

AUTOMATIC IMAGE QUALITY IMPROVEMENT FOR VIDEOCONFERENCING

Cuizhu Shi^{*1}, Keman Yu², Jiang Li² and Shipeng Li²

¹Inst. of Image Communication & Information Processing, Shanghai Jiaotong Univ., Shanghai, China
²Microsoft Research Asia, Beijing, China

ABSTRACT

In videoconferencing, the image quality is significantly affected by the illumination condition. Unsatisfactory illumination conditions may lead to underexposure or overexposure of the area of interest, in particular a human face. To resolve this issue, we propose a solution to automatically improve image quality by correcting exposure and enhancing contrast. Our work is characterized by a method for automatically building a skin-color model and a novel contrast enhancement approach. Some techniques that can reduce the computational cost are also introduced. Experimental results show that obvious improvement in image quality is achieved while the computation overhead is very small. The proposed solution can be integrated into videoconferencing systems and is especially suitable for scenarios where low-complexity computing is required.

1. INTRODUCTION

Due to the rapid development of network services and video compression techniques in recent years, videoconferencing applications have been widely used on PCs as well as dedicated conferencing terminals. In videoconferencing, the image quality is not only determined by the available network bandwidth and video compression techniques, but also significantly influenced by the illumination condition. Although most cameras, such as PC digital cameras, can automatically adjust the exposure level according to the illumination condition, they cannot identify the area of interest in an image, often a human face. Therefore, the automatically adjusted exposure level may be unsatisfactory in the area of interest.

Figure 1 shows two typical cases. Image (a) represents the scenes that possess a bright background. In this image, the face area is obviously underexposed because the camera's exposure level is influenced by the bright background. On the other hand, a dark background also affects the performance of the automatic exposure, which is manifested by image (b) with an overexposed face.



Figure 1: An underexposed face (a) and an overexposed face (b).

Therefore, it is meaningful to correct the exposure level based on the area of interest to improve the image quality.

However, the existing widely used videoconferencing tools, such as NetMeeting, do not provide such an enhancement.

To automatically improve the quality of digital images, many algorithms have been developed. The automated global enhancement method [3] identifies the visually important regions within a still image and then applies re-exposure based on these regions. The adaptive exposure correction method [4] is designed for video applications. This method consists of three phases. First, a skin-color model is built offline from a training data set. Second, human face areas are detected using the model at run-time. Third, an exposure correction process is carried out on the entire image based on the intensity information of the face area.

Some issues appear when the method of [4] is applied in practice. First, since the accurate human skin-color distribution vastly varies in different scenes, there is not a predefined model that is universally effective in various environments [5]. Second, we found that unsatisfactory illumination conditions also result in blurry images, and this effect cannot be effectively eliminated by correcting exposure alone.

We propose a solution to automatically improve image quality in videoconferencing. The procedure is similar to that of [4] but differs from it in two aspects: i) the skin-color model is automatically built at run-time to adapt to scene changes, and ii) a novel contrast enhancement method is integrated into the exposure correction process to further eliminate the blurry effect. Moreover, some techniques that can reduce the computational cost are developed.

The rest of the paper is organized as follows. Section 2 describes the method of building a skin-color model and detecting a human face. In Section 3, the exposure correction method and the contrast enhancement technique is presented. Some experimental results are shown in these two sections. Section 4 introduces techniques we use to speed up the entire algorithm. Finally, we conclude this paper and give future directions in Section 5.

2. BUILDING SKIN-COLOR MODEL AND DETECTING HUMAN FACE

2.1. Automatically building skin-color model

Among various face detection techniques which use different information, including color, shape, texture and motion, we take a color-based approach because of its simplicity and robustness. The histogram [6] and Gaussian models [4] [7] are two commonly used color-based methods. We employ a Gaussian model for its good convergence characteristic. We use a two-dimensional Gaussian model to describe the distribution of

* The work presented in the paper was carried out at Microsoft Research Asia.

human skin colors. Color space YCbCr is used in our work. The center of the Gaussian model is defined by $\mu = (\mu_{Cb}, \mu_{Cr})^T$, which is the mean vector for skin pixels, and its shape is determined by the covariance matrix Σ . The conditional probability $P(X|\mu, \Sigma)$ of a pixel with color vector X belonging to the skin color class is defined as follows.

