INTEGRATING IMAGING AND VISION FOR CONTENT-SPECIFIC IMAGE ENHANCEMENT

Gianluigi Ciocca, Claudio Cusano, Francesca Gasparini and Raimondo Schettini
DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione)
Università degli Studi di Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy

Keywords: Automatic image enhancement, white balancing, contrast enhancement, edge sharpening, image classification, image annotation, redeye removal, face detection, skin detection.

Abstract: The quality of real-world photographs can often be considerably improved by digital image processing .In this article we describe our approach, integrating imaging and vision, for content-specific image enhancement. According to our approach, the overall quality of digital photographs is improved by a modular, image enhancement procedure driven by the image content. Single processing modules can be considered as autonomous elements. The modules can be combined to improve the overall quality according to image and defect categories.

1 INTRODUCTION

The great diffusion of digital cameras and the widespread use of the internet have produced a mass of digital images depicting a huge variety of subjects, generally acquired by non-professional photographers using unknown imaging systems under unknown lighting conditions. The quality of these real-world photographs can often be considerably improved by digital image processing. Since interactive processes may prove difficult and tedious, especially for amateur users, an automatic image enhancement tool would be most desirable. There are a number of techniques for image enhancement, including global and local correction for color balancing, (Buchsbaum, 1980), (Cardei, 1999), (Barnard, 2002), contrast enhancement (Tomasi, 1998), (Moroney, 2000) and edge sharpening (Kashyap, 1994), (Polesel, 2000). Other techniques merge color and contrast corrections, such as all the Retinex like algorithms (Land, 1977), (Rahman, 2004), (Rizzi, 2003), (Meylan, 2004). Rarely, traditional enhancement algorithms available in the literature are driven by the content of images (Naccari, 2005). Our interest is related to the design of content-aware image enhancement for amateur digital photographs. The underlying idea is that global and/or local image classification makes it possible to set the most appropriate image enhancement strategy according to the content of the photograph. To this end, we have pragmatically designed a modular enhancing procedure integrating imaging and vision techniques. Each module can be considered as an autonomous element, related to color, contrast, sharpness and defect removal. These modules can be combined in a complete unsupervised manner to improve the overall quality, according to image and defect categories. The proposed method is modular so that each step can be replaced with a more efficient one in future work, without changing the main structure. Also, the method can be improved by simply inserting new modules. The overall procedure is shown in Figure 1, while the single processing modules are described in the following Sections. The initial global image classification makes it possible to further refine the localization of the color regions requiring different types of color and sharpness corrections. The following color, contrast, and edge enhancement modules may exploit image annotation, together with further image analysis statistics (in same cases locally adaptive as well). Red eye removal is the only specific module we have developed for defect correction in digital photographs. Other modules related to different acquisition and/or compression artifacts are under development. In order to achieve a more pleasing result, a further processing module boosts the colors of typical regions such as human skin, grass, and sky. As a final step we also propose a self-adaptive image cropping module exploiting

Ciocca G., Cusano C., Gasparini F. and Schettini R. (2007).
INTEGRATING IMAGING AND VISION FOR CONTENT-SPECIFIC IMAGE ENHANCEMENT.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 192-199
DOI: 10.5220/0002065201920199

