

Reference-free Image Quality Evaluation for Digital Film Restoration

Majed Chambah

IUT de Reims-Châlons-Charleville, rue des Crayères, BP 1035, 51687 Reims CEDEX 2, France
Email: chambah@univ-reims.fr

Motion pictures represent an important cultural heritage. As films age they became prone to all kinds of defects such as dust, scratches, vinegar syndrome and dye fading, some of which are chemically irreversible processes. The quality of a restored movie tends to be estimated subjectively by experts. On the other hand, objective quality metrics do not necessarily correlate well with perceived quality. Digital restoration methods are now being developed which can augment traditional photochemical restoration techniques, or even address problems that are out of the reach of traditional methods. Also digital restoration has the advantage of not affecting the original material, since it works on a digital copy. Digital restoration techniques are becoming increasingly automated but restoration evaluation of their efficacy remains still a rarely tackled issue. This paper outlines some reference-free image quality metrics for photographs and digital film restoration. The metrics can assess some intricate defects in frames and sequences such as colour-faded films with more than one colour cast (two or more dominant colours). They permit the evaluation process to be speeded up. These measures can be used also to characterise an image sequence before its processing in order to automatically fine tune the parameters of the restoration techniques.

Introduction

Motion pictures represent an important cultural heritage. As films age they became prone to all kinds of defects such as dust, scratches, vinegar syndrome and dye fading. Several film defects such as dye fading and vinegar syndrome are chemically irreversible processes [1,2] and often a degraded version is the only available record of a film, hence the importance of digital film restoration. In addition to the fact that it can tackle artefacts that are out of reach of traditional photochemical restoration techniques, digital film restoration presents the advantage of not affecting the original material, since it works on a digital copy.

Several digital restoration techniques [3–10] have emerged during the last decade and became increasingly automated, but restoration evaluation remains still a rarely tackled issue [11,12].

Some caution must be exercised when doing quality assessment in the field of restoration. The material to be restored is not coming from a stable technological process that would produce the same basic quality before any degradation. The cinematographic technological methods have not changed notably since the beginning of cinema exploitation, keeping the same basic camera and projector movements, the same format size, the ubiquitous 35 mm film, and similar laboratory procedures, but the quality has always evolved.

Film quality has advanced in a stepwise fashion over time. For example, in the early days of colour film, the high-quality Technicolor process using two or three strips of film tended

not to be adopted for economic reasons. As a result there was actually a loss in quality, not only in resolution because of the combination of the three colour layers onto one strip of film, but also in the gamut of colour covered by the system, because of the complicated chemical process used.

It is therefore probably not a good choice to use absolute reference metrics when assessing the quality of a specific film. Thus, reliable automatic methods for visual quality assessment are needed in the field of digital film restoration. Ideally, a quality assessment system would perceive and measure image or video impairments just like a human being would do.

After presenting some important aspects of digital colour film restoration evaluation, this paper proposes some reference-free image quality metrics for photographic and digital film restoration. The proposed metrics can assess some intricate defects in frames and sequences such as colour-faded films with more than one colour cast (two or more dominant colours), allowing the evaluation process to be speeded up. These measures can be used also to characterise an image sequence before processing in order to automatically fine tune the parameters of the restoration techniques.

Defects Affecting Films

The goal here is not to present an exhaustive list of defects: indeed there is more than one way to present such a list. The point of view of the librarian does not coincide with the point of view of the image processing specialist. In order to avoid a lengthy and possibly controversial enumeration, a structured approach better suited for the problem at hand is employed.

First, I will describe the mechanical and photochemical origin of the film and the associated degradation. I will cover then electronic handling of images and how it affects their quality.

Defects Affecting Photochemical Materials

A film consists of a photochemical emulsion deposited on a base. The base must adhere firmly to the emulsion, staying chemically neutral for this emulsion and flexible. It must not be susceptible to deformation and must resist all the environmental and mechanical constraints that the technology imposes on the film.

Mechanical Degradation

Dust, dirt and thin scratches: Dust causes thin scratches on film and small particles of dust may penetrate the film base or the emulsion.

Elongated scratches: Protruding pieces of metal or small defects on metal surfaces, mainly in the camera during shooting, or later in other equipment can cause elongated vertical scratches which run across many frames.

Jitter (Image vibrations): Repeated loading, unloading, winding and rewinding of the film strips can also damage film holes that are supposed to guarantee a stable repeated positioning of the images during projection.

Missing parts and missing frames: Severe mistreatment of the film strip or repeated careless manipulation may cause tears and breaks of the strip.

Chemical Degradation of the Base Layer

Vinegar syndrome: This is degradation that affects film material using a cellulose acetate base, involving the hydrolysis of acetate groups which results in the formation of acetic acid (vinegar). The main problem during earlier stages is a deformation of the support which causes a variable local blurring effect. The problem can be addressed by a digital restoration method [13].

Chemical Degradation of the Emulsion

Contrast saturation: A common problem seen in old black and white films is the strong saturation of the black and of the white areas with a severe loss of middle tones.