$$P(X | \mu, \Sigma) = \frac{\exp[-\frac{d(X)}{2}]}{(2\pi) \times |\Sigma|^{-\frac{1}{2}}} \quad (1)$$

where $d(X)$ is the Mahalanobis distance from X to μ and is defined as:

$$d(X) = (X - \mu)^T \Sigma^{-1} (X - \mu) \quad (2)$$

It has been widely accepted that human skin colors cluster in a small range in the chrominance space. However, it also has been found that a predefined Gaussian model is not always effective in various environments [5]. In particular, the skin-color model derived from a person cannot accurately represent the skin color distribution for others. Moreover, when the scene changes, which commonly occurs in mobile video communication, the model for the same person also becomes unreliable. Therefore, an adaptive skin-color model is required.

Generally, the Gaussian model is built from a manually selected pixel set, but in our solution the skin-color model is built at run-time using automatically selected training data. In general, the pixels that have a vector X close to the center of Gaussian distribution μ are highly probable to be skin pixels. We design a two-step method to build the required model. First, we collect many test sequences for various persons and environments, and manually select some skin pixels from each sequence to calculate the Gaussian distribution center respectively, and then use the results of the collected sequences to compose a range of skin colors, e.g. [100, 135] for μ_{Cb} and [125, 160] for μ_{Cr} . Second, we use this range to select training pixels and build a Gaussian model at run-time.

In order to provide a correct guide for the face detection process, we send the model to a reliability examination phase before applying it to face detector. The examination method is based on the observation that $\mu_{Cr} + \mu_{Cb}$ and $\mu_{Cr} - \mu_{Cb}$ also cluster in a small range. The statistical results of 65 test sequences are shown in Figure 2. These sequences are selected from standard MPEG-4 test video clips and video clips captured from various real scenes using PC cameras. To improve the effectivity of the examination method, we slightly enlarge the ranges shown in this figure. Then, we use the ranges, [230, 280] for $\mu_{Cr} + \mu_{Cb}$ and [4, 45] for $\mu_{Cr} - \mu_{Cb}$, to examine the models. If a model's center falls out of these ranges, it is regarded as unreliable and will not be applied in the face detector. We will introduce how to handle this case in the next section. In practice, this examination method is efficient in distinguishing between reliable and unreliable models.

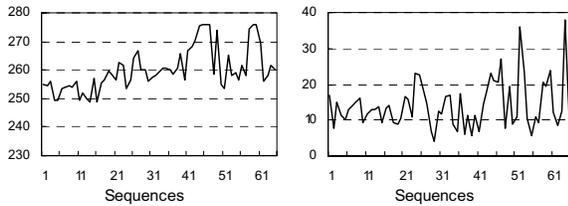


Figure 2: The range of $\mu_{Cr} + \mu_{Cb}$ (a) and $\mu_{Cr} - \mu_{Cb}$ (b).

2.2. Detecting human face

Once a reliable skin-color model has been created, it can be applied to the human face detection process. In this process, all pixels in the input image are scanned. A pixel is regarded as a skin pixel only if its chrominance vector X satisfies the following criterion.

$$P(X | \mu, \Sigma) > THS_P \quad (3)$$

where THS_P is a threshold for the probability of a pixel being a skin pixel. This criterion can be represented by:

$$d(X) < (-2) \log(THS_P \times 2\pi \times |\Sigma|^{-\frac{1}{2}}) = THS_DIST \quad (4)$$

where THS_DIST is a Mahalanobis distance threshold. Because this value is fixed for an image, it can be calculated before the scan process.

After finishing the scan process, we can obtain the mean luminance value (referred to as *AvgGray*) of those pixels that are regarded as belonging to the face. Consequently, this value can be used as reference in the exposure correction process. To avoid flickers that may be caused by large differences of *AvgGray* between adjacent images, we limit the differences in a moderate range, ± 8 for instance.

