Colour Stability in Live Image Capturing

Guy K. Kloss, Napoleon H. Reyes and Ken A. Hawick

Computer Science
Institute of Information & Mathematical Sciences
Massey University at Albany, Auckland, New Zealand
Email: {G.Kloss | N.H.Reyes | K.A.Hawick}@massey.ac.nz

Abstract. Digital colour cameras are dramatically falling in price, making them affordable for ubiquitous appliances in many applications. An attempt to use colour information reveals a significant problem that usually escapes our awareness. Due to the adaptive nature of the human visual system in most cases we do not recognise most changes in illumination characteristics, a camera however will measure scenes under changing illumination differently. An image's colour depends on three basic factors. The physical content of the scene, the scene's illumination, and the capturing characteristics of the camera. Attempts to deduce object colour from the images will need to cope with the influence of the illumination and the camera's characteristics. Furthermore a large variety of colour spaces are available to describe colour. Differences between them and their fitness to quantify colour are discussed. Also the respective suitability for colourimetric computations are covered. This paper tries to establish a basic understanding of the intricacies behind the processes involved in capturing images and recognising colour – from light as a stimulus to the colour sensed values in cameras. The myriad of factors affecting colour perception by machine vision systems and the different technologies that evolved to address those issues are discussed. The goal is to outline a novel approach fusing common industrial best practices with dynamic adaptation capabilities needed for robustly monitoring colour in real-time. First positive results towards improving colour based reasoning on adaptable colour spaces are stated as an outlook for further development directions.

Keywords: Chromatic adaptation; colour management; colour vision; computer vision; digital video imaging; artificial intelligence; colour constancy; colour spaces.

1 Introduction

In most cases pictures of scenes do not exactly match the original capture. The pictured scene does not match from colour measurements or even from an appearance standpoint.

Besides the artistic or documentary purpose of photography and filming, with increasing computing capabilities digital cameras are increasingly used for analytical image capturing. Images are processed to extract information. This information may be based on shapes, (3D) objects (through multiple cameras or frames), colour, etc. Analyses of the artistic or aesthetic image content is of lesser interest, but rather precision is wanted.

The following scenario will provide an impression of what an image processing system may need to be able to cope with: A scene during the day is observed by a camera in a room with a window (no artificial light) and an overcast sky (quite "neutral" daylight illumination). The weather clears up to a spotless blue sky, and the light changes to a more bluish shade and is brighter. In the evening sun rays directly fall in through the window onto the scene to increase the light intensity further and "paint the scene" in yellow/orange shades while adding hard shadows. During dusk, as the sun sets, a person turns on fluorescent light in the room with yet another shade and light intensity; additionally the spectrum of the fluorescent light is composed differently. Ideally the scene's colour composition is detected properly (after digital image processing), regardless of the camera images' brightness, dynamic range, and colour shifts due to the different lights' colour compositions. Our eyes tend to present a "filtered" representation by adapting to the conditions quite well, the raw camera images do not. We perceive a white piece of paper within the scene as white, regardless whether viewed in the direct sun or candle light.

This argumentation also leads to the insight that models of human vision may not merely be derived by measurement only. One theory for demystifying the human vision system is to try to understand it by gaining an understanding of the *computational* problems to be solved to obtain equivalent results.

Next we will discuss (digital) cameras as measuring devices for chromaticity. In the following Sect. 2 some background colour theory for quantitative handling of colour information in images is established. The next Sect. 3 introduces solutions for coping with problems faced by changing illumination conditions. Sect. 4 discusses various existing techniques on solving these problems adaptively and projected enhancements for the general live image capturing. Finally in Sect. 5 we focus on an implementation for real world problems.

2 Colours in the Digital World

2.1 Cameras as Measuring Devices

"What we observe is not nature itself,
but nature exposed to our method of questioning."
– Werner Heisenberg, *Physics and Philosophy* (1958)

To understand problems in quantifiable colour capturing, we first have to achieve a better understanding of the image capturing process within digital cameras [1,2]. This process can be dissected into three distinct stages: The physical capturing of the image on the sensor, the rendering of the scene, and finally the encoding of the scene representation.

