TOWARDS INTENT DEPENDENT IMAGE ENHANCEMENT
State-of-the-art and Recent Attempts

Marco Bressan, Gabriela Csurka and Sebastien Favre
Xerox Research Centre Europe, 6, ch. de Maupertuis, 38240 Meylan, France

Keywords: Intent-based image enhancement, semantic content, image understanding, photographic aesthetics.

Abstract: Image enhancement is mostly driven by intent and its future largely relies on our ability to map the space of intentions with the space of possible enhancements. Taking into account the semantic content of an image is an important step in this direction where contextual and aesthetic dimensions are also likely to have an important role. In this article we detail the state-of-the-art and some recent efforts in for semantic or content-dependent enhancement. Through a concrete example we also show how image understanding and image enhancement tools can be brought together. We show how the mapping between semantic space and enhancements can be learnt from user evaluations when the purpose is subjective quality measured by user preference. This is done by introducing a discretization of both spaces and notions of coherence, agreement and relevance to the user response. Another example illustrates the feasibility of solving the situation where the binary option of whether or not to enhance is considered.

1 INTRODUCTION

Considering both digital cameras and camera phones, it is estimated that almost 400 billion images will be captured in 2007 only (Hoffenberg, 2006). Regardless of the final medium where the images will be managed, shared and visualized, the quality expectations of consumers are likely to grow steadily. It is currently very easy for users to integrate their own content into workflows such as online photofinishing or content-sharing communities. The variability of content type, mainly due to the democratization of the production and distribution tools, together with the increased quality expectations, results in a demand for automated or semi-automated image enhancement tools that can help reducing user interaction.

New features such as automatic color balance or red-eye correction are now standard components in mainstream image editing applications. New companies offer products focused exclusively on automatic image enhancement to software vendors, camera and phone manufacturers, printing providers or directly to end-users. Most of the current offering follows a classical approach to image enhancement where some kind of degradation, which has to be compensated, is assumed.

Acquisition conditions, user expertise, compression algorithms or sensor quality, can seriously degrade the final image quality and image enhancement attempts to compensate for this degradation by altering image features for subsequent analysis, distribution or display. Examples include contrast and edge enhancement, noise filtering for a wide variety of noise sources, sharpening, exposure correction, colour balance adjustment, automatic cropping or correction of shaky images. Some of these features, such as noise, can be objectively defined and others, such as contrast, can be inspired by human perception. Still, in most cases, the final judgment over the performance of an enhancement algorithm is subjective. For example, while some people might prefer to see the shadowed details made visible by some local contrast approach, others will appreciate the sensation of depth caused by the original shadows.

Enhancement is mostly driven by intent. The intention of a photographer to depict a scene will value those photographs or those enhancement operations that lead to a more faithful representation of the captured scene. The intention of a designer or an advertiser could be to enhance an image in a way optimal for transmitting a message, e.g. an emotion.

Bressan M., Csurka G. and Favre S. (2007).
TOWARDS INTENT DEPENDENT IMAGE ENHANCEMENT - State-of-the-art and Recent Attempts.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 208-218
DOI: 10.5220/0002068202080218

The intention of a person including a photo of a baby in a family album can be to simply highlight a facial expression, at the cost of leaving degradations untouched or even highlighting them. The intent of a photofinishing operation is to automate image enhancement in order to please the largest possible audience. Their intent driven enhancement issues (e.g enhancing shadow details) could be solved by user evaluations, indicating which approach is preferred by the majority of users.

In general, the future of image enhancement relies on our ability to map the space of intents with the space of possible enhancements. While we can assume the space of enhancements to be reasonably well defined, the challenge will be for long to model intentions. For this purpose, the present model based on image degradations is insufficient and we need to extend its scope to multiple semantic, aesthetic and contextual dimensions. Recently, we have witnessed the first efforts in this direction. These efforts have mostly focused on the semantic dimension and in many cases were pushed by the advances in object and scene recognition towards understanding image content. The assumption here is that semantic content drives intention. Also, the fact users are starting to share content and express their preferences online can be understood as instances from which this mapping can be learnt and some recent approaches focus on this fact. These are probably the first approaches to take into account the aesthetic dimension. Working on large amounts of available user preference data is a very promising direction for understanding intent space.

This article details first some recent efforts on image enhancement to then focus on the particular problem of semantically dependent enhancement. In this case, the concept space is defined by the set of semantic categories and, eventually, the relationship among these categories. To illustrate our approach with experiments, instead of considering the complete space of possible image enhancements, we restrict our approach to the variations that might be generated from a particular image enhancement approach. Having defined our semantic categories and enhancement space, the mapping is learnt from user preference evaluations.

The remainder of this paper is organized as follows: section 2 describes the prior art in semantic content dependent image enhancement including a state-of-the-art of image enhancement (section 2.1) and semantic image understanding (section 2.2). In section 3 we present a semantic content dependent image enhancement (SCDIE) system and conclude the paper in section 4.

2 PRIOR ART

There have always been enhancements designed for specific types of images such as remote sensing imagery, medical imaging or document images. In this last field, recent approaches propose different enhancements depending on whether the document is classified as text, block diagrams, road maps, computer generated images or user photos (Allen et al., 2004; Ichikawa and Miyasaka, 2005). There are also enhancements which are specific to the output device (Furuki and Yamada, 2006) especially those related to mobile phones (Quelard, 2004).

In the field of photography probably the first enhancements clearly linked with semantic content were human skin and sky: skin dependent exposure correction (Battiato et al., 2003), skin defect detection and correction (Hillebrand et al., 2003) sky detection based image orientation correction (Luo, 2003) or sky-based color enhancements (Luo and Etz, 2002; Zafarifar and de With, 2006). Skin and sky detection typically require low-level image understanding since detection strongly relies on color analysis techniques.