Dye fading: The complex chemical composition of the emulsion in colour film is quite susceptible to the influence of radiation, temperature and humidity. Colour fading is caused by chemical changes in the dyes used in colour film. Many older films have taken on a distinct colour cast, caused by the rapid fading of one or two dyes. Colour negative film, colour slide film, colour print material, interpositives and colour motion-picture print film are all affected in the same way. The fading of one or two chromatic layers of the film results in a drab image with poor saturation and an overall colour cast.

Film reproduction degradation

Cinematographic films are reproduced by optical duplication mechanisms. Even when conducted with special care, this operation can result in deterioration in resolution.

Halo: During the printing operations a number of other degradations may occur. Improper adjustment in lighting can cause halos or uneven distribution in intensity in each frame.

Fog: Incomplete isolation from extraneous light sources may cause a fogging effect.

Flicker: Any mechanical misalignment during the optical printing may cause the phenomenon called flicker, in which the overall image intensity varies from frame to frame.

Digitisation

Image processing tools are applied in the digital domain, with image digitisation consisting of a sampling process followed by a digital encoding process. In the case of scanned photographic images there are two sources of noise: film grain noise and quantisation noise.

Traditional image capturing devices process images in linear manner. If we exclude the two extreme parts of the sensitivity scale of an imaging device, the measurements and therefore the numbers involved are roughly linearly proportional to the incoming light intensity. This is true for the photocell, for the now obsolete camera tubes and for CCDs.

In contrast cinema and photography have always used logarithmic scales and measurements. Film analysis is carried out through density measures, the logarithm of the inverse of transmission, and consequently the input scale used is the relative logarithm of exposure. Lens apertures and filter measurements are also expressed on logarithmic scales. A large number of film scanners provide their result on such scales, in order to mimic the photographic grading process within the digital domain.

The third type of processing is gamma compensation. This was defined in the early days of television technology in order to compensate for non-linear rendition in cathode ray tubes. The gamma value of a properly adjusted CRT is between 2.2 and 2.5. All video cameras incorporate such processing, and both analogue and digital video signals carry gamma-compensated values. The importance of the measuring system must not be underestimated because it is necessary to know which one has been used when it comes to deciding the weightings of quality measurements.

Digital Film Restoration Techniques

The restoration process starts with the conversion of each frame to a digital image using a high-resolution film scanner. The size of a digital colour image is up to 45 MB per frame (for 4000×3000 resolution). The digitised images are processed by workstations, with the restored images being finally recorded onto a new photochemical material.

Digital film restoration techniques differ from traditional image processing techniques mainly in terms of the large size of the processed images (high definition) that requires more computing power and storage capacity, and by the high number of images being processed (24 images per second). As an example, a hour and a half of a colour movie in 2K resolution represents $2048 \text{ lines} \times 1536 \text{ pixels} \times 3 \text{ channels} \times 129\,600 \text{ images} = 1200 \text{ billion pixels}$. Each pixel is encoded with 2 bytes (10 or 12 bits). The total storage capacity required for such a colour movie would therefore be 2.4 terabytes.

Usually, restoration techniques proceed in two steps. In the first step defects are detected and in the second step corrections are applied.

Several techniques for film restoration were developed especially for black and white films. Morris used so-called stochastic modals for interpolating [14]. Buisson and Mueller detected and corrected defects such as dust, jittering, flickering and scratches by using morphology detectors, movement detection and interpolation, and reconstruction methods [4,15]. Decencière Ferrandière has described morphological tools and spatiotemporal mathematical approaches to defect detection and correction [9]. Joyeux and Besserer developed techniques for scratch detection and tracking [5,6]; they used Bayesian correction techniques that gave nearly invisible defects on high-definition images. Kokaram has applied autoregressive (AR) filters to restore TV films (video definition) [16].

Gschwind restored faded colour film by establishing fading models based on accelerated bleaching experiments [17]. Chambah used assisted methods, semi-automatic colour constancy techniques and non-uniform saturation enhancement approaches to correct drab colours and colour casts caused by dye fading when no *a priori* information was available [8,10]. Rizzi and Chambah used unsupervised automated techniques based on perceptual principles to balance colours of images in order to correct colour casts induced by dye fading [3].

Image Quality Evaluation

Image quality evaluation techniques can be classified into two kinds: subjective and objective. These evaluations can be done with *a priori* information (with reference) or without use of *a priori* information (reference free).

Subjective Quality Evaluation

Subjective experiments, which to date are the only widely recognised method of determining actual perceived quality, are complex and time-consuming, both in their preparation and execution [18]. Subjective evaluation is formalised with defined procedures [19]. For instance, ITU-R Recommendation BT.500-11 suggests standard viewing conditions, criteria for observer and test scene selection, assessment procedures, and analysis methods [10]. The following paragraphs summarise some of the most common procedures [20].

