A Global Enhancement Pipeline For Low-cost Imaging Devices

Sebastiano Battiato, Alfio Castorina, Mirko Guarnera and Paolo Vivirito[1]

Abstract — *The paper describes a suitable algorithms pipeline able to enhance the picture quality in terms of both measured and perceived quality. The overall pipeline is mainly devoted to improve image acquired by low cost imaging sensors, typically present in consumer devices (i.e. mobile phone, web-cams, PDA, etc). A series of ad-hoc heuristics and techniques have been applied, taking into account main compression artifacts, chromatic and geometric distortions, etc.. Experimental results show effectiveness of the proposed pipeline.*

Index Terms — Adaptive sharpness, exposure correction, de-blocking, JPEG, white balancing.

I. INTRODUCTION

Typical consumer devices acquire digital images by properly using digital sensors (e.g. CCD/CMOS). Quality improvement is obtained by increasing the resolution of the sensor or by using more sophisticated image-processing algorithms, [1], [2], [3] and [4]. The overall IGP (Image Generation Pipeline) aimed to reconstruct the final high quality image is complex and is devoted to exploit all the information acquired by sensor to achieve the "best" possible image [5]. For low cost imaging sensors (e.g. VGA, CIF resolution) the particular working conditions are not able to generate high quality images; several artifacts can be easily detected: compression artifacts due to the heavy compression, chromatic and geometric distortions and aberrations, color casting, etc.. In this paper a suitable global post-enhancement pipeline, able to overcome such problems, is proposed. Each algorithm is applied sequentially and has been properly tuned. Some of them use classical solutions, while new techniques are proposed for white-balancing, adaptive-sharpness, contrast adjustment and directional smoothing. Experiments over a large data set of images acquired using real low-cost sensors have showed the real improvement obtained in term of perceived quality.

The rest of the paper is organized as follows: section 2 describes in detail each step of the proposed pipeline; section 3 shows experimental results and the effectiveness of the proposed techniques, while section 4 closes the paper tracking directions for future works.

[1] All the authors are with the STMicroelectronics in the Advanced System Technology Division (AST Lab – Catania), 95121 Catania , ITALY - E-mail: {Sebastiano.Battiato,Alfio.Castorina,Mirko.Guarnera, Paolo.Vivirito}@st.com).

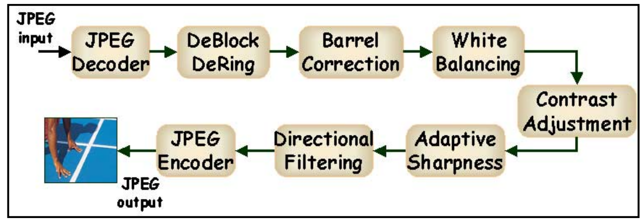


Figure 1 - Schematic description of the proposed global enhancement pipeline.

II. THE OVERALL PIPELINE

In the following the technical details, relative to each one of the enhancement steps used in the proposed pipeline, are described (see Figure 1).

A. De-blocking and De-ringing

Classical JPEG compression engines split input images in 8x8 non-overlapping blocks, each of them is subjected to a DCT transform followed by quantization which is an irreversible information loss of the original image content: thus unpleasant artifacts are not unlikely to appear. This becomes most evident in band constrained information transmission (e.g. digital images acquired by mobile engines) where low bit rate working modes are desirable and realized through coarse quantization. Main JPEG compression artifacts are blocking and ringing. The former originates from independent quantization of adjacent blocks irrespective of existing spatial correlations between them and is visually characterized by pseudo-edges across blocks boundaries while the latter derives from damaged mid-frequency coefficients and results in noticeable distortions around image edges. In order to attenuate both artifacts, filters derived from the post-processing algorithms adopted in the MPEG4 [6] codec engine specifically were adapted to work on JPEG compressed images. Main differences regarded the de-blocking filter where this meant operating on a block by block basis, skipping over the macroblock subdivision of MPEG sequences to evaluate the specific content and classify blocks as smooth and non smooth. Classification was done as described in [6] but discrimination thresholds were referred to the quantization step adopted on the DC coefficient instead of the QP factor. Smooth blocks were lowpass filtered across boundaries while

Manuscript received June 18, 2003

Figure 2 - (a) distorted image; (b) barrel corrected image (some distortion lines are plotted over the image).