At a higher-level there has been strong research on face enhancement such as face makeup simulation (Utsugi, 2003), skin smoothing, eye and teeth whitening filters (Simon and Matraszek, 2006), facial skin color-based color saturation, white balance and overall density correction (Mutza, 2006) or adjusting lightness, contrast, and/or the color levels of the image based on the detected faces (Lin et al., 2002). Many red-eye detection and correction make use of face detection as a preprocessing stage for reducing the number of candidate regions (Gaubatz and Ulichney, 2002; Gasparini and Schettini, 2005).

The attention received by approaches focused on face enhancement is not surprising. Faces, specially known faces, are common fixation points when observing a scene (Buswell, 1935; Henderson and Hollingworth, 1999). Enhancing faces makes a lot of sense: they are likely to receive much attention. For closely related reasons, photographs with people are more frequent than without. To estimate this importance we applied a face detector to approximately 130000 images randomly selected from a photofinishing workflow. Roughly two out of every three images contained at least one face. On the other side face detection and recognition have for long represented a challenge for the vision community and available solutions and algorithms are readily available (Zhao et al., 2003; Yang et al., 2002).

General object and scene detection and recognition approaches have also been proposed and, in the general case, current performance is considerably below the case of faces. Recent competitions (Everingham et al., 2005; Everingham et al., 2006) show that the problem is difficult though important advances were made recently (see section 2.2). Although they mainly focus on object class recognition and detection, some of them are able to succesfully handle scene categories too like: Indoor, Outdoor, Beach, Mountain, City, Suburb, Road, Underwater, Sunset, etc (Li and Wang, 2003; Barnard et al., 2003; Carbonetto et al., 2004; Quelhas et al., 2005; Perronnin et al., 2006; Bosch et al., 2006).

There were some recent attempts to combine such categories with enhancement. For example (Chambah et al., 2004) propose a typical enhancement of underwater images. On the other hand (Gallagher and Bruehs, 2006) proposes a system where an improvement parameter of sharpening or noise reduction is generated from the belief map indicating the likelihood that respective pixels are representative of faces, flash, sky, or vegetation. Similarly, (Gasparini and Schettini, 2004; Fredembach et al., 2003) identify regions as probable skin, sky, sea or vegetation in order to avoid color cast removal which is intrinsic to those categories, e.g. the blue of the sky.

Most of the efforts exposed link the semantic content with the enhancement under the assumption that semantic content can guide the processing. Nevertheless, other aspects of the image can lead the enhancement process, such as image aesthetic or originality which in contrast with classical image quality are highly subjective measures. Recently, (Datta et al., 2006) attempted to infer them with some visual features using machine learning with a user preferences gathered from a peer-rated online photo set.

Image understanding and enhancement may be also combined with meta-data information. In (Oberhardt et al., 2003), red eye correction and detection counts on the knowledge the flash was triggered at time of capture. Also for correcting eyes (Sadovsky et al., 2004) uses the information stored in the Exchangeable Image File Format (EXIF). This kind of information is also suggested to improve image categorization (Lin and Tretter, 2005; Boutell and Luo, 2007). Unfortunately, the presence of meta-data is not always ensured, mainly due to the variability encountered in those imaging scenarios where end users can directly integrate into workflows. This is likely to change, as standards are agreed and cameras are able to include valuable meta-data such as geographic location or web-retrieved information (O'Hare et al., 2005).

The challenge of intent-based enhancement is to estimate the function that maps an image and an intention to an enhancement. If we take into account only semantic information and a single label is considered per image, e.g. indoors, then the space of intents can be modelled with a single discrete dimension which corresponds to the categories. Dimensionality increases as we consider multiple categories, locality constraints, additional information such as meta-data, etc. On the enhancement side, the types of processing that can be applied to an image are restricted. In the following sections we complete this prior art with a list of typical enhancement techniques (section 2.1) and with state-of-the-art image understanding techniques that can be used for modelling the semantic information on an image (section 2.2).

2.1 Image Enhancement

Image enhancement techniques are applied to obtain a resulting image which is more suitable than the original for a specific objective. Visual quality is a sample objective but, depending on the application, quality might not be the main purpose of enhancement, e.g. medical imaging.

Enhancement algorithms can be global, where the parameters controlling the enhancement approach are the same over the whole image; or local, where the parameters can vary spatially and are generally based on the local characteristics of an image. Many enhancements require user interaction for setting or controlling some of its parameters. In this case, the enhancement is called manual. When all parameters can be set without interaction, based on general considerations or on image content, the enhancement is called automatic. Automatic enhancements frequently come with a preprocessing stage which estimates the parameters of the actual enhancement using image statistics. For instance, an unsharp mask where the filter values are different on regions specified by the user can be considered a local manual enhancement. A common technique for enhancing images is through Tone Reproduction Curves (TRCs) which are global mappings of luminance or chrominance channels. The case where the mapping depends on the image region is referred to as Tone Reproduction Operator (TRO).

The most common enhancement techniques are sharpening, exposure correction, color balance and saturation adjustment, contrast and edge enhancement, blocking artifact reduction and noise reduction. There are many more enhancements focused on specific problems such as redeye correction, automatic cropping, or glass glare removal.

Sharpness refers to the presence of crisp edges and fine details in an image. Basic sharpening filter on images (Rosenfeld and Kak, 1982; Gonzalez and

Woods, 1992) can work in many cases but, since they are usually high-pass filters they can also enhance noise. To overcome this problem specific sharpening algorithms robust to noise have been proposed, based on weighted median filters (Fischer et al., 2002), non-linear reaction-diffusion (Saito et al., 2003) or locally adaptive filters (Polesel et al., 2000).

Exposure refers to the average of the global distribution of intensity along the dynamic range of the image. Making the image darker or lighter can bring details from the shadows or give depth to the colors of the photograph. The automatic setting of exposure, a feature present in most digital cameras, can yield unrealistic results and exposure correction attempts to overcome this problem. The most common approach to correcting exposure is to apply gamma correction to the image intensity. For instance, (Eschbach and Fuss, 1999) propose a method to determine the gamma parameter automatically from the histogram of the input image.