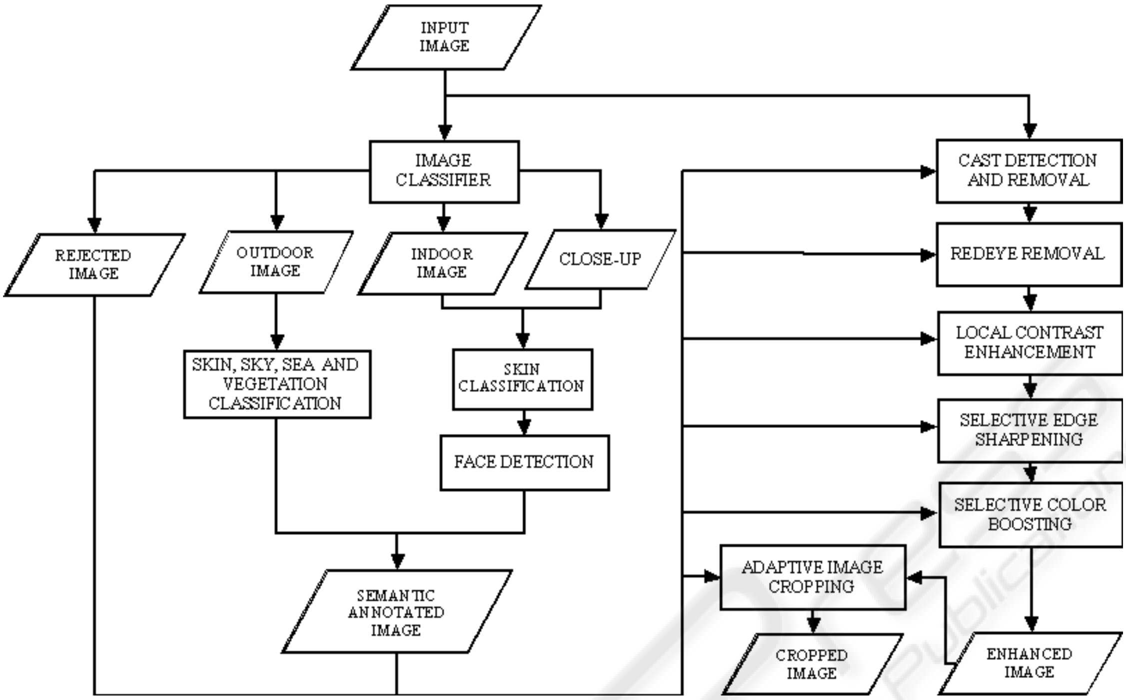

Figure 1: Workflow of our modular procedure for content- specific image enhancement.

both visual and semantic information. This module can be useful for users looking at photographs on small displays that require a quick preview of the relevant image area.

2 IMAGE CLASSIFICATION

Image classification makes it possible to use the most appropriate image enhancement strategy according to the content of the photograph. To this end we developed an automatic classification strategy (Schettini, 2004) based on the analysis of low-level features, that is, features that can be automatically computed without any prior knowledge of the content of the image.

We considered the classes outdoor, indoor, and close-up which correspond to typologies of images that require different enhancement approaches in our image processing chain. The indoor class includes photographs of rooms, groups of people, and details in which the context indicates that the photograph was taken inside. The outdoor class includes natural landscapes, buildings, city shots and details in which the context indicates that the photograph was taken outside. The close-up class includes portraits and photos of people and objects in which the context provides little or no information in regards to where the photo was taken. Examples of images of these classes are depicted in Figure 2.

We adopted a decision forest classifier: an ensemble of decision trees constructed according to the CART (Classification And Regression Trees) methodology. The features we used are related to color (moments of inertia of the color channels in the HSV color space, and skin color distribution), texture and edge (statistics on wavelets decomposition and on edge and texture distributions), and composition of the image (in terms of fragmentation and symmetry). To fully exploit the fact that trees allow a powerful use of high dimensionality and conditional information, we take all the features together and let the training process perform complexity reduction, and detect any redundancy. Each decision tree has been trained on bootstrap replicates of a training set composed of about 4500 photographs manually annotated with the correct class. Given an image to classify, the classification results produced by the single trees are combined applying the majority vote rule. To further improve the accuracy of the classifier and to avoid doubtful decisions, we introduced an ambiguity rejection option in the classification process: an image is “rejected” if the confidence on the classification result is below a tuneable threshold.

Figure 2: Classification results: examples of outdoor image (left), indoor image (middle), close-up (right).