Figure 3 - (a) input image; (b) white balanced image.

for non-smooth blocks (where details must be preserved) the DC offset mode was used.

B. Geometric Distortion

Usually low cost cameras exhibits significant lens distortion, especially radial distortion. Geometric distortion describes the difference between the observed and real position of a point in the space. Geometric distortion is typically modeled as a 5-vector *k* plus an offset vector that accounts for "de-centering" of the lens and other mechanical components. The distortion function is however dominated by radial components: thus it is sufficient to use the first two terms only (see [7] for more details). The radial symmetrical distortion can be corrected using a suitable polynomial relation:

$$\breve{x} = x(1+k_1r^2+k_2r^4), \qquad (1)$$
$$\breve{y} = y(1+k_1r^2+k_2r^4),$$

where $k1>>k2$, (x, y) and $(\breve{x}, \breve{y})$ are the observed and ideal corrected image coordinates. Results can be further improved employing bilinear interpolation.

C. Automatic White Balancing

The role of white balancing is to compensate color-cast problems in scenes taken under non-white illumination. The algorithm here adopted can be seen as an improved variation from gray world algorithms [8], where the three channel energies are forced to be equal. Channel energies ce_c are computed by means of the following:

$$ce_c = \frac{\sum_{i=0}^{N-1} p_{ci} \cdot \min(p_{ri}, p_{gi}, p_{bi}) \cdot \alpha(T, p_{ri}, p_{gi}, p_{bi})}{\sum_{i=0}^{N-1} \min(p_{ri}, p_{gi}, p_{bi}) \cdot \alpha(T, p_{ri}, p_{gi}, p_{bi})} \quad c \in \{r, g, b\}, \qquad (2)$$

where p_{ci} is the value of the *c*-channel of pixel *i*, *N* is the number of total pixels, and a is a 4-ary operator whose value is 0 if all channels are greater than an user fixed threshold *T*, 1 otherwise. Using the weighted average with respect to minimum values de-correlated computing will be avoided, while the α discards useless near saturation values. Once the channel energies are computed, 3 multiplicative gains, g_r, g_g, g_b, are obtained in such a way to force the channel energies to be equal to the maximum one. Correction will be done as

$$p_{ci} = p_{ci} \cdot g_c \quad c \in \{r, g, b\}, \quad i = 0 \ldots N-1 \qquad (3)$$

Figure 3 shows an example of corrected image.

D. Exposure Correction

Acquiring scenes where automatic exposure settings are employed and no skilled settings are at hand, errors (i.e. wrong exposure) are not unlikely to occur: thus a post-processing step to try to further improve the already acquired image would be desirable. The algorithm adopted, a loose variation of that described in [9], employs a feature driven correction curve. The term feature driven stands for the tone adjustments based on the specific relevant image content. Of course there's no universally accepted definition of what the relevant image content is, that's why two heuristically scenarios have been drawn:

1. zones where people are present, thus suggesting a portrait;

2. zones of high energy content, (i.e. focus, contrast) thus avoiding flat areas.

The two scenarios are mutually exclusive, in the sense that the image is analyzed and if a relevant percentage (following a user fixed threshold) are reliably classified as skin the first operating mode is used (see [10] for more details), otherwise the second one. Skin detection has been done using the techniques described in [11], [12]. When the second criterion is used image is first divided in *NxM* disjoint and equally sized blocks, and block energy content is estimated using mean laplacian value (e.g. focus) and histogram deviation (e.g. contrast). Since energy measures are usually higher in well-exposed regions, in order to avoid excluding bad exposed blocks, analysis is made at an intermediate overexposed level. Corrections are done forcing the average (*avg*) gray value of relevant regions (the skin pixels or the relevant blocks) to occupy mid gray tonal range (≅ 130).
Offset from desired exposure is computed as:

$$\Delta = f^{-1}(130) - f^{-1}(avg), \tag{4}$$

while correction of all pixels Y channel is done using:

$$Y'(x,y) = f(f^{-1}(Y(x,y)) + \Delta) \ . \tag{5}$$

In the afore described formulas *f* is a kind of camera response function that could be parametrically described and estimated using the techniques in [13]. Once Y has been corrected, changes are projected in the RGB space with the following:

$$\begin{aligned} R' &= 0.5\cdot\left(\frac{Y'}{Y}\cdot(R+Y)+R-Y\right) \\ G' &= 0.5\cdot\left(\frac{Y'}{Y}\cdot(G+Y)+G-Y\right) \\ B' &= 0.5\cdot\left(\frac{Y'}{Y}\cdot(B+Y)+B-Y\right) \end{aligned} \tag{6}$$

where *R,G,B,Y* are input values and *R',G',B',Y'* the corrected counterparts. More details on (6) can be found in [14].