The human visual system ensures the perceived color of objects remains relatively constant under varying illumination and reflectance conditions, e.g. color constancy. When imaging devices are tailored to common illuminants, e.g. D65, they can introduce strong color casts when the scene has another light source. The problem of adjusting the colors to resemble perceptual response is called color balance or white balance and generally consists in a global enhancement. The two most common approaches for color balance are based on two very simple assumptions to estimate the color cast. The Gray World approach assumes the average chrominance on any given image is approximately gray (Evans, 1951). The Perfect Reflector algorithm assumes we can find a specular surface on the image reflecting the actual color of the light source. Closely related, are white point (Eschbach and Fuss, 1999) and black point (Adams et al., 2003) approaches. Much research on this field has derived from the effort of developing automatic color constancy algorithms (Barnard et al., 2002) for machine vision purposes.

Saturation refers to the vividness of colored objects in an image. A color with more gray is considered less saturated, while a bright color, one with very little gray in it, is considered highly saturated. The saturation of a color can affect the emotional reaction to an image. Colors that have low saturations are often seen as dull and boring, but can also be thought of as restful and peaceful. Highly saturated colors, on the other hand, are more vibrant and emotionally aggressive. Therefore, color saturation is an important element in an intent based enhancement system. In a classical automatic enhancement approach, where neither the image content nor the users intent is known, the system detects and modifies these extremes bringing the image saturation to a generally acceptable level. An alternative to the direct modification of the saturation value in HSV space, is to interpolate or extrapolate between the original image and a black-and-white version of the image (Haeberli and Voorhies, 1994). Even if there exist automatic saturation enhancement techniques (Eschbach and Fuss, 1999), they must be rather conservative as saturation preferences vary a lot between individuals and depends often on the semantic content of the image.

Contrast refers to the efficient use of the dynamic range. Improved contrast should make image details more evident to a human observer. Contrast enhancement can be achieved via global approaches (Tumblin and Rushmeier, 1993; Eschbach et al., 1995). Spatially uniform contrast enhancement approaches however fail to model perceptual attributes where locality is an important characteristic. Depending on the aggressiveness of the approach the images can appear washed-out or artificial. Limitations due to the global nature of this technique are observed in those images where luminance is uniformly distributed over the whole range.

Local approaches through TROs have also been proposed (Zuiderveld, 1994; Devlin et al., 2002; DiCarlo and Wandell, 2001; Fattal et al., 2002). A complete evaluation of TRO performance focused on high dynamic range display appears in (Ledda et al., 2005). More complex approaches rely on generative models to recover the reflectance typically using edge preserving filters to avoid halo effects (Chiu et al., 1993; Tumblin and Turk, 1999; Durand and Dorsey, 2002).

Blocking artifacts are the result of coding, resizing or compressing the image. A traditional approach to reducing blocking artifacts is to low-pass filter the pixels directly adjacent to the block boundaries. Using a Gaussian spatial domain filter (Reeve and Lim, 1984) is very fast; however, it cannot reduce artifacts that are not confined to pixels next to block boundaries. To overcome this problem linear block boundary filters (Avril and Nguyen-Trong, 1992) or separable anisotropic Gaussian filters perpendicular to the block boundary (Tzou, 1988) were proposed. The drawback of these techniques is that they do not adapt to local characteristics of the signal and change a high frequency artifact for a low frequency one. Therefore, (Ramamurthi and Gersho, 1986; Meier et al., 1999), propose edge preserving space-variant region-based filters and (Xiong et al., 1997; Kim et al., 1998) wavelet transform to smooth blocking effects while preserving edges. In an automatic approach it is important also to be able to estimate the blockiness of an image (Minami and Zakhor, 1995; Tan and Ghanbari, 2000; Fan and de Queiroz, 2003) to adjust the level of correction and avoid unnecessary degradation.

Imperfect instruments, problems with the data acquisition, transmission and compression can all be sources of noise on the image. Random image noise corresponds generally to visible grain or particles present in the image which are generally caused by the electronic noise in the input device sensor and circuitry (e.g. scanner, digital camera). Intensity spikes, speckle or salt and pepper noise will only affect a small number of image pixels. They are caused by flecks of dust on the lens or inside the camera, dust or scratches on scanned photography or film, faulty CCD elements, “hot pixels” occurring with long exposures with digital camera, etc. Banding noise can be introduced when the data is read from the digital sensor (e.g. scanner streaks) and scratches on the film will appear as additional artifacts on the images. One method to remove noise is by convolving the original image with a mask (e.g. Gaussian). Its drawback is the blurring of edges. In contrary, a properly designed median filter is very good at removing salt and pepper noise preserving image detail. Promising denoising results can be achieved using a wavelets (Portilla et al., 2003), anisotropic diffusion (Perona and Malik, 1990), and bilateral filtering (Tomasi and Manduchi, 1998). A recent survey of different techniques can be found in (Motwani et al., 2004).

Image blur is a form of bandwidth reduction typically caused by relative motion between the camera and the original scene or by an optical system that is out of focus. It can affect the totality or part of an image and many cameras today have built in solutions to stabilize image capture. There are different techniques available for solving the restoration problem from blind de-convolution methods (Zhang et al., 2000; Stern et al., 2002) to approaches that combine power-laws with wavelet domain constraints (Jalobeanu et al., 2002; Neelamani et al., 2004). A method to remove the specific blur due to camera shake was proposed in (Fergus et al., 2006). The automatic implementation of these techniques requires the estimation of the level of blur or motion blur for the different image regions, prior to correction.

An example of a completely automatic system including most of the mentioned enhancements is Xerox’s Automatic Image Enhancement (XAIE) (Xerox, 2006). This approach is composed of two stages: a decision mechanism stage and an apply stage. Given an image, the (low-cost) decision mechanism stage determines whether a particular enhancement will or will not be applied on an image and this is typically done anticipating image improvement or degradation. For this purpose, the decision mechanism uses statistics such as noise measures or luminance and chrominance distributions on a low resolution version of the image. Details for particular components of the XAIE decision mechanism can be found in (Eschbach and Fuss, 1999; Bressan et al., 2007). The settings for the decision mechanism are largely based on user preference evaluations. The output of all decision mechanisms are combined and fed to the apply stage. This architecture also presents advantages in terms of computational costs since only those enhancements that can benefit the image are actually applied.