Double Stimulus Continuous Quality Scale (DSCQS): Viewers are shown multiple sequence pairs consisting of a reference and a test sequence, which are quite short (typically 10 seconds). The reference and test sequence are presented twice in alternating fashion, with the order of the two chosen randomly for each trial. Subjects are not informed about which sequence is the reference or the test. They rate each of the two separately on a continuous quality scale ranging from 'bad' to 'excellent'. Analysis is based on the difference in rating for each pair, which is often calculated from an equivalent numerical scale from 0 to 100.

Double Stimulus Impairment Scale (DSIS): Unlike in the DSCQS method, the reference is always shown before the test sequence, and neither is repeated. Subjects rate the amount of impairment in the test sequence on a discrete five-level scale ranging from 'very annoying' to 'imperceptible'.

Single Stimulus Continuous Quality Evaluation (SSCQE): This is one of the most common reference-free procedures. Instead of seeing separate short sequence pairs, viewers watch a programme of typically 20–30 minutes duration that has been processed by the system under test; the reference is not shown. Using a slider, the subjects continuously rate the instantaneously perceived quality on the DSCQS scale from 'bad' to 'excellent'.

Objective Quality Evaluation

Objective quality evaluation uses metrics to evaluate image quality. Objective evaluation is automated, hence it costs less than a subjective evaluation, plus it can be done in real time since it needs no user interaction. Objective quality metrics can be classified into the following categories [18].

Full-reference (FR) metrics: FR metrics perform a direct comparison between the image or video under test and a reference or original. They thus require the entire reference content to be available, usually in uncompressed form, which is quite an important restriction on the usability of such metrics. Another practical problem is the alignment of the two, especially for video sequences, to ensure that the frames and image regions being compared actually correspond. Basic fidelity metrics such as mean-squared error (MSE), peak signal-to-noise ratio (PSNR), or colour difference (ΔE) belong to this class.

No-reference (NR) or free reference metrics: NR metrics look only at the image or video under test and have no need for reference information. This makes it possible to measure the quality of any visual content, anywhere in an existing compression and transmission system. The difficulty here lies in telling apart distortions from regular content, a distinction humans are able to make from experience and context.

Reduced-reference (RR) metrics: RR metrics are midway between no-reference and full-reference metrics. They extract a number of features from the reference image or video (e.g. spatial detail, amount of motion). The comparison with the image/video under test is then based only on those features. Additionally, image metadata as available with some file formats (e.g. EXIF, JPEG2000) can also be used. This makes it possible to avoid some of the pitfalls of pure no-reference metrics.

Image Quality Evaluation in Recent Studies

Lately, many studies and publications have emerged in the field of image quality evaluation. These works are influenced by the advent of digital image and video and new image coding needs (compression, transmission, etc.). In fact, the advent of digital video compression, storage and transmission systems has exposed fundamental limitations of techniques and methodologies that have traditionally been used to measure video performance. Traditional performance parameters have relied on the constancy of a video system's performance for different input scenes. Thus, one could inject a test pattern or test signal (e.g. a static multi-burst), measure some resulting system attribute (e.g. frequency response), and be relatively confident that the system would respond similarly for other video material (e.g. video with motion). A great deal of research has been performed to relate the traditional analogue video performance parameters (e.g. differential gain, differential phase, short-time waveform distortion, etc.) to perceived changes in video quality. While the recent advent of video compression, storage and transmission systems has not invalidated these traditional parameters, it has certainly made their connection with perceived video quality much more tenuous. Digital video systems adapt and change their behaviour depending upon the input scene. Therefore, attempts to use input scenes that are different from what is actually used in service can result in erroneous and misleading results. Variations in subjective performance ratings as large as 3 quality units on a subjective quality scale that runs from 1 to 5 (1 = lowest rating, 5 = highest rating) have been noted in tests of commercially available systems [21]. This explains why most recent works in the field of image quality evaluation deal with compression introduced impairment evaluation [21–26].

No-reference quality assessment is a relatively new research direction, with promising applications but little progress. Objective quality assessment is a very complicated task, and even full-reference quality assessment methods have had only limited success in making accurate quality predictions. Researchers therefore tend to break up the problem of NR quality assessment into smaller, domain-specific problems by targeting a limited class of artefacts, the most common being the blocking-artefact, which is usually the result of block-based compression algorithms running at low bit rates [26]. Therefore, subjective image quality assessment still remains the most used and the most efficient approach in fields such as printing quality [27,28] and digital film and picture restoration [3,11].

Issues in Restoration Evaluation

The Code and the Image

In the traditional photochemical world, in order to evaluate the state of a film for restoration or to assess its quality of conservation, it is only necessary to have a look at it. A quick examination of the base gives an indication of its ageing conditions. A projection is the only way to judge the quality. The film itself is the recording support (the base), the recording method

and medium (the emulsion), the storage medium, the viewing support and the reproduction support.

The digital film is dissociated from the recording, viewing and storing methods and supports. The digital code is not the viewed image; it is only the algorithmic potentiality of an image, when arranged and transmitted for display. The examination of the code itself does not indicate if there are defects, because the code itself is built to stay unaffected through multiple copies by virtue of its algorithmic nature. It is only because the digitally coded image is the representation (translation) of a bi-dimensional arrangement exhibiting such and such regularities that it is possible to speak of defects and to evaluate a quality.