E. Adaptive Sharpness

Low cost imaging sensors tend to exhibit, moving away from the center, an increasing degree of out-of-focus. In order

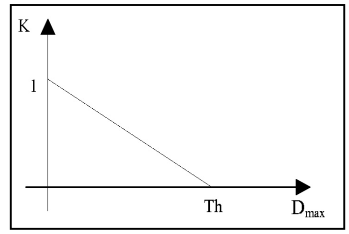


Figure 4 - Relationship between D_{max} and *k* coefficient.

to remove such blurring effect a simple linear combination between [15] and a classical un-sharp masking is used. Such joint approach reduces noise and ringing artifacts related to the linear equalization described in [15]:

$$PSE(u,v) = \frac{1}{OTF(u,v) * \left[1 + 0.414\left(\frac{p}{|\,OTF(u,v)\,|}\right)^2\right]} . \tag{7}$$

where *p* represents the 3 Db threshold of the restoration filter and *OTF(u,v)* is the "Optical Transfer Function" of the optics system. Taking into account the radial blurring effect the input image is split in several concentric areas. Each area is then processed with properly derived filters, (e.g. the taps number depends on the blur magnitude). Space invariant algorithms of edge enhancement are typically related to noise and ringing artifacts. It is possible to reduce side effects combining adaptively restored image with the original one. Since lens blur is a low-pass filter, low textured areas are little interested in blur process, then it is possible to obtain an image as a weighted combination of blurred and restored ones. More specifically, homogeneous areas can be selected from blurred image, while textured from restored one. Thus, the final "restored" image is obtained by the following relation:

$$F(x,y) = k * G(x,y) + (1-k) * R(x,y), \tag{8}$$

where *F* is resulting image, *G* is the original image, *R* is the linear filtered image and $0 < k < 1$ is locally derived taking into account the 8-neighborhood of the (*x, y*) pixel. In order to evaluate presence of texture we calculate differences between local pixel and its neighboring. Let be *Dmax* the maximum of local differences. Figure 4 shows relationship between *Dmax* and *k* coefficient, where *Th* is empirically fixed to 10. It's also possible to relate the value of *Th* to the luminance value of local pixel, using consideration closer to the Human Vision System.

F. Directional Smoothing

Even if the de-blocking and de-ringing processing is able to remove most disturbing compression artifacts, frequencies that are completely lost during quantization cannot adequately reproduced. This can somehow result in a zig-zagged appearance around image edges. Such unpleasant artifact is also introduced by "poor" sampling of low-resolution images. The adaptive sharpness filter

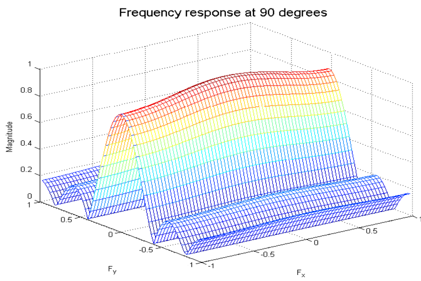


Figure 5 - Frequency response at 90 degrees.

described above, could in some cases, amplify these problems. In order to produce a more pleasant final appearance a directional smoothing filter was adopted, aimed to low-pass filter the image following edges orientations, thus avoiding excessive blurring. Edges orientations are identified using the following:

$$or(x, y) = \arctan\left(\frac{S_y(x, y)}{S_x(x, y)}\right), \quad (9)$$

where Sy and Sx are the vertical and horizontal Sobel filtered values at given pixel location. Each pixel is filtered around the prominent (max magnitude) direction in a 3*x*3 surround. Filtering is carried out by means of an elliptical shaped gaussian filter given by:

