SKIN-AWARE LOCAL CONTRAST ENHANCEMENT

Tarik Arici and Salih Dikbas
School of Electrical and Computer Engineering, Georgia Institute of Technology
Atlanta, Georgia 30332–0250
Email: {tariq, salih}@ece.gatech.edu

ABSTRACT

Local contrast enhancement (LCE) gives a more lively look to an image or video. With a larger difference of a pixel's luma value from its local mean, the eye needs less time to adjust to that local region for a better contrast sensitivity. Since the human eye is trained and conditioned to recognize a natural looking face, the faces in a locally enhanced image may look unnatural even though other parts of the enhanced image is visually more attractive than the original image. To solve this problem, a low-complexity skin-aware local contrast enhancement algorithm is proposed. The proposed Skin-Aware Local contraSt enhAncement (SALSA) algorithm avoids creating false edges. Experiment results show that SALSA produces natural looking face and non-facial skin regions in the locally enhanced image.

Index Terms— Image enhancement, skin, video signal processing

1. INTRODUCTION

Contrast enhancement techniques are widely used to increase the visual image quality. Global contrast enhancement (GCE) techniques remedy problems that manifest themselves in a global fashion such as excessive/poor lightning conditions in the source environment. GCE algorithms enhance the image using a single mapping derived from its histogram and maps the original dynamic range to a bigger dynamic range. On the other hand, local contrast enhancement tries to enhance the visibility of local details in the image by amplifying the difference of a pixel's luma value from its local mean. Locally enhanced images look more attractive than the originals because with higher local contrast the human eye needs less amount of time to adjust to the local mean of the luminance [1].

A local contrast enhancement method that uses an adaptive first-order recursive (YENI) filter to find the local mean [2] and amplifies the deviations from that local mean is used. The YENI filter is a low-pass filter with very low computational and memory requirements that are especially tailored for real-time applications. The first-order delay coefficient is spatially adapted to preserve edges while smoothing out the details on the objects. Enhancement is done by amplifying the detail image (i.e. difference of the original image with the low-pass filtered image) and adding it back to the low-pass filtered image. Since the YENI filter is essentially an edge preserving low-pass filter, the detail image does not have large energy components around the edges and over/under shooting is successfully avoided.

Human intelligence is highly trained to recognize human faces and skin. Starting from the early infancy phases, humans learn to recognize other people with their faces [3], and communicate indescribable feelings and thoughts with simple facial mimics. As Cicero said, "Everything is in the face". Hence, special care must be taken for face and skin regions of an image before displaying. Faces in local contrast enhanced images may sometimes look unnatural although other regions look visually more attractive than the original. There is a clear need for detecting skin and non-skin regions and applying a lighter level of local contrast enhancement to skin regions, if not some smoothing.

Pixel-based skin detection algorithms utilize the fact that skin colors are clustered in the color space [4–6]. Although there are small variations with respect to race and illumination conditions, skin color lies inside a definite shaped color space region (i.e. skin locus) [7]. Variations on the skin color due to race and lightning conditions are explained by intensity variations in the chrominance components. Therefore, color space based skin color detection methods are robust to lightning conditions. Another reason to discard the luminance component is to decrease the computational complexity through dimensionality reduction.

Skin color modeling based skin detection methods have 90% true positive, and 20% false positive rate on average depending on the model and (to some extent) the color space used [8]. Since our goal is to discriminate skin pixels for a lighter level of enhancement, the true positives compared to the false positives are given more importance. To increase the performance of skin detection algorithms, more information is needed. One such information is the skin color homogeneity based on the assumption that skin regions in general consist of smooth skin patches. This information is especially useful in nudity detection [9]. With the help of the homogeneity assumption true positive and false positive rates can be improved moderately (e.g. to 96%, and 4.5% respectively [6]).

In this paper, SALSA: a low-complexity human skin aware local contrast enhancement method is proposed. SALSA utilizes the likelihood of a pixel to be a skin pixel in order to modulate the amplification gain of the detail image for local contrast enhancement. The skin color distribution is modeled with a Gaussian distribution on the YCbCr color space. Contrary to pixel-based skin detection algorithms that utilize thresholds, a soft discrimination between skin and non-skin pixels by using the skin likelihood is used. This way, unwanted false edges caused by false classification is prevented. Furthermore, to deal with isolated non-skin pixels in skin regions (or the opposite), edge information is used to impose correlation between skin likelihoods of neighboring pixels. Thus, a skin likelihood map consists of contiguous skin regions corresponding to actual skin regions. Experiment results show that SALSA produces natural looking face and non-facial skin regions in the enhanced image while keeping the same level of enhancement on non-skin regions.

