



Provided by the author(s) and NUI Galway in accordance with publisher policies. Please cite the published version when available.

Title	Facial enhancement and beautification for HD video cameras
Author(s)	Corcoran, Peter
Publication Date	2011
Publication Information	Florea, C. and Capaùtaù, A. and Ciuc, M. and Corcoran, P. (2011) Facial enhancement and beautification for HD video cameras Consumer Electronics (ICCE), 2011 IEEE International Conference on
Publisher	IEEE
Link to publisher's version	http://dx.doi.org/10.1109/ICCE.2011.5722839
Item record	http://hdl.handle.net/10379/3623

Downloaded 2022-08-14T21:12:19Z

Some rights reserved. For more information, please see the item record link above.



Facial Enhancement and Beautification for HD Video Cameras

Corneliu FLOREA^{a,b}, Adrian CAPĂȚĂ^a, Mihai CIUC^{a,b}, and Peter CORCORAN^c

^aTessera România, Calea Șerban-Vodă nr. 133, Sector 4, Bucharest, Romania

^bUniversitatea "Politehnica" București, Bd. I. Maniu 1-3, Bucharest, Romania

^cNational University of Ireland Galway, College of Engineering & Informatics, Galway Ireland

Abstract – Influenced by the widespread adoption of HDTV modern consumer imaging devices have begun to feature full HD video capture. An interesting social consequence is that many small imperfections in human faces are captured with full-HD video. When used for portrait imaging the resulting video is frequently unsatisfactory to the user. Hence, manufacturers of imaging devices require practical, real-time solutions which can mitigate facial details emphasized at HD resolution while preserving the overall quality of face-portrait images. The high bandwidth and processing requirements of full-HD video make this a challenging task. In this paper we explain a practical algorithm, now incorporated in a number of consumer devices. Major challenges and their solutions are explained. Attention is given to real-time optimizations of the algorithm which is designed with a view to its suitability for partial or full hardware implementation.

INTRODUCTION

In the last few years there has been a widespread adoption of flat-screen HD TV technology by the consumer. In many cases these TVs are capable of 1080p full-HD video display even where a full-HD TV service is not yet available. Given these developments, the manufacturers of many imaging devices have started to introduce full-HD video capture to differentiate their products, particularly at the high-end of the market. In theory such features are very desirable to the consumer and enable them to derive benefit from their full-HD TV set even where full-HD TV service is still absent.

In practice, however, the level of facial detail which is captured by full-HD video when faces are close to the imaging appliance is noticeably enhanced over 720p video. Furthermore the ability of such imaging devices to optimize image capture for the face region enables a very detailed capture of the "true" appearance of the face. Most consumers, however, would prefer a gentler presentation of their "true" appearance which is kinder to the many small imperfections of a typical face. In fact most of us have become accustomed to the lower resolution of conventional video which provided an inherent "soft focus".

Now the device manufacturers who added full-HD to their devices are faced with the problem of adapting the capture of facial regions to suit the expectations of consumers. The high bandwidth and processing requirements of full-HD video introduce a further complication which makes this quite a challenging problem. This paper presents a new approach to skin enhancement for full HD video. Section II explains the main approaches to facial enhancement; in Section III a brief presentation of the proposed method for video skin enhancement is given; Section IV presents preliminary results and evaluation tools.

A. Approaches to Facial Enhancement

Various techniques of facial enhancement are well known from the professional photography industry. Typically two

major possibilities are generally considered. A first category refers to geometrical warping faces such as to increase the attractiveness of the face look [1]. A second category deals with enhancing the face skin appearance, i.e., concealing undesired skin flaws (e.g., blemishes, pimples, red-spots etc.), attenuating wrinkles, and giving the skin a smoother, thus younger look, while preserving main features, e.g., eyes, mouth, eyebrows [2].

In a practical realization geometrical face warping requires the determination of numerous control points that define the exact locations of important features (e.g., eyes, nose, mouth, chin, cheeks, hairline, etc.). Also, there is significant computational cost, associated with the geometrical operations and re-rendering of the face region. Finally we note that the memory requirements to store a full resolution facial region and the memory bandwidth to support the required image manipulations imply a dedicated face-processing subsystem of similar costs and complexity to the main image acquisition subsystems. Eventually we may be able to parallelize such algorithms in hardware, but for today's imaging devices this approach is impractical and we will not consider it further.

On the other hand, skin enhancement is amenable to an automated implementation, thanks to the powerful face detection and skin tone identification algorithms existing in today's digital cameras. These algorithms are already optimized for speed and have a relatively low computational cost [3].

VIDEO FACE BEAUTIFICATION

The most important constraint imposed to a solution dedicated to video enhancement is to process the whole amount of data (30-60 frames per second at 1920 x 1080 resolution) in real time, in an embedded device. Such a huge task may be dealt with only if the processing is accelerated by using application specific integrated circuits (ASIC). In order to keep the circuit area low and to avoid extra usage of the memory bandwidth, all operations must use very light processing on neighboring pixels.

A typical portrait-oriented skin beautification algorithm consists of the following steps [3]:

- Face detection and tracking
- Skin detection
- Skin enhancement

An efficient algorithm must use pixel-oriented solutions for all of these intermediate parts.