Digital colour cameras nowadays are equipped with a light sensitive sensor array (CMOS or CCD sensor). It contains a matrix of single sensors that are (e. g. through filters) responsive to specific ranges of radiation in the visible light spectrum. From each sensor a reading is acquired after the exposure.

The firmware of the camera then processes the readings. Sensor responses are rendered into the destination image. This stage compensates for improper illumination, white balance, etc. Manufacturers intend to achieve by rendering a "most visually pleasing representation" of the scene for the customer. Therefore, in most cases sensor readings *will not* be passed through, but altered. Every camera handles this scene rendering differently (in an unknown way). Some professional camera models offer output of "raw" images. These eliminate the internal rendering process, however, still leaving colour measurements influenced by the illuminant and the camera's specific sensor characteristics.

Finally the rendered image representation is encoded into the image format, usually in a standard compliant colour space as sRGB. Unfortunately, this standard compliant colour space encoding suggests a colour compliant handling of the imagery. Due to the rendering process this is however not the case.

2.2 Colour Theory

The origins of colour lay in the distribution of the intensities in the *visible spectrum*. The human eye can perceive light in the range for wavelengths of about 400 to 700 nm.

In the most general case light is reflected on the surface of bodies (Fig. 1). Light is emitted from a source (natural, artificial, or due to reflectance from other objects) and incides on a surface. A fraction of the light is reflected towards the eye or an observant detector (camera). A lens focuses the incoming light onto a light sensitive photo receptor.

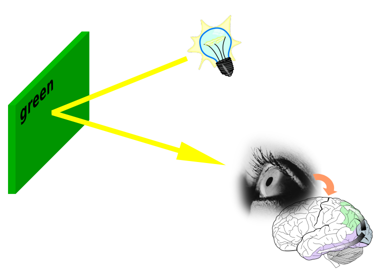

Fig. 1: A simple model of image formation: light from a source is reflected on a (coloured) object's surface. A fraction of the (coloured) light is reflected towards the eye or camera.

Colour Sensing The incident light's spectral power distribution (SPD) (Fig. 2, top left) is modified upon reflection on the surface. The reflectance is characterised by a wavelength dependent function (Fig. 2, top right).

In the human eye the sensitive surface is the retina containing receptors, the so called rod cells and cone cells. Rod cells are very light sensitive and used for night vision, they only detect lightness. Cone cells contain a colour pigment on their tip making them responsible for colour vision. These cells can be distinguished into three types, the *short (S), medium (M),* and *long* wavelength *(L)* sensitive cells. Similarly cameras possess sensors detecting three different basic colours *red (R), green (G),* and *blue (B).* The resulting SPD after reflection is sensed either in the eye or camera as described by the three colour channel's sensitivity functions in Fig. 2 (bottom diagrams).

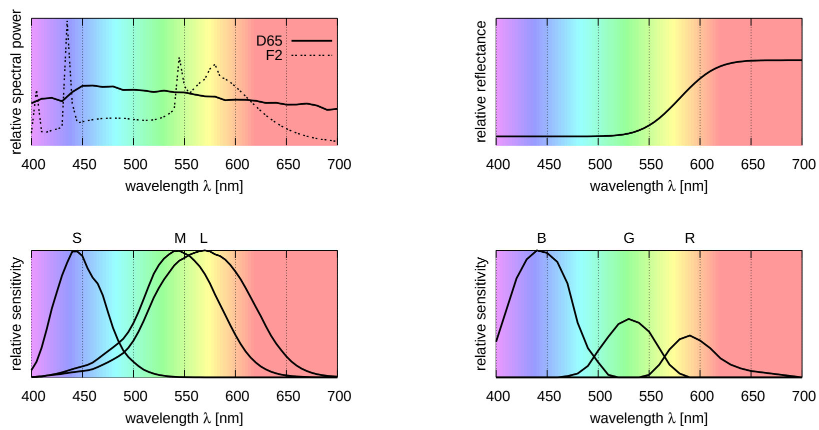

Fig. 2: Spectral power distribution of CIE D_{65} daylight (6504 K) and F2 cool white fluorescent illuminant (4200 K), as specified by the CIE (top left, see Sect. 2.2 for a description of CIE). A bright red surface's spectral reflectance (top right). Relative sensitivities of the cone cells within the human eye for the S, M, and L cells ($S(\lambda)$, $M(\lambda)$, $L(\lambda)$) (bottom left). Relative sensitivities of a typical digital camera's R, G, and B sensors ($R(\lambda)$, $G(\lambda)$, $B(\lambda)$) (bottom right).

