

(51) International Patent Classification:
G06T 5/00 (2006.01)(21) International Application Number:
PCT/SG2012/000204(22) International Filing Date:
6 June 2012 (06.06.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/498,386 17 June 2011 (17.06.2011) US(71) Applicant (for all designated States except US): **NANYANG TECHNOLOGICAL UNIVERSITY** [SG/SG];
50 Nanyang Avenue, Singapore 639798 (SG).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **KAKARALA, Ramakrishna** [US/SG]; c/o Nanyang Innovation and Enterprise Office (NIEO) Blk 1, Unit 213, Innovation Centre, 16 Nanyang Drive, Singapore 637722 (SG).(74) Agent: **WATKIN, Timothy, Lawrence, Harvey**; Marks & Clerk Singapore LLP, Tanjong Pagar, P O Box 636, Singapore 910816 (SG).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

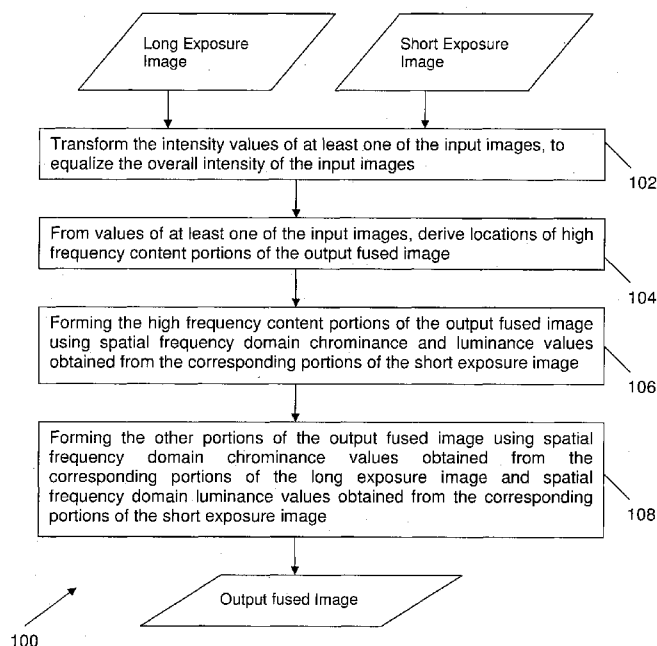
Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))

(54) Title: A METHOD AND SYSTEM FOR FUSING IMAGES



(57) Abstract: The present invention relates to a method for fusing images, for example, first and second input images to form an output fused image. The input images are photographic images of the same scene captured successively and with different exposure times. The method comprises (i) transforming intensity values of at least one of the input images, to equalize the overall intensity of the input images; (ii) deriving, from values of at least one of the input images, locations of high frequency content portions of the output fused image; (iii) forming the high frequency content portions of the output fused image using spatial frequency domain chrominance and luminance values from the second input image; and (iv) forming other portions of the output fused image using spatial frequency domain chrominance values from the first input image and spatial frequency domain luminance values from the second input image.

A Method and System for Fusing Images

Field of the invention

- 5 The present invention relates to a method and system for fusing images. The method and system may be used for image stabilization or high dynamic range (HDR) capture.

Background of the Invention

10

JPEG

- JPEG refers to the 1992 image compression standard by the Joint Photographic Experts Group. The JPEG algorithm compresses an image by converting
15 intensity values in the image to luminance (Y) and chrominance (C_b , C_r) values, computing the Discrete Cosine Transform (DCT) coefficients of these values, quantizing the DCT coefficients and compressing the quantized DCT coefficients by using run-length encoding in a "zig-zag" scan. The DCT coefficients may be referred to as spatial frequency domain luminance (Y) and
20 chrominance (C_b , C_r) values.

- The JPEG file interchange format (JFIF) is the standard file format for storing JPEG images. In this format, the DCT coefficients of the luminance (Y) channel are placed in 8x8 blocks where each luminance (Y) block corresponds to an 8x8
25 pixel region in the image. The DCT coefficients of the chrominance (C_b , C_r) channels are also placed in blocks, with these blocks being stored separately from the luminance (Y) blocks.

- For each 16x16 pixel region of an image in the 4:4:4 color format, the DCT
30 coefficients of the luminance (Y) channel are stored as four 8x8 blocks. In this format, the DCT coefficients of each chrominance (C_b , C_r) channel are also stored as four 8x8 blocks. However, it is well known that the human visual

system is more sensitive to details in the luminance (Y) channel than in the chrominance (C_b , C_r) channels. JPEG takes advantage of this by sub-sampling the two chrominance (C_b , C_r) channels for the 4:1:1 color format. In particular, in each 16x16 pixel region of an image in the 4:1:1 color format, the DCT coefficients of the luminance (Y) channel are stored as four 8x8 blocks whereas the DCT coefficients of each chrominance (C_b , C_r) channel are sub-sampled by a factor of two in both the horizontal and vertical directions. This gives rise to two chrominance (C_b , C_r) blocks (one C_b block and one C_r block), each of size 8x8.

A macro-block comprises the luminance (Y) and chrominance (C_b , C_r) blocks of a 16x16 pixel region in an image. For example, in the 4:4:4 color format, a macro-block comprises four 8x8 luminance (Y) blocks and eight 8x8 chrominance (C_b , C_r) blocks whereas in the 4:1:1 color format, a macro-block comprises four 8x8 luminance (Y) blocks and two 8x8 chrominance (C_b , C_r) blocks.

More details of JPEG image compression and file storage may be found in Pennebaker and Mitchell (1993).

Image fusion

The need for image fusion arises in many applications, including digital image stabilization and HDR image acquisition (i.e. HDR capture). For image stabilization, the motion blur that occurs in an image may be corrected by fusing this image with an image of the same scene but captured with a shorter exposure time (Tico and Pulli (2009)). This fusion combines the higher signal-to-noise ratio (SNR) provided by the image with the longer exposure time and the sharp details provided by the image with the shorter exposure time to achieve image stabilization that is otherwise available only through opto-mechanical means. For HDR photography, a set of images with varying exposure times or ISO settings (Hasinoff et al (2010)) are typically fused to

capture a wide range of scene luminance. Otherwise, with only a single exposure, the resulting image is likely to contain saturated regions or dark features of interest.

5 Image fusion may be performed either in the spatial domain, by combining the relevant regions of two or more images, or in a transform domain such as the wavelet or Fourier transform, by combining the relevant transform components. To date, published research in image fusion, while acknowledging computational complexity as important, generally places greater emphasis on
10 the resulting image quality, rather than on the computational complexity or suitability for embedded systems. Extensive overviews of fusion techniques drawn from various fields may be found in Mitchell (2010) and Stathaki (2008). A few examples of published research in the digital camera literature for image fusion are as follows.

15 In an attempt to reduce ghosting artifacts when fusing images of moving objects, Khan et al (2006) proposed employing a weighting function based on the probability of a pixel belonging to a moving object to attenuate the contribution of such a pixel to the fused image. However, Khan's proposed
20 method requires multiple iterations to converge. An iterative method is also proposed in Lu et al (2009) whereby the method is used for HDR image fusion and relies on de-convolution to correct motion blur.

Similarly, to reduce artifacts when fusing input images with different exposure
25 times, Reinhard et al (2005) proposed using the variance across the input images to determine the likelihood that a pixel will contribute to ghosting artifacts in the fused image. However, due to slight misalignments and errors in camera calibration, the variance at the pixel level is often noisy and thus, affects the results of Reinhard's proposed method.

30 Techniques to enhance the dynamic range of monochromatic and color images and videos by using optical flow estimation as a means of per-pixel registration

have also been proposed by Bogoni (2000) and Kang et al (2008). Furthermore, Jacobs et al (2008) proposed using a local entropy-based method on different low dynamic range images to detect motion over a sequence of images.

5 Gallo et al (2009) proposed selecting a reference image from a stack of images whereby the images are of the same scene but with different exposure times. Gallo et al further proposed detecting, from each of the remaining images in the stack, parts that can be combined with regions in the reference image without causing artifacts or ghosting. The resultant HDR image in Gallo's proposed
10 method is a collage of these detected parts and the regions in the reference image.

Tico and Pulli (2009) and Tico et al (2010) proposed a method for fusing a pair of images with different exposure times whereby the method works in the
15 wavelet domain and is useful for either image stabilization or HDR capture.

All of the above-mentioned proposed methods require at least the following: two or more entire images to be kept in memory, two or more passes over each pixel in at least one of the images, and significant computation (such as a
20 wavelet transform) that is normally not required for a typical digital camera's image processing pipeline. While computing power onboard digital cameras has been steadily increasing, the camera processors that are currently available still require several seconds to perform HDR image fusion even when ghosting artifact reduction is not carried out. This long processing time is due to the large
25 amount of RAM memory access required for storing and retrieving images (which are usually in mega-pixels), spatial filtering and intensity transformations. Though it is still possible to implement some of the above-mentioned proposed methods in a mobile device, as demonstrated in a proof-of-concept by Gelfand et al (2010), it is difficult to implement them efficiently. There is also no
30 published work describing a minimally complex fusion algorithm suitable for low-cost embedded cameras.

In several of the published research techniques including the above-mentioned ones, the proposed methods comprise two key steps: image registration and tone-mapping. Thus, it is worth noting Sen et al (2008) and Akil (2011) which describe implementations of image fusion techniques on FPGAs and GPUs, respectively. Further relevant background material includes the operations of a digital camera's embedded image processing "pipeline" as described by Nakamura (2005).

Summary of the invention

The present invention aims to provide a new and useful method and system for fusing images.

In general terms, the present invention proposes combining luminance and chrominance values from different images to form an output fused image.

Specifically, a first aspect of the present invention is a method for fusing a first input image with a second input image to form an output fused image, the first and second input images being photographic images of the same scene captured successively, the exposure time of the first input image being longer than the exposure time of the second input image, the first and second input images having equal pixel dimensions and each comprising respective intensity values for each pixel and for each of multiple colors, the method comprising: (i) transforming the intensity values of at least one of the input images, to equalize the overall intensity of the input images; (ii) deriving, from values of at least one of the input images, locations of high frequency content portions of the output fused image; (iii) forming the high frequency content portions of the output fused image using spatial frequency domain chrominance and luminance values obtained from the corresponding portions of the second input image; and (iv) forming other portions of the output fused image using spatial frequency domain chrominance values obtained from the corresponding portions of the first input

image and spatial frequency domain luminance values obtained from the corresponding portions of the second input image.

The invention may alternatively be expressed as a computer system for performing such a method. This computer system may be integrated with a device for capturing images such as a digital camera. The invention may also be expressed as a computer program product, such as one recorded on a tangible computer medium, containing program instructions operable by a computer system to perform the steps of the method.

Brief Description of the Figures

Embodiments of the invention will now be illustrated for the sake of example only with reference to the following drawings, in which:

Fig. 1 shows a flow diagram of a method for fusing images according to an embodiment of the present invention;

Fig. 2 shows an example of implementing the method of Fig. 1;

Fig. 3 shows a scatter plot of mean intensity values in two images with different exposure times and two curves (a sigmoidal curve and a gamma curve);

Figs. 4(a) - (f) each show a scatter plot of mean intensity values in two images with different exposure times and a sigmoidal curve;

Figs. 5(a) - (b) show a first set of input images and Figs. 5(c) - (e) show the results of applying the example of Fig. 2 on the first set of input images;

Figs. 6(a) - (d) show enlarged portions of the results of Figs. 5(c) and 5(e);

Figs. 7(a) - (b) show a second set of input images and Figs. 7(c) - (e) show the results of applying the example of Fig. 2 on the second set of input images;

Figs. 8(a) - (b) show enlarged portions of the results of Figs. 7(c) and 7(e);

Fig. 9 shows how a percentage of chrominance blocks replaced in a step of the example of Fig. 2 varies with an EOB location threshold for two different nighttime scenes; and

Figs. 10(a) – (b) show the results of applying prior art methods on the first set of input images in Figs. 5(a) – (b).

Detailed Description of the Embodiments

Method 100

10

Referring to Fig. 1, the steps are illustrated of a method 100 which is an embodiment of the present invention, and which fuses images.

The input to method 100 may be a pair of photographic images of the same scene captured successively but with different exposure times. In this document, the first input image with a longer exposure time is referred to as a long exposure image whereas the second input image with a shorter exposure time is referred to as a short exposure image. Typically, the long exposure image has a higher SNR than the short exposure image whereas the short exposure image comprises sharper details than the long exposure image. Note that the “first input image” is not necessarily acquired first. In other words, the “second input image” may be acquired before the “first input image”.

The input images have equal pixel dimensions. Furthermore, each input image comprises a plurality of pixel regions (for example, 8x8 pixel regions or 16x16 pixel regions) whereby each pixel region in turn comprises a plurality of pixels. Each pixel has respective intensity values for each of multiple colors. For example, the intensity values may be in the Red, Green and Blue (RGB) color space.

30

Method 100 works on the following assumptions.

Firstly, it is assumed that both input images are acquired one after another within a certain time frame (for example, in immediate succession) such that there is a negligible need for registration between the images. For example, the input images may be acquired in the automatic exposure bracketing (AEB) mode which ensures minimal motion between the successive frames captured.

Secondly, it is assumed that the exposure ratio between the input images is known prior to acquiring the input images. This assumption is reasonable for cameras operating in the AEB mode, in which the exposure ratios are set beforehand. Typically the exposure ratios are in powers of two, but their values may be programmable. Note that in this document, the well-known logarithmic exposure value (EV) notation of describing the ratio of exposure times of two images is used. EV(0) represents the exposure time determined by the camera's auto-exposure routine, and relative to this exposure time, EV(Δ) represents an exposure time that is 2^Δ larger. For example, EV(+1) represents an exposure time which is twice the exposure time represented by EV(0), whereas EV(-1) represents an exposure time which is half the exposure time represented by EV(0).