These regularities, defined *a priori*, are those of a photochemical image recorded by a specific camera, reproduced by some laboratory process, kept on a certain base and digitised on a specific modern system. All the various devices involved have their characteristics, and the processes themselves will have taken place at certain dates, and all this may be known or not.

The question of how to evaluate the quality of a digitised film may be addressed by considering the control strategy, microstructures and macrostructures concerned.

Control Strategy

The first consequence of the separation of the code from the support is the necessity for a control strategy quite different from the traditional one.

It seems evident that the digitisation process must be fully analysed and all its characteristics documented in full details. As stated previously, coding systems are different and may give rise to quite different quantisation noises for example. Earlier techniques may not be known and may be only approximately dated, but this stage at least must be known.

The basic characteristics are the sampling method and dimension, the encoding system (linear, logarithmic or gamma) and the original quantisation. This may include also knowledge of the characteristics of the film being digitised. The nature of the film is important because negative and positive films have quite different contrast properties. The colour process is also interesting; it gives an indication of the possible colour gamut. The sampling of a piece of film without any image impressed on it gives an indication of the density of the base, the absolute minimum density. Lastly a sampling of a moderately dense flat image gives an indication of the grain size.

Moreover, if a quick look at the film base is enough to judge the conservation quality of one spool, there is nothing comparable in the digital world. It seems therefore imperative to do a search through all the images of the film or at least through a subset offering all the conditions for completeness of control. This is the condition for being able to know precisely the full range of the individual characteristics, amount of noise, grain characteristics, maximum contrast range, largest colour gamut, etc.

Degradation or Artistic Distortions?

A complete and thorough control strategy cannot solve everything. Another difficulty arises from the artistic nature of cinematographic work. Contrary to an industrial controlled environment, the artistic nature of the film work can generate considerable variations in the characteristics of the recorded images, which can cause problems in quality assessment. An example will illustrate this point clearly.

Sometimes when undertaking quality control on numerous sample images it is possible to find a certain part of a film to be affected by a specific colour problem. This part might

exhibit different grain and noise values to the rest of the film, but the scene compositions do not differ from those of any other scenes. The contrast values may be slightly lower and the colour saturation may have a general blue dominant colour. Is this a defect? A careful look at the sequence could show that it is a group of 'day for night effect' scenes in the film. But surely it could have been a bad copy or a bad transfer. How to solve this puzzle?

One possible answer to this question is to look at the structure of a film. If the transitions between normal quality scenes and scenes having this specific distortion are happening at the same location as scene transitions, it is almost certainly an artistic choice.

Microstructures and Macrostructures

We have seen that the structure of the film could guide the quality evaluation process. The Technicolor process example given in the introduction will serve to further illustrate the conclusion of the previous section. The decision not to adopt the highest quality Technicolor process resulted in a loss of quality in resolution and also in the colour gamut of colour films. Here we can see here two different kinds of qualities which must be treated separately.

It is certain that digital processing may give useful indications when considering the characteristics of noise, grain size, maximum gradient contours and measures of this sort. Those are the measurable microstructures of the images. But when it comes to different qualities like colour gamut, contrast and luminance, we are considering qualities that are relating to perception and to artistic expression. We must exercise some caution while measuring these characteristics. Their evaluation requires a knowledge of the technical representation systems and of their possible artistic or semantic usage.

Without having to investigate fully the semantic aspects of film, it is probably enough to accord a lower priority to the interpretation of these aspects of the visible construction of the film. The editing, the scene content to some extent and the camera movements are often macrostructural aspects that may help prevent misinterpretations. It is interesting to note that these macrostructures do not seem difficult to detect automatically.

Cinema Image Quality Evaluation

In the sphere of cinema, image quality is judged visually. In fact, experts and technicians judge and determine the quality of the film images during the calibration (post production) process. As a consequence, the quality of a restored movie is also estimated subjectively by experts.

On the other hand, objective quality metrics do not necessarily correlate well with perceived quality [18,29]. Plus, some measures assume that there exists a reference in the form of an 'original' to compare with, which prevents their usage in digital restoration, where often there is actually no reference with which to compare. This is why subjective evaluation is the most used approach and probably remains the most efficient. Nevertheless, subjective assessment is expensive and time consuming, so reliable automatic methods for visual quality assessment have their part to play in digital film restoration.

Ideally, a quality assessment system would perceive and measure image or video impairments just like a human being would. The following two approaches can be taken.

- *The psychophysical approach*, which is based on models of the human visual system [30]. This approach is usually based on the modelling of visual effects, such as colour appearance, contrast sensitivity and visual masking. Due to their generality, these metrics can be used in a wide range of applications; the downside to this is the high complexity of the underlying vision models. Besides, the visual effects modelled are best understood at the threshold of visibility, whereas image distortions are often supra-threshold.