$$\begin{aligned} f(x, y, \alpha) &= he^{-\frac{\tilde{x}^2}{2\sigma_x^2}-\frac{\tilde{y}^2}{2\sigma_y^2}}, \\ \tilde{x} &= x\cos(\alpha) - y\sin(\alpha), \\ \tilde{y} &= x\sin(\alpha) + y\sin(\alpha), \end{aligned} \quad (10)$$

where σ_x, σ_y are the variances along the two dimensions, h is a normalization constant and α is the orientation angle (16 possible orientations were used). Figure 4 shows the frequency response of a kernel at 90 degrees. In order to avoid excessive filtering of fine details, effectiveness can be attenuated proportionally to orientation disorder in the same surround used for orientation detection (low disorder is expected around relevant edges and high disorder in textured areas). Disorder was computed using the absolute differences between the gradient orientations of the central pixel and its surround and expanding (by means of a *min* filter) the obtained values. Once disorder is computed filtering is carried out by means of the following blending:

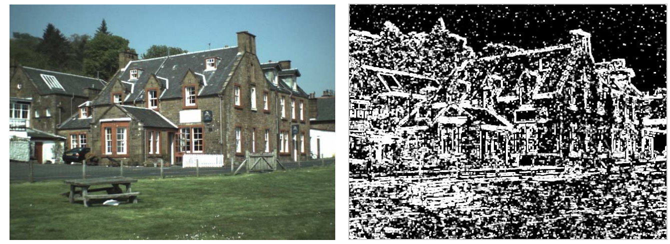


Figure 6 - Original image and the α values (*T1=2*, *T2=5*).

$$p_{out}(x, y) = \alpha \cdot p_{filtered}(x, y) + (1-\alpha) \cdot p_{unfiltered}(x, y) \quad (11)$$

where $p_{out}(x,y)$, $p_{filtered}(x,y)$ and $p_{unfiltered}(x,y)$ are respectively the final, filtered and original pixel value in position *(x,y)*, and α is computed with the following step-like function :

$$\alpha = \begin{cases} 1 & \text{if } \alpha < T1 \\ \max\left(0, \frac{T1 - d(x, y)}{T2 - T1} + 1\right) & \text{otherwise} \end{cases} \quad (12)$$

, where $d(x,y)$ is the disorder in position *(x,y)* and *T1,T2* could be used to model the filter strength. Figure 6 shows the α values at different locations for the image on the left. It can be seen that filtering has been more effective across relevant edges.