Thirdly, it is assumed that the short and long exposure images vary only in their exposure times, and not in their aperture and ISO settings.

If any of the above assumptions do not hold, further steps may be implemented to for example, register the input images, determine the exposure ratio between the input images and/or adjust the input images to compensate for any differences in their aperture and ISO settings. For instance, an equivalent exposure ratio between the input images may be computed based on a formula which is dependent on the design of the camera used to capture the input images and which is determined during the design cycle of the camera.

As shown in Fig. 1, method 100 fuses the short and long exposure images to form an output fused image using steps 102 - 108. In step 102, the intensity values of at least one of the input images are transformed to equalize the overall intensity of the input images. In step 104, from values (for example, luminance values) of at least one of the input images, locations of high frequency content portions of the output fused image are derived. Next in step 106, the high frequency content portions of the output fused image are formed using chrominance and luminance values obtained from the corresponding portions of the short exposure image. Then in step 108, the other portions of the output fused image are formed using chrominance values obtained from the corresponding portions of the long exposure image and luminance values obtained from the corresponding portions of the short exposure image. If necessary, method 100 may further comprise converting all the intensity values (which may be transformed intensity values from step 102) of one or both the input images to luminance and chrominance values. Method 100 may also further comprise writing the luminance and chrominance values of one of the input images to a file. Note that the fusion in steps 106 – 108 works in the spatial frequency domain. In other words, the luminance and chrominance values mentioned above with respect to these steps are spatial frequency domain luminance and chrominance values.

Note that although Fig. 1 shows the input to method 100 comprising only two images, method 100 may be extended to fuse more than two images. In this case, the input images are preferably captured in rapid succession so that there remains a negligible need for image registration. For example, it may be optimal to capture the images with the camera's AEB mode.

Example of implementing Method 100

Fig. 2 shows an example of implementing method 100. In this example, the long exposure image is acquired before the short exposure image.

The example in Fig. 2 comprises steps 202 – 214. In step 202, the intensity values of the long exposure image are converted to luminance (Y_l) and chrominance (C_{bl} , C_{rl}) values. These luminance (Y_l) and chrominance (C_{bl} , C_{rl}) values are then written to a file in step 204. Step 206 implements step 102 of method 100. Specifically, in step 206, the intensity values of the short exposure image are transformed to form a transformed short exposure image. Next, in step 208, the transformed intensity values of the transformed short exposure image are converted to luminance (Y_{bs}) and chrominance (C_{bbs} , C_{rbs}) values. Steps 210 and 212 implement steps 104 – 108 of method 100. Specifically, in step 210, the luminance (Y_l) values of the long exposure image are overwritten with the luminance (Y_{bs}) values of the transformed short exposure image to form an initial fused image (i.e. an overwritten long exposure image). Next in step 212, artifacts are removed from the initial fused image to form an output fused image by overwriting some of the chrominance (C_{bl} , C_{rl}) values of the initial fused image with corresponding chrominance (C_{bbs} , C_{rbs}) values of the short exposure image. In step 214, the luminance and chrominance values of the output fused image are converted to intensity values for each of multiple colors.

Steps 202 – 214 will now be described in more detail.

Step 202: Convert intensity values of the long exposure image to luminance (Y_l) and chrominance (C_{bl} , C_{rl}) values

In step 202, the intensity values in the long exposure image are converted to values in the YC_bC_r space. In particular, the intensity values are converted to luminance (Y_l) and chrominance (C_{bl} , C_{rl}) values.

Step 204: Write the luminance (Y_l) and chrominance (C_{bl} , C_{rl}) values of the long exposure image to a file

Next in step 204, the luminance (Y_l) and chrominance (C_{bl} , C_{rl}) values of the long exposure image are written to a file which may be a file in secondary storage such as a flash memory.

5 Step 206: Transform the intensity values of the short exposure image to form a transformed short exposure image

In step 206, the intensity values of the short exposure image are transformed to form a transformed short exposure image. This equalizes the overall intensity of
10 the input images and the camera response between the short and long exposure images.

In particular, in step 206, the intensity values of the short exposure image are transformed to boost the luminance (i.e. brightness) of the short exposure
15 image so as to match that of the long exposure image.

The luminance boosting may be done by first estimating a compensating function that matches the response of the camera used to acquire the input images, and then transforming the intensity values in the short exposure image
20 based on the compensating function. In particular, the transformation may comprise replacing the intensity values in the short exposure image with corresponding values of the compensating function. Following Tico and Pulli (2009), the compensating function may be referred to as a brightness transfer function (BTF).

25

The BTF may be a pre-generated function, which may be estimated by first plotting the intensity values of a long exposure image against corresponding intensity values of a short exposure image, and then applying basic curve fitting to the plot (Reinhard et al (2005)). Alternatively, the BTF may be calculated by
30 determining intensity values which are present in both the long and short exposure images, and locally smoothing the mean of these intensity values (Tico and Pulli (2009)). Both these methods of estimating the BTF require two

passes over the short exposure image, the first to estimate the BTF and the second to boost the short exposure image.

Alternatively, the BTF may be in the form of a sigmoidal function (or more specifically, a parametric sigmoidal curve). The sigmoidal function may be expressed in the form shown in Equation (1). This is similar to the form proposed in Zhang et al (2006). However, note that in Zhang et al (2006), the sigmoidal function is proposed for tone-mapping and not for BTF estimation.

$$f(x; M, \alpha, \beta) = \frac{M}{1 + \beta e^{-\alpha x}} \quad (1)$$

In Equation (1), the monotonically increasing function $f(x; M, \alpha, \beta)$ represents the sigmoidal function and x represents an intensity value in the short exposure image. As $x \rightarrow \infty$, this monotonically increasing function $f(x; M, \alpha, \beta)$ tends towards M where M is the maximum possible intensity value in the images. The value of M ranges from 255 to 4095, depending on the bit-depth of the pixel data in the images. The slope of the function $f(x; M, \alpha, \beta)$ is controlled by α whereas the location of the function $f(x; M, \alpha, \beta)$ in the x - y plane is controlled by a shift parameter β .

Through experimentation with different cameras and input images having various EV differences, the inventors of the present invention have found that setting $\beta = 1$ in Equation (1) allows the sigmoidal function to best match the experimental data. With $\beta = 1$, the function $f(x; M, \alpha, \beta)$ in Equation (1) may be expressed as a scaled function $f_s(x; M, \alpha)$ as shown in Equation (2).

$$f_s(x; M, \alpha) = 2f(x; M, \alpha, 1) - M \quad (2)$$

For the slope of the sigmoidal function to be proportional to α , $f_s(0)$ and $f_s'(0)$ are set as $f_s(0) = 0$ and $f_s'(0) = M\alpha/2$.

The inventors have also found that the camera response for images with different exposure times varies primarily as a function of the EV differences between these images. Specifically, it was found that setting α according to Equation (3) allows the sigmoidal function to best match the experimental data.

$$\alpha = \frac{2(\Delta EV + 1)}{M} \quad (3)$$

10

In Equation (3), ΔEV is the difference in exposure times between the input images. In other words, the sigmoidal function for use as the BTF in step 206 is representative of a correspondence between the intensity values of the input images based on a difference in the exposure times ΔEV between these input images. Note that with Equation (3), the initial slope at the origin may be expressed as $f_s'(0) = (\Delta EV + 1)$, and f_s has an asymptotic value equal to M as $x \rightarrow \infty$.

As mentioned above, each pixel of the input images has respective intensity values for each of multiple colors. A BTF curve may be estimated for each of the multiple colors. In one example, the intensity values of the short and long exposure images are in the RGB space and three BTF curves are estimated, one for each channel of the RGB space i.e. the Red channel, Green channel and Blue channel. The intensity values of the short exposure image in each channel are then boosted using the corresponding BTF curve estimated for the channel.

Fig. 3 shows a scatter plot of mean intensity values in the Green channel (i.e. mean Green channel values) of two images with different exposure times, EV(0) and EV(-2). In particular, the mean Green channel values in the image with the

30

longer exposure time EV(0) are plotted against the mean Green channel values in the image with the shorter exposure time EV(-2). Each mean Green channel value is computed over an 8x8 pixel block and the images are captured with a Canon 400D camera.

5

Fig. 3 further shows the scaled sigmoidal function f_s of Equation (2) (with the α value computed according to Equation (3)) plotted on the same axes as the scatter plot. This is shown as the sigmoidal curve 302. As shown in Fig. 3, there is a good fit between the sigmoidal curve 302 and the mean Green channel values in the two images.

It is worth comparing the sigmoidal function to other parametric functions which may be used in step 206 for luminance boosting or brightness compensation. Referring to Fig. 3 again, a gamma curve 304 with gamma (γ) = 2.2 is plotted on the same axes as the scatter plot. The gamma curve 304 may be expressed in the form shown in Equation (4). As shown in Fig. 3, the gamma curve 304 rises too steeply for low mean Green channel values, and levels off too quickly for high mean Green channel values. Thus, estimating the BTF using the gamma curve 304 will result in an under fitting between the BTF and the mean Green channel values.

20

$$V_{out} = V_{in}^{\gamma} \quad (4)$$

Figs. 4(a) – (f) each shows a scatter plot of mean Green channel values of two images captured with different exposure times ($\Delta EV = 1, 2, 4$). The input images for Figs. 4(a) – (f) are captured using different camera makes. In each of Figs. 4(a) – (f), the mean Green channel values of the image with a shorter exposure time are plotted against the mean Green channel values of the image with a longer exposure time. Figs. 4(a) – (f) further show sigmoidal curves based on the scaled sigmoidal function f_s on the same axes.

30

In particular, each of Figs. 4(a), (b) and (c) shows a sigmoidal curve and a scatter plot for an EV(-2) image and an EV(0) image (i.e. $\Delta EV = 2$). The images for Figs. 4(a), (b) and (c) are captured using a Nikon D40, a Pentax K-7 and a Canon 400D respectively. Fig. 4(d) shows a sigmoidal curve and a scatter plot for an EV(-2) image and an EV(+2) image (i.e. $\Delta EV = 4$) captured using a Casio EX-F1 camera. Fig. 4(e) shows a sigmoidal curve and a scatter plot for an EV(-1) image and an EV(0) image (i.e. $\Delta EV = 1$) captured using an Olympus E-450 camera. Fig. 4(f) shows a sigmoidal curve and a scatter plot for an EV(-2) image and an EV(+2) image (i.e. $\Delta EV = 4$) captured using a Sony DSC-W30 camera.

As shown in Figs. 4(a) – (f), the sigmoidal curves fit the mean Green channel values of the images reasonably well. This shows that the BTF estimated using the sigmoidal function is a reasonably good match to the camera response for several makes of recently-manufactured digital SLR cameras including the Canon 400D, Pentax K7, Olympus E-450, Nikon D40, and point and shoot cameras such as the Sony DSC-W30 and Casio EX-F1.

Estimating the BTF based on the sigmoidal function is advantageous as it requires only a single pass over the intensity values of both the short and long exposure images. Such a BTF works in a similar manner as the BTF computed in Tico and Pulli (2009), but is estimated with a much lower memory requirement and computational time.

Furthermore, because the sigmoidal curve is parametric and there is an assumption that the exposure ratio between the input images (and thus, ΔEV) is known prior to acquiring the input images, the values of the sigmoidal function (e.g. the scaled sigmoidal function f_s) corresponding to all possible intensity values may be pre-computed prior to acquiring the short exposure image. A look-up table (LUT) comprising these values may then be formed. In this case, as each intensity value in the short exposure image is read, the LUT is

accessed and from the LUT, a value of the sigmoidal function corresponding to the intensity value in the short exposure image is determined. The intensity value of the short exposure image is then transformed by replacing it with its corresponding value of the sigmoidal function as determined from the LUT.

- 5 Using the LUT to transform the intensity values helps to reduce computational time as only one LUT transformation per intensity value is required.

In step 206, the transformed intensity values of the short exposure image may be further transformed based on a tone map.

10

This tone-mapping to the luminance boosted short exposure image may be performed if the captured images comprise daylight scenes and the luminance boosted short exposure image comprises large regions appearing nearly saturated. Such a situation usually occurs when ΔEV between the short and long exposure images is large. Note that it is possible to determine whether an image comprises nighttime or daylight scenes by looking at the exposure time of the image while the shot is framed in the viewfinder, or by looking at the meter reading if the camera has one.

15

- 20 The tone-mapping may be performed by applying a global tone map across the luminance boosted short exposure image i.e. the same mapping is applied to each pixel of the luminance boosted short exposure image. Furthermore, the tone map may also be implemented as a LUT whereby the LUT comprises values of the tone map corresponding to all possible transformed intensity values (derived from transforming all possible intensity values using the BTF). Similarly, for each transformed intensity value in the luminance boosted short exposure image, the LUT is accessed and from the LUT, a value of the tone map corresponding to the transformed intensity value in the luminance boosted short exposure image is determined. This transformed intensity value is then further transformed by replacing it with its corresponding tone map value as determined from the LUT.
- 25
- 30

The tone mapping may be performed simultaneously with the luminance boosting for efficiency. In one example, the tone map values are combined with the sigmoidal function values in a single LUT. Direct correspondence between the tone map values and all possible intensity values in the input short exposure image may thus be determined from this single LUT. The transformation of the intensity values in the input short exposure image may be performed using the single LUT in a manner similar to that as described above. In particular, for each intensity value in the input short exposure image, the single LUT is accessed and from the single LUT, a tone map value corresponding to the intensity value in the input short exposure image is determined. This intensity value is then transformed by replacing it with its corresponding tone map value as determined from the single LUT.