To correctly guide the exposure correction process, we further examine the detection results with the following criteria.

$$\frac{1}{Num_{face\ pixels}} \sum [d(X)]^2 < THS_DIST^2 \times 0.75 \quad (5)$$

$$Num_{face\ pixels} > \alpha \times Num_{all\ pixels} \quad (6)$$

where Σ represents the sum rather than the covariance matrix, X is the vector of the selected facial pixels, α is set to the minimal acceptable percentage of the face regions within an image.

If the detection results cannot pass the examination, or if the skin-color model is regarded as invalid in the previous phase, the calculation for *AvgGray* is skipped, and then the *AvgGray* of the previous image will be adopted in the exposure correction process. Since the adjacent images in a video clip are relatively similar in general, it is highly probable that the previous result is more reliable than the current one that did not pass the examination.

Figure 3 shows that the proposed method efficiently detects human faces in different sequences that are selected from MPEG-4 test video clips and clips captured from real scenes.

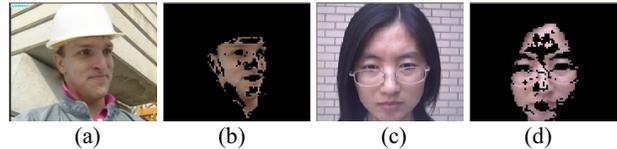


Figure 3: Original images (a) (c) and detection results (b) (d).

3. EXPOSURE CORRECTION AND CONTRAST ENHANCEMENT

In this section, we first introduce a general exposure correction process, and then describe a novel contrast enhancement method and how to integrate it with the exposure correction process. The *AvgGray* acquired in the face detection phase is used as a reference point in the correction process.

3.1. Exposure correction

An exposure-density function described in [3] [4] can be used to simulate a camera's activities. It estimates how incoming light intensities S (referred to as exposure) are transformed by the camera's sensor to pixel values I as shown in the following equation.

$$I = f(S) = \frac{255}{1 + e^{-A \times S}} \quad (7)$$

where the constant A is used to control the contrast level. Two curves shown in Figure 6 (a) demonstrate the relationship between S and I with different values of A .

The key idea of the exposure correction method is to adjust the average exposure of the area of interest towards the ideal exposure, which corresponds to the gray level of 128, and then compute the final pixel values using Equation (7) with corrected exposures. The correction process is performed on the entire image to preserve the image's harmony in the following way.

(1) Calculate the difference between the ideal exposure and the average exposure in the area of interest.

$$Diff = f^{-1}(128) - f^{-1}(AvgGray) \quad (8)$$

(2) For each pixel, re-expose it as follows.

$$S = f^{-1}(I) + Diff \quad (9)$$

$$I' = f(S) \quad (10)$$

As shown in Figure 4, the quality of the typical underexposed image (a) and overexposed image (b) is obviously improved. It shows that the above process is efficient in correcting the exposure of the face and preserving the image's harmony. Moreover, Figure 5 shows the transform of the gray scale histogram from image (c) to (d) in Figure 4. We can see that the majority of the gray values of the face area and the entire image are transformed towards the mid-tone. This is consistent with our expectation.

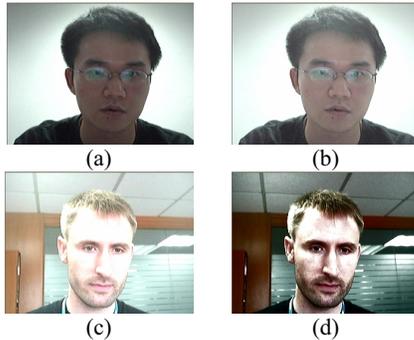


Figure 4: Underexposed image (a), overexposed image (c) and their corrected results (b) (d).

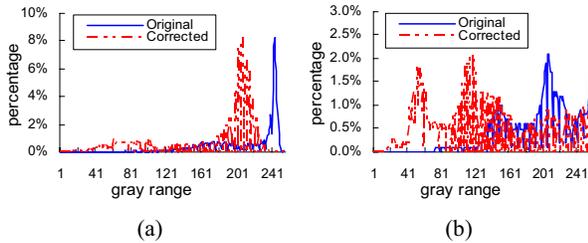


Figure 5: Histogram transform for face area (a) and the entire image (b).

3.2. Contrast enhancement

In practice, we found that unsatisfactory illumination conditions can also cause a blurry effect, and this issue cannot be effectively solved by only correcting exposure. We propose a novel contrast enhancement approach to further eliminate the blurry effect.