- *The engineering approach*, where metrics make certain assumptions about the types of artefacts that are introduced by a specific compression technology or transmission link. Such metrics look for the strength of these distortions in the video and use their measurements to estimate the overall quality.

Based on the latter approach, a few studies of restoration quality (mainly on black and white films) are emerging [12], in order to characterise the impairments to detect (e.g. dust, flickering and scratches). Only limited success has been achieved. This is due to factors such as the absence of a comparison reference, the difficulty in characterising precisely the impairments affecting films, the high definition of images that makes defects very visible, the spatiotemporal dimension of the images to restore, and to the lack of correlation between the metrics and perceived quality. In fact, a metric may indicate that a scratch is less prominent after its correction, but perceptually an ill corrected scratch offends more than the original scratch, because we have become accustomed to seeing scratched movies. This example illustrates the complexity of some perception mechanisms and the difficulty in setting measures that correlate these mechanisms.

Some Reference-free Quality Metrics

This work has concentrated on evaluation tools and measures that corroborate perceived quality (and hence perceived impairments). We addressed a particular film defect, dye fading. As we have mentioned above, the effect of dye bleaching is seen as an overall colour cast and a loss of contrast, saturation and chromatic diversity. To evaluate colour fading impairment and colour correction results, objective tools and measures were devised to evaluate colour cast, chromatic diversity and naturalness of aspect.

Objective Tools

The notion of image quality is very closely aligned to the notion of image naturalness. As said by Ralph M Evans, the former Eastman Kodak company scientist, ‘a good quality image is one that does not offend you’ [31]. Many studies have focused on natural image descriptors, often for a specific field [32,33]. In the present study some objective tools to lessen subjectivity have been investigated in order to increase the reliability of subjective evaluation.

When considering correction techniques for colour fading, it was noticed that a ‘naturally’ colour-corrected image correlates with the similarity of the RGB histogram shape between the original (faded) image and the corrected one [3]. In fact, many artefacts such as noise and grain may be introduced by colour correction techniques. Such artefacts are visible on the RGB histogram shape, due to equalising effects. This tool can help in estimating the naturalness aspect of the restored image. A good restoration preserves the general original shape while a distorted histogram, such as an equalised one, reveals an unnatural aspect of the image (Figure 1).

Since faded images are involved, the most important evaluation is related to colour cast and chromatic diversity. Our evaluation is based on the hue polar histogram (HPH). It represents on a hue circle all the hues present in the image. Each hue is represented by a spoke (radius) joining the centre to the hue value. The histogram is made easier to read by plotting each spoke with the colour it represents. Hence the HPH summarises the chromatic diversity of the image. The length of the spokes is proportional to the size of the population having the same hue. It

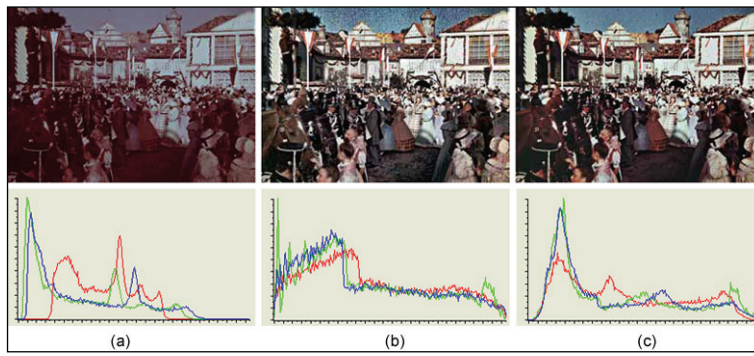
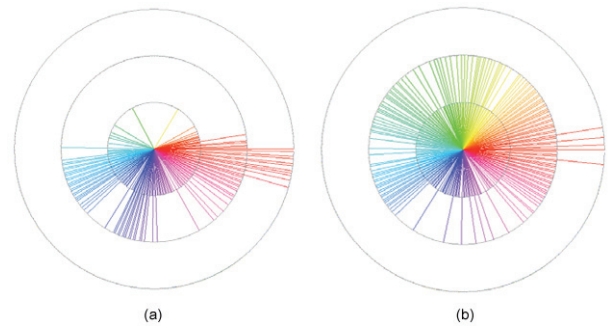


Figure 1 (a) Original faded image, (b) improperly restored image and (c) well restored image, together with their associated histograms

Figure 2 Hue polar histograms of (a) original faded image from Figure 1 and (b) well restored image from Figure 1



shows the predominant hues of the image. Figure 2 illustrates the usefulness of the HPH in evaluating the predominant colour and the chromatic diversity of the image.

Objective Metrics

Objective metrics can be obtained using these tools and these in turn permit a reference-free objective evaluation. The metrics used are subjected to standard data analysis and statistical techniques (principal component analysis and circular statistics).

The ACE model

Automatic Colour Equalisation, or ACE, is an algorithm for automatic digital image enhancement. It was developed following studies of mechanisms of the human visual system, and is able to enhance images and correct colour without any *a priori* knowledge of the image itself.