3 IMAGE ANNOTATION

3.1 Outdoor Image Annotation

For the detection of semantically meaningful regions in outdoor photographs, we developed a method which is capable of automatically segmenting the images by assigning the regions to seven different classes: sky, skin, vegetation, snow, water, ground, and buildings (Cusano, 2005), as depicted in Figure 3. Briefly, the process works as follows: the images are processed by taking a fixed number of partially overlapping image subdivisions (tiles) for each pixel that contain it, each of which is then independently classified by a multi-class Support Vector Machine (SVM). The results are used to assign the pixel to one of the categories. Before submitting a tile to the classifier we computed a description of it in terms of low-level features. As feature vectors we used a joint histogram which combines color distribution with gradient statistics. For classification, we used a multi-class SVM, constructed according to the "one per class" strategy. Seven SVM have been trained to discriminate between the different classes. The discriminating functions of the single classifiers are compared to obtain the output of the combined classifier.

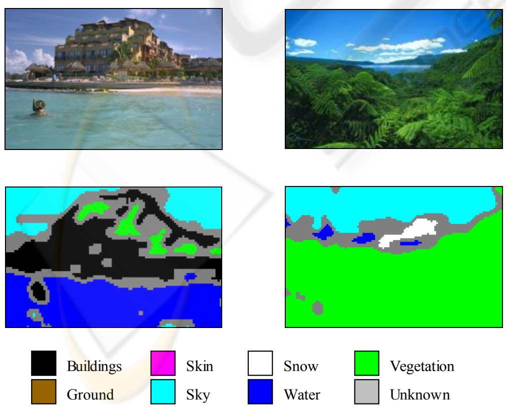

Figure 3: Examples of annotated outdoor images.

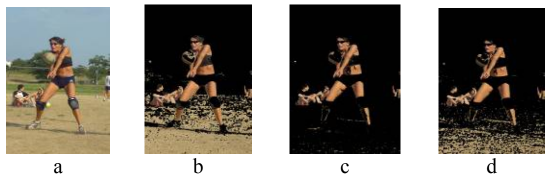

Figure 4: a: original image; b: segmented skin with a recall strategy. c: segmented skin with a precision strategy; d: segmented skin with a trade-off strategy.

3.2 Indoor and Close-ups Image Annotation: Skin Detection

Many different methods for discriminating between skin pixels and non-skin pixels are available. The simplest and most often applied method is to build an "explicit skin cluster" classifier which expressly defines the boundaries of the skin cluster in certain color spaces. The underlying hypothesis of methods based on explicit skin clustering is that skin pixels exhibit similar color coordinates in a properly chosen color space. This type of binary method is very popular since it is easy to implement and does not require a training phase. The main difficulty in achieving high skin recognition rates, and producing the smallest possible number of false positive pixels, is that of defining accurate cluster boundaries through simple, often heuristically chosen, decision rules. We approached the problem of determining the boundaries of the skin clusters in multiple color spaces by applying a genetic algorithm. A good classifier should have high recall and high precision, but typically, as recall increases, precision decreases. Consequently, we adopted a weighed sum of precision and recall as the fitness of the genetic algorithm. Keeping in mind that different applications can have sharply different requirements, the weighing coefficients can be chosen to offer high recall or high precision or to satisfy a reasonable trade-off between these two scores according to application demands (Gasparini, 2006), as illustrated in Figure 4. In the following applications addressing image enhancement, we adopted the boundaries evaluated for recall oriented strategies.

3.3 Indoor and Close-ups Image Annotation: Face Detection

Face detection in a single image is a challenging task because the overall appearance of faces ranges widely in scale, location, orientation and pose, as well as in facial expressions and lighting conditions (Rowley, 1998) and (Yang, 2002). Our objective therefore was not to determine whether or not there are any faces, but instead to evaluate the possibility

