001001721.aspx.

[PRM00] POLESEL A., RAMPONI G., MATHEWS V. J.: Image enhancement via adaptive unsharp masking. *IEEE Transactions on Image Processing 9* (2000), 505–510.

[RWPD05] REINHARD E., WARD G., PATTANAIK S., DEBEVEC P.: *High Dynamic Range Imaging*. Morgan Kaufmann, 2005.

[SP09] SCHMITT F., PRIESE L.: Sky detection in CSC-segmented color images. In *Proc. Computer Vision Theory and Applications (VISAPP)* (2009), vol. 2, pp. 101–106.

[TA05] TAO L., ASARI V. K.: Adaptive and integrated neighborhood-dependent approach for nonlinear enhancement of color images. *Journal of Electronic Imaging* (2005).

[TYS09] TAO L., YUAN L., SUN J.: Skyfinder: Attribute-based sky image search. *ACM Transactions on Graphics 28*, 3 (July 2009), 68:1–68:5.

[vdWSV07] van de Weijer J., SCHMID C., VERBEEK J.: Using high-level visual information for color constancy. In *Proc. IEEE ICCV* (2007).

[VJ04] VIOLA P., JONES M. J.: Robust real-time face detection. *International Journal of Computer Vision 57* (2004), 137–154.

[WYW*10] WANG B., YU Y., WONG T.-T., CHEN C., XU Y.-Q.: Data-driven image color theme enhancement. *ACM Transactions on Graphics 29*, 6 (December 2010), 146:1–146:10.