Step 208: Convert the transformed intensity values of the transformed short exposure image to luminance (Y_{bs}) and chrominance (C_{bbs} , C_{rbs}) values

In step 208, the transformed intensity values of the transformed short exposure image are converted to values in the YC_bC_r space. In particular, these transformed intensity values are converted to luminance (Y_{bs}) and chrominance (C_{bbs} , C_{rbs}) values.

Step 210: Overwrite the luminance values (Y_l) of the long exposure image with the luminance values (Y_{bs}) of the transformed short exposure image to form an initial fused image

In step 210, the luminance values (Y_l) of the pixels in the long exposure image are overwritten with the luminance values (Y_{bs}) of the corresponding pixels in the transformed short exposure image to form an initial fused image. In other words, the chrominance values (C_{bl} , C_{rl}) in the long exposure image are merged with the luminance values (Y_{bs}) in the transformed short exposure image. The pixel dimensions of the initial fused image are equal to that of the short and long exposure images.

Note that the long exposure image was written to a file in step 204, and in step 210, this same file stores the initial fused image since the luminance (Y_{bs}) and chrominance (C_{bl} , C_{rl}) values are combined by overwriting the values in the long exposure image.

5

Step 212: Remove artifacts from the initial fused image to form an output fused image

Since the short and long exposure images are taken at different times, the initial fused image is likely to contain motion artifacts such as ghosting and color bleeding. This is because there is likely to be a mismatch between the luminance (Y_{bs}) values from the transformed short exposure image and the chrominance (C_{bl} , C_{rl}) values from the long exposure image, especially at regions with high frequency content, for example, edges.

15

In step 212, the artifacts are removed from the initial fused image to form an output fused image. The output fused image also has the same pixel dimensions as the short and long exposure images.

Specifically, the artifact removal is performed in step 212 by first detecting regions comprising significant high frequency content (representing for example, edges or texture) in the short exposure image. This also derives the locations of the high frequency content portions in the output fused image since these high frequency content portions correspond to the detected regions in the short exposure image. In particular, this may be done based on the luminance (Y_{bs}) values in the short exposure image, for example, by detecting pixels in the short exposure image whose luminance (Y_{bs}) components comprise high frequency content. Next, in portions of the initial fused image corresponding to the detected regions in the short exposure image (also corresponding to the high frequency content portions of the output fused image), the chrominance (C_{bl} , C_{rl}) values of the pixels in the initial fused image are overwritten with the

30

chrominance (C_{bbs} , C_{rbs}) values of the corresponding pixels in the short exposure image.

The above therefore forms an output fused image with its high frequency
5 content portions comprising luminance (Y_{bs}) and chrominance (C_{bbs} , C_{rbs}) values from the transformed short exposure image, and its other portions comprising luminance (Y_{bs}) values from the transformed short exposure image and chrominance (C_{bl} , C_{rl}) values from the long exposure image.

10 Step 214: Convert luminance and chrominance values in the output fused image to intensity values for each of multiple colors

In step 214, the luminance and chrominance values in the output fused image are converted to intensity values for each of multiple colors, for example,
15 intensity values in the RGB space.

Implementation of method 100 on a pixel region-by-pixel region basis

Method 100 may be implemented on a pixel region-by-pixel region basis
20 through a series of operations.

In particular, method 100 may be performed on a pixel region-by-pixel region basis through a series of operations whereby in each operation, a pixel region in the output fused image is formed. The final output fused image is a collage of all
25 the pixel regions formed in the operations.

In a more specific example, the following steps are performed to implement the example in Fig. 2 on a pixel region-by-pixel region basis.

30 In a first operation, a first pixel region in the long exposure image (here called a "first-long-region") is processed in steps 202 – 204. In step 202, intensity values in this first-long-region are converted to luminance (Y_l) and chrominance (C_{bl} , C_{rl})

values and in step 204, these luminance (Y_l) and chrominance (C_{bl} , C_{rl}) values are written to a file. Next, the intensity values of the corresponding pixel region of the short exposure image are transformed in step 206. Then, a first pixel region of the transformed short exposure image (here called a "first-short-region") corresponding to the first-long-region is processed in step 208. In particular, intensity values in the first-short-region are converted to luminance (Y_{bs}) and chrominance (C_{bbs} , C_{rbs}) values. Luminance (Y_l) values of the first-long-region are then replaced with the luminance (Y_{bs}) values of the first-short-region in step 210 to form a first pixel region in the initial fused image (here called a "first-fused-region"). Next in step 212, it is determined if the first-short-region comprises high frequency content by for example, determining if the luminance (Y_{bs}) components in this region comprise high frequency information. This also indicates if the pixel region of the output fused image to be formed in this operation comprises high frequency content. If so, the chrominance (C_{bbs} , C_{rbs}) values in the first-short-region are used to overwrite the chrominance (C_{bl} , C_{rl}) values in the first-fused-region to form a first pixel region of the output fused image (here called a "first-output-region"). Finally, in step 214, the luminance and chrominance values in the first-output-region are converted to intensity values for each of multiple colors.

20

In subsequent operations, the remaining pixel regions in the output fused image are formed in the same manner. In other words, for each pixel region in the output fused image, if the corresponding pixel region in the short exposure image comprises high frequency content (indicating that the pixel region of the output fused image comprises high frequency content), the pixel region of the output fused image is formed using the chrominance (C_{bbs} , C_{rbs}) and luminance (Y_{bs}) values obtained from the corresponding pixel region of the short exposure image. Otherwise, the pixel region of the output fused image is formed using the chrominance (C_{bl} , C_{rl}) values obtained from the corresponding pixel region of the long exposure image and luminance (Y_{bs}) values obtained from the corresponding pixel region of the short exposure image.

30

Note that step 206 may alternatively be performed for the whole of the short exposure image in the first operation. The entire transformed short exposure image may be stored in a buffer for use in the subsequent operations and in this case, a memory buffer the size of the short exposure image is required.

- 5 Similarly, steps 202 – 204 may be performed on the entire long exposure image in the first operation and in this case, these steps do not need to be performed in the subsequent operations.

10 Furthermore, since intensity values of the short exposure image may be read in a row-by-row manner from an image sensor of the camera used to capture the input images, steps 206 – 214 may be performed on a row-by-row basis with the processing of each row beginning once the row is read. Using such standard design optimization, the amount of memory required may be reduced.

15 Implementation of method 100 in the JPEG domain

Method 100 may be implemented in the JPEG domain. In this case, method 100 may be referred to as a transform domain method.

- 20 For example, steps 202 – 214 of Fig. 2 may be implemented in the JPEG domain in the digital camera used to capture the input images as follows.

25 Steps 202 – 204 are performed by processing the long exposure image as per normal in the digital camera. This involves converting (or compressing) the long exposure image to the JPEG format and then writing the JPEG formatted long exposure image to a JPEG file, for example, a file in secondary storage such as flash memory.

30 Steps 206 – 208 are performed by transforming the intensity values of the short exposure image in the manner as described above and then, converting (or compressing) the transformed short exposure image to the JPEG format. A JPEG encoder of the digital camera is modified to transform the intensity values

of the short exposure image in step 206 (implementing step 102 of method 100).

Since the JFIF places the luminance (Y) blocks separate from the chrominance (C_b, C_r) blocks, the addresses of the luminance (Y) and chrominance (C_b, C_r) blocks in a JPEG file may be calculated and overwritten selectively. Therefore, instead of writing a separate JPEG file for the transformed short exposure image, the file write operation in the digital camera may be modified to perform steps 210 – 212 (implementing steps 104 – 108 of method 100). This is performed in the JPEG encoder.

In particular, in steps 210 – 212, parts of the JPEG file comprising the long exposure image are overwritten with relevant portions of the JPEG formatted transformed short exposure image as follows.

In step 210, the luminance (Y_l) blocks of the long exposure image are overwritten with the luminance (Y_{bs}) blocks of the transformed short exposure image to form an initial fused image in the JPEG format. In other words, the luminance of the initial fused image is that of the transformed short exposure image.

In step 212, artifacts are detected and removed from the initial fused image in the JPEG domain to form an output fused image. This is done by using a JPEG-based detector which exploits the JPEG's built-in frequency analysis using DCT to detect regions comprising high frequency content. As mentioned above, the quantized DCT coefficients are compressed by using run-length encoding in a "zig-zag" scan. As part of the JPEG syntax, an End-of-Block (EOB) symbol occurs in every 8x8 luminance (Y) block and indicates the location of the last non-zero AC coefficient (of the DCT coefficients) in the 64-coefficient zig-zag scan. Thus, it is possible to determine if a pixel region in an image comprises high frequency content based on whether the EOB symbol in the corresponding luminance (Y) block occurs early or late in the scan.

Specifically, in step 212, for each luminance (Y_{bs}) block in the transformed short exposure image, the location of the EOB symbol in the block is determined from the DCT coefficients (i.e. spatial frequency domain luminance values) of the block. If the EOB symbol in the luminance (Y_{bs}) block occurs after a pre-determined EOB location threshold, high frequency content (for example, an edge) is deemed to have been detected in the pixel region comprising the luminance (Y_{bs}) block. In other words, a portion comprising the luminance (Y_{bs}) block in the short exposure image and thus, its corresponding portion in the output fused image are identified as high frequency content portions. In this case, the luminance (Y_{bs}) block is classified as an edge block. Otherwise the luminance (Y_{bs}) block is classified as a smooth block. Next, in pixel regions in the initial fused image corresponding to pixel regions comprising the edge blocks in the transformed short exposure image (also corresponding to high frequency content regions in the output fused image), chrominance values (C_{bl} , C_{rl}) in the initial fused image are overwritten with corresponding chrominance (C_{bbs} , C_{rbs}) values in the transformed short exposure image.

Next in step 214, the output fused image in the JPEG format is decompressed. This is performed in the JPEG decoder.

The compressed information (i.e. the quantized DCT coefficients) of the images may be determined on an 8x8 pixel region-by-pixel region basis in the digital camera. In this case, the overwriting in steps 210 – 212 may be performed on an 8x8 block-by-block basis. More specifically, in step 212, if a luminance (Y_{bs}) block of an 8x8 pixel region in the transformed short exposure image is classified as an edge block, the chrominance (C_{bbs} , C_{rbs}) blocks of this 8x8 pixel region are used to overwrite the chrominance (C_{bl} , C_{rl}) blocks of a corresponding 8x8 pixel region in the initial fused image. Since the image fusion is done by overwriting in steps 210 – 212, RAM memory beyond that required for three 8x8 blocks (comprising one 8x8 Y block, one 8x8 C_b block and one 8x8 C_r block) is not needed in this case.

Alternatively, the compressed information may be determined on a 16x16 pixel region-by-pixel region basis in the digital camera and the overwriting in steps 210 – 212 may be performed on a macro-block-by-macro-block basis. In particular, in step 212, if the number of luminance (Y_{bs}) blocks classified as edge blocks in a macro-block of the transformed short exposure image is greater than a pre-determined threshold, the chrominance (C_{bbs} , C_{rbs}) blocks of the macro-block are used to overwrite the chrominance (C_{bl} , C_{rl}) blocks of a corresponding macro-block in the initial fused image. For example, if the images are encoded in the standard 4:1:1 color format, the overwriting may be performed as follows: for every macro-block of the short exposure image, if any of the four luminance (Y_{bs}) blocks in the macro-block is an edge block, the two chrominance (C_{bbs} , C_{rbs}) blocks in the macro-block are used to overwrite the two chrominance (C_{bl} , C_{rl}) blocks in a corresponding macro-block of the initial fused image. Similarly, since the image fusion is done by overwriting in steps 210 – 212, RAM memory beyond that required for one macro-block (comprising one 16x16 Y block (i.e. four 8x8 Y blocks), one 8x8 C_b block and one 8x8 C_r block) is not needed in this case.

Alternative Ways of Implementing Method 100

Alternative ways of implementing method 100 (other than the example shown in Fig. 2) are possible within the scope of the invention as will be clear to a skilled reader. Some of these alternative ways are described below.

Alternative ways of equalizing the overall intensity of the input images

In the example of Fig. 2, the overall intensity of the input images is equalized by transforming the intensity values of the short exposure image.

Alternatively, the intensity values of the long exposure image may be transformed instead. Specifically, the luminance of the long exposure image may be reduced to match that of the short exposure image. In this case, the

luminance values from the transformed long exposure image are used as the luminance values in the output fused image instead. As compared to the output fused image obtained using the example of Fig. 2, the output fused image obtained in this case is likely to have a lower luminance (i.e. it is likely to be darker). This is because the output fused image in this case is formed from the input short exposure image (which tends to have a relatively low luminance) and the transformed long exposure image (with a luminance reduced to match that of the short exposure image). This may be useful for some applications whereby a darker output fused image is preferred. For example, a night time scene may be made to look realistically darker with such an output fused image. Furthermore, since the details in the luminance (Y_l) channel of the long exposure image are less sharp as compared to the details in the luminance (Y_{bs}) channel of the short exposure image, the sharpness of the output fused image obtained in this case is likely to be lower than that obtained using the example of Fig. 2. This may also be useful for some applications, for example, an application aiming to provide a particular artistic effect.