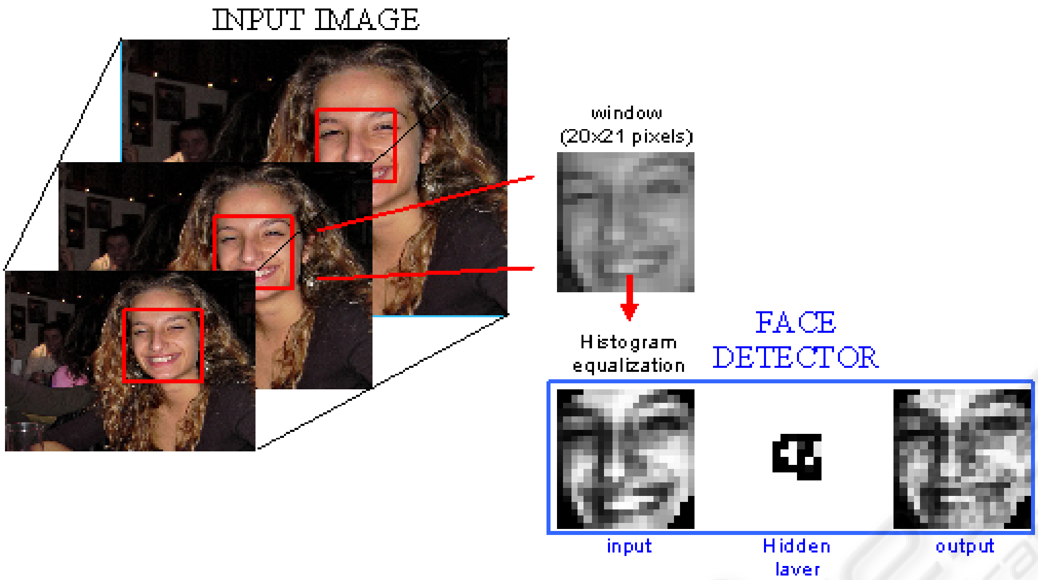

Figure 5: The neural network: the pyramid of the scaled input image, a sampled pixel window, preprocessing, consisting in histogram equalization of the oval part inside the window, and finally the application of the neural network.

of having a facial region. To do so, we have trained an autoassociative neural network (Gasparini, 2005) to output a score map reflecting the confidence of the presence of faces in the input image. It is a three layer linear network, where each pattern of the training set is presented to both the input and the output layers, and the whole network has been trained by a backpropagation sum square error criterion, on a training set of more than 150 images, considering not only face images (frontal faces), but non-face images as well. The network processes only the intensity image, so that the results are color independent. To locate faces of different sizes, the input image is repeatedly scaled down by a factor of 15%, generating a pyramid of subsampled images. Figure 5 illustrates this preprocessing and the application of the neural network with a sample image. The output is obtained with a feedforward function, and the root mean square error ε, between output and input is calculated. The performance of the network is evaluated analysing the True Positive, versus the False Positive varying the root mean square error ε. A score map of the input image is obtained collecting the likeliness that each single window in the pyramid contains a facial region evaluated as 1-*FP*(ε).

4 AUTOMATIC WHITE BALANCING

Traditional methods of color balancing do not discriminate between images with true cast (i.e. a superimposed dominant color) and those with predominant colors, and are applied in the same way to all images. This may result in an undesirable distortion of the chromatic content with respect to the original scene. To avoid this problem we developed a reliable and rapid method for classifying and removing a color cast in a digital image. (Gasparini, 2004). A multi-step algorithm classifies the input images as i) no-cast images; ii) evident cast images; iii) ambiguous cast images (images with feeble cast, or for which whether or not the cast exists is a subjective opinion), iv) images with a predominant color that must be preserved, v) unclassifiable images. The whole analysis is performed by preliminary image statistics for color distribution in the CIELAB color space. To avoid the mistaken removal of an intrinsic color, regions previously identified by image annotation as probably corresponding to skin, sky, sea or vegetation, are temporarily removed from the analyzed image. If an evident or ambiguous cast is found, a cast remover step, which is a modified version of the white patch algorithm, is applied. Since the color correction is calibrated on the type of the cast, an incorrect choice for the region to be whitened is less likely, and even ambiguous images can be processed without color distortion. In Figure