The proposed approach embodies the principle of gray scale stretching in digital image processing. It differs from the previous techniques, such as the curvelet transform method [1] and the genetic algorithm [2], in two aspects: i) it re-exposes the image using the exposure-density function to achieve enhancement in contrast, and ii) it can be integrated with the exposure correction process.

The contrast enhancement process described below is individually performed following the exposure correction process, and how to integrate these two processes will be introduced in next section. As mentioned in last section, the constant A is used to control the contrast level. The proposed approach is represented by the following equations.

$$S = f_1^{-1}(I) \quad (11)$$

$$I' = f_2(S) \quad (12)$$

where f_1^{-1} and f_2 are the reverse and forward computation of Equation (4). f_1^{-1} and f_2 use A_1 and A_2 as their parameters respectively, and A_1 is smaller than A_2 . Figure 6 (a) shows two exposure curves with different values of A , $A_1 = 0.85$ and $A_2 = 1.20$. Equation (11) and (12) correspond to mapping points in the A_1 curve to the points in the A_2 curve in the exposure field.

Figure 6 (b) illustrates how original gray values I are transformed to the final pixel values I' using the above two equations. From this figure, we can see that the slope of the middle region is larger than that of the saturation regions at both ends. Consequently, the gray scale range of the middle region is stretched. The gray values of the area of interest which has been corrected are exactly in this region.

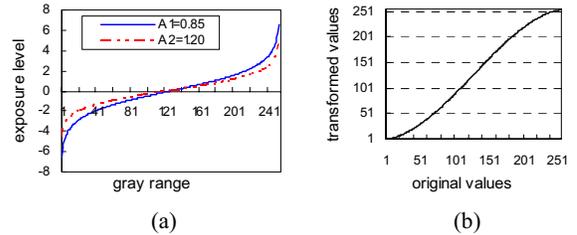


Figure 6: Exposure curves with different A (a) and gray value transform (b).

Figure 7 gives a subjective evaluation of an image captured in a real scene using a PC digital camera. We can see that the exposure corrected image (b) looks better than the original image (a) but still has obvious blurry effects, while the enhancement result (c) looks much clearer and sharper than the other two images, thus the visual effect is also better.

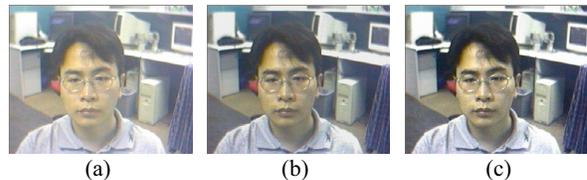


Figure 7: Original image (a), exposure correction result (b) and contrast enhancement result (c).

Next, we evaluate the objective enhancement results using the Detail Variances (DV) and the Background Variances (BV) as contrast measures. DV and BV are obtained in the following way [2]. First, for each pixel in the image, the variance of the gray values in the neighboring pixels within a 5×5 area are calculated. Second, if the variance is larger than a threshold, e.g. 100 here, the pixel is classified to the foreground; otherwise it is classified to the background. Then, DV is defined as the average variance of the foreground pixels and BV is the average variance of the background pixels. The contrast enhancement is regarded as effective when DV is increased while BV is not significantly changed. Table 1 compares the DV and BV obtained from the original images, the exposure corrected images and the contrast enhanced results. The results show that the proposed method is efficient in enhancing contrast.

Table 1: Comparison of DV and BV.

Images	Original		Corrected		Enhanced	
	DV	BV	DV	BV	DV	BV
Fig. 1 (a)	148.01	0.65	170.20	0.69	204.67	0.80
Fig. 1 (b)	173.62	0.96	309.97	1.05	366.58	1.40
Fig. 3 (a)	251.53	0.95	288.75	0.95	341.73	1.16
Fig. 3 (c)	217.48	1.17	292.65	1.15	372.17	1.42
Fig. 7 (a)	68.58	0.99	86.06	1.00	139.07	1.99

3.3. Integrating exposure correction and contrast enhancement methods

By combining Equations (9), (10), (11) and (12), we get:

$$S = f_1^{-1}(I) + Diff \quad (13)$$

$$I' = f_2(S) \quad (14)$$

The final results are the same as performing exposure correction and contrast enhancement individually, while computation is substantially reduced.