In another alternative, the intensity values of both the input images (i.e. both the short and long exposure images) are transformed. For example, the luminance of the long exposure image may be reduced whereas the luminance of the short exposure image may be boosted so that both images have a predetermined desired luminance. In this case, the luminances of both the transformed short and long exposure images are fused to form the luminance of the output fused image. This may be done by combining the luminance values of these two images in a manner similar to how the chrominance values are combined in step 212 of the example in Fig. 2. Although an output fused image obtained in this manner may not appear nice in the traditional sense and may comprise some image distortion, such an output fused image may still be desirable for some applications, for example, an application aiming to achieve a particular artistic effect.

In both the above cases, the transformation of the intensity values in the long exposure image may also be performed using the sigmoidal function. In particular, this may be done using the inverse of the sigmoidal function since this function is monotonically increasing. For example, representing the transformation of the short exposure image as $y_s = f(x_s)$ where $f(\bullet)$ is the sigmoidal function and may be the scaled function $f_s(x; M, \alpha)$ in Equation (2), x_s denotes intensity values in the input short exposure image and y_s denotes transformed intensity values in the transformed short exposure image, the transformation of the long exposure image may be represented as $x_l = f^{-1}(y_l)$ where y_l denotes intensity values in the input long exposure image and x_l denotes transformed intensity values in the transformed long exposure image.

Note that it is rare to have applications in which it is preferable to transform the intensity values of the long exposure image or the intensity values of both the input images as described above. More often than not, it is preferable to equalize the overall intensity of the input images by transforming the intensity values of the short exposure image and using the luminance values of the transformed short exposure image for the output fused image (as in the example of Fig. 2).

Alternative order of acquiring the input images

In the example of Fig. 2, method 100 is implemented by acquiring the long exposure image before the short exposure image.

Alternatively, method 100 may be implemented by acquiring the short exposure image before the long exposure image.

For example, method 100 may be implemented with the following steps:

Step 1: Acquire the short exposure image.

- Step 2: Transform the intensity values of the short exposure image to equalize the overall intensity of the input images.
- Step 3: Convert the transformed intensity values of the transformed short exposure image to luminance (Y_{bs}) and chrominance (C_{bbs} , C_{rbs}) values.
- 5 Step 4: Write the Y_{bs} , C_{bbs} , C_{rbs} values of the transformed short exposure image to a file.
- Step 5: Acquire the long exposure image.
- Step 6: Convert the intensity values of the long exposure image to luminance (Y_{bl}) and chrominance (C_{bl} , C_{rl}) values.
- 10 Step 7: In high frequency content portions, leave the luminance (Y_{bs}) and chrominance (C_{bbs} , C_{rbs}) values of the transformed short exposure image as they are. In other regions, leave the luminance (Y_{bs}) values of the transformed short exposure image as they are, but overwrite the chrominance (C_{bbs} , C_{rbs}) values of the transformed short exposure image with chrominance (C_{bl} , C_{rl}) values of the long exposure image.
- 15

However, there is a disadvantage in implementing method 100 by acquiring the short exposure image before the long exposure image. This is elaborated below.

20

It is preferable to detect the high frequency content portions from the luminance (Y_{bs}) blocks of the short exposure image (as done in the example of Fig. 2) than from the luminance (Y_l) blocks of the long exposure image. This is because the details in the long exposure image tend to be less sharp than the details in the short exposure image. This may be due to the presence of more motion artifacts in the long exposure image from for example, a greater amount of shaking during the longer exposure time.

25

If the above-mentioned more preferable way of detecting the high frequency content portions is adopted, then implementing method 100 by acquiring the short exposure image before the long exposure image is likely to require more computational effort and/or memory. This is because in this case, while

30

processing the long exposure image, it is necessary to refer to the stored short exposure image to detect the high frequency content portions or a map indicating where these high frequency content portions are, so as to determine which chrominance (C_{bl} , C_{rl}) values of the long exposure image are needed. A
5 more detailed explanation is provided below.

There are usually two separate blocks in a camera processor, namely the JPEG generation block for computing the DCT coefficients of the luminance (Y) and chrominance (C_b , C_r) channels, and the file storage block for storing the DCT
10 coefficients in a JPEG file in memory. In a typical camera, data is processed one way through the JPEG generation block first and then through the file storage block. If the long exposure image is acquired before the short exposure image, this one way flow structure is preserved in the processing of both the long and short exposure images. This is preferred because a longer amount of
15 time is usually required to access the file storage block (storage memory) than the JPEG generation block. On the other hand, if the short exposure image is acquired before the long exposure image, then while processing the long exposure image, there is a need to refer to the stored short exposure image i.e. there is a need to access the file storage block. Thus, the one way flow
20 structure is disrupted, and more computational time and effort may be required.

It is possible not to disrupt the one way flow structure by detecting the high frequency content portions while processing the short exposure image and by constructing and storing a map indicating the locations of these portions in the
25 JPEG generation block. For example, the map may be in the form of a matrix with bits "1" and "0", each bit indicating if a corresponding 8x8 pixel region in the short exposure image comprises high frequency content (thus, the size of the matrix may be the size of the short exposure image divided by 8 in each direction e.g., if the short exposure image has a size of 800x800 pixels, the
30 matrix may be a 100x100 matrix). In this case, access to the file storage block is not required while processing the long exposure image. However, the amount of

memory required is increased. Furthermore, increasing the amount of data stored in the JPEG generation block is usually not preferred.

Alternative ways of detecting high frequency content portions

5

In the example of Fig. 2, the high frequency content portions are detected using the luminance (Y_{bs}) blocks of the short exposure image. As mentioned above, this is preferable. Specifically, in the example of Fig. 2, such detection is performed on the transformed short exposure image. However, it is also possible to perform such detection on the input short exposure image or the initial fused image.

Alternatively, it is also possible to detect the high frequency content portions using the luminance (Y_l) blocks of the input long exposure image (or a transformed long exposure image). However, as mentioned above, the details in the long exposure image tend to be less sharp than the details in the short exposure image. Hence, this alternative is not preferred.

The high frequency content portions may also be detected from the chrominance (C_b , C_r) blocks of any one of the following: the input short exposure image, a transformed short exposure image, the input long exposure image, a transformed long exposure image, the initial fused image. However, this alternative is not preferred either because the chrominance (C_b , C_r) blocks are not reliable indicators of visible high frequency content. This is because the human visual system tends to be less sensitive to edges in the chrominance (C_b , C_r) channels than to edges in the luminance (Y) channel.

The high frequency content portions may also be detected from the intensity values or transformed intensity values (e.g. values in the RGB color space) in any one of the following: the input short exposure image, a transformed short exposure image, the input long exposure image, a transformed long exposure image. In fact, this may give more reliable results than detecting the high

frequency content portions using the luminance (Y) or chrominance (C_b , C_r) blocks. However, it is more computationally intensive to detect the high frequency content portions from the intensity values or the transformed intensity values because a substantial amount of additional computation is required. In contrast, minimal additional computational effort is required if the high frequency content portions are detected using the luminance (Y) or chrominance (C_b , C_r) blocks in JPEG. This is because the detection may be performed during the JPEG compression stage as the DCT coefficients are calculated (using for example, the EOB symbol location as described above).

Alternative ways of generating and storing data

In the example as shown in Fig. 2, all the luminance (Y) and chrominance (C_b , C_r) values of both the input images are generated. Furthermore, all the luminance (Y) and chrominance (C_b , C_r) values for the entire long exposure image are stored.

Alternatively, it is possible to generate only the necessary luminance (Y) and chrominance (C_b , C_r) values of each input image. In the event that all the luminance (Y) and chrominance (C_b , C_r) values are generated for an input image, it is also possible to store only the necessary luminance (Y) and chrominance (C_b , C_r) values of the input image.

For example, in steps 202 – 204 of the example of Fig. 2, it is possible to omit the generation and storage of the luminance (Y_l) values of the long exposure image since these values are not used for forming the output fused image later on. However, since the long exposure image is acquired before the short exposure image and the high frequency content portions are detected using the luminance (Y_{bs}) blocks from the short exposure image in this example, all the chrominance (C_{bl} , C_{rl}) values of the long exposure image have to be generated and stored because it is necessary to refer to the later acquired short exposure image to determine which of these chrominance (C_{bl} , C_{rl}) values are required.

Omitting the generation and storage of unnecessary luminance (Y) and chrominance (C_b , C_r) values may help to reduce the amount of memory and/or computational effort required. However, performing this omission may not be always preferable due to the following reasons.

5

Firstly, to perform the above-mentioned omission, several modifications to existing camera designs have to be made. In contrast, for the example in Fig. 2, only a few minor modifications have to be made to the design of existing cameras (specifically, the JPEG generation blocks) to, for example, overwrite portions of the stored long exposure image with relevant portions of the short exposure image. The JPEG generation blocks in existing cameras tend to be highly optimized and thus, it is desirable to minimize the amount of modifications made to them. Therefore, in some cases, the tradeoff of increased memory requirement to store unnecessary luminance (Y) and chrominance (C_b , C_r) values for a reduced amount of modifications made to a JPEG generation block in an existing camera may be worth it, and may be preferred by camera hardware designers.

Secondly, in an existing camera, luminance (Y) and chrominance (C_b , C_r) blocks in a JPEG formatted image are read off in a raster scan order (starting from blocks at the top left of the image) and are stored in spatial order according to the JFIF. Thus, even if the camera is configured to write only the necessary chrominance (C_b , C_r) and luminance (Y) blocks of an input image to a JPEG file, when an unnecessary chrominance (C_b , C_r) or luminance (Y) block is read, the camera will still have to write a filler pattern comprising bits "1" and/or "0" to the part of the JPEG file originally meant for storing this unnecessary block since this part of the JPEG file cannot be left blank. The filler patterns in the JPEG file will be subsequently overwritten by relevant values from the other input image. Therefore, even though it is possible to omit the storage of unnecessary chrominance (C_b , C_r) and luminance (Y) blocks, the reduction in computational effort and memory requirement achieved by this omission, though present, may not be substantial. Thus, it may be preferable to

retain the regularity of the existing camera's operations and make fewer modifications to the existing camera design than to perform the conditional generation and storage of chrominance (C_b , C_r) and luminance (Y) blocks.

- 5 If the preferred way of detecting the high frequency content portions from the luminance (Y_{bs}) blocks of the short exposure image is adopted, the option to omit the generation and storage of unnecessary chrominance (C_b , C_r) and luminance (Y) blocks provides an advantage to acquiring the short exposure image before the long exposure image. This is because, in this case, since it is possible to determine which of the chrominance (C_{bbs} , C_{rbs}) values in the short exposure image are required before acquiring the long exposure image, it is possible to generate and store only these required chrominance (C_{bbs} , C_{rbs}) values. In contrast, if the long exposure image is acquired first, it is necessary to generate and store all the chrominance (C_{bl} , C_{rl}) values of the long exposure image because the later-acquired short exposure image is required for determining which of these chrominance (C_{bl} , C_{rl}) values are actually unnecessary.

However, as mentioned above, acquiring the short exposure image before the long exposure image has a disadvantage in that either the one way flow structure from the JPEG generation block to the file storage block has to be disrupted or more data has to be stored in the JPEG generation block. In the design of camera hardware, it is usually more preferable to retain the regularity of operations so as to avoid the use of the "if" operation (e.g. "if (condition) do processing") which may cause the total processing time to become non-deterministic. It is also usually more preferable to minimize the amount of data stored in the JPEG generation block. Therefore, in most cases, the disadvantage in acquiring the short exposure image before the long exposure image outweighs its advantage. This is especially since, as mentioned above, the reduction in computational effort and/or memory requirement achieved by omitting the generation and storage of unnecessary luminance (Y) and

chrominance (C_b , C_r) blocks may not even be substantial. Thus, acquiring the long exposure image first is still the preferable option in most cases.

Experimental Results

5

Method 100 (using the specific example in Fig. 2) was applied on test images of a variety of scenes. These test images were captured using hand-held cameras in both day light and low light. The AEB in continuous drive mode (bracketing the exposure time) was used and all other settings such as ISO and the F number were held constant. The test images and results may be found on the website: <http://picasaweb.google.com/ramya.hebbalaguppe/JRTIP> where the images are shown in color and at a higher resolution to enable a closer visual inspection.

10

15

Figs. 5(a) – (e), 6(a) – (d), 7(a) – (e) and 8(a) – (b) illustrate experimental results obtained by applying method 100 (using the example in Fig. 2) on some of the test images. These results are obtained using an empirically determined EOB location threshold of 15 (out of 64). Furthermore, the default matrices described in Pennebaker and Mitchell (1993) are used as the quantization matrices. Note that although the following includes description on the color information in the images (for example, “bleeding” of the colors or discoloration), such color information is not visible in the grayscale version of the images. However, such color information is visible in the color version of the images at the above-mentioned website.

20

25

In particular, Figs. 5(a) – (e) illustrate the results obtained by applying method 100 for digital image stabilization.

Specifically, Figs. 5(a) and (b) show a first set of input images to method 100.

30

These images are hand-held images of a basketball court scene captured with two different exposure times under low light. Fig. 5(a) shows the long exposure image with exposure time $EV(0)$ (i.e. normal exposure) whereas Fig. 5(b) shows

the short exposure image with exposure time $EV(-2)$ (i.e. deliberately underexposed by a factor of 4). Note that the captured short and long exposure images contain some artifacts. For example, in the long exposure image in Fig. 5(a), motion blur artifacts are present around the lady's legs on the left and the child on the right.