2.2 Image Understanding

Image understanding refers to a set of operations that transforms pictorial inputs into commonly understood descriptions. Even if we are far from a complete automatic understanding/description of the image, huge advances were made in the last few years to successfully assign keywords to an image based on its high-level content. These techniques can analyze the whole scene or focus on objects within the image. Systems are considered *generic* when their technology is independent of the classes or object types. The main difficulty of such generic systems is that they have to handle not only the variations in view, imaging, lighting and occlusion, typical of the real world, but also intra-class variations typical of the semantic classes, e.g. types of chairs.

The most common tasks are recognition, classification or detection. Recognition concerns the identification of particular object instances. Object and scene classification are the tasks of assigning one or more general tags to an image. Detection is the problem of determining if one or more instances of an object occur in an image and, typically, estimate locations and scales of the detected instances. From the perspective of image enhancement, classification and detection are considered more relevant than recognition. By far, the “object” that received most of the attention for detection and recognition has been faces (Yang et al., 2002).

The first multi-class categorization approaches were based on image segmentation. Their aim was labelling relatively homogeneous image segments with keywords such as sky, grass, tiger, water, rocks. To do this they used statistical models to learn a sort of dictionary between individual image blobs (segments) and a set of predefined keywords (Barnard et al., 2003; Carbonetto et al., 2004; Chen and Wang, 2004; Li et al., 2004).

Motivated by an analogy to bag-of-words based learning methods for text categorization, a large set of bag of visual word (BOV) based approaches emerged recently. Similarly to text, an image is characterized by a histogram of visual word counts. In contrast to text categorization where a dictionary is available, one of the challenges for images is that the visual vocabulary has to be built automatically from the training set. To do this, first some image descriptors are extracted from the image. Those descriptors are generally based on texture, color, shape, structure or their combination and are extracted locally on regions of interest (ROI). The ROI can be obtained by image segmentation as above, by applying specific interest point detectors (Csurka et al., 2004; Quelhas et al., 2005), by considering a regular grid (Carbonetto et al., 2004; Fei-Fei and Perona, 2005) or simply random sampling of image patches (Marée et al., 2005; Novak et al., 2006). All features extracted are then mapped to the feature space and clustered to obtain the visual vocabulary. Often a simple K-means is used, however Gaussian Mixture Models (Farquhar et al., 2005) or Self Organization Maps (Lefebvre et al., 2006) can also be used to obtain a soft clustering, inline with the continuous nature of visual words.

Given a test sample, each feature vector is assigned to its closest visual word in the previously trained vocabulary or to all visual words in a probabilistic manner in the case of a stochastic model. The histogram is computed by accumulating the occurrences of each visual word. Finally, the histogram is fed to a classifier, for example K nearest neighbor (Bosch et al., 2006), probabilistic latent semantic classifier (Quelhas et al., 2005) or support vector machines (Csurka et al., 2004).

Though most of the mentioned approaches use a single visual vocabulary generally built on the whole training set, this is not always the best option: Very good performance is achieved when category labels are used during the estimation of the visual vocabulary. While (Farquhar et al., 2005) agglomerate category-specific vocabularies into a single vocabulary, (Perronnin et al., 2006) propose to adapt the visual vocabulary (universal) trained on the whole training set to each class using class-specific images. An image is then characterized by a set of bipartite histograms - one per class - where each histogram describes whether the image content is best modeled by the universal vocabulary, or the corresponding class vocabulary.

One of the drawbacks of the BOV approach is that it considers the image as a “bag” of independent visual word instances. (Sivic et al., 2005) tried to overcome this by building a visual vocabulary of features describing the co-occurrences of visual words. (Fergus et al., 2003; Leibe et al., 2004; Crandall and Huttenlocher, 2006) proposed building generative models that take into account relatively strong geometric constraints between image patches. However, this requires the alignment and segregation of different views of objects in the dataset. (Csurka et al., 2005) propose to incorporate geometric information based on scale, orientation and closeness of the keypatches in a boosting framework. The selected weak classifiers are combined with the original BOV classifier. In (Carbonetto et al., 2004) geometry has been included through generative MRF models of neighboring relations between segmented regions. (Sudderth et al., 2006; Fidler et al., 2006) proposes hierarchical learning of generic parts and feature combinations. The above approaches aim to handle mainly object classes and showed performance improvements for classes such as cars, planes, faces. In contrary, (Boutell et al., 2006) propose generative models for outdoor scene configurations, consisting of regions’ identities (beach, field, mountain, street, suburban, open-water) and their spatial relations (above, far above, below, far below, beside, enclosed, and enclosing). However, the improvement achieved by theses systems over the BOV approaches is relatively modest compared to the increased computational cost.

3 SCDIE

Semantic Content Dependent Image Enhancement (SCDIE) is the result taking into account semantic content for image enhancement. Classical enhancement is mainly based on dimensions that model low-level quality measures. SCDIE also considers semantic dimensions and this extension allows for more precise models, e.g. an overexposed sunny landscape. SCDIE maps this sample to a particular enhancement which is then used to improve image quality. We call the domain of this mapping ”Intent Space” (Ω_I) and the target ”Enhancement Space” (Ω_E). Figure 1 illustrates this approach for the case of SCDIE. Notice that other dimensions can be naturally incorporated into this model.

Such a system can be built with the following components.

1. an image quality measure component which analyzes low-level features related with quality in the image.
2. an image understanding component which assigns one or more semantic labels to an image. Notice that we assume quality and understanding are independent: image labels do not depend on quality.

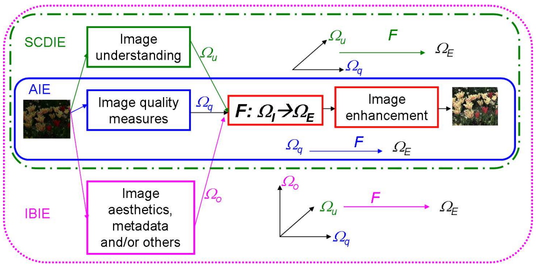

Figure 1: The clssical AIE, extended to SCDIE and possible extentions to other dimensions of intent based image enhancement (IBIE).