4. COMPLEXITY REDUCTION

In general, reducing the complexity of a computing process is significant when the computing process is used in real-time video communications. In this section, we introduce some techniques that can speed up the execution of the entire algorithm.

In the experiments of applying the entire algorithm without complexity reduction to a typical underexposed sequence and overexposed sequence (illustrated by Figure 1) in QVGA format, the average frame rates achieved on a PC with a 1.4 GHz Pentium 4 processor are only about 28 fps. We investigated the computation distribution of the entire algorithm and found that the integrated exposure correction and contrast enhancement processes (Equation (13) and (14)) consumed the majority of the execution time. Thus, the primary goal is clearly to accelerate the computation of the two equations (13) and (14).

Equation (13) can be greatly accelerated by simply using a lookup table, whereas the computation of Equation (14) cannot be replaced by directly indexing a table, because S is obtained by the nonlinear Equation (13). However, because f_2 is monotonic, i.e. I' increases with S , the sign of $Diff$ provides the correction direction. We create another table using f_2^{-1} , and then use input I as the starting point to search in the lookup table in the direction indicated by the sign of $Diff$. The index value corresponding to

the element that is most approximate to S is just the result we want. The result usually can be found by a few comparisons. Moreover, a table mapping I to I' can be established at run-time, so that the above process can be replaced by directly indexing this table. By using these tables, the complexity of Equation (13) and (14) is significantly reduced.

Taking the temporal similarity of video clips into account, skin-color model building and face detection can be performed in a defined interval, such as every five frames. Besides this, if the *AvgGray* acquired in face detection is very close to the ideal level, falling in the range of [124, 132] for instance, the remaining processes can be skipped.

In this way, we can process more than 800 frames per second in the above experimental condition.

5. CONCLUSIONS

We propose a solution to automatically improve image quality in videoconferencing.

Our solution consists of three phases. First, a skin-color Gaussian model is automatically built at run-time to provide adaptations to scene changes. Second, human face areas are detected using the model. Third, an integrated exposure correction and contrast enhancement process is carried out based on the gray information of the face areas. Moreover, some techniques are developed to reduce the computational cost. Experimental results show that obvious improvement in image quality, especially for the areas of interest, is achieved while the computation overhead is very small. The proposed solution can be integrated into videoconferencing systems and is especially suitable for scenarios where low-complexity computing is required.

Future directions may include refining the skin-color model and developing face detection methods that provide more accurate and reliable guides for the successive processes.

6. ACKNOWLEDGMENT

The authors thank Steve Lin for proofreading the paper.

7. REFERENCES

- [1] J.L. Starck, F. Murtagh, E.J. Candès, and D.L. Donoho, "Gray and Color Image Contrast Enhancement by the Curvelet Transform", *IEEE Trans. On Image Processing*, vol. 12, no. 6, pp. 706-717, Jun. 2003.
- [2] F.Saitoh, "Image contrast enhancement using genetic algorithm", *Proc of IEEE International Conference on SMC 1999*, vol.4, pp. 899-904, Oct. 1999.
- [3] S.A. Bhukhanwala and T.V. Ramabadram, "Automated Global Enhancement of Digitized Photographs", *IEEE Trans. On Consumer Electronics*, vol. 40, no. 1, 1994.
- [4] Messina, G., Castorina, A., Battiato, S. and Bosco, A., "Image quality improvement by adaptive exposure correction techniques", *Proc. of ICME 2003*. pp. 549-552, July 2003.
- [5] M. Soriano, B. Martinkauppi, S. Huovinen and M. Laaksonen, "Adaptive skin color modeling using the skin locus for selecting training pixels", *Pattern Recognition* 36 (2003), 681-690.
- [6] M.J. Jones and J.M. Rehg, "Statistical color models with application to skin detection", *IEEE Proc. of CVPR 1999*, pp. 274-280, Jun. 1999.
- [7] M.H. Yang and N. Ahuja, "Detecting human faces in color images", *Proc. of ICIP 1998*, pp. 127-130, Oct. 1998.