Colour as detected by a camera is the integral response over the visible range. As the SPD is condensed into a tuple (R, G, B), all spectral information is lost. Further processing based on the tristimulus values depends on assumptions of boundary conditions, usually of the illuminat and its colourimetric stability.

Furthermore, it needs to be well understood, that the eye *does not* perceive colour by a red, green, and blue component. The additive RGB primaries have been chosen at a time for technical reasons – as their distribution across the visible spectrum produces a wide-gamut colour representation – not because they match the eye's response in a particularly accurate way.

Colour Spaces In 1931 the CIE (Commission Internationale de L'Éclairage) standardised a quantified representation of human colour vision. It uses a set of imaginary *red, blue,* and *green* primaries, that – when combined – cover the full gamut of human colour vision [3]. The standard CIE RGB (determined experimentally) contains negative values. Certain pure monochromatic spectral colours could not be matched entirely using the three selected primaries. Therefore, it was necessary to move the red light in the experiment over to the side of the monochromatic light, resulting in a negative sign.

The visual impression of a colour is based on three factors (Sect. 2.2): the properties of the light, surface reflectance, and sensor. Experiments determined the perception properties of an "averaged" human observer as a baseline for comparative measurements, standardising the third factor. To deduce the reflectance properties of a surface (its colour) quantitatively, also the light needs to be standardised. A CIE illuminant D_{50} (daylight at 5000 K "colour temperature[1]," similar to D_{65} illuminant in Fig. 2) was the declared default for the CIE colour spaces.

To avoid negative values a linear transformation was introduced, yielding the XYZ coordinate system. The human visible gamut entirely fits into positive values of X, Y, and Z. Y was defined to reflect the overall lightness of the colour tristimulus. The humans' capability to see (in colour) is the fundamental reason for us to be interested in quantifying colour. CIE XYZ was the first general attempt, based on the common denominator of human vision in contrast to technical oriented solutions (as RGB).

Colour space for technical purposes (RGB) and visually oriented purposes (XYZ) are available. Neither of the two are without their flaws: they are not perceptually uniform. A simple way to introduce more visual linearity into RGB components is the concept of *gamma correction.* This exponentially distorts the linearity of the colour channels yielding an R'G'B' colour space. Commonly used systems today are highly ambiguous in this respect, as it is commonly not distinguished whether the channels in "RGB" are linear or gamma corrected, and what gamma value is used.

Device Independent Colour Spaces Even greater flaws in perceptional non-linearity of RGB and XYZ exist. Just noticeable colour differences are much greater in some regions (of the RGB or XYZ colour space) than in others.

Additionally, each device is different in its colour sensing characteristics. Cameras often represent images in RGB, YUV, or other colour spaces, and

[1] The colour temperature of a light source is determined by comparing its chromaticity with a theoretical, heated black-body radiator. The Kelvin temperature at which the heated black-body radiator matches the colour of the light source is that source's colour temperature.

screens for example use an RGB input. But the displaying properties (e. g. through the phosphorus) are usually different from those of the capturing device. Therefore, it is important to distinguish between device dependent (RGB, YUV, HLS, CMYK, etc.) and device independent colour spaces that have been "normalised" (e. g. CIE XYZ).

A perceptionally linearised, device independent colour system is needed. One of them has been established in 1912 by the American artist Albert H. Munsell [4]. He was the first to separate hue, value, and chroma into perceptionally uniform and independent dimensions. In 1976 the CIE specified simultaneously two colour spaces as perceptionally uniform versions of the XYZ colour space that today have superseded most other models, these namely are the L*a*b* and the L*u*v* colour spaces.