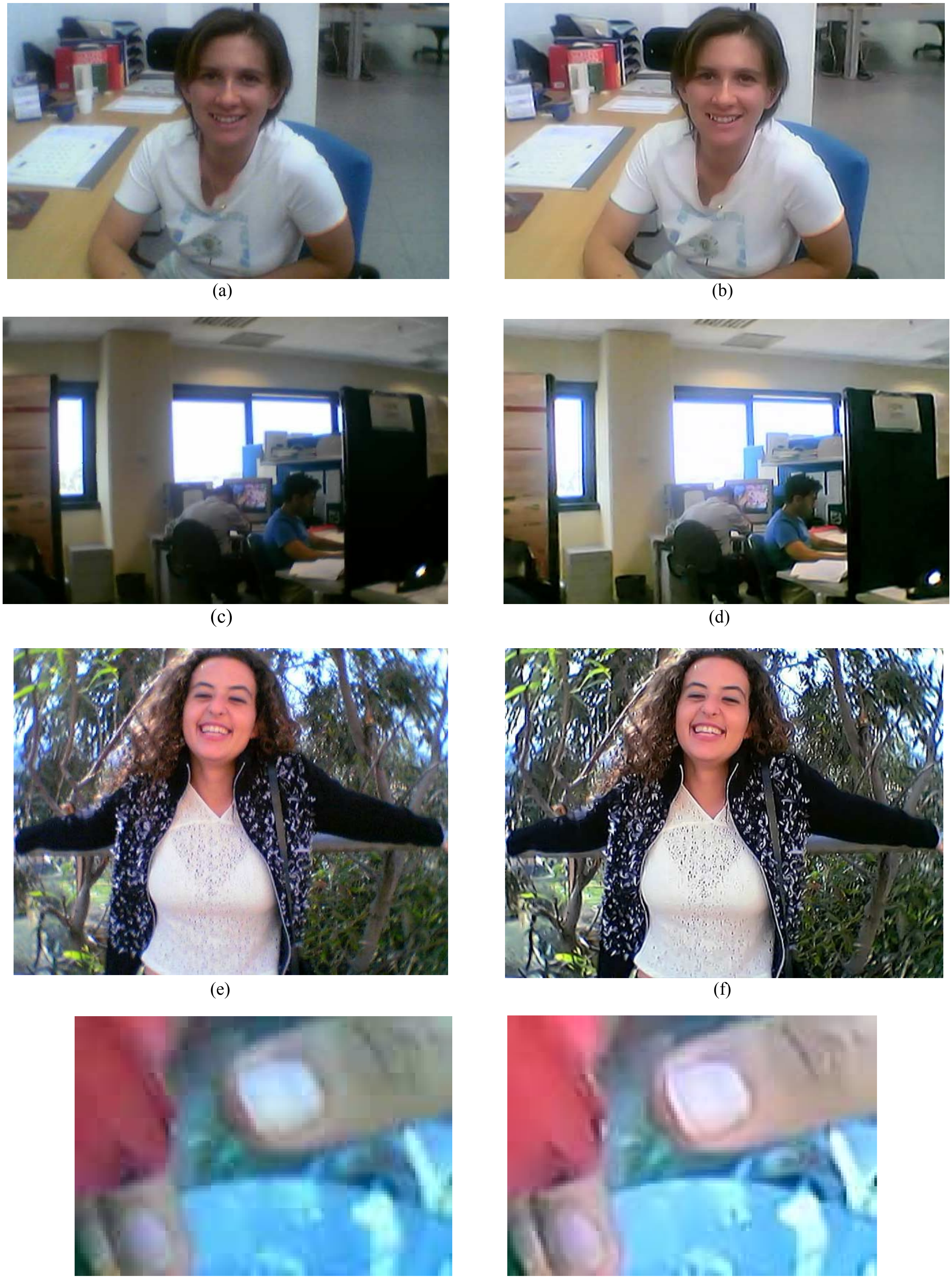
III. EXPERIMENTAL RESULTS

In order to evaluate the real performances of the proposed enhancement pipeline, a set of 300 images has been used. The input images acquired in different working conditions by a typical VGA sensor are compressed using the "high quality" JPEG setting of the corresponding device (a consumer mobile imaging phone) where the typical compression ratio obtained is about 30:1. Experiments confirm the improvement in term of perceived quality (Color, Contrast/Sharpness, Artifact removal, etc.). Some examples are reported in Figure 7.

IV. CONCLUSIONS AND FUTURE WORKS

A suitable algorithms pipeline able to enhance the picture quality in terms of both measured and perceived quality has been proposed. The overall pipeline is mainly devoted to improve image acquired by low cost imaging sensors, typically present in consumer devices (i.e. mobile phone, web-cams, PDA, etc...).

A series of ad-hoc heuristic and techniques have been applied, taking into account main compression artifacts, chromatic and geometric distortions, etc.. Experimental results, realized on a large dataset of input images, show effectiveness of the proposed pipeline. Future works will include the possibility to evaluate numerically the quality improvement obtained using some no-reference assessment quality metric [16].



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 7 - (a), (c), (e), (g) Images acquired by mobile phone camera; (b), (d), (f), (h) Output images.

REFERENCES

[1] A.Bosco, M. Mancuso, S. Battiato, G. Spampinato "Temporal Noise Reduction of Bayer Matrixed Video Data", Proceedings of IEEE ICME'02 International Conference on Multimedia and Expo 2002, pp.681-684, Lausanne, Switzerland, August 2002;

[2] A. Bosco, M. Mancuso, S. Battiato, G. Spampinato "Adaptive Temporal Filtering for CFA Video Sequences", Proceedings of IEEE ACIVS'02 Advanced Concepts for Intelligent Vision Systems 2002, pp. 19-24, Ghent University, Belgium, September 2002;

[3] S. Battiato, A. Castorina, M. Mancuso "High Dynamic Range Imaging: Overview and Application", Accepted for publication: SPIE Journal of Electronic Imaging, November 2002;

[4] G. Messina, S. Battiato, M. Mancuso, A. Buemi, "Improving Image Resolution by Adaptive Back-Projection Correction Techniques", IEEE Transaction on Consumer Electronics 2002, vol.48, no.3, pp.400-408, August 2002;