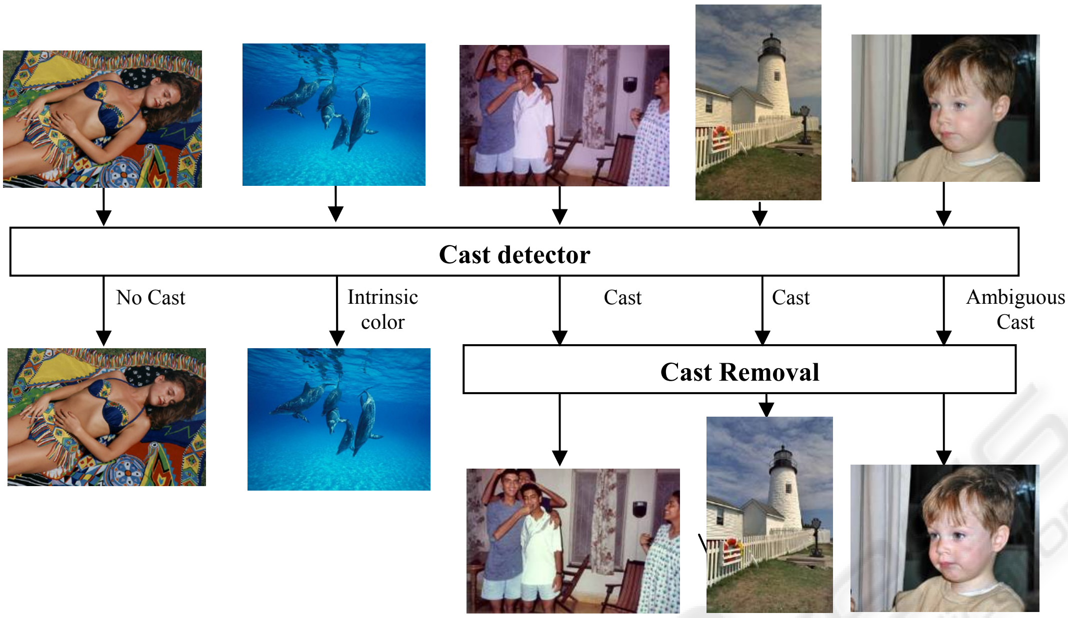

Figure 6: Our automatic color balancing procedure.

6 some examples of images processed by our color balancing procedure are shown.

5 REDEYE REMOVAL

Figure 7: Top row: original images, bottom row, the effect of our redeye removal procedure.

The redeye effect is a well known problem in photography. It is often seen in amateur shots taken with a built-in flash, but the problem is also well known to professional photographers. Redeye is the red reflection of the blood vessels in the retina caused when a strong and sudden light strikes the eye. Fixing redeye artifacts digitally became an important skill with the advent of digital technologies, which permit users to acquire digitalized images either directly with a digital camera or by converting traditional photos from scanners. Also, the widespread use of small devices with built-in flashes, including cell phones and handheld computers, produces a large number of digital photographs that are potentially affected by redeye. Currently, many image processing software applications in the market offer redeye removal solutions. Most of them are semi-automatic or manual solutions. The user has to either click on the redeye or draw a box containing the redeye before the redeye removal algorithm can find the redeye pixels and correct them (Benati, 1998), (Patti, 1998), (Hardeberg, 2002). A typical problem with most of these algorithms is poor pupil segmentation that leads to unnatural redeye correction. Even with user interaction, these algorithms sometimes correct redeye pixels too aggressively, darkening eyelid areas, or too conservatively, leaving many redeye pixels uncorrected. The proposed method (Gasparini, 2005) is modular so that each step can be removed and substituted with a more efficient one in future work, without changing the main structure. Also, it can be improved by simply inserting new modules. In our enhancement chain it follows our color balancing algorithm. This phase not only facilitates the subsequent steps of processing, but also improves the overall appearance of the output image. Like several redeye removal algorithms, the method we developed looks for redeye within the most likely face regions. The localization of these candidate regions is obtained by combining, through a scoring process, the results of a color-based face detector based on skin segmentation and the face detector based on a multi-resolution neural network, working only on the intensity channel. In the final phase, red eyes are automatically corrected by exploiting a novel algorithm that has been designed to remove the unwanted effects while maintaining the natural appearance of the processed eyes.