In the following, local contrast enhancement using the YENI filter is discussed. Then, the skin color model used is discussed along with how the skin likelihood is obtained using the edge information. Next, the amplification gain modulation using this likelihood is described. Finally, experimental results are presented and conclusions are drawn.

2. LCE USING THE YENI FILTER

The enhanced image pixel $y(m,n)$ is obtained from the input image pixel $x(m,n)$ as

$$y(m,n) = \mu(m,n) + [1 + g(m,n)][x(m,n) - \mu(m,n)], \quad (1)$$

where $\mu(m,n)$ is the local mean, $g(m,n)$ is the enhancement gain, m is the row number, and n is the column number.

YENI filter finds the local mean by averaging two opposite direction non-linear filters operating row-wise. Row-wise implementation is done to save computational time and memory resources. A gain function is designed to suppress noise visibility in smooth regions and amplify local details.

The local mean $\mu(m,n)$ at row m and column n is the output of YENI filter, which is the average of two different filter outputs given by

$$\mu(m,n) = \frac{\mu_F(m,n) + \mu_B(m,n)}{2}. \quad (2)$$

where $\mu_F(m,n)$ and $\mu_B(m,n)$ are the outputs of the two opposite direction filters that run horizontally on a single row. The first filter runs from left to right and is referred to as the forward filter. The forward filter outputs $\mu_F(m,n)$. The second filter runs from right to left and is similarly referred to as the backward filter. The backward filter outputs $\mu_B(m,n)$. The two filters are single pole infinite-length impulse response (IIR) filters at any given pixel location. The input-output relationship for the forward filtered $\mu_F(m,n)$ is

$$\mu_F(m,n) = \lambda(m,n)\mu_F(m,n-1) + [1-\lambda(m,n)]x(m,n), \quad (3)$$

and for the backward filtered $\mu_B(m,n)$ is

$$\mu_B(m,n) = \lambda(m,n)\mu_B(m,n+1) + [1-\lambda(m,n)]x(m,n), \quad (4)$$

where $\lambda(m,n)$ is the edge adaptive delay coefficient.

The adaptation of $\lambda(m,n)$ to the edge information is crucial for preventing the smoothing of edges. Considering that $\lambda(m,n)$ is the weight of the previous output, a stronger $\lambda(m,n)$ increases the low-pass characteristic of the filter. Hence, when an edge is encountered, $\lambda(m,n)$ must be decreased so that the edge will be preserved in the output. The used edge signal is $|\mu_F(m,n-1) - x(m,n)|$ for the forward filter, and $|\mu_B(m,n+1) - x(m,n)|$ for the backward filter. Both of the edge signals are the differences between the original pixel value and the previous filter output. Using these edge signals, $\lambda(m,n)$ is obtained using

$$\lambda(m,n) = [1 - \frac{|\mu_F(m,n-1) - x(m,n)|}{255}]^\alpha \quad (5)$$

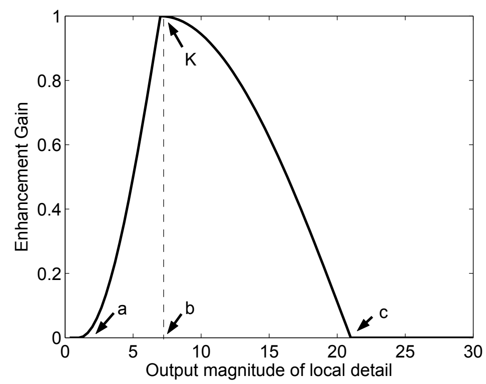

Fig. 1. Enhancement gain function

for the forward filter, and using

$$\lambda(m,n) = [1 - \frac{|\mu_B(m,n+1) - x(m,n)|}{255}]^\alpha \quad (6)$$

for the backward filter.

Here, 255 is used for the maximum possible pixel value. As can be observed from (5) and (6), strong edges reduce $\lambda(m,n)$ more, hence the low-pass characteristic of the filter is lessened. Typical α values range from 5 to 9.