ACE implementation follows the scheme shown in Figure 3. A first stage encompasses a chromatic spatial colour recomputation, while a second stage, dynamic tone reproduction scaling, configures the output range to implement accurate tone mapping. The first stage

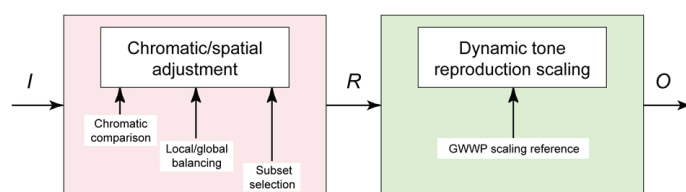


Figure 3 Basic ACE scheme

performs a sort of contrast enhancement, weighted by pixel distance. The result is a local-global filtering. The second stage maximises the image dynamic range, normalising the global lightness. No user supervision, statistics nor data preparation is required to run the algorithm.

In Figure 3, I is the input image, R is an intermediate result and O is the output image. The first stage, the chromatic/spatial adaptation, produces an output image R in which every pixel is recomputed according to the image content, approximating the visual appearance of the image. Each pixel of the output image R is computed separately for each chromatic channel.

The second stage maps the intermediate pixel array R into the final output image O . In this stage a simple dynamic maximisation can be made (linear scaling), and different reference values can be chosen in the output range to map into levels of grey the relative lightness appearance values of each channel. A more detailed description of the algorithm has been published [34].

An important property of ACE is its so-called quasi-idempotence. This means that if we apply ACE again on its own output it produces very little further effect. In other words, the first filtering is responsible for almost all the visual normalisation and the model produces a stable output.

Test Setup

For these experiments, three sets of test images were considered. The first set consists of different versions of a single outdoor image (Figure 4). The different versions were prepared using image processing software by altering of contrast and colour balance settings. The second set of images is of a Pantone colour checker taken with a digital camera using different white-balance and exposure settings (Figure 5). The third set consists of a variety of images: outdoor, indoors, night scenes, faded movie frames (Figure 6).

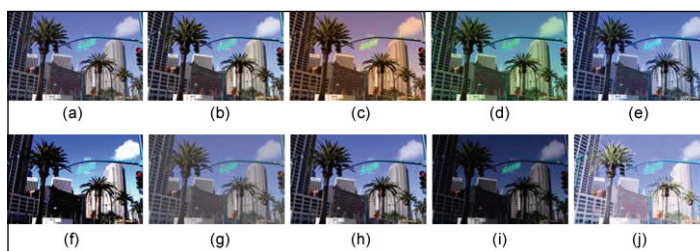


Figure 4 Images used in first test set: (a) original, (b) colours revived, (c) reddish illuminant, (d) greenish illuminant, (e) bluish illuminant, (f) high contrast, (g) low contrast, (h) darker shadows, (i) darker image, (j) lighter image

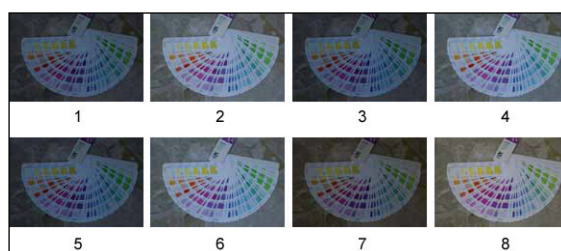


Figure 5 Images used in second test set (for key see Table 2)



Figure 6 Images used in third test set (for key see Table 3)

Since the proposed reference-free image quality measure is the difference between the image and its ACE-filtered version, all of the images have been filtered using a standard parameter configuration. Fine tuning of the measure was beyond of the scope of this preliminary work that aims at a qualitative and not quantitative result, hence no parameter tuning has been carried out in this test. Moreover, as this study deals with differential measures, absolute values are less significant than variation across images.

As a difference measure, ΔE in CIELAB space under illuminant D65 was chosen and computed between each original image and its ACE-filtered version (referred to as differential ACE filtering or DAF). Also two subjective quality evaluations were run on each set of images. The first subjective evaluation is a single stimulus quality evaluation (SSQE). Viewers were shown the original images one after the other. The ACE-filtered images were not shown. The subjects rated the perceived quality of the images instantaneously on a 5-grade scale from 'very poor' to 'very good'.

The second subjective evaluation was a double stimulus quality scale (DSQS). Viewers were shown multiple image pairs consisting of an original and an ACE-filtered sequence. The original and ACE-filtered images were presented twice in alternating fashion, with the order of the two chosen randomly for each trial. Subjects were not told which image was the original or the ACE-filtered version. They were asked to choose the better image between the two and to rate it from 'much worse' to 'much better' on a 5-grade scale. The equivalent numerical scale from 1 to 5 was then computed.