Face detection and tracking modules signal if there are faces in an image and provide face rectangles which contain face pixels. There are various solutions in the literature but practically all state-of-art face detectors employ principles originally developed by Viola and Jones [4].

For the skin detection step, we propose a two-stage

process. The first step involves a rough analysis with simple conditions. One suitable solution is the set of conditions proposed by Peer et. al [6]. A pixel being described in terms of its R , G , and B components is labeled as a skin pixel if:

$$(R > 95) \& (G > 40) \& (B > 20) \& (\max\{R, G, B\} - \min\{R, G, B\} > 15) \& (|R - G| > 15) \& (R > G) \& (R > B) \quad (1)$$

For the second stage we use thresholding on the PCA principal axes derived from skin color, see [3] for details. This technique has increased computational burden so processing is distributed over two or more consecutive video frames. In the first frame of a sequence, skin pixels inside the face rectangle are selected using Eq. (1). The mean and covariance matrices of skin color are computed. The eigenvectors of the covariance matrix represent axes of a new 3D coordinates system. In the next frame, skin pixels, selected by Eq. (1), are projected on these axes and thresholds are computed to retain a certain percentage of pixels (e.g., 95%). In the last frame of a sequence, a refined skin map is built by representing all the frame pixels on the newest axes and comparing them with the most recently determined thresholds. Enhancement is then applied only for those pixels classified as skin.

The process of skin enhancement is mainly a smoothing process [6]. However the sharpness of the principle facial features (mouth, eyes) must be preserved. Thus smoothing is selective and more intensified in uniform areas. In practice an implementation of the Lee filter [7] preserves edges and thus facial features. The output value at some location, identified as skin, is computed as:

$$f_{out} = \left(\frac{\sigma_{local}}{\sigma_{skin}} \right)^2 \overline{f_{in}} + \left(1 - \left(\frac{\sigma_{local}}{\sigma_{skin}} \right)^2 \right) f_{in} \quad (2)$$

where f_{in} is the input pixel value, $\overline{f_{in}}$ is the average computed on the initial image on a squared window which is also used for local variance (σ_{local}). Finally, σ_{skin} is the local variance corresponding to a uniform area of skin. A set of typical results is shown in Figure 1.



Fig.1. Original frame (top left) and corresponding processed images: refined skin map (top right), Lee filter variation map (bottom left) and the final enhanced facial image.

RESULTS AND EVALUATION

The implementation of the solution we have described is split into two parts. The pixel oriented operations, such as skin classification, are suitable for hardware implementation whereas the frame statistics and corresponding PCA computations are better performed on a general purpose CPU, which can also coordinates the overall processing. With this hybrid architecture, real-time constraints for HD video can be met.

For qualitative analysis we used the following test metrics:

- Correction style and quality (subjective metric): using the original video as reference, we asked several test groups of observers to score the enhanced video on a scale from 0 (very bad) to 5 (very good) when compared with the original (with reference rating of 3). The mean opinion score was 4.6.
- Skin Correction Accuracy (objective metric): here we used a set of around 30 videos having the skin manually marked. The following error rates have been found:
 - The False Positive rate (i.e., the percentage of non-skin pixels wrongly “beautified”) is 2.04 %.
 - The False Negative rate (i.e., percentage of skin pixels not corrected) is 3.8 %.

Both the objective and subjective measures indicate that the quality provided by the present algorithm is good enough to satisfy reasonable demands for accuracy and quality.

There are still some constraints for this solution to work correctly: faces must be larger than 40 x 40 pixels, white balance levels must be accurate and if a frame is not correctly exposed then noise effects prevent this technique working correctly.

Further details and experimental tests will be given at the conference presentation and in a full length paper.

REFERENCES

- [1] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski, “Data-driven enhancement of facial attractiveness,” *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2008)*, vol. 27, Aug. 2008.
- [2] K. Arakawa, and K. Nomoto, “A system for beautifying face images using interactive evolutionary computing”, in *Proceedings of 2005 IEEE Intelligent. Symposium. on Intelligent Signal Processing and Communication Systems, ISPACS 2005*, p. 9-12, Dec. 2005.
- [3] M. Ciuc, A. Capata, V. Mocanu, C. Florea, A. Pososin, and P. Corcoran “Automatic Face and Skin Beautification Using Face Detection”, U.S. Patent, No. 20100026831A1.
- [4] P. Viola, and M. Jones, “Robust Real-Time Face Detection” in *International Journal of Computer Vision*, vol. 57, No. 2, May 2004, p. 137-154.
- [5] P. Peer, J. Kovac, and F. Solina, “Human skin color clustering for face detection”. Submitted to EUROCON 2003 - International Conference on Computer as a Tool 2003.
- [6] M. Ciuc, A. Capata, C. Florea, “Objective measures for quality assessment of automatic skin enhancement algorithms”, in *Proc. of IS&T/SPIE EI 2010, San Jose, U.S.A, January 2010, Paper 7529-23 of Conference 7529*.
- [7] J.S. Lee “Digital image enhancement and noise filtering by use of local statistics” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 2, pp. 165–168, Mar 1980.