2.1. Transfer function of the YENI filter

For ease of notation, the original pixel at the n^{th} column of row m is denoted as $x_m(n)$. Then, each row of the original image is a 1-D signal. From (3) and (4), the transfer function of the forward and backward filters at a locality with $\lambda(m,n) = \lambda$ are derived as below

$$H_F(w) = \frac{1-\lambda}{1-\lambda e^{-jw}} \quad (7)$$

$$H_R(w) = \frac{1-\lambda}{1-\lambda e^{+jw}}, \quad (8)$$

respectively. Here, it is implicitly assumed that λs for the two filters are equal since ideally edge information at the same locality must be the same. The stability condition for the IIR filters (i.e. the poles must lie inside the unit circle) is satisfied by (5) and (6) since $0 \leq \lambda \leq 1$.

From (7) and (8), it can be seen that the phases of both filters not being zero causes a phase shift in the filtered output. In fact, the forward filter lags and the backward filter leads the input signal $x_m(n)$. However, the frequency response of the local mean filter using (2), (7), and (8) can be obtained as below

$$H(w) = (1-\lambda)\frac{1-\lambda\cos(w)}{1-2\lambda\cos(w)+\lambda^2}, \quad (9)$$

which has a zero phase. Thus, the 1-D local mean filter applied row-wise and given by (2) does not shift the 1-D input signal column-wise.

2.2. Enhancement Gain Function

There are two important design goals for the enhancement gain function: avoiding noise visibility especially in smooth regions and preventing intensity saturation to minimum and maximum possible intensity values (e.g. 0 and 255 for 1 byte per channel source format). To deal with these problems, the enhancement gain depends on the magnitude of the detail in that locality.

To avoid noise visibility in smooth regions, the gain should also be small when the magnitude of the local detail is small. As the details increase the gain should also be increasing. However, increasing the gain function continuously may lead to saturation. This can be prevented by reducing the enhancement gain after some point. Considering these specifications, a gain function is designed that is an upward shifted cosine evaluated in the 3^{rd} quadrant for [a-b], and a cosine evaluated in the 1^{st} quadrant for [b-c], where a, b, and c are levels for the local detail's magnitude as shown in the example gain function given in Fig. 1. Here, a, b, c, and K are chosen as 1, 7, 21, 1, respectively, where K is the maximum achievable gain that determines the strength of the enhancement signal.

3. SKIN AWARE LOCAL CONTRAST ENHANCEMENT

Skin color models can be divided into two groups in general: parametric and non-parametric models. Non-parametric models are mainly histogram-based models learned from training data sets. They require memory and may not generalize depending on the representativeness of the training set. However, parametric modeling interpolates the training set and generalizes better. The interpolation is a desired feature for the objective since a soft discrimination between skin and non-skin pixels is wanted.

3.1. Gaussian skin color model

Modeling skin color distribution with a Gaussian model is done before in [10–12]. The joint probability density function of the color vector, $\mathbf{c}$, of a pixel given that it is a skin pixel is defined as:

$$p(\mathbf{c}|s=1) = \frac{1}{2\pi|\mathbf{\Sigma}|^{1/2}} e^{\frac{1}{2}(\mathbf{c}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{c}-\mu)}, \quad (10)$$

where $s = 1$ indicates a skin pixel, μ is mean vector, and $\mathbf{\Sigma}$ is covariance matrix. The Gaussian parameters estimated in [5] for the YCbCr color space are used

$$\mu = \begin{bmatrix} \mu_b \\ \mu_r \end{bmatrix} = \begin{bmatrix} 108.15 \\ 152.00 \end{bmatrix} \quad (11)$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_\mathbf{b}^\mathbf{2} & \sigma_\mathbf{br} \\ \sigma_\mathbf{br} & \sigma_\mathbf{r}^\mathbf{2} \end{bmatrix} = \begin{bmatrix} 55.77 & -58.66 \\ -58.66 & 85.27 \end{bmatrix} \quad (12)$$

Since linear combinations of Gaussian variables are also Gaussian variables, to find the Gaussian parameters for other colorspaces, such as the YUV for analog composite video, all one needs to do is to transform the above estimates using the transformation matrix T between the two colorspaces (i.e. the transformed estimates are $T\mu, T\Sigma T^T$).

Using Bayesian formula the posterior probability of a color vector to come from a skin pixel (i.e. the skin likelihood) is

$$p(s=1|\mathbf{c}) = \frac{p(\mathbf{c}|s=1)p(s=1)}{p(\mathbf{c})} \quad (13)$$

The prior probability of skin $p(s = 1)$ will be a constant depending on the type of the image or video. To save computation, color vectors are assumed to be uniformly distributed on the color space (i.e. all colors are equally likely with $p(\mathbf{c}) = \frac{1}{224^2}$). Hence,

$$p(s=1|\mathbf{c}) = k\, p(\mathbf{c}|s=1). \quad (14)$$

$k = 1$ is used in the following discussions and the experimental results.