CIE L*u*v* has been a popular colour space for lighting, video, and photographic applications. The linearisation of this space has been achieved pragmatically by means of various projection transformations to approximate the perceived difference between colours. In contrast the values in the CIE L*a*b* space are computed through four transformation steps, which are intuitively logical when examined separately, and have been derived through a combination of perceptionally sensible operations.

It is considered sensible to always use CIE L*a*b* to represent surface colours, and not CIE L*u*v*. Validation studies [5] of several colour appearance models have been performed. Overall, the CIE L*a*b* typically performed well against more complex models and occasionally was as good as the best; CIE L*u*v* was often the worst.

3 Chromatic Adaptation

Sect. 2.2 states that the colours received from cameras (e. g. in RGB) are device dependent and perceptionally not uniform. To arrive at a device independent colour representation (as CIE L*a*b*), a transformation has to be performed. This is commonly achieved by using a transformation profile as standardised by the International Color Consortium (ICC). These *ICC profiles* describe a mapping of device colour to independent colour, and they are determined through a calibration process capturing a (larger) number of precisely known colour patches (see activity diagram in Fig. 3).

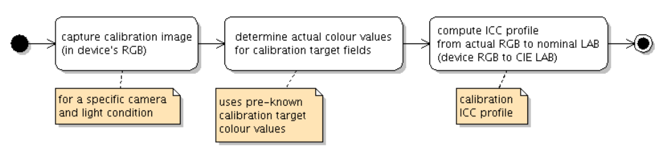

Fig. 3: Static calibration of camera setup to derive an ICC profile for a camera under given light conditions.

As outlined in Sect. 2.2, measurements of colours with cameras are also dependent on illumination properties. They are only standardised and directly usable together with CIE D_{50} light. For alternative illuminants, transformations need to be performed on the colour tristimuli. This process is a computational replication of the human eye's psycho-physical perception phenomenon. It is called a *chromatic adaptation transformation* (CAT). A CAT needs to be performed whenever a colour tristimulus is to be viewed under another but its reference white (illuminant). After transforming the XYZ colour tuple to the cone response domain (LMS, see Sect. 2.2), it is scaled with the relative shift of the white point. Several variations of this so called *von Kries* transformation are in use. The most commonly accepted today being the (linear) Bradford transformation [6,7], using a 3×3 matrix multiplied with the XYZ colour vector.

Colour tuples are only meaningful in conjunction with their respective illuminant's white point. The accuracy becomes progressively less accurate as the colour is farther away from neutral greys, and the more the illuminants differ from each other (or from the standardised CIE daylight SPDs) [6].

To use cameras as sensible quantitative measuring devices for colour, the camera's colour mapping characteristics (in the form of e. g. an ICC profile) and the current illumination conditions must be known. The camera's image then is first normalised (using its specific ICC profile) and then can be chromatically adapted from the current illumination to a standardised light source. Colour information acquired this way can be analysed quantitatively without severe colour aberrations.

4 Adaptive Solutions

Under common conditions the biggest problem is the fact that the light source in a scene is not known and/or not constant. As we know from our own experience, our eye can adapt quite well. This is true regardless of the commonly encountered light source. A lemon does not look greenish under a blue sky or orange at the reddish light at dusk, even though the light our eye senses is of that shade. We do not perceive the object alone, but the object within its context, and the "scene rendering" algorithm (see Sect. 2.1) in our brain performs an adaptation.

A similar solution is desired for technical systems using a camera. Cameras do some of that in the rendering process already, but unfortunately not towards the goal of creating stability in colour, but rather to create a "pleasant" balance in the representation. The exact process itself is unknown and not constant across cameras by different vendors.

4.1 Current Chromatic Adaptation Attempts

Besides sensible colour quantisation for technical applications these are still missing: algorithms to estimate the properties of the (current) illumination and methods to adapt the ICC profile or other transformations in a changing environment.