Experimental Results

First Test Set

Table 1 gives the DAF values for each version of the image and the rank of each image according to the DAF values, as well as the ranking of each image according to its visual quality as rated by the test subjects. The images having the worse quality were given lower DAF rankings and the better quality images were highly ranked according to DAF. In order to reach a more precise judgment about the adequacy of the DAF rank according to the visual range, the correlation between DAF rank and visual rank was computed to be 0.63. This was judged to be a relatively high score, confirming the validity of this preliminary work. However, it is worth stressing that the visual rank is subjective and may vary according to the users involved.

Table 1 Results of first test set

Image	DAF value	Rankings	
		DAF	Visual
Original	21.11	5	2
Colours revived	17.87	2	1
Reddish illuminant	18.80	3	7
Greenish illuminant	19.55	4	8
Bluish illuminant	33.62	8	5
High contrast	45.65	10	9=
Low contrast	17.41	1	3
Darker shadows	23.15	6	4
Darker image	44.49	9	9=
Lighter image	26.63	7	6

Second Test Set

The image used for the second test was of a Pantone folder lying on a marble surface acquired under sunlight coming through a window. At each white balance setting used, a correctly metered image was shot together with one underexposed by 0.7 EV. Table 2 reports all the combination together with the DAF rank between the original and the ACE-filtered version.

The lower DAF values were associated with the images that were correctly exposed. Also images with less colour cast had the lower DAF values. The lowest DAF value was associated with the image with the best colour balance, i.e. the one taken with automatic white balance and correctly exposed.

Table 2 Results of second test set

Image no	White balance	Exposure value	DAF value
1	5500 K	-0.7	37.78
2	5500 K	0	22.14
3	Neon	-0.7	38.20
4	Neon	0	23.95
5	Sunlight	-0.7	38.54
6	Sunlight	0	23.69
7	Automatic	-0.7	34.65
8	Automatic	0	21.90

Third Test Set

As stated above, users were asked to compare each ACE-filtered image to the original image (unfiltered) on a 5-grade scale from much worse (-2) to much better (+2). In order to provide a more precise judgment about the adequacy of the DAF measure compared with the visual judgment rating of the original images, the correlation between DAF and visual judgment rating was computed and the value was found to be -0.40, in other words not very high.

Discussion

The correlation between DAF and various reference-free measures has been investigated and discussed elsewhere [35,36], where the conclusion was drawn that DAF is only loosely correlated with these other measures. These studies assessed more quantitatively the comparability of these reference-free measures, including DAF, and their relationships with visual judgment ratings. A procedure of stepwise regression gave a model based on three variables to predict the visual rating: DAF and two other reference-free metrics [35,36]. The result confirmed that DAF is a complementary measure to the other developed measures and it is useful in enhancing predictions of visual quality.

We applied the same procedures with DAF and visual comparison (between original and ACE-filtered image) rating (Table 3, 4th column). The results obtained were not significant, partly because the visual ratings reported by the test subjects varied widely throughout the experiment. While some users judged the ACE-filtered images to be of better quality, some others found the original image to be superior (that is why the mean of the visual comparison is close to zero and the standard deviation was relatively high for some images).

Table 3 gives the mean of these evaluations (columns 3 and 4) along with the DAF values (column 2). The fully restored clip is shown in Video 1.

Table 3 Results of third test set

Original image	DAF value	Visual judgment of original image ^a	Visual comparison between ACE and original ^b
1	41.76	2.95	0.53
2	14.69	3.55	0.56
3	36.01	2.73	-0.56
4	23.74	2.91	0.56
5	14.45	2.64	0.22
6	49.09	1.91	0.89
7	39.36	2.09	1.33
8	23.42	3.64	-0.31
9	65.57	1.64	0.89
10	36.08	3.41	-0.17
11	11.25	2.32	1.22
12	17.39	2.50	1.50
13	16.90	2.18	0.39
14	16.22	2.05	1.56
15	59.67	2.02	0.56
16	30.07	1.98	1.67

a 1	Very poor	b -2	Much worse
2	Poor	-1	Worse
3	Acceptable	0	Similar
4	Good	1	Better
5	Very good	2	Much better

Video 1



Conclusion

The issue of image quality assessment in the field of digital film restoration field has been addressed in this paper. The defects that affect films and the issues of quality assessment in the field of film restoration have been reviewed, and some image quality measures for detecting and quantifying colour faded materials have been discussed.

Various reference-free metrics have shown their efficiency in helping to assess images objectively, and to speed up restoration quality estimation. The differential ACE filtering (DAF) metric is based on the idea that the worse the quality of a given image the greater will be its correction by the ACE algorithm, which enhances images in a perceptual way and gives an output that is an estimate of our visual perception of a scene. The results show that the metric presented correlated reasonably well with the perceived visual quality as judged by users. This fact is very important especially since the experiments have showed that DAF has a role to play along with other reference-free quality metrics, and is complementary to these measures.