[5] M. Mancuso, S. Battiato, "An introduction to the digital still camera technology", ST Journal of System Research, Vol. 2, No. 2, pp. 1-9, December 2001;

[6] ISO/IEC JTC1/SC29/WG11 N 2502:Final Draft of International Standard MPEG-4;

[7] Z. Zhang. "Flexible Camera Calibration By Viewing a Plane From Unknown Orientations", International Conference on Computer Vision (ICCV'99), Corfu, Greece, pages 666-673, September 1999;

[8] A. Rizzi, C. Gatta, D. Marini "Color Correction between Gray World and White Patch", Proceedings of SPIE Electronic Imaging 2002, Vol. 4662, San Jose' CA USA, January 2002;

[9] S.A. Bhukhanwala, T.V. Ramabadran, "Automated Global Enhancement Of Digitized Photographs", IEEE Transactions on Consumer Electronics, Vol. 40, No. 1, 1994;

[10] S. Battiato, A. Bosco, A. Castorina, G. Messina, "Automatic Global Image Enhancement by Skin Dependent Exposure Correction", In Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing – NSIP 2003 – Grado, Italy - June 2003;

[11] B.D. Zarit, B.J. Super, and F.K.H. Quek, "Comparison of five colour models in skin pixel classification", Proc. Of Int. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, IEEE Computer Society, Corfu, Greece, pp. 58-63, 1999;

[12] J.Yang, W. Lu, and A. Waibel, "Skin-colour modeling and adaptation", Technical Report CMU-CS-97-146, School of Computer Science, Carnegie Mellon University, 1997;

[13] S. Mann, "Comparametric Equations with Practical Applications in Quantigraphic Image Processing", IEEE Transactions on Image Processing, Vol. 9, No. 8, 2000;

[14] S. Sakaue, A.Tamura, M.Nakayama, S.Maruno, "Adaptive gamma processing of the video cameras for the expansion of the dynamic range", IEEE Transactions on Consumer Electronics, Vol. 41, No. 3, pp.555-561, August 1995;

[15] P. Vivirito, S. Battiato, S. Curti, M. La Cascia, R. Pirrone, "Restoration of Out of Focus Images Based on Circle of Confusion Estimate", In Proceedings of SPIE 47th Annual Meeting 2002 – Applications of Digital Image Processing – Vol. 4790, Seattle, Washington, USA, July 2002;

[16] Z. Wang, A. C. Bovik, "A Universal Image Quality Index", IEEE Signal Processing Letters, Vol. 9, No.3, pp.81-84, March 2002.

Sebastiano Battiato received his Italian degree in Computer Science in 1995 and PhD in Applied Mathematics and Computer Science in 1999 both at Catania University. He attended several international workshops/conferences in the imaging field where some results of his research has been presented and then published in the relative Proceedings. A few papers have been also published in International Journals. Since November 1999 has been working at STMicroelectronics in AST Catania Lab in the Digital Still Camera Group as Algorithm Dev. Team Leader. His current research interests are in the area of Image Processing, Compression (Jpeg, Jpeg2000), Watermarking, and Analysis/Synthesis of Texture.

Alfio Castorina received his Italian degree in Computer Science in 2000 at the University of Catania doing a thesis about watermarking algorithms for digital images. Since September 2000 he has been working in STMicroelectonics in the AST Digital Still Camera Group as System Engineer. His current activities include high dynamic range imaging, post-processing image enhancement techniques and Jpeg2000.

Mirko Guarnera received his Italian degree in Electronic Engineering from the University of Palermo (Italy) in 1998. In 1999, he joined the AST Digital Still Camera group of STMicroelectronics. Since the beginning of 2002 he is leading a team devising innovative applications and services for the next generation of Mobile Imaging appliances. He is also attending the PhD in Information Technology at University of Messina. His researches concern imaging and multimedia.

Paolo Vivirito received his Italian degree in Electronic Engineering from the University of Palermo (Italy) in 2001 doing a thesis about restoration of out-of-focus images using inverse filtering techniques. Since January 2002, he joined the AST Digital Still Camera group of STMicroelectronics. His current activities include post-processing image enhancement and Wavefront Coding technologies.