Barnard has established a comprehensive overview of various algorithms for illuminant estimation and chromatic adaptation and evaluated their relative strengths and weaknesses [8]. All analyses were conducted in device dependent colour spaces. Additionally, colour differences were quantified as Euclidean distances in non-linearised colour spaces. However, colour comparisons should be performed in a suitable colour space. For this usually a ΔE value derived from CIE L*a*b* is considered to be current good practise [5]. This does not reduce the quality/effort of the analyses, but shows a direction for further research in this field.

Porikli has taken a completely different approach [9]. He also stays entirely in the camera's own colour space. The imaging characteristics are mapped directly between two distinct cameras or illumination conditions. For this initially a direct mapping through histogram cross-correlated of the colour channels is performed. Therefore, the sensor responses are directly mapped for known camera or lighting combinations. This circumvents the problem of accumulative errors in a series of transformations. However, at the cost of necessary individual pair mappings for each setup, and no gain in additional knowledge on conditions. Currently only results for the case of scenes with similar colour composition have been published. Individual channels can be transformed with complex arbitrary mapping functions. But as applied on each single channel, no shearing or warping of the colour space is possible. Still this method may prove to be useful if applied in other colour matching attempts.

The methods described above depend on a static scene analysis (to deduce illuminant information) or on a single initial calibration of a distinct pairing. Many other methods are known to be widely used that are based on a priori knowledge of scene properties, as distinct colours of object markings (e. g. ball colour, team markings, goal marks, etc. in robot soccer).

4.2 Research Attempt

In general cases objects with a priori known colouring are not usable. Specific objects cannot be expected to often reoccur in a scene. But we can take advantage of the fact, that in most scenes are changing slowly. Therefore, the background can be expected to be more stable than a commonly quickly changing foreground. By detecting changes in the background, it is possible to derive an adapted CAT through modifying an ICC profile. In live image capturing changes between consecutive frames are generally small, thus a fairly robust statistical background model can be established.

Our current research aims at developing practical solutions for detecting changes in the current illumination and adapting the CAT accordingly. This is limited to the common case of slow changing backgrounds as stated. In cases of failure due to a sudden strong illumination change (e. g. someone operates a light switch) we can assume a tolerable error induced by re-using a background segmentation mask from a previous frame. If, however, the scene changes rapidly, the analysis can be pre-pended by a suitable approach as outlined by Barnard [8] to determine an estimation of a scene's illumination.

Furthermore, it needs to be examined, whether the approach by using colour histogram cross-correlations [9] can be modified towards moving from a direct mapping of two device dependent representations into a device independent intermediate representation suitable for sensible image analysis.

Lastly the image perception process of humans comprises of a complex layering of physical, neural, and psychological contributions. Most possibly various means of Artificial Intelligence (AI) can be helpful in increasing the efficiency and robustness of the adaptation process. It is for example foreseeable, that in an application distinct illuminations are predominant, and certain changes are re-occurring. A system employing AI can "learn" these conditions and changes to adapt more quickly to them, as responses are available from "experience" already. For these cases e. g. Neural Networks, Genetic Algorithms, or Artificial Immune Networks (combining features of learning and response-memory) could prove to be suitable. In more robust comparison and decision making (e. g. in the background/foreground segmentation) Fuzzy Logic may also be beneficial.

For tests of the suitability of this system verifications against the best candidates of Barnard [8] should be tested. For a fair comparison an attempt to improve these algorithms from Barnard towards using device independent colour spaces should be made. Furthermore, Porikli's histogram based cross-correlation model function should be similarly tested towards usability in device independent, linearised colour spaces, and in more dynamic environments than stated in [9].

5 Implementation

Even though this system is intended to be used on live camera image capturing, the computational overhead should not be the limiting overall factor. Illumination changes tend to happen slowly, and cases of quick changes are detectable. Therefore, a potentially computationally expensive adaptation of the ICC profile is tolerable, and it can be updated at intervals. In case of strong sudden changes, the system will suffer a brief period to adapt to the changed situation, but this also happens to humans in day to day situations. If the overall system gains robustness, this situation of momentary "colour blindness" can be tolerated.