An example image and its skin likelihood map is given in Fig. 2(a) and Fig. 2(b). Isolated non-skin pixels in skin regions and skin pixels in non-skin regions can be seen. This is partly due to the Gaussian model's failure (e.g. the non-skin pixels on the face) and use of the color information only to detect the skin. For example pixels on the edge's of the shirt have skin color because slow chrominance transitions imposed by the composite analog video broadcast standards (e.g. NTSC, PAL) cause hue change artifacts and in this case the hue changes from shirt's red color to skin color on the edges. To deal with the Gaussian model's failure, the edge information is used to impose spatial correlation between the skin likelihoods.

As described in Section 2, λ of the YENI filter (a number between 0 and 1) is updated with the edge information using (5) such that lambda decreases with the edge strength. To deal with isolated skin likelihoods (SLs), the skin likelihood is correlated in a locality with a recursive estimation that adapts the amount of correlation according to the edge strength. No correlation across the edges is imposed.

Since λ is between 0 and 1, the probability of edge is assumed to be $p(E = 1) = (1 - \lambda)$; then, the probability of no edge is $p(E = 0) = \lambda$. Then, SL in a locality, $\bar{p}(s = 1)$, can be written in terms of its conditional probabilities:

$$\bar{p}^{new}(s=1) = \bar{p}^{old}(s=1|E=0)p(E=0) + p(s=1|\mathbf{c})p(E=1), \quad (15)$$

where $\bar{p}^{new}$ is the updated local SL using the old local SL ($\bar{p}^{old}$) and current pixel's SL. In other words, the old local SL is used if there is no edge and current pixel's SL is used if there is an edge. To decrease the estimation variance of the local SL, a weighted average using the current pixel's SL is calculated as

$$\bar{p}^{old} = \frac{7}{8}\bar{p}^{old} + \frac{1}{8}p(s=1|\mathbf{c}). \quad (16)$$

Substituting (16) and edge probabilities in (15) gives

$$\bar{p}^{new} = \frac{7}{8}\lambda\bar{p}^{old} + (1 - \frac{7}{8}\lambda)p(s=1|\mathbf{c}). \quad (17)$$

An example map for the correlated skin likelihood is given in Fig. 2(c).

Using this skin likelihood, the enhancement gain can be modulated so that a skin pixel will be enhanced less than a non-skin pixel. This way, the natural look of the face can be preserved while other parts of the image are enhanced. This modulation can be done using the updated local SL ($\bar{p}^{new}$) (17) in (1) as below

$$\begin{aligned} y(m,n) &= \mu(m,n) + [1 + (1 - \bar{p}^{new}) g(m,n)] \\ & \quad [x(m,n) - \mu(m,n)]. \end{aligned} \tag{18}$$

Thus, amplification of the local detail will decrease with its skin likelihood. Other modulation functions can also be used such as the square root of skin likelihood if less amount of skin enhancement is desired. Furthermore, for a better visual quality one can smoothen skin regions by contracting the deviation from local luminance mean instead of amplifying it with a gain. This can simply be done by making the gain negative for highly likely skin regions as below

$$\begin{aligned} y(m,n) &= \mu(m,n) + [1 + \frac{k - \bar{p}^{new}}{k} g(m,n)] \\ & \quad [x(m,n) - \mu(m,n)]. \end{aligned} \tag{19}$$

where the used gain modulation function ($\frac{k-\bar{p}^{new}}{k}$) is negative for $\bar{p}^{new} > k$ and 1 for $\bar{p}^{new} = 0$.

4. EXPERIMENTAL RESULTS

Video captured from an NTSC broadcast is used for performance evaluation. Fig. 2(c) shows that by correlating the SLs, isolated likelihoods are lessened and smoothed. Fig. 2(c) gives an imaged enhanced using LCE only. Unnatural looking skin regions can be seen on the nose, sides of the mouth and on the right side of the neck. These artifacts on the face and the neck are almost removed using SALSA as shown in Fig. 2(e) and Fig. 2(h). It is important to note that there is no loss of enhancement on non-skin regions. Smoothing the skin regions looks more natural as given in Fig. 2(f) and Fig. 2(i), which are smoothed using $k = 0.75$. However, this comes with a cost and the level of enhancement on the other image part is not as strong as SALSA without smoothing.