Fig. 5(c) shows the initial fused image (after decompression) from step 210. Specifically, the initial fused image of Fig. 5(c) comprises the chrominance (C_{bl} , C_{rl}) values of the long exposure image and the luminance (Y_{bs}) values of the transformed short exposure image. As shown in Fig. 5(c), the initial fused image contains several artifacts, for example, around the lady's legs and the child's hands. Furthermore, the color of the initial fused image "bleeds". Note that in this case, the luminance of the short exposure image is boosted in step 206 by estimating the BTF using the scaled sigmoidal function in Equation (2).

Fig. 5(d) shows a replacement map for overwriting the chrominance (C_{bl} , C_{rl}) values in the initial fused image. Specifically, the specks shown in Fig. 5(d) represent the 8×8 pixel regions in the initial fused image whose chrominance (C_{bl} , C_{rl}) values are to be overwritten.

Fig. 5(e) shows the output fused image after removing artifacts from the initial fused image by overwriting the chrominance (C_{bl} , C_{rl}) values in the initial fused image based on the replacement map in Fig. 5(d). As shown in Fig. 5(e), there are fewer artifacts (i.e. less noise) in the output fused image as compared to the initial fused image and the output fused image appears sharper. Furthermore, the output fused image has clearer colors (for example, the grass is greener in the output fused image than in the initial fused image).

Fig. 6(a) shows a first enlarged portion of the initial fused image in Fig. 5(c) whereas Fig. 6(b) shows a corresponding enlarged portion of the output fused image in Fig. 5(e). As shown in Fig. 6(a), discoloration due to artifacts is present on the lady's legs in the initial fused image. After the artifact removal in step

212, the discoloration is removed and the skin tone of the lady's legs is restored in the output fused image as shown in Fig. 6(b). Fig. 6(c) shows a second enlarged portion of the initial fused image in Fig. 5(c) whereas Fig. 6(d) shows a corresponding enlarged portion of the output fused image in Fig. 5(e). Similarly,
5 as shown in Figs. 6(c) and (d), the artifacts on the baby's shirt and legs in the initial fused image are removed after the artifact removal in step 212. Thus, as shown in Figs. 6(a) – (d), step 212 is able to effectively remove ghosting artifacts.

10 Figs. 7(a) – (e) shows the results of applying method 100 for HDR imaging.

In particular, Figs. 7(a) and (b) show a second set of input images to method 100. These images show a person at a covered walkway in Nanyang Technological University and are captured using a hand-held camera in daylight. More specifically, Fig. 7(a) shows the long exposure image with exposure time EV(+2) (i.e. an over-exposed image) whereas Fig. 7(b) shows
15 the short exposure image with exposure time EV(-2) (i.e. deliberately under-exposed by a factor of 4). As shown in Fig. 7(a), although most of the details are visible in the long exposure image, the edges in the long exposure image are not sharp. In contrast, as shown in Fig. 7(b), the edges (for example, those
20 of the walkway) in the short exposure image are sharp. Furthermore, the intensity of the road on the right of the walkway is saturated in the long exposure image whereas that in the short exposure image is not (hence, the divider line on the road in the short exposure image remains visible). However,
25 the long exposure image has a higher SNR and more color information as compared to the short exposure image, particularly at areas where the luminance is low. For example, the lights on the ceiling of the covered walkway are not visible in the short exposure image whereas they are visible in the long exposure image. Furthermore, the person's face is not fully luminated and is
30 hence hardly visible in the short exposure image, whereas it is clearly visible in the long exposure image.

Fig. 7(c) shows the initial fused image obtained after fusing the luminance (Y_{bs}) values of the transformed short exposure image with the chrominance (C_{bl} , C_{rl}) values of the long exposure image in step 210. As shown in Fig. 7(c), ghosting artifacts are present in the initial fused image as halos around the person.

5 Furthermore, the color of the initial fused image "bleeds". Note that in step 206 in this case, the luminance of the short exposure image is boosted by estimating the BTF using the scaled sigmoidal function in Equation (2) and tone-mapping is applied to the luminance boosted short exposure image to form the transformed short exposure image. The tone-mapping is done using a
10 global tone map with the value of α in Equation (3) limited to $6/M$. This global tone map is implemented as a LUT and is applied to the luminance boosted short exposure image to improve the contrast in the bright regions on the road. This tone-mapping allows the short exposure image, and thus, the initial fused image in Fig. 7(c), to appear more pleasing to the viewer. This can be seen
15 from Figs. 7(a) – (c) whereby the brightness of the initial fused image is visibly different from that of both input images.

Fig. 7(d) shows the replacement map for the artifact removal in step 212. The specks in Fig. 7(d) represent the 8×8 pixel regions in the initial fused image
20 corresponding to the regions having strong edges in the short exposure image (as detected using the EOB symbol locations and a EOB location threshold of 15).

Fig. 7(e) shows the output fused image after performing step 212 with the replacement map in Fig. 7(d). As shown in Fig. 7(e), the output fused image is sharp, has a high SNR and has more color information than the short exposure
25 image. For example, unlike in the short exposure image, the lights on the ceiling of the walkway are visible in the output fused image. There is also less color "bleeding" in the output fused image. Furthermore, the output fused image
30 has fewer artifacts, for example, ghosting artifacts around the person in the initial fused image are removed.

The removal of the artifacts from the initial fused image is more clearly illustrated in Figs. 8(a) and (b). In particular, Fig. 8(a) shows an enlarged portion of the initial fused image in Fig. 7(c) whereas Fig. 8(b) shows a corresponding enlarged portion of the output fused image in Fig. 7(e). As shown in Figs. 8(a) and (b), step 212 is effective in removing artifacts (this is more apparent in the color version of the images at the above-mentioned website).

As shown in Figs. 5(a) – (e), 6(a) – (d), 7(a) – (e) and 8(a) – (b), method 100 works well on a variety of scenes, and in both day light and low light. It can further be seen that while the presence of noise complicates the selection of chrominance blocks to be overwritten for artifact removal, using an EOB location threshold of 15 allows step 212 to perform well on a variety of scenes taken at night with varying exposure times and ISO sensitivities.

Method 100 (using the example in Fig. 2) is further evaluated using quantitative measures of the SNR and sharpness of the output fused image as follows.

In method 100, the short exposure image is fused with the long exposure image. Therefore, the SNR of the output fused image is at best equal to that of the input long exposure image, and at worst equal to that of the input short exposure image.

The SNR of an image may be expressed as the ratio between the mean and standard deviation (StdDev) of the intensity values in the image as shown in Equation (5).

$$\text{SNR of an image} = \frac{\text{Mean(Intensity values in the image)}}{\text{StdDev(Intensity values in the image)}} \quad (5)$$

The percent improvement in SNR of the output fused image, relative to the SNR of the input short exposure image, may be calculated according to Equation (6) whereby SNR(fused) represents the SNR of the output fused image,

SNR(short) represents the SNR of the input short exposure image and SNR(long) represents the SNR of the input long exposure image.

$$\text{SNR percent improvement} = 100 \times \frac{\text{SNR(fused)} - \text{SNR(short)}}{\text{SNR(long)} - \text{SNR(short)}} \quad (6)$$

5

Similarly, the sharpness of the output fused image is at best equal to that of the input short exposure image, and at worst equal to that of the input long exposure image.

10 The percent improvement in sharpness achieved by the image fusion in method 100 may also be calculated. There are many measures of sharpness, such as the effective bandwidth measure (Fishbain et al (2008)) and the blur metric (Crete-Roffet et al (2007)). The blur metric (denoted as Blur) is relatively simple and effective, and thus, will be used here. Blur computes blurriness as a
15 continuous index in the range [0, 1], with 0 denoting the sharpest possible image and 1 the blurriest. The complement of blurriness is the sharp metric (denoted as Sharp) which measures sharpness.

The percent improvement in sharpness of the output fused image, relative to
20 that of the input long exposure image, may be calculated according to Equation (7) where $\text{Sharp} = 1 - \text{Blur}$, Sharp(fused) represents the sharpness of the output fused image, Sharp(long) represents the sharpness of the input long exposure image and Sharp(short) represents the sharpness of the input short exposure image.

25

$$\text{Sharp percent improvement} = 100 \times \frac{\text{Sharp(fused)} - \text{Sharp(long)}}{\text{Sharp(short)} - \text{Sharp(long)}} \quad (7)$$

Note that even the best of published sharpness measures merely correlate with our perception of sharpness, and may not be truly reflective of how sharp the
30 image appears. Therefore, the sharpness measures in Equation (7) may

occasionally give false readings. In the rare case where the denominator on the right hand side of Equation (7) is negative, it is replaced by 1.0. The sharpness measures are evaluated on the grayscale version of the images. Furthermore, following standard practice in sharpness measurements, the images are resized to reflect the viewing conditions which may be for example, the conditions on a camera's preview screen.

Table 1 shows the quantitative evaluation of the improvement in SNR and sharpness achieved by method 100. In particular, Table 1 shows the SNR percent improvement and the Sharp percent improvement of the output fused image when method 100 (using the example in Fig. 2) is applied on input images of five different scenes.

Scene	%SNR increase	% Sharpness increase
Basketball	67.0	91.1
Walkway	6.3	95.3
Cathedral	46.9	88.1
Cafe	2.2	0.4
Monas	2.5	81.4

Table 1

From Table 1, it can be seen that method 100 improves the SNR by an average of 25% for the five scenes tested. Method 100 also improves the sharpness by an average of 71%. For the "Cafe" scene, the sharpness of the long exposure image is calculated as being slightly higher than that of the short exposure image. This is likely to be a measurement error in the blur metric, which is in turn, probably due to the fact that there is only a slight amount of subject motion between acquiring the input images.

Note that although the same EOB location threshold and quantization matrices are used to obtain the results above, these may be further optimized to achieve

better results. For example, the EOB location threshold may be varied in the range from 1 to 64 to adjust the sensitivity of the JPEG-based detector employed in step 212.

5 Fig. 9 shows how the percentage of chrominance blocks replaced in the initial fused image in step 212 varies inversely with the EOB location threshold for two different night time (or low light) scenes. These scenes comprise a basketball court scene (see graph 902) and a cathedral scene (see graph 904). The basketball court scene is shown in Figs. 5(a) – (e) whereas the cathedral scene
10 is not shown in the figures.

The results in Fig. 9 are obtained by using default JPEG quantization matrices as described in Pennbaker and Mitchell (1993). As shown in Fig. 9, with the default quantization matrices, the percentage of chrominance blocks replaced in
15 step 212 is very small (less than 3.5%). In other words, nearly all of the chrominance values in the output fused image are those of the high SNR input long exposure image. Note that with input images of a different scene or with different quantization matrices, graphs different from those shown in Fig. 9 may be obtained. However, it has been observed that with the default quantization
20 matrices, the percentage of chrominance blocks replaced in step 212 usually remains very small for images captured at night time regardless of the scene or the quantization matrices used.

Advantages of method 100

25

The following describes some advantages of method 100.

As shown in the results above, method 100 works well for both image stabilization and HDR capture. In particular, method 100 is capable of
30 combining the uniform regions of the long exposure image (having a high SNR) with the detailed regions of the short exposure image (having sharp details),

thereby reducing noise while providing sharp details to obtain a fused image with a reasonably high image quality.

Furthermore, method 100 is efficient in both computational effort and memory.

5 Thus, it can be implemented at low cost in real time and as part of a digital camera's hardware image processing engine.

The efficiency in computational effort and memory of method 100 is partly achieved due to the possibility of performing the steps of method 100 in a
10 pipelined manner. For example, if the BTF in step 206 is estimated using the sigmoidal function, method 100 (implemented using the example in Fig. 2) requires only a single pass over each input image. In particular, steps 206 – 214 may be performed in a pipelined fashion so that the processing of each pixel of the short exposure image through these steps starts once the pixel is read from
15 the camera sensor. In each of these steps, there is no need to buffer more than the minimum number of pixels required to compute the necessary operations in the step. The same applies for steps 202 – 204 for the long exposure image. The computational complexity of method 100 may be further reduced by implementing the BTF in the form of a LUT as described above.

20

In addition, the fusion of the short and long exposure images is achieved in method 100 (implemented using the example in Fig. 2) by overwriting data in the long exposure image with relevant data in the short exposure image. During the fusion process i.e. steps 210 – 212, it is only necessary to store the
25 overwritten long exposure image and there is no need to store any part of either the input long exposure image or the input short exposure image. Due to this and the single pass nature of method 100, at any point in time, method 100 does not need memory beyond that required for storing a single image. In fact, if method 100 is performed on a pixel region-by-pixel region basis as described
30 above, method 100 needs only the memory required for storing a single pixel region (for example, a 16x16 pixel region corresponding to a macro-block in the JPEG format).

Method 100 is also computationally efficient as it is able to achieve a good image quality without the need for image registration. Usually, even if two images are taken in immediate succession, combining the luminance of one of the images with the chrominance of the other image is still likely to lead to artifacts. These artifacts may be caused by either camera motion (due to hand shake) or object motion between the two shots capturing the images. Such motion artifacts are usually compensated by means of image registration (see Tico and Pulli (2009)) which uses for example, cross-correlation to shift the later acquired image to align with the earlier acquired image. On the other hand, method 100 removes artifacts by making use of the observation that motion artifacts are most noticeable at edges or other regions comprising high frequency content, and that the human visual system is less sensitive to motion artifacts in the chrominance domain than in the luminance domain. For example, by replacing chrominance values of the initial fused image with chrominance values of the transformed short exposure image in regions with high frequency content, method 100 implemented using the example in Fig. 2 is able to obtain a reasonably good image quality without performing image registration.