Due to the common slow changes in lighting conditions, algorithms determining the dynamic nature of changing chromatic adaptation may be decoupled. Tasks may be performed in separate threads or processes, possibly even on a separate system, freeing resources for the live, real time image processing. Fig. 4 (top) outlines in an UML activity diagram the implementation of the general image processing loop. The process graphed underneath shows the loop for the ICC profile adaptation. Images for analysis need to be relayed to the ICC profile modification, whereas the image processing loop needs to receive occasionally updated ICC profiles. The implementation is designed as a Service Oriented Architecture to assist the ease of decoupling and distribution of these sub-processes to designated systems.

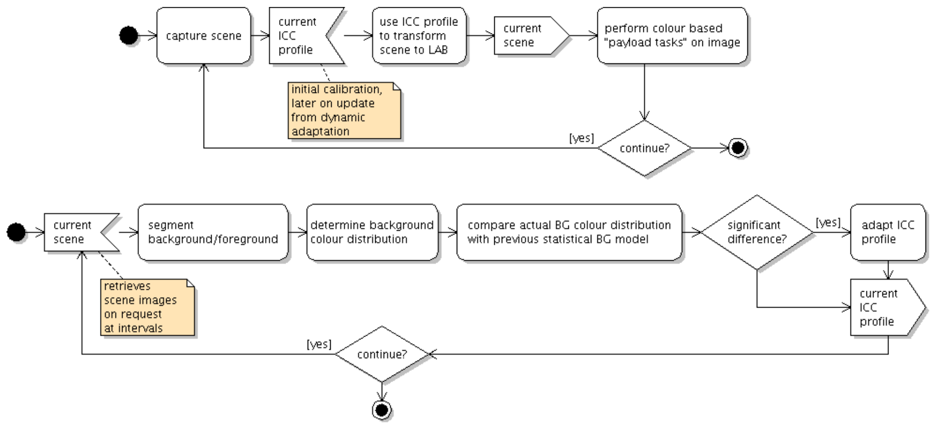

Fig. 4: Image processing (top) and ICC profile adaption processes (bottom). Organised in a service oriented architecture, requesting images/ICC profiles on demand.

5.1 Verification

Colour Indexing provides a method for recognising objects based on their distribution of colours [10]. Barnard [8] already verified the improvements of his chromatic adaptation efforts using this method. As some of Barnard's algorithms are to be used for algorithmic comparisons, we will also use Colour Indexing for further verification. This should make it possible to judge the quality of algorithms statistically and to eliminate a human bias.

5.2 Benefits

Dynamic Chromatic Adaptation promises to be valuable in all fields of colour based decision making. Colour measurements become more comparable, therefore, all dependent processes should gain robustness. We are expecting improvements in our robot soccer implementation, eliminating tedious and error prone manual parameter tweaking by gaining a self-calibrating system. Additionally constancy in longer running games and setups, as well as the ability for robots to move between differently lit environments (in- or outdoors, robot rescue, etc.) will be gained. Industrial partners will benefit from improved Colour Indexing quality for object identification, and others are aiming at more stability for colour based object sorting systems.

5.3 Current State

In a development with the industrial partners we are currently migrating camera based colour processing from using the camera's raw RGB towards L*a*b* colour spaces. This transformation still "blindly" assumes sRGB as the device's output colour encoding with no compensation for the illuminant.

Due to the perceptional linearisation of the colour space used, the background/foreground segmentation (also using chromaticity information) has already stabilised significantly. Additionally, the Colour Indexing algorithm [10] used is currently in the stage of refactoring. Colour histograms for chromaticity are computed from values in the a^*b^* plane, and grey scale histograms from the L values. These histograms so far visually reveal a much more "natural" appearance of colour distributions, with actual results to follow upon completion of the code refactoring.

The current impressions of the migration path away from device dependent (RGB) colour spaces towards independent and linearised L*a*b* suggest that we have made the right decision in adopting industrial standard colour handling methodologies. Although the colour space transformation is still based on a static model disregarding any illumination influence. These modifications will also allow us to extend image processing in a modular way, without the need to re-invent best practices for device dependent colour handling.