5. CONCLUSIONS

Contrast enhancement is a technology used in medical imaging and consumer electronics display applications. Image enhancement for Television displays makes use of LCE techniques. However, these techniques can sometimes produce unnatural images especially on human faces. To deal with this problem, modulating the enhancement gain with the skin likelihood is proposed. Spatial correlation is imposed on the likelihoods to reduce isolated skin and non-skin pixels. Furthermore, a light level of smoothing for the face is also proposed while enhancing other parts of the image.

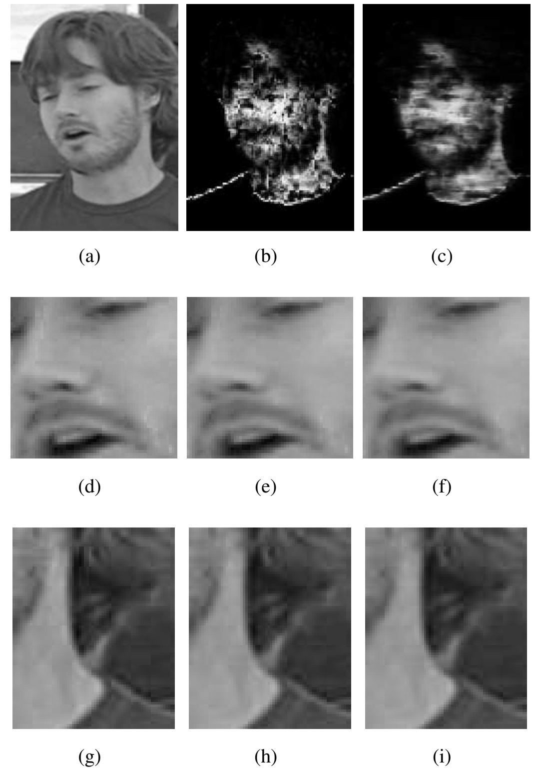

Fig. 2. (a) Original image, (b) Uncorrelated skin likelihood map, (c) Correlated skin map, (d) LCE of face, (e) SALSA of face, (f) SALSA with smoothing of face (k=0.75), (g) LCE of neck, (h) SALSA of neck, (i) SALSA with smoothing of neck (k=0.75)

6. REFERENCES

[1] Mark J. T. Smith and Alen Docef, *A Study Guide for Digital Image Processing*, Scientific Publishers, Georgia, 1999.

[2] T. Arici and Y. Altunbasak, "Image local contrast enhancement using adaptive non-linear filters," in *IEEE International Conference on Image Processing (ICIP-06)*, Oct. 2006, pp. 2881–2884.

[3] J. H. Langlois and L. A. Roggman, "Attractive faces are only average," *Psychological Science*, vol. 1, pp. 115–121, 1990.

[4] T. Rzeszewski, "A novel automatic hue control system," *IEEE Transactions on Consumer Electronics*, vol. 21, pp. 155–162, 1975.

[5] L. Gu and D. Bone, "Skin color region detection in MPEG video sequences," in *10th International Conference on Image Analysis and Processing (ICIAP'99)*, 1999, pp. 898–903.

[6] S. L. Phung, D. Chai, and A. Bouzerdoum, "Adaptive skin segmentation in color images," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, Apr. 2003, vol. 3, pp. 353–356.

[7] J. B. Martinkauppi, M. N. Soriano, and M. V. Laaksonen, "Behavior of skin color under varying illumination seen by different cameras at different color spaces," in *Proc. SPIE ,Machine Vision Applications in Industrial Inspection IX*, Apr. 2001, vol. 4301, pp. 102–112.

[8] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proc. Graphicon*, Sept. 2003, pp. 85–92.

[9] C.Y. Jeong, J.S. Kim, and K.S. Hong, "Appearance-based nude image detection," in *17th International Conference on Pattern Recognition (ICPR-04)*, Aug. 2004, vol. 4, pp. 467–470.

[10] A. Senior, R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 696–706, 2002.

[11] B. Menser and M. Wien, "Segmentation and tracking of facial regions in color image sequences," in *SPIE Visual Comm. and Image Processing (VCIP-00)*, June 2000, vol. 4067, pp. 731–740.

[12] J. Ahlberg, "A system for face localization and facial feature extraction," Report no.LiTH-ISY-R2172, Linkoping University.