6 LOCAL CONTRAST ENHANCEMENT

The original dynamic range of a scene is generally constrained into the smaller dynamic range of the acquisition system. This makes it difficult to design a global tone correction that is able to enhance both shadow and highlight details. Several methods for adjusting image contrast, (Tomasi, 1998), (Moroney, 2000), (Meylan, 2004), (Rahman, 2004), (Rizzi, 2003), have been developed in the field of image processing for image enhancement. In general, it is possible to discriminate between two classes of contrast corrections: global and local corrections. With global contrast corrections it is difficult to accommodate both lowlight and highlight details. The advantage of local contrast corrections is that it provides a method to map one input value to many different output values, depending on the values of the neighbouring pixels and this allows for simultaneous shadow and highlight adjustments.

Our contrast enhancement method is based on a local and image dependent exponential correction (Capra 2006). The simplest exponential correction, better known as gamma correction, is common in the image processing field, and consists in elaborating the input image through a constant power function. This correction gives good results for totally underexposed or overexposed images. However, when both underexposed and overexposed regions are simultaneously present in an image, this correction is not satisfactory. As we are interested in a local correction, the exponent of the gamma correction used by our algorithm is not a constant. Instead, it is chosen as a function that depends on the point to be corrected, on its neighbouring pixels and on the global characteristics of the image. This function is also chosen to be edge preserving to eliminate halo artifacts. Usually it happens, especially for low quality images with compression artefacts, that the noise in the darker zones is enhanced. To overcome this undesirable loss in the image quality, a further step of contrast enhancement was added. This step consists of a stretching and clipping procedure, and an algorithm to increase the saturation. An example of this processing is shown in Figure 8.

Figure 8: Left, original image. Right, final image processed by our whole contrast enhancement procedure.

7 SELECTIVE EDGE ENHANCEMENT

Digital images are often corrupted by artifacts due to noise in the imaging system, digitization, and compression. Smoothing is a widely used technique to obtain more visually pleasing images, and several methods have been proposed in the literature to reduce edge blurring when smoothing is applied. Among the edge sharpening techniques, the unsharp masking approach is widely used to improve the perceptual quality of an image. Even though unsharp masking is simple and produces good results in many applications, its main drawback is that it does not distinguish between significant and non-significant high frequencies, such as noise, and thus all these high frequencies are added with the same weight. As a result, the algorithm applied to the original low quality image also enhances noise, digitization effects and blocking artifacts. We developed a new approach for selective edge enhancement (Gasparini, 2005) able to perform image smoothing, which not only preserves but also enhances the salient details in images. Our algorithm is based on the consideration that there is a strong relationship between biological vision and image rendering. In particular, the image rendering process is more successful interpreting the original scene and applying the appropriate transformations. The key idea is to process the image locally according to topographic maps obtained by a neurodynamical model of visual attention, overcoming the tradeoff between smoothing and sharpening typical of the traditional approaches. In fact, only high frequencies corresponding to regions that are non-significant to our visual system are smoothed while significant details are sharpened.

8 SELECTIVE COLOR BOOSTING

In many cases, global modification of the image colors will not result in correct reproduction. For some objects whose colors are well known, preferred color reproduction may be required (Hunt, 1977). Several authors, e.g. (Naccari, 2005), have suggested selective color correction for objects having a typical color such as human skin, grass, or sky in order to achieve a more pleasing result. Kanamori and Kotera (Kanamori, 1991) in particular, have suggested a smooth selective change in hue and saturation color attributes. Although effective, the method they have developed requires a great deal of practice: for each image to be processed, the color set to be changed, and the degree of change itself, must be specified in numbers. Taking this as our point of departure, we have developed a soft color cluster editor which allows the user to correct or modify the image colors as they appear, in a simple and effective way, until a satisfactory reproduction is obtained (Schettini, 1995), (Boldrin, 1999). A soft color cluster is composed of colors that are similar to a selected color centroid so that the farther a color lies from the centroid, the less it will be changed in editing. In order to effectively define the cluster we have exploited the best medium for color communication, sighting, and the fact that computer-driven displays allow the user to select and view the colors forming composite images on the screen in real time. Visual interaction allows the user to select the color centroids, and to define and edit soft color clusters without considering their internal representation, physical qualities, or names.