JPEG is the preferred compression method for the vast majority of digital cameras in use today. Calculations, including the DCT, run-length coding and entropy coding (Wallace (1992)), are performed by embedded software and hardware blocks residing within a camera's onboard image processor. Therefore, implementation of method 100 in the JPEG domain can benefit from the considerable optimization and design that is available for JPEG. Furthermore, by performing method 100 in the JPEG domain, several calculations which are already part of the JPEG algorithm can be reused. For example, to implement steps 202 – 214 of Fig. 2 in the JPEG domain, the processing of the long exposure image need not be altered, and the only principal modifications that need to be made to a typical digital camera are (1) a modification to the JPEG encoder and (2) a modification to the file save operation. Furthermore, several characteristics of the JPEG file format may be

exploited. Firstly, the overwriting in steps 210 – 212 exploits the fact that in the JPEG format, the luminance (Y) blocks are stored separately from the chrominance (C_b , C_r) blocks. Secondly, step 212 exploits the built-in spatial frequency analysis already available in JPEG to remove artifacts from the initial fused image. In particular, step 212 uses the EOB symbol locations in the DCT coefficients from JPEG DCT calculations. Note that while it is well known that DCT coefficients may be used to detect pixel blocks containing strong edges (see Pennebaker and Mitchell (1993) and Kakarala and Bagadi (2009)), to date, there has been no prior art suggesting the use of the EOB symbol for detecting the edges.

Method 100 is superior to several methods proposed in published research. For example, since method 100 (implemented using the example in Fig. 2) requires only a single pass over each input image, it is more efficient and requires less computation effort than the methods of Lu et al (2009), Tico and Pulli (2009) and Tico et al (2010). These prior art methods require at least three passes on every input image to complete the operations of registration, intensity matching, and fusion (two passes for BTF estimation and image boosting, in addition to implementing chrominance fusion in the wavelet domain). Furthermore, the method of Lu et al (2009) is iterative in nature, requiring 6-15 iterations in practice, and therefore clearly not capable of real-time application on digital cameras. The method of Tico and Pulli (2009) uses a wavelet transform on both the short and long exposure images. This requires a significant overhead which cannot be reused for image compression. In contrast, method 100 performed in the JPEG domain is able to reuse several calculations which are already part of the JPEG algorithm.

Figs. 10(a) – (b) respectively show output fused images obtained by applying the methods proposed in Lu et al (2009) and Tico et al (2010) on the first set of input images with the basketball scene shown in Figs. 5(a) – (b). These results may also be found at <http://picasaweb.google.com/ramya.hebbalaguppe/JRTIP>.

As shown in Fig. 10(a), the output fused image of Lu et al (2009), which takes 187 seconds to compute, contains considerable noise and ghosting artifacts. As pointed out in Lu et al (2009), their method is not suited to be used on input images taken over a time frame with subject motion.

5

On the other hand, as shown in Fig. 10(b), the output fused image of Tico et al (2010), which takes roughly 60 seconds to compute for 10-megapixel images, is of a reasonably good quality. Specifically, the SNR percent improvement achieved by Tico et al (2010) is 66% as computed using Equation (6). This is roughly the same as the SNR percent improvement achieved by method 100 (implemented using the example in Fig. 2). However, note that the output fused image from the method of Tico et al (2010) as shown in Fig. 10(b) is obtained using 3 input images rather than 2.

10

Although the methods proposed by Tico and Pulli (2009) and Tico et al (2010) are also able to achieve a reasonably good image quality, their computational complexity is high as they require operations such as registration, intensity equalization, deconvolution, and wavelet-based image fusion. Implementing any of these operations on a mobile device is not trivial. This is because with these operations, a powerful processor is required in order to produce the results without necessitating a long wait on the part of the user. In contrast, method 100 requires minimal computation: e.g. using the example in Fig. 2, method 100 only requires operations such as accessing a LUT, detecting the position of the EOB symbol in each luminance block, and selectively overwriting the file storing the long exposure image. In particular, it requires only 0.8 seconds to perform method 100 (using the example in Fig. 2) on 2-megapixel images with a laptop computer comprising a 1.6 GHz dual-core processor. This is much faster than the methods proposed in Lu et al (2009) and Tico et al (2010). Moreover, if method 100 (using the example in Fig. 2) is implemented as part of the processing pipeline in a typical digital camera, it is only necessary to modify the JPEG encoder and the file write operation in the camera.

20

25

30

Note that as with any image fusion method including those having much higher complexity, there is a tradeoff between the image dynamic range and the extent of visibility of the artifacts. Artifacts from output fused images of method 100 may still be noticeable if the images are enlarged and scrutinized closely.

- 5 Nevertheless, method 100 uses minimal computation time and is able to achieve a relatively good performance with the amount of artifacts being small enough such that they are hardly noticeable on prints or small screens typically used by, for example, mobile devices.

References

- [1] Akil M, Grandpierre T, Perroton L (2011) Real-time dynamic tone-mapping operator on GPU. J. Real-Time Image Processing, Published online first: 11 Feb 2011, doi: 10.1007/s11554-011-0196-7.
- [2] Bogoni L (2000) Extending dynamic range of monochrome and color images through fusion. In: 15th International Conference on Pattern Recognition (ICPR), IEEE, vol 3, pp 7–12.
- [3] Roffet-Crete F, Dolmiere T, Ladret P, Nicolas M (2007) The blur effect: perception and estimation with a new no-reference perceptual blur metric. In: Proc. SPIE Vol. 6492: Human Vision and Electronic Imaging XII.
- [4] Gallo O, Gelfand N, Chen W, Tico M, Pulli K (2009) Artifact-free high dynamic range imaging. In: IEEE Intl. Conf. Computational Photography (ICCP), pp 1–7.
- [5] Gelfand N, Adams A, Park SH, Pulli K (2010) Multi-exposure imaging on mobile devices. In: ACM Multimedia, pp 823–826.
- [6] Fishbain, B, Yaroslavsky L, Ideses I, Roffet-Crete F (2008) No-reference method for image effective-bandwidth estimation. In: Proc. SPIE Vol. 6808: Image Quality and System Performance V.
- [7] Hasinoff SW, Durand F, Freeman WT (2010) Noise-optimal capture for high dynamic range photography. In: CVPR, pp 553–560.
- [8] Jacobs K, Loscos C, Ward G (2008) Automatic high dynamic range image generation for dynamic scenes. IEEE Computer Graphics and Applications, vol 28, pp 24–33.
- [9] Kakarala R, Bagadi R (2009) A method for signaling block-adaptive quantization in baseline sequential jpeg. In: IEEE TENCON, pp 1–6.
- [10] Kang S, Uyttendaele M, Winder S, Szeliski R (2008) System and Process for generating high dynamic range video. U.S. Patent 7,382,931.
- [11] Khan E, Akyuz A, Reinhard E (2006) Ghost removal in high dynamic range images. In: IEEE Intl. Conf. Image Processing (ICIP), pp 530–533.
- [12] Lu PY, Huang TH, Wu MS, Cheng YT, Chuang YY (2009) High dynamic range image reconstruction from hand-held cameras. Computer Vision and

Pattern Recognition (CVPR), IEEE Computer Society Conference on pp 509–516.

[13] Mitchell HB (2010) Image fusion: theories, techniques and applications. CRC Press.

5 [14] Nakamura J (2005) Image sensors and signal processing for digital still cameras. CRC Press.

[15] Pennebaker WL, Mitchell JL (1993) The JPEG still image data compression standard. Kluwer Academic Publishers.

10 [16] Reinhard E, Ward G, Pattanaik S, Debevec P (2005) High dynamic range imaging: Acquisition, display and image-based lighting. Morgan Kaufmann Publishers.

[17] Sen M, Hemaraj Y, Plishker W, Shekhar R, Bhattacharya SS (2008) Model-based mapping of reconfigurable image registration on FPGA platforms. J. Real-Time Image Processing 3 (3), pp 149- 162.

15 [18] Stathaki T (2008) Image fusion: algorithms and applications. Academic Press.

[19] Tico M, Pulli K (2009) Image enhancement method via blur and noisy image fusion. In: Proc. Int. Conf. on Image Processing (ICIP), IEEE, pp 1521 – 1524.

20 [20] Tico M, Gelfand N, Pulli K (2010) Motion-blur free exposure fusion. In: Proc. Int. Conf. on Image Processing (ICIP), IEEE, pp 1521 – 1524.

[21] Wallace GK (1992) The jpeg still picture compression standard. IEEE Trans Consumer Electronics 38(1).

25 [22] Zhang X, Jones RW, Baharav I, Reid DM (2006) System and method for digital image tone mapping using an adaptive sigmoidal function based on perceptual preference guidelines. US Patent 7,023,580.

Claims

1. A method for fusing a first input image with a second input image to form an output fused image, the first and second input images being photographic
5 images of the same scene captured successively, the exposure time of the first input image being longer than the exposure time of the second input image, the first and second input images having equal pixel dimensions and each comprising respective intensity values for each pixel and for each of multiple colors, the method comprising:
 - 10 (i) transforming the intensity values of at least one of the input images, to equalize the overall intensity of the input images;
 - (ii) deriving, from values of at least one of the input images, locations of high frequency content portions of the output fused image;
 - (iii) forming the high frequency content portions of the output fused
15 image using spatial frequency domain chrominance and luminance values obtained from the corresponding portions of the second input image; and
 - (iv) forming other portions of the output fused image using spatial frequency domain chrominance values obtained from the corresponding portions of the first input image and spatial frequency domain luminance values
20 obtained from the corresponding portions of the second input image.
2. A method according to claim 1, wherein step (i) comprises transforming the intensity values of the at least one of the input images based on a sigmoidal
25 function.
3. A method according to claim 2, wherein the sigmoidal function is representative of a correspondence between the intensity values of the first and second input images based on a difference in the exposure times between the first and second input images.

4. A method according to claim 2 or 3 wherein transforming the intensity values of the at least one of the input images based on the sigmoidal function comprises, for each intensity value of the at least one of the input images,
accessing a look-up table comprising values of the sigmoidal function
5 corresponding to respective intensity values;
from the look-up table, determining a value of the sigmoidal function corresponding to the intensity value of the at least one of the input images; and
replacing the intensity value of the at least one of the input images with
its corresponding value of the sigmoidal function determined from the look-up
10 table.
5. A method according to any one of claims 2 – 4, wherein step (i) further comprises further transforming the transformed intensity values of the at least one of the input images based on a tone map.
- 15
6. A method according to claim 5, wherein further transforming the transformed intensity values comprises for each transformed intensity value of the at least one of the input images,
accessing a look-up table comprising values of the tone map
20 corresponding to respective transformed intensity values;
from the look-up table, determining a value of the tone map corresponding to the transformed intensity value of the at least one of the input images; and
replacing the transformed intensity value of the at least one of the input
25 images with its corresponding value of the tone map determined from the look-up table.
7. A method according to claim 5 or 6, wherein the intensity values of the at least one of the input images are transformed based on the sigmoidal function
30 and the tone map simultaneously.

8. A method according to any one of the preceding claims, wherein step (i) comprises transforming the intensity values of the second input image.
9. A method according to any one of the preceding claims, wherein steps (iii)
5 – (iv) comprise:
 overwriting spatial frequency domain luminance values of the first input image with the spatial frequency domain luminance values obtained from the second input image to form an initial fused image; and
 in portions of the initial fused image corresponding to the high frequency
10 content portions of the output fused image, overwriting spatial frequency domain chrominance values with the spatial frequency domain chrominance values obtained from the corresponding portions of the second input image.
10. A method according to any one of the preceding claims, wherein step (ii)
15 comprises deriving the locations of the high frequency content portions of the output fused image from the spatial frequency domain luminance values of one of the input images.
11. A method according to claim 10, wherein the locations of the high
20 frequency content portions of the output fused image are derived from the spatial frequency domain luminance values of the second input image.
12. A method according to any one of the preceding claims, wherein the method is implemented in the JPEG domain.
- 25
13. A method according to claim 12, wherein the first and second input images each comprises spatial frequency domain luminance values in luminance blocks, and step (ii) further comprises for each luminance block of one of the input images:
30 determining a location of an End-of-Block symbol in the luminance block;
 and

if the location of the End-of-Block symbol occurs after a pre-determined threshold, identifying a portion of the output fused image corresponding to a portion comprising the luminance block in the one of the input images as a high frequency content portion.

5

14. A method according to any one of the preceding claims, wherein a JPEG encoder of a digital camera is modified to perform step (i).

15. A method according to any one of the preceding claims, wherein a file write operation of a digital camera is modified to perform steps (ii) – (iv).

10

16. A method according to claim 1, wherein the first and second input images each comprises a plurality of pixel regions with each pixel region comprising a plurality of the pixels and wherein steps (ii) – (iv) are performed on a pixel region-by-pixel region basis through a series of operations, whereby in each operation, a pixel region of the output fused image is formed by:

15

determining, from values of the corresponding pixel region of at least one of the input images, if the pixel region of the output fused image comprises high frequency content; and

20 if so, forming the pixel region of the output fused image using spatial frequency domain chrominance and luminance values obtained from the corresponding pixel region of the second input image;

otherwise, forming the pixel region of the output fused image using spatial frequency domain chrominance values obtained from the corresponding pixel region of the first input image and spatial frequency domain luminance values obtained from the corresponding pixel region of the second input image.

25

17. A method according to any one of the preceding claims, wherein the first input image has a higher signal-to-noise ratio than the second input image and the second input image comprises sharper details than the first input image.