References

1. J M Reilly, *Storage Guide for Colour Photographic Materials*, The New York State Program for the Conservation and Preservation of Library Research Materials (Albany, NY: New York State Library, 1998).
2. *Preservation and Reuse of Motion Picture Film Material for Television*, Guidance for Broadcasters Tech 3289-E (European Broadcasting Union, March 2001).
3. M Chambah, A Rizzi, C Gatta, B Besserer and D Marini, Proc. SPIE/IS&T Electronic Imaging 2003, Santa Clara, CA, **5008** (2003) 138–149.
4. O Buisson, B Besserer, S Boukir and F Helt, *Deterioration Detection For Digital Film Restoration*, Proc. Conf. on Computer Vision and Pattern Recognition (CVPR 1997).
5. L Joyeux, S Boukir, B Besserer and O Buisson, *Image Vision Computing*, **19** (8) (2001) 503–516.
6. B Besserer and C Thiré in *Computer Vision – ECCV 2004* (Berlin: Springer, 2004) 264–275.
7. P Kornprobst, R Deriche and G Aubert, Proc. 5th European Conf. on Computer Vision, **2** (1998) 548–562.
8. M Chambah, B Besserer and P Courtellemont, Proc. IS&T Conf. on Graphics, Imaging, and Vision CGIV 2002, Poitiers (April 2002) 613–618.
9. E Decenci ere Ferrandi ere, PhD thesis, Ecole Nationale Sup erieure des Mines de Paris (1997).
10. M Chambah, B Besserer and P Courtellemont, *Machine Graphics Vision*, **11** (2/3) (2002) 363–395.
11. M Chambah, Proc. 8th World Multi-Conf. on Systemics, Cybernetics and Informatics (SCI 2004) (Colour Image Processing and Applications Invited Session), Orlando, FL (2004).
12. E Decenci ere, Proc. IEE Seminar on Digital Restoration of Film and Video Archives (Jan 2001) 1/1–1/6.

13. F Helt and V La Torre Proc. IEE Seminar on Digital Restoration of Film and Video Archives (Jan 2001) 4/1–4/7.
14. R D Morris, PhD thesis, University of Cambridge (1995).
15. H Mueller-Seelich, W Plaschzug and K Glaz, Proc. SPIE, **3309** (1998) 287–296.
16. C Kokaram, PhD thesis, University of Cambridge (1993).
17. R Gschwind, F S Frey and L Rosenthaler, Proc. SPIE, **2421** (1995) 57–63.
18. S Süssstrunk and S Winkler, Proc. IS&T/SPIE Electronic Imaging 2004: Internet Imaging V, **5304** (2004) 118–131. <http://infoscience.epfl.ch/record/33957>.
19. M Stokes and T White, Proc. IS&T/SID 6th Colour Imaging Conf. (1998) 258–262.
20. ITU-R Recommendation BT.500-11 (Geneva: International Telecommunication Union, 2002).
21. M Pinson and S Wolf, National Telecommunications and Information Administration (NTIA) Report 02-392 (June 2002).
22. L Lu, Z Wang, A C Bovik and J Kouloheris, Proc. IEEE Int. Conf. on Multimedia and Expo, **1** (2002) 61–64.
23. Z Wang and A C Bovik, Proc. International Conf. on Multimedia Processing and Systems (Aug. 2000). <http://www.cns.nyu.edu/~zwang/files/papers/icmps.pdf>
24. H R Sheikh, Z Wang, L K Cormack and A C Bovik, Proc. 36th Annual Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA (Nov 2002).
25. M Pinson and S Wolf, *IEEE Transactions on Broadcasting*, **50** (3) (Sept 2004) 312–322.
26. Z Wang, H R Sheikh and A C Bovik, Proc. IEEE International Conf. on Image Processing, Rochester, NY (Sept 2002) 477–480.
27. L C Cui, Proc. SPIE Electronic Imaging Conf. (Image Quality and System Performance), San José CA, **5294** (Jan 2004).
28. C Cui, Proc. 8th World Multi-Conf. on Systemics, Cybernetics and Informatics (SCI 2004) (Colour Image Processing and Applications Invited Session), Orlando, FL (2004).
29. Z Wang, A C Bovik and L Lu, Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (May 2002).
30. S Winkler, PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Switzerland (2000).
31. R Govil, Proc. IS&T Image Processing, Image Quality, Image Capture, Systems Conf., Portland, OR (1998) 1–3.
32. H R Sheikh, A C Bovik, and L K Cormack, Proc. 37th IEEE Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA (Nov 2003).
33. S N Yendrikhovski, F J Blommaert and H de Ridder in *Colour Imaging: Vision and Technology*, Ed. L W MacDonald and M R Luo (New York: Wiley and Sons, 2000) 363–382.
34. A Rizzi, L Rovati, D Marini and F Docchio, *J. Electronic Imaging*, **12** (3) (July 2003) 431–441.
35. M Chambah, C Saint Jean and F Helt, Proc. SPIE/IS&T Electronic Imaging Conf. (Image Quality and System Performance II), San José, CA (Jan 2005) 220–231.
36. M Chambah, C Saint Jean, F Helt and A Rizzi, Proc. SPIE/IS&T Electronic Imaging Conf. (Image Quality and System Performance III), San José, CA (Jan 2006) 245–256.