Different image categories (indoor, outdoor and close-ups) usually require different color corrections, therefore different image training sets have been defined and interactively corrected by a panel of specialized users. For each image class, we considered only satisfactory image matches, and all the colors that have been modified more then a given threshold are used to train a feed forward neural network. The implicit mapping coded in the trained neural networks can be applied to correct the colors of the processed images.

9 ADAPTIVE IMAGE CROPPING

Some of the efforts that have been put on image adaptation are related to the ROI coding scheme introduced in JPEG 2000 (Christopoulos, 2000). Most of the approaches for adapting images only focused on compressing the whole image in order to reduce the data transmitted. Few other methods use an auto-cropping technique to reduce the size of the image transmitted (Chen, 2003), (Suh, 2003). These methods decompose the image into a set of spatial information elements (saliency regions) which are then displayed serially to help users' browsing or searching through the whole image. These methods are heavily based on a visual attention model technique that is used to identify the saliency regions to be cropped. We designed a self-adaptive image cropping algorithm exploiting both visual and semantic information (Ciocca, 2007). Visual information is obtained by a visual attention model, while semantic information relates to the automatically assigned image genre and to the detection of face and skin regions. The processing steps of the algorithm are firstly driven by the classification phase and then further specialized with respect to the annotated face and skin regions.

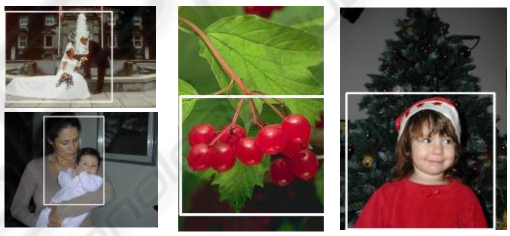

Figure 9: Examples of cropping areas selected by our algorithm.

10 CONCLUSIONS

We have described here our approach, integrating imaging and vision, for content-specific image enhancement. The key idea is that the most appropriate enhancement strategy can be applied if the photographs are semantically annotated. All of our image processing methods take into account the content of the photograph to drive the image enhancement. We have collected a variety of images for evaluation purpose. Different algorithms and/or parameter settings have been quantitatively evaluated whenever possible, or subjectively evaluated by pair wise comparison. To this end we have developed a web based system that makes it possible the comparison and ranking of different processing results by different users. Our results indicates that the proposed solution has several features in terms of effectiveness, friendliness and robustness that make it an ideal candidate to be included within software for the management and enhancement of digital photo albums by non expert, amateur photographers.

REFERENCES

Buchsbaum, G., 1980, A spatial processor model for object color perception, *Journal of Franklin Institute* 310, pp 1-26.

Cardei, V., Funt, B., Barnard, K., 1999, White Point Estimation for Uncalibrated Images, *Proceedings of the IS&T/SID Seventh Color Imaging Conference, Scottsdale, USA*, pp. 97 100.

Barnard, K., Cardei, V., Funt, B., 2002 A Comparison of Computational Color Constancy Algorithms-Part I: Methodology and Experiments with Synthesized Data, *IEEE Transactions on Image Processing* 11 (9), pp. 972-983.

Tomasi, C., Manduchi, R.,1998, Bilateral filtering for gray and color images, *Proc. IEEE Int. Conf. on Computer Vision,* pp. 836-846.

Moroney, N., 2000, Local colour correction using non-linear masking, *IS&T/SID Eighth Color Imaging Conference*, pp. 108-111.

Land, E., (1997), The Retinex Theory of Color Vision, *Scientific American* 237, pp.108-129.

Rahman, Z., Jobson, D., Woodell, G. ,2004, Retinex processing for automatic image enhancement, *Journal of Electronic Imaging*, 13, (1), pp. 100-110.

Rizzi, A., Gatta, C., Marini, D., 2003, A new algorithm for unsupervised global and local color correction, *Pattern Recognition Letters*, 24, pp. 1663-1677.

Meylan L., S¨usstrunk, S., 2004 Bio-inspired image enhancement for natural color images*, in IS&T/SPIE Electronic Imaging 2004. The Human Vision and Electronic Imaging IX*, 5292, pp. 46-56.