30

18. A method according to any one of the preceding claims, wherein the first input image is acquired before the second input image.

19. A computer system having a processor arranged to perform a method
5 according to any one of claims 1 – 18.

20. A computer program product such as a tangible data storage device,
readable by a computer and containing instructions operable by a processor of
a computer system to cause the processor to perform a method according to
10 any one of claims 1 – 18.

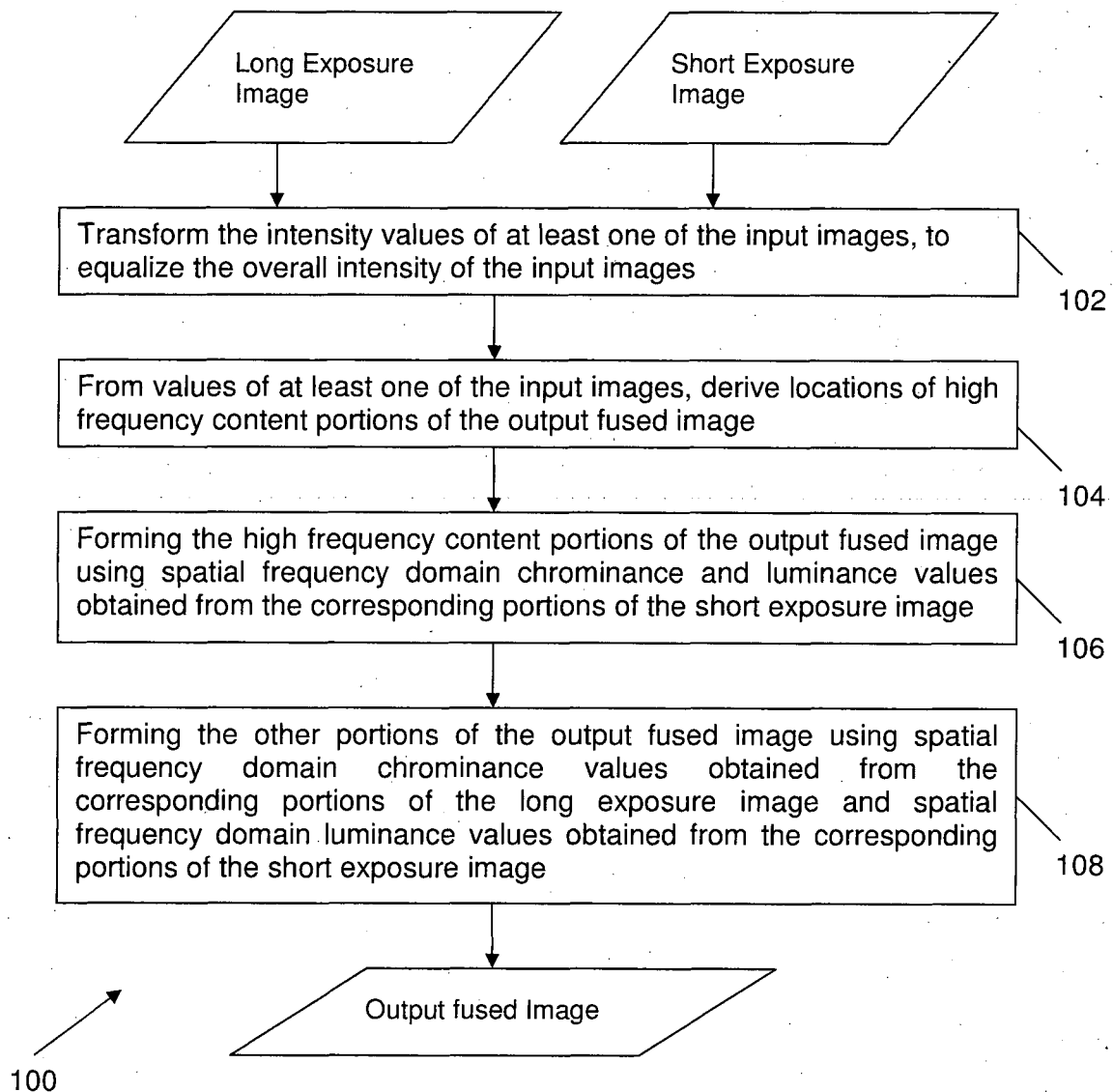


Fig. 1

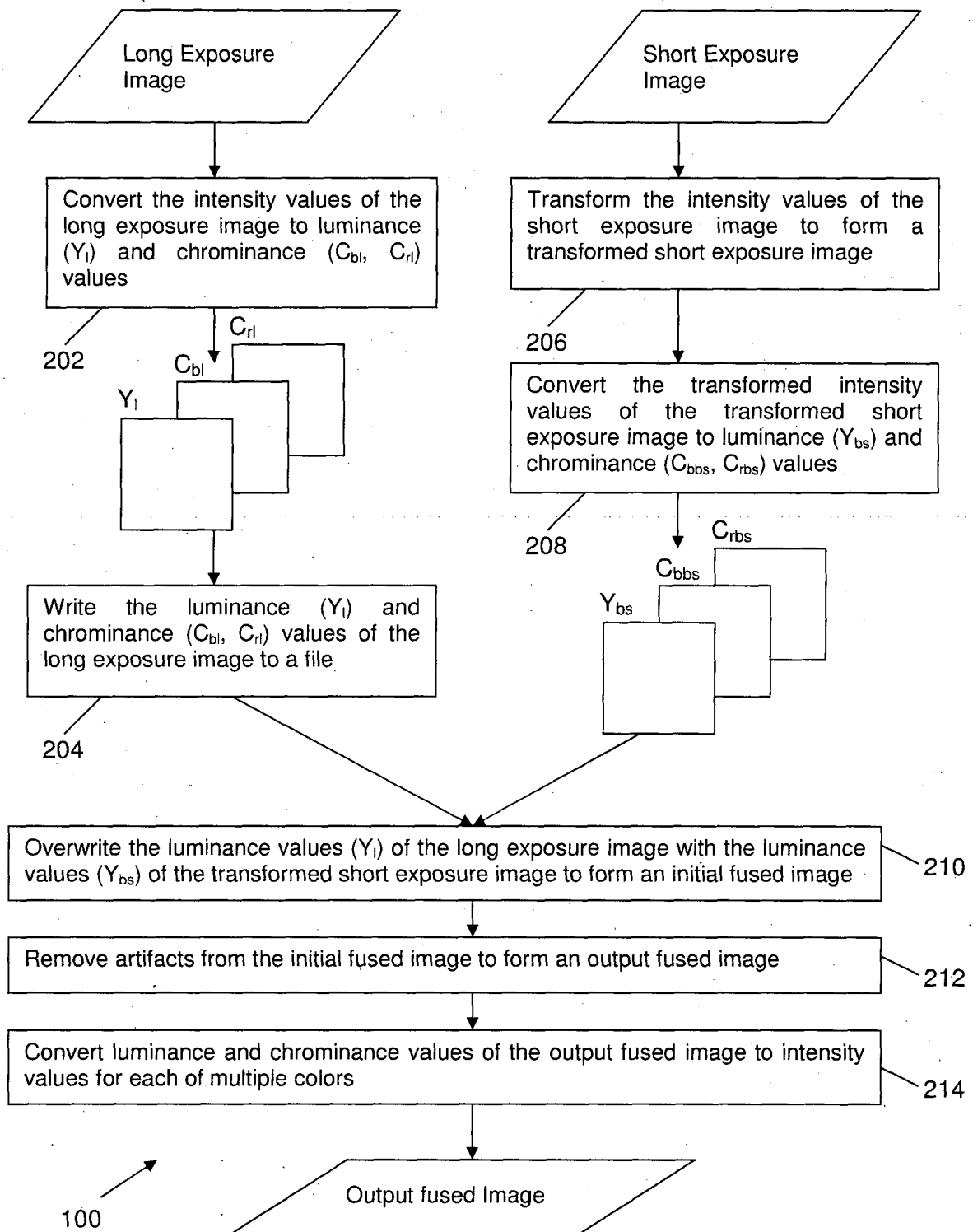


Fig. 2

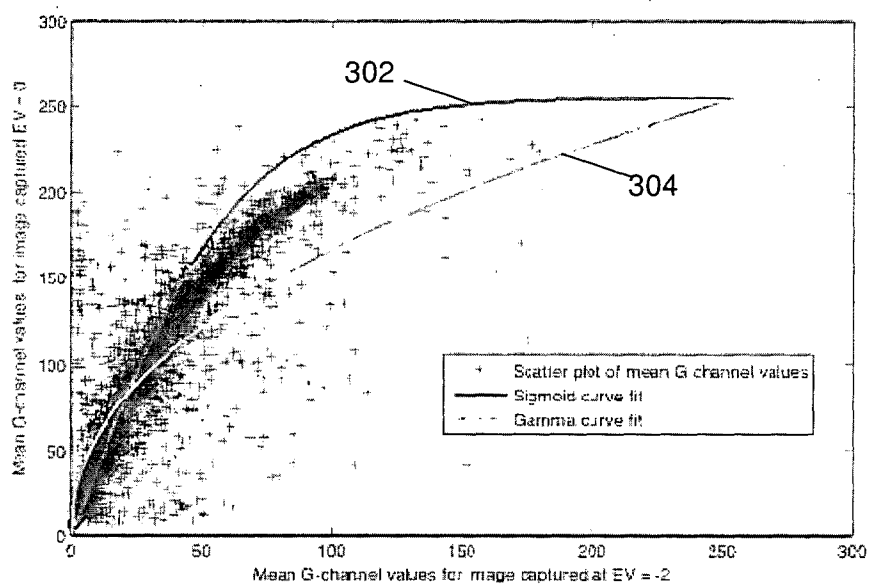


Fig. 3

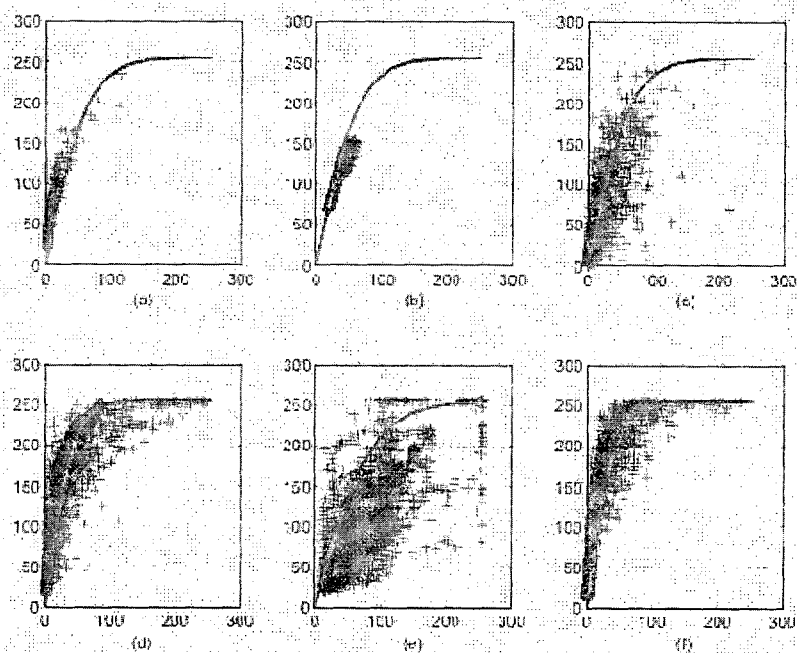
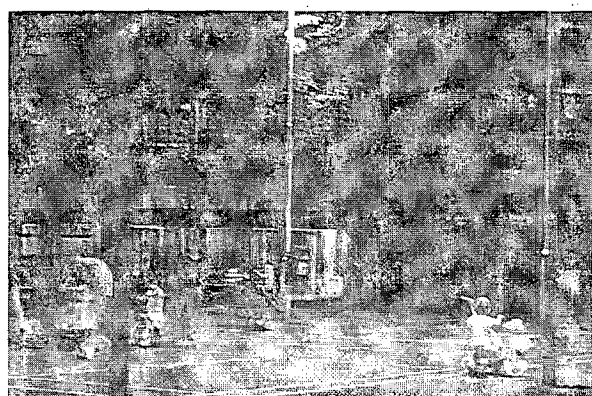
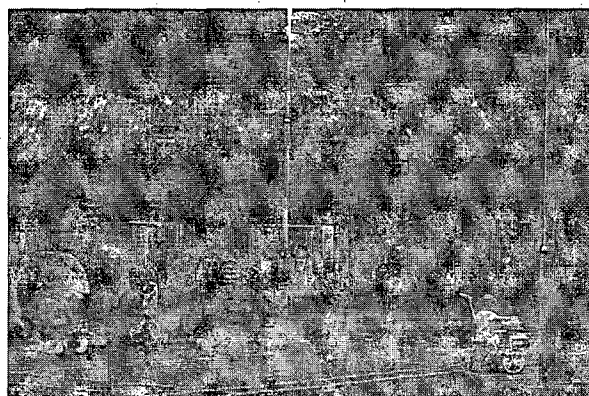