This useful assumption holds most of the time but not always since it depends on the labels we consider.

3. a function that maps the space spanned by image quality and understanding output into the space of possible enhancements.
4. an image enhancement component in charge of applying the resulting enhancement.

We now specify the components in the design of a particular instance of the system. For the image quality component we take into account the statistics which are computed in the decision mechanism stage of XAIE (see section 2.1), e.g. luminance and chrominance distributions and statistics, noise, edge levels, blocking artifacts, etc. The understanding component uses a BOV-based multi-label categorizer trained on 8 categories: Urban, Portrait, Flowers, Interiors, Landscape, Snow and Sky. These categories were chosen to be representative of images found in typical imaging scenarios.

For the space of possible enhancements Ω_E we choose to use the topology provided by XAIE. We consider 7 different enhancements dimensions included in XAIE: contrast, exposure, shadow details, saturation, color balance, noise reduction and sharpness and discretize them into three intensity levels or modes: low, default and high. Depending on the enhancement, the three bins can have a different interpretation, e.g. dark, default and light for exposure correction. With this approach, the discretization of Ω_E yields $3^7 = 2187$ bins.

XAIE already estimates the mapping function in the decision stage of the algorithm (see section 2.1). The decision stage uses only the image quality components in Ω_I to determine the best enhancement in Ω_E. The decision mechanism does not take into account the semantic content. Since we assumed independence between quality and understanding, we can extend the mapping provided by the decision mechanism by linking semantic categories with enhancement modes (semantic decision) provided the decision mechanism decides to apply the enhancement (quality decision). The mapping between categories and enhancements is learnt through user evaluations.

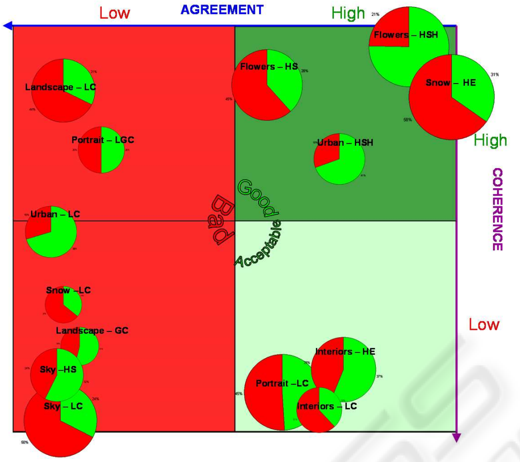

Figure 2: Pairs of (class,enhancement) plotted in the agreement/coherence map. Green means it was preferred while red means it was considered to have a bad effect. The size of the circle was proportional with the relevance. Attention should be paid to large circles with dominant color (red or green) in the rightmost and uppermost hemispheres.

Given a category, choosing the preferred enhancement is untractable even in the case Ω_E is discretized. So we first restrict the space of enhancements by assuming independence between enhancements and by screening out improbable mappings. The independence assumption reduces the image comparisons to $3 * 7 = 21$ and the screening leaves out around two thirds of these possibilities. Examples of enhancements left out by the screening process were high saturation on portraits or sharpness to sky images. No category had more than six candidate enhancements after the screening.

A representative set of images was printed from each category and, for each image, all possible enhancements were printed on a single A3 size page. Participants had to choose for each image the worst and best enhancement. To analyze a given enhancement on a given category, we considered the following criteria:

- Agreement: measures either positive or negative agreement among different user test participants on a given image:

$$\frac{1}{N^I}\sum_{i=1}^{N^I}\left(g\left(\frac{N_b^i+N_n^i/2}{N^U}\right)+g\left(\frac{N_w^i+N_n^i/2}{N^U}\right)\right)$$

where $g(x) = xlog_2(x)$, N^U is the number of users, N^I is the number of different image considered. N_b^i, N_w^i and N_n^i is the number of users that chose

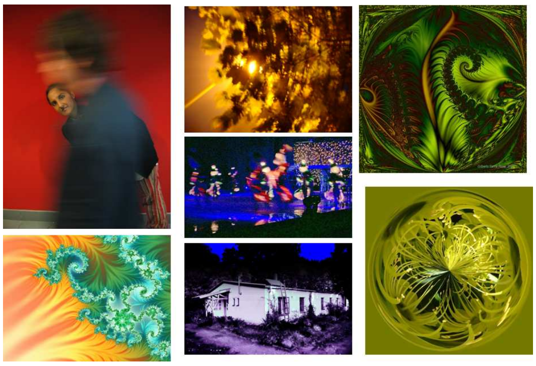

Figure 3: Some examples of special effect images.

the enhancement as being the best, the worst or none of them for a given image i. Agreement measures the entropy of the distribution of preference for all user for a single image.

- Coherence: measures uniformity of opinion across all images of a single category for any given participant

$$\frac{1}{N^U}\sum_{u=1}^{N^U}\left(g\left(\frac{N_b^u+N_n^u/2}{N^I}\right)+g\left(\frac{N_w^u+N_n^u/2}{N^I}\right)\right)$$

where N_b^u, N_w^u and N_n^u is the number of image for which the user u considered the enhancement as being the best, the worst or non of them. Coherence measures the entropy of the distribution of the preference of a single user across all images.

- Relevance: Percentage of times the enhancement was chosen as either the best or the worst.

Figure 2 shows the pairs of (class,enhancement) in an agreement/coherence map, allowing the following interpretation. High agreement and high coherence is good, meaning that everyone agrees and opinion is consistent for all category images. Low agreement and high coherence is bad because it indicates that the appreciation of the enhancement is highly subjective and dependant on user preference for categories. High agreement and low coherence is neutral, probably indicating that the image set for that category was poorly chosen. Obviously, low agreement and low coherence is bad.