6 Conclusion

This paper has established a sound overview of the physical and perceptional basics of colour capturing and perception. It distinguishes between the illuminant's influence, the surface reflection, and the camera's perception properties. Various methods to quantise colour through tristimulus description have been discussed (colour spaces). In order to obtain sensible results in a quantitative analysis from tristimulus based colour spaces, various requirements towards the employed colour space need to be considered. The colour space needs to be independent of the specific capturing device (for comparability), and it needs to be linearised across the perceptible colour gamut. A way to arrive at this is to transform the captured device dependent colour space (e. g. RGB) into the device independent CIE L*a*b* colour space by means of an ICC profile obtained through camera calibration.

The focus in colour handling has been set on standardised and industry approved colour spaces and methods. On this basis possibilities to obtain robust colour information from live image capturing have been discussed, either by modifying existing algorithms or deriving new calibration algorithms, that can be adapted to illumination changes as they appear in real time during a live observation.

Some promising approaches for possible solutions through tracking of background changes after a background/foreground segmentation, as well as histogram based cross-correlation model functions have been stated. The proposed research is limited to the common case of scenes with slow changes in distinct background properties, but possible solutions for cases violating this constraint were mentioned.

An outline towards the implementation of established best practices for the field of fast live image capturing has been presented. Based on these, an outlook towards potential candidates for the goal of providing robust dynamic chromatic adaptation has been given. To avoid bottle necks in processing performance, the dynamic modification is decoupled from the processing pipeline of the live capturing into a separate thread or process employing a Service Oriented Architecture. Positive advancements in chromaticity based reasoning after changing to an L*a*b* based analysis colour space could be observed already.

It is important to note that it is not possible to establish a *precise* description of the scene's chromaticity. We have, however, presented a novel approach for joining device independent colour management – as it is an accepted long standing industry standard – with dynamically adaptive characteristics to reliably compensate for changes in the observed environment in real time. It tries to compensate for dynamic changes that have not been able to account for, by providing an estimation of the true chromaticity previously impossible to obtain through direct measurement using common cameras.

Acknowledgements

We would like to thank the companies Baggage Sortation Management Ltd. and ecomgroup Ltd. for their support and cooperation. They gave the opportunity to develop and test the algorithms and software for live systems, and to validate its practical robustness through a real world validation.

References

1. Green, P.: International Color Consortium White Paper #17: Using ICC profiles with digital camera images. `http://www.color.org/ICC_white_paper_17_ICC_profiles_with_camera_images.pdf` (2005) last visited Oct. 2007.
2. Green, P.: International Color Consortium White Paper #20: Digital photography color management basics. `http://www.color.org/ICC_white_paper_20_Digital_photography_color_management_basics.pdf` (2005) last visited Oct. 2007.
3. Glynn, E.F.: Chromaticity Diagrams Lab Report. Technical report, efg's Computer Lab and Reference Library (1999)
4. Landa, E.R., Fairchild, M.D.: Charting Color from the Eye of the Beholder. American Scientist **93** (2005) 436–438
5. Fairchild, M.D.: Color Appearance Models. 2 edn. Wiley-IS&T, Chichester, UK (2005)
6. Lindbloom, B.J.: Chromatic Adaptation Evaluation. Technical report, Bruce J. Lindbloom (2007)
7. Süsstrunk, S., Holm, J., Finlayson, G.D.: Chromatic adaptation performance of different RGB sensors. IS&T/SPIE Electronic Imaging 2001: Color Imaging **4300** (2001) 172–183
8. Barnard, K.: Practical Colour Constancy. PhD thesis, Simon Fraser University, School of Computing (1999)
9. Porikli, F.: Inter-Camera Color Calibration using Cross-Correlation Model Function. In: In Proceedings of IEEE International Conference on Image Processing, Barcelona (2003)
10. Swain, M.J., Ballard, D.H.: Colour Indexing. International Journal of Computer Vision **7** (1991) 11–32