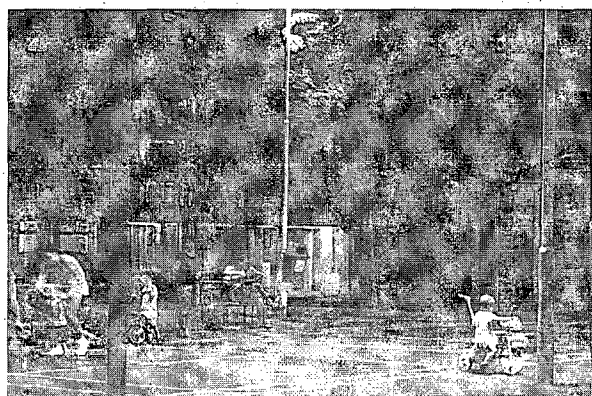
Fig. 4



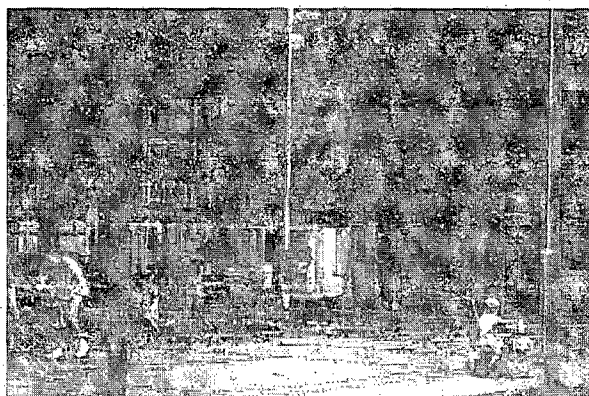
(a)



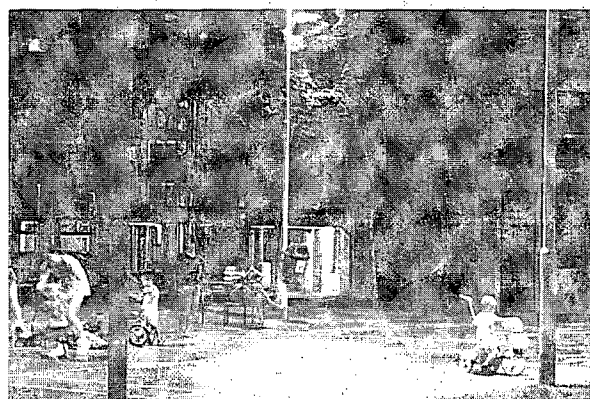
(b)



(c)



(d)



(e)

Fig. 5



(a)



(b)



(c)

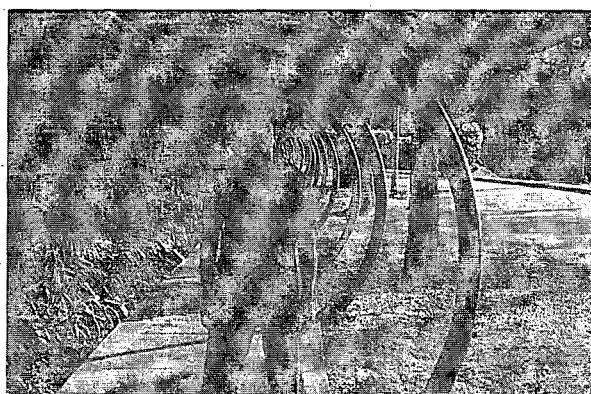


(d)

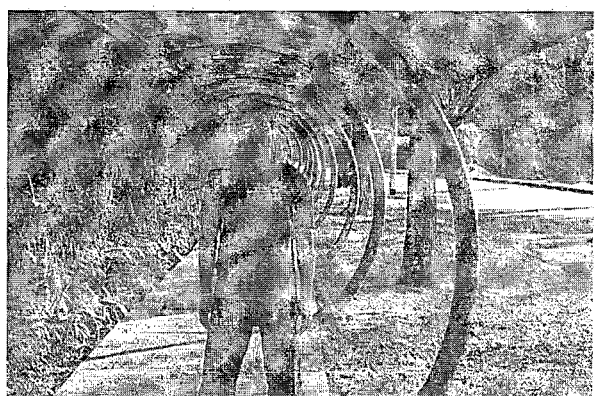
Fig. 6



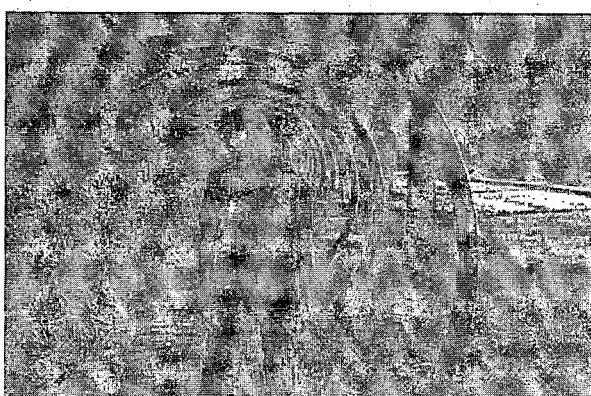
(a)



(b)



(c)



(d)



(e)

Fig. 7

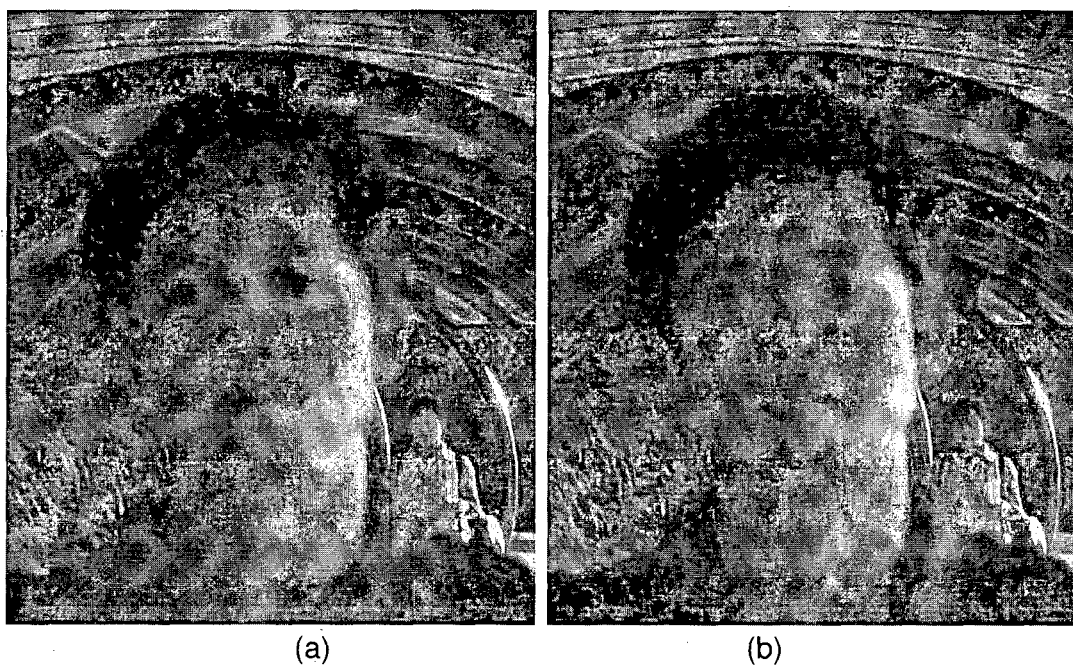


Fig. 8

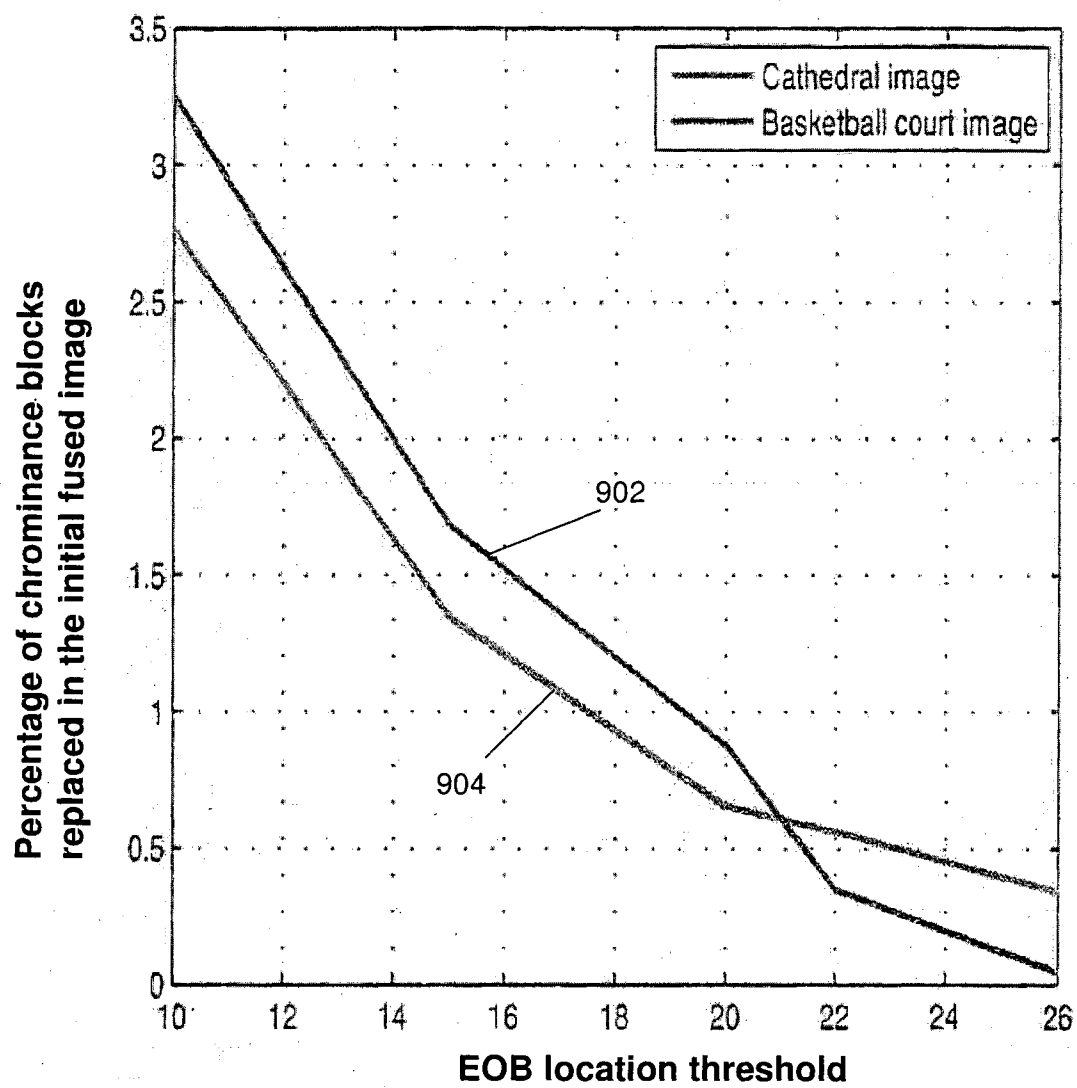


Fig. 9



(a)



(b)

Fig. 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2012/000204

A. CLASSIFICATION OF SUBJECT MATTER

G06T 5/00 (2006.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 EPODOC, WPI, ESPACE, Google Patents. Search terms: image, fuse, exposure, intensity, chrominance, luminance, and other similar terms.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

☒ Further documents are listed in the continuation of Box C☒ See patent family annex

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
14 August 2012

Date of mailing of the international search report
15 August 2012

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
 PO BOX 200, WODEN ACT 2606, AUSTRALIA
 Email address: pct@ipaustalia.gov.au
 Facsimile No.: +61 2 6283 7999

Authorized officer

Kim Ung
 AUSTRALIAN PATENT OFFICE
 (ISO 9001 Quality Certified Service)
 Telephone No. 0399359621

INTERNATIONAL SEARCH REPORT		International application No.
C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		PCT/SG2012/000204
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2009/0268963 A1 (KANG et al.) 29 October 2009 Abstract; paragraphs [0012]-[0027]; Figures 1-8.	
A	US 2010/0103194 A1 (CHEN et al.) 29 April 2010 Abstract; paragraphs [0009]-[0029]; Figures 1-11.	
A	US 2008/0316354 A1 (NILEHN et al.) 25 December 2008 Abstract; paragraphs [0005]-[0040]; Figures 1-3.	
A	US 2010/0183071 A1 (SEGALL et al.) 22 July 2010 Abstract; paragraphs [0001]-[0025]; Figures 1-12.	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/SG2012/000204

This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	Publication Date	Publication Number	Publication Date
US 2009/0268963 A1	29 Oct 2009	KR 20090112131 A	28 Oct 2009
		US 2009268963 A1	29 Oct 2009
		US 8135235 B2	13 Mar 2012
US 2010/0103194 A1	29 Apr 2010	CN 101394487 A	25 Mar 2009
		CN 101394487 B	14 Sep 2011
		GB 2464574 A	28 Apr 2010
		GB 2464574 B	29 Dec 2010
		US 2010103194 A1	29 Apr 2010
US 2008/0316354 A1	25 Dec 2008	BR PI0607206 A2	03 Nov 2010
		CN 101116324 A	30 Jan 2008
		CN 101116324 B	09 Mar 2011
		EP 1689164 A1	09 Aug 2006
		EP 1689164 B1	19 Dec 2007
		JP 2008529435 A	31 Jul 2008
		MX 2007009328 A	21 Sep 2007
		RU 2007132891 A	10 Mar 2009
		US 2008316354 A1	25 Dec 2008
		US 7916181 B2	29 Mar 2011
		WO 2006082186 A1	10 Aug 2006
US 2010/0183071 A1	22 Jul 2010	CN 102282838 A	14 Dec 2011
		EP 2380344 A1	26 Oct 2011
		US 2010183071 A1	22 Jul 2010
		US 2012050474 A1	01 Mar 2012
		WO 2010082692 A1	22 Jul 2010

End of Annex