From this evaluation, enhancement improvements were suggested for three out of the eight selected categories: flowers, sky and urban.

3.1 Special Effects

In this section, we show a second example of SCDIE subsystem, which tries to handle the problem of detecting images with special effects in order to automatically turn off the enhancement process for those images. Such an automatic subsystem is interesting in a printing workflow, where the growing number of manipulated photos, images with unusual viewpoint, lighting or artificial images of non-professional photographers (see examples in Figure 3) are most often mixed with the rest of their album images. It is understood that photographers generally prefer no further changes in there artificially manipulated images.

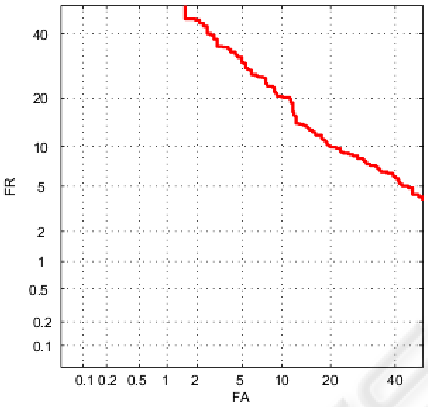

Figure 4: False Alarm Rates (FA) and False Rejection Rates (FR) of the BOV as special effect occurence detector plotted on a DET curve.

In the context of SCDIE, this can be translated as follows. The image understanding component labels those images with *Special Effects* and the decision mechanism F maps this label to *No enhancement*. To test the feasibility of such a system, we trained a BOV (Perronnin et al., 2006) on images with and without special effects collected from two independent sources. Images from the first source (1160 with effect and 966 without) served as training data and images from the second source (536 with effect and 524 without) were used as test data. Figure 4 shows the DET (Detection Error Tradeoff) Curve (Martin et al., 1997) of the system acting as special effect occurence detector. To the best of our knowledge, these are the first results reported for such a task. Furthermore, special effects generally belong to clearly defined categories: blur, painting/stylization, artificial lighting, etc. We are currently experimenting if training our system on these, better defined subcategories as well as the subcategories of natural photos (indoors, outdoors, portrait, city, landscape, etc.) can improve the current performance (EER =13.7) of our system.

4 CONCLUSION

The future of image enhancement relies on our ability to map the space of intents with the space of possible enhancements for a given image. The main challenges are given by modelling the space of intents and estimating the actual mapping. When the objective is quality for user preference, user evaluations can be a way of estimating the mapping.

The fact users are sharing their content and posting their opinions online provides a unique opportunity for understanding visual preference more in depth. From the variable facets of this visual preference: perceptual, semantic content, aesthetic or contextual, we mainly focused in this paper on the particular problem of semantically dependent image enhancement. To illustrate our approach with experiments, instead of considering the complete space of possible image enhancements, we restrict our approach to the variations that might be generated from a particular image enhancement approach and learnt a mapping between our semantic categories and enhancement space from user preference evaluations.

Finally, a simple example scenario is presented, showing how an SCDIE system can handle the problem of detecting images with special effects in a printing workflow in order to automatically turn off the enhancement process for those images.

REFERENCES

Adams, J. E., Hamilton, J. F., Gindele, E. B., and Pillman, B. H. (2003). Method for automatic white balance of digital images. US Patent 6573932, Kodak.

Allen, D. J., Carley, A. L., and Levantovsky, V. (2004). Method of adaptively enhancing a digital image. US Patent 6807313, Oak Technology, Inc. (Sunnyvale, CA).

Avril, C. and Nguyen-Trong, T. (1992). Linear filtering for reducing blocking effects in orthogonal transform image coding,. *J. Electronic Imaging*, 1(2).

Barnard, K., Duygulu, P., D. Forsyth, N. de Freitas, D. B., and Jordan, M. (2003). Matching words and pictures. *J. of Machine Learning Research*, 3.

Barnard, K., Martin, L., Coath, A., and Funt, B. (2002). A comparison of computational color constancy algorithms. *IEEE Trans. on Image Processing*, 11(9).

Battiato, S., Bosco, A., Castorina, A., and Messina, G. (2003). Automatic global image enhancement by skin dependent exposure correction. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*.

Bosch, A., Zisserman, A., and Munoz, X. (2006). Scene classification via pLSA. In *ECCV*.

Boutell, M. and Luo, J. (2007). Beyond pixels: Exploiting camera metadata for photo classification. pattern recognition. *Special Issue on Image Understanding for Digital Photos*. to appear.

Boutell, M., Luo, J., and Brown, C. (2006). Factor-graphs for region-based whole-scene classification. In *CVPR Workshop on Semantic Learning Applications in Multimedia*.

Bressan, M., Dance, C. R., Poirier, H., and Arregui, D. (2007). LCE: (automatic) local contrast enhancement. In *SPIE, Electronic Imaging*.

Buswell, G. (1935). *How People Look at Pictures*. Chicago University Press, Chicago.

Carbonetto, P., de Freitas, N., and Barnard, K. (2004). A statistical model for general contextual object recognition. In *ECCV*.

Chambah, M., Semani, D., Renouf, A., Coutellemont, P., and Rizzi, A. (2004). Underwater color constancy: enhancement of automatic live fish recognition. In *SPIE Electronic Imaging, Science and Technology*, volume 5293.

Chen, Y. and Wang, J. Z. (2004). Image categorization by learning and reasoning with regions. *JMLR*, 5.

Chiu, K., Herf, K., Shirley, M., Swamy, P., Wang, S., and Zimmerman, K. (1993). Spatially nonuniform scaling functions for high contrast images. In Kaufmann, M., editor, *Proc. Graphics Interface '93*.

Crandall, D. and Huttenlocher, D. (2006). Weakly supervised learning of part-based spatial models for visual object recognition. In *ECCV*.

Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C. (2004). Visual categorization with bags of keypoints. In *ECCV Workshop on Statistical Learning for Computer Vision*.

Csurka, G., Willamowski, J., Dance, C., and Perronnin, F. (2005). Incorporating geometry information with weak classifiers for improved generic visual categorization. In *Int. Conf. on Image Analysis and Processing*.

