

AUTOMATED AND COMPUTATIONALLY INEXPENSIVE EXPOSURE FUSION FOR MOBILE DEVICES

Tomislav Kartalov, Zoran Ivanovski

*Faculty of Electrical Engineering and Information Technologies,
"Ss. Cyril and Methodius" University in Skopje,
Rugjer Bošković bb, P.O. box 574, 1001 Skopje, Republic of Macedonia
kartalov@feit.ukim.edu.mk*

Abstract: This paper presents a fully automated, computationally inexpensive and high quality exposure fusion algorithm, intended for use on mobile or handheld devices. A utilization of the device's view-finder screen video feed data is proposed, in order to increase the overall performance of the exposure fusion, both in static scenes and in scenes with moving objects. Several novel ideas are implemented in order to make the whole procedure fully automated, working without need for any intervention, or parameter adjustment by the end-user. The performed experimental tests show an efficient performance and high quality results, both in visual and objective terms.

Key words: exposure fusion; computational efficiency; mobile platform; image decomposition; motion estimation

АВТОМАТИЗИРАНА И ПРЕСМЕТКОВНО ЕФИКАСНА ЕКСПОЗИЦИСКА ФУЗИЈА ЗА МОБИЛНИ УРЕДИ

Апстракт: Во овој труд е претставен автоматизиран и пресметковно ефикасен алгоритам за експозициска фузија со висок квалитет на излезните слики, наменет за употреба во мобилни и рачни уреди со ограничена пресметковна моќ и ограничен мемориски простор. Предложена е употреба на видео-поток наменет за корисничкиот екран на уредите, со цел да се подобри квалитетот на експозициската фузија. Алгоритмот работи за слики со статички сцени, но и за слики со подвижни објекти. Повеќе нови идеи се имплементирани со цел да се автоматизира целата постапка и крајниот корисник да може да работи без никаква интервенција или нагудување на параметри. Извршените експериментални тестови укажуваат на тоа дека предложениот алгоритам работи ефикасно и генерира резултати со висок субјективен и објективен квалитет.

Клучни зборови: експозициска фузија; пресметковна ефикасност; мобилна платформа, декомпозиција на слика, естимација на движење

1. INTRODUCTION

However advanced, all the modern imaging devices have limited physical capabilities that can often be surpassed by the photometric quantities of the recorded real world scenes. In the case of a scene with high radiance dynamic range (HDR), the imaging device can be set to capture only fragment of the scene area with an optimal exposure (low dynamic range – LDR), and the rest of the scene objects would be either too bright or too dark in the final digital image.

To solve this problem, instead of one, two or several LDR input images are recorded from the scene, each with a different exposure setting, capturing different intervals of the radiance range. Afterwards, the information present in the input images is combined, to obtain an output image with higher visual quality, where all the objects are properly exposed and easily distinguishable.

Limited only to the professionals possessing knowledge and expensive equipment a few decades back, nowadays these image combining procedures are available almost to everyone holding a camera, following the advancements in both the

algorithms and the hardware. Various software tools for image processing support them, and the scientific community constantly provides new and better solutions [1–19].

The two adopted concepts are the true HDR processing, and the exposure fusion (EF) processing. The true HDR processing [1], [2], [8], [38], yields HDR output image with luminance ranging in several orders of magnitude that is not suitable for reproduction on the standard electronic view screens. Hence, the HDR image has to be range-compressed, or tone-mapped to a standard, 8-bit, LDR image [3], while preserving local contrast as much as possible. The earliest HDR works were based on reconstruction of the real radiance in the scene, using no a priori data for the input images [1] or using the information about their exposure times to recover the camera response function [2]. Early works focused mainly on the image combining methods and assumed perfectly aligned input images. The modern state of the art algorithms shift their center of attention towards the real world usage, in scenes with moving objects, or non-static cameras. In [8], the HDR image synthesis equation is introduced for patch-based energy minimization, which allows non-aligned input images and moving objects in the scene.

The EF processing [4]–[7], [9], [10] does not involve an intermediate HDR stage, and fuses the most informative areas from the input LDR images directly to a single LDR output image, not considering the real radiance interdependences (lighter, darker) of the objects in the scene. The modern EF algorithms use large number of concepts to perform the actual fusing, few of the notable being the multi-scale decomposition approach [4], [6], [10], the maximum entropy approach [5], the global energy optimization approach [7], and the luminance segmentation approach [9].