Naccari, F., Battiato, S., Bruna, A., Capra, A. Castorina, A., 2005, Natural Scene Classification for Color Enhancement, *IEEE Transactions on Consumer Electronics,* 5,(1), pp.234-239.

Kashyap, R.L. 1994, A robust variable length nonlinear filter for edge enhancement and noise smoothing, *Proc. of the 12th IAPR International Conference on Signal Processing*, 143-145.

Polesel, A., Ramponi, G., Mathews, V.J., 2000, Image Enhancement Via Adaptive Unsharp Masking, *IEEE Transactions on Image Processing*, 9, (3), pp.505-510.

Schettini, R., Brambilla, C., Cusano, C., Ciocca, G., 2004, Automatic classification of digital photographs based on decision forests, International Journal of Pattern Recognition and Artificial Intelligence, 18, (5), pp. 819-845.

Cusano, C., Gasparini, F., Schettini, R., 2005, Image annotation for adaptive enhancement of uncalibrated color images, *Springer, Lecture Notes in Computer Sciences*, 3736, pp. 216-225.

Yang, M. H, Kriegman, D.J., Ahuja, N., 2002, Detecting Faces in Images: A Survey, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24, (1), pp. 34-58.

Rowley, H., Baluja, S., Kanade, T., 1998, Neural Network-Based Face Detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 20, (1), pp.23-28.

Gasparini, F., Schettini, R., 2004, Color Balancing of Digital Photos Using Simple Image Statistics, *Pattern Recognition*, 37, pp. 1201-1217.

Benati, P., Gray, R., Cosgrove, P., 1998, Automated detection and correction of eye color defects due to flash illuminatio, US patent 5,748,764 5 May.

Patti, A., Kostantinides, K., Tretter, D., Lin, Q., 1998, Automatic Digtal Redeye Reduction, Proceedings of the IEEE International Conference on Image Processing, 3, pp 55-59.

Hardeberg, J. Y., 2002, Digital red eye removal', Journal of Imaging Science and Technology, 46 (4), pp. 375-381.

Gasparini, F., Schettini, R., 2005, Automatic redeye removal for smart enhancement of photos of unknown origin, *Springer, Lecture Notes in Computer Sciences*, 3736, pp. 226-233.

Capra, A., Castorina, A., Corchs, S., Gasparini, F., Schettini, R., 2006, Dynamic range optimization by local contrast correction and histogram image analysis, 2006 IEEE International Conference on Consumer Electronics, Las Vegas, USA, (in print).

Gasparini, F., Corchs, S., Schettini, R., 2005, Adaptive edge enhancement using a neurodynamical model of visual attention, *In Proc. IEEE International Conference on Image Processing,* 3, pp. 972-975,

Hunt, R. W. G., 1977, Colour reproduction by photography, *Reports on Progress in Physics*, 40, pp. 1071-1121.

Schettini, R., Barolo, B., Boldrin, E., 1995, A soft color cluster, *Editor, Image Process. Commun*. 1(1) pp. 17-32.

Kanamori, K., Kotera, H., 1991, A method for selective color control in perceptual color space, *J. Imaging Technol*. 35(5) pp. 307-316.

Boldrin, E., Schettini, R., 1999, Cross-media color matching using neural networks, *Pattern Recognition*, 32, pp. 465-476.

Christopoulos C., Skodras A., Ebrahimi T., 2000, The JPEG2000 still image coding system: an overview, IEEE Trans Cons Elect. vol. 46, no. 4, pp. 1103–1127,.

Chen L., Xie X., Fan X., Ma W., Zhang H.J., Zhou H.Q., 2003, A visual attention model for adapting images on small displays, *Multimedia Systems*, 9, pp. 353–364.

Suh B., Ling H., Bederson B.B., Jacobs D.W., 2003, Automatic Thumbnail Cropping and its Effectiveness, *In Proc. UIST'03*, pp. 95-104,.

Ciocca, G., Cusano, C., Gasparini, F., Schettini, R., 2007, Self Adaptive Image Cropping for Small Displays, *Proc. IEEE Int. Conference on Consumer Electronics (ICCE),* Las Vegas, USA, Jan. 10-14, to be published.