Datta, R., Joshi, D., Li, J., and Wang, J. (2006). Studying aesthetics in photographic images using a computational approach. In Leonardis, A., Bischof, H., and Pinz, A., editors, *ECCV*.

Devlin, K., Chalmers, A., Wilkie, A., and Purgathofer, W. (2002). Star: Tone reproduction and physically based spectral rendering. In *State of the Art Reports, Eurographics*.

DiCarlo, J. and Wandell, B. (2001). Rendering high dynamic range images. In *SPIE: Image Sensors*, volume 3965.

Durand, F. and Dorsey, J. (2002). Fast bilateral filtering for the display of high dynamic range images. *ACM Trans. on Graphics 21*, 3.

Eschbach, R. and Fuss, W. (1999). Automatic enhancement of scanned photographs. In *EI Color Imaging: Device Independent Color, Color Hardcopy and Graphic Arts IV (ei16)*.

Eschbach, R., Waldron, B., and Fuss, W. (1995). Us patent 5340502: Image-dependent luminance enhancement. Xerox Corporation.

Evans, R. M. (1951). Method for correcting photographic color print. US Patent 2571697, Kodak.

Everingham, M., Gool, L. V., Williams, C., and Zisserman, A. (2005). The pascal visual object classes challenge results. http://www.pascal-network.org/challenges/VOC/voc2005/results.pdf.

Everingham, M., Zisserman, A., Williams, C., and Gool, L. V. (2006). The pascal visual object classes challenge 2006. http://www.pascal-network.org/challenges/VOC/voc2006/results.pdf.

Fan, Z. and de Queiroz, R. (2003). Identification of bitmap compression history: Jpeg detection and quantizer estimation. *IEEE Trans. on Image Processing*, 12(2).

Farquhar, J., Szedmak, S., Meng, H., and Shawe-Taylor, J. (2005). Improving “bag-of-keypoints” image categorisation. Technical report, University of Southampton.

Fattal, R., Lischinski, D., and Werman, M. (2002). Gradient domain high dynamic range compression. *ACM Trans. on Graphics 21*, 3.

Fei-Fei, L. and Perona, P. (2005). A Bayesian hierarchical model for learning natural scene categories. In *CVPR*, volume 2.

Fergus, R., Perona, P., and Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. In *CVPR*.

Fergus, R., Singh, B., Hertzmann, A., Roweis, S., and Freeman, W. T. (2006). Removing camera shake from a single image. In *SIGGRAPH*.

Fidler, S., Berginc, G., and Leonardis, A. (2006). Hierarchical statistical learning of generic parts of object structure. In *CVPR*.

Fischer, M., Parades, J., and Arce, G. (2002). Weighted median image sharpeners for the world wide web. *IEEE Trans. On Image Processing*, 11(7).

Fredembach, C., Schröder, M., and Süsstrunk, S. (2003). Region-based image classification for automatic color correction. In *IS&T Color Imaging Conference*.

Furuki, I. and Yamada, K. (2006). Image enhancement device and image enhancement method of thermal printer. US Patent Application 20050168561, Mitsubishi Denki Kabushiki Kaisha.

Gallagher, A. and Bruehs, W. (2006). Method and system for improving an image characteristic based on image content. US Patent 20060228040, Eastman Kodak Company (Rochester, NY).

Gasparini, F. and Schettini, R. (2004). Color balancing of digital photos using simple image statistics. *Pattern Recognition*, 37.

Gasparini, F. and Schettini, R. (2005). Automatic redeye removal for smart enhancement of photos of unknown origin. In *Int. Conf. on Visual information systems*.

Gaubatz, M. and Ulichney, R. (2002). Automatic red-eye detection and correction. In *ICIP*.

Gonzalez, R. C. and Woods, R. (1992). *Digital image processing*. Addison-Wesley Pub. Comp, Inc., Reading, MA.

Haeberli, P. and Voorhies, D. (1994). Image processing by linear interpolation and extrapolation. *IRIS Universe Magazine, Silicon Graphics*, 28.

Henderson, J. and Hollingworth, A. (1999). High-level scene perception. *Annu. Rev. Psychol.*, 50.

Hillebrand, G., Miyamoto, K., and et al (2003). Skin imaging and analysis systems and methods. US Patent 6571003,The Procter & Gamble Company.

Hoffenberg, S. (2006). Changing cameraphone user behaviour. Half-day seminar at Photokina.

Ichikawa, T. and Miyasaka, T. (2005). Web print system with image enhancement. US Patent 6914694, Seiko Epson Corporation (Tokyo, JP).

Jalobeanu, A., Blanc-Fraud, L., and Zerubia, J. (2002). Estimation of blur and noise parameters in remote sensing. In *Int. Conf. on Acoustics, Speech and Signal Processing*.

Kim, N., Jang, I. H., Kim, D., and Hong, W. H. (1998). Reduction of blocking artifact in block-coded images using using wavelet transform. *IEEE Trans. Circuits and Systems*, 8(3).

Ledda, P., Chalmers, A., Troscianko, T., and Seetzen, H. (2005). Evaluation of tone mapping operators using a high dynamic range display. In *Proc. ACM SIGGRAPH ’05*.

Lefebvre, G., Laurent, C., Ros, J., and Garcia, C. (2006). Supervised image classification by som activity map comparison. In *ICPR*.

Leibe, B., Leonardis, A., and Schiele, B. (2004). Combined object categorization and segmentation with an implicit shape model. In *ECCV Workshop on Statistical Learning for Computer Vision*.

Li, J. and Wang, J. Z. (2003). Automatic linguistic indexing of pictures by a statistical modeling approach. *PAMI*, 25:9.

Li, Y., Bilmes, J. A., and Shapiro, L. G. (2004). Object class recognition using images of abstract regions. In *ICPR*.

Lin, Q., Atkins, C., and Tretter, D. (2002). Image enhancement using face detection. US Patent Application 20020172419, Hewlett Packard Company.

Lin, Q. and Tretter, D. (2005). Camera meta-data for content categorization. US Patent 6977679, Hewlett Packard Company.

Luo, J. (2003). Determining orientation of images containing blue sky. US Patent 6512846, Eastman Kodak Company (Rochester, NY).

Luo, J. and Etz, S. (2002). A physical model based approach to detecting sky in photographic images. *IEEE Trans. on Image Processing*, 11(3).

Marée, R., Geurts, P., Piater, J., and Wehenkel, L. (2005). Random subwindows for robust image classification. In *CVPR*, volume 1.

Martin, A., Doddington, G., Kamm, T., Ordowski, M., and Przybocki, M. (1997). The DET curve in assessment of detection task performance. In *EUROSPEECH*.

Meier, T., Ngan, K. N., and Crebbin, G. (1999). Reduction of blocking artifacts in image and video coding. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(3).

Minami, S. and Zakhor, A. (1995). An optimization approach for removing blocking effects in transform coding. *IEEE Trans. Circuits and Systems for Video Technology*, 5(4).

Motwani, M., Gadiya, M., Motwani, R., and Harris, F. C. (2004). A survey of image denoising techniques. In *Global Signal Processing Expo and Conference*.

Mutza, D. (2006). New fujifilm image intelligence: The next generation of automatic image quality optimization. In *International Congress of Imaging Science*. Fuji Photo Film (USA).

Neelamani, R., Choi, H., and Baraniuk, R. (2004). Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Trans. on Signal Processing*, 52.

Novak, E., Jurie, F., and Triggs, B. (2006). Sampling strategies for bag-of-features image classification. In *ECCV*.

Oberhardt, K., Taresch, G., and et al (2003). Method for the automatic detection of red-eye defects in photographic image data. US Patent Applications 20030044178,Milde & Hoffberg, L.L.P.

O'Hare, N., Gurrin, C., Lee, H., Murphy, N., Smeaton, A. F., and Jones, G. J. (2005). My digital photos: where and when? In *Annual ACM international conference on Multimedia*.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *PAMI*, 12(7).

Perronnin, F., Dance, C., Csurka, G., and Bressan, M. (2006). Adapted vocabularies for generic visual categorization. In *European Conf. on Computer Vision*.

Polesel, A., Ramponi, G., and Mathews, V. J. (2000). Image enhancement via adaptive unsharp masking. *IEEE Trans. On Image Processing*, 9(3).

Portilla, J., Strela, V., Wainwright, M. J., and Simoncelli, E. P. (2003). Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11).

Quelard, S. (2004). Image quality improvement for a cmos mobile phone digital camera. Technical report, KTH Stockholm Royal Innstitute of Technology.

Quelhas, P., Monay, F., Odobez, J.-M., Gatica-Perez, D., Tuytelaars, T., and Gool, L. V. (2005). Modeling scenes with local descriptors and latent aspects. In *ICCV*.

Ramamurthi, B. and Gersho, A. (1986). Nonlinear space-variant postprocessing of block coded images. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34.

Reeve, H. C. and Lim, J. S. (1984). Reduction of blocking effects in image coding. *Optical Engineering*, 23(1).

Rosenfeld, A. and Kak, A. (1982). *Digital picture processing*. Academic Press, Inc., New York.

Sadovsky, V., Yuan, P., Ivory, A. S., and Turner, R. (2004). Automatic analysis and adjustment of digital images upon acquisition. US Patent Application 20040258308, Microsoft.

Saito, T., Harada, H., Satsumabayashi, J., and Komatsu, T. (2003). Color image sharpening based on nonlinear reaction-diffusion. In *ICIP*.

Simon, R. and Matraszek, W. (2006). Method and system for enhancing portrait image that are processed in a batch mode. US Patent Application 7050636, Eastman Kodak Company (Rochester, NY).

Sivic, J., Russell, B., Efros, A., Zisserman, A., and Freeman, W. (2005). Discovering objects and their locations in images. In *ICCV*.

Stern, A., Kruchakov, I., Yoavi, E., and Kopeika, N. (2002). Recognition of motion-blurred images by use of the method of moments. *Applied Optics*, 41.

Sudderth, E., Torralba, A., Freeman, W., and Willsky, A. (2006). Learning hierarchical models of scenes, objects, and parts. In *ICCV*.

Tan, K. T. and Ghanbari, M. (2000). Blockiness detection for mpeg-2-coded video. *IEEE Signal Processing Letters*, 7.

Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *ICCV*.

Tumblin, J. and Rushmeier, H. (1993). Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6).

Tumblin, J. and Turk, G. (1999). Lcis: A boundary hierarchy for detail-preserving contrast reduction. In *SIGGRAPH*.

Tzou, K. H. (1988). Post-filtering of transform-coded images. In *SPIE: Applications of Digital Image Processing XI*, volume 974.

Utsugi, R. (2003). Method of correcting face image, makeup simulation method, makeup method makeup supporting device and foundation transfer film. US Patent 6502583, DRDC Limited (Tokyo, JP); Scalar Corporation (Tokyo, JP).

Xerox (2006). Xerox's Automatic Image Enhancement System. http://www.xerox.com/downloads/usa/en/f/FILE_PROD_AIE_Brochure.pdf.

Xiong, Z., Orchard, M. T., and Zhang, Y. Q. (1997). A deblocking algorithm for jpeg compressed images using overcomplete wavelet representations. *IEEE Trans. Circuits and Systems for Video Technology*, 7(4).

Yang, M.-H., Kriegman, D., and Ahuja, N. (2002). Detecting faces in images: A survey. *PAMI*, 24(1).

Zafarifar, B. and de With, P. H. N. (2006). Blue sky detection for picture quality enhancement. In *Advanced Concepts for Intelligent Vision Systems*.

Zhang, Y., Wen, C., and Zhang, Y. (2000). Estimation of motion, parameters from blurred images. *Pattern Recognition Letters*, 21.

Zhao, W., Chellappa, R., Phillips, P., and Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Comput. Surv.*, 35.

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In Press, A., editor, *Graphic Gems IV*.