Similar to the HDR processing, the EF algorithms recently shifted towards real world usage scenarios, where the static camera assumption is not always true, and the recorded scene can contain moving objects, leading to ghosting artifacts in the fused image. Numerous contributions were published in the last decade [11]–[19], which are focused solely on the problems of image misalignment and ghosting artifacts. The solutions are usually implemented as a preprocessing step, prior to the HDR or EF process.

Some of the algorithms consider only the spatial alignment of the images recorded by a non-static camera from a static scene [11], [12]. Others,

assume static camera and correct the object motion in the scene [18], [19], and some of them engage both problems simultaneously [13] – [17].

Whereas all of the mentioned algorithms work well under their design scenarios, none of them is specifically designed for usage on a mobile device, where certain demands are in place regarding processing power, memory consumption, and procedure automation. Most of the algorithms are too complex or iteratively based, some utilize computationally expensive filtering [5], [10], lack non-supervised operation [6], [10], or assume static scenes and perfectly aligned input images [4]–[7], [9], condition that is almost never met if the algorithm is implemented on a handheld mobile device. Also, the total processing time must be kept relatively short otherwise the end-users would rarely opt for this procedure. Although in some of the algorithms fairly short computing times are published for their optimized codes, those results are always obtained on a desktop PCs, and often low-resolution input images are used for testing.

Several main principles should be followed, in order to design and build an EF algorithm that can be implemented on a mobile device:

- 1) The whole procedure should be performed completely automatically, without any end-user intervention. There should be no externally adjustable parameters. A "single click" approach should be adopted, where the mobile device captures, stores, and processes all the necessary input data for the exposure fusion, and delivers single output image.
- 2) The procedure should be implemented in an efficient and computationally inexpensive way, such that it should be easily executed on any modern mobile hardware (e.g. ARM based platform, or similar).
- 3) The handheld usage cancels the assumption of the input images perfect spatial alignment. Inclusion of a global motion compensation preprocessing step is mandatory. This step should not increase in a significant way the total computational load of the EF procedure.
- 4) In most of the usage scenarios, the static scene assumption will not be valid, and object motion will be present. Implementing full frame motion estimation and compensation is not feasible, due to its high computational load. Instead, the fusion algorithm should be made robust to such object motion, implementing intelligent and adaptive selec-

tion criteria, carefully choosing the parts from the input images that will be included in the output image.

- 5) The procedure should be universal, and independent on the capturing device properties. Any image statistics or parameters needed for the fusion, should be calculated directly from the captured images, and should be assumed a priori unknown.

In this paper a complete exposure fusion procedure is proposed. It is designed specifically for usage on a mobile platform devices. The proposed algorithm works in the Luminance/Chrominance color space, and its core is based on the multi-scale image decomposition into the Laplacian pyramid [26]. Considering the above stated principles, several novel solutions are implemented, among them, accelerated computing employing integer operations, lowpass filtering using only additions and binary shifts, forward pyramid implementation, automatic calculation of the optimal pyramid height, adaptive selection criteria, viewfinder data-stream utilization, incomplete pyramid decomposition, reduced motion vector set for input image alignment, and overall decision whether the fusion process is necessary for the recorded scene.

The experiments show that both visually and in terms of objective fusion quality measures, the proposed EF procedure obtains comparable, and often superior results, evaluated against the state-of-the-art fusion methods.

Computationally wise, the proposed EF procedure works in a fast and efficient way. A Java/C implementation of the algorithm core tested on an Android mobile devices, requires as little as 60 ms to finish the fusion of a VGA sized image.

The rest of the paper is organized as follows. The proposed EF procedure is explained in detail in Sections 3 – 5. The experimental validation of the work is given in Section 6, and Section 7 concludes the paper.

2. PRIOR CONNECTED WORK

In the recent few years, several contributions, [20]–[25], gradually led to the development of all the parts of the proposed algorithm. Firstly, the procedure was intended to work with two input images, the overexposed (OV), obtained with positive exposure bias and the underexposed (UN) obtained using an equivalent (in absolute terms) negative exposure bias. The optimal exposure bias

values that would result in capturing the highest radiance range from the recorded scene, was deduced in [20]. Later, an automated EF algorithm for mobile platforms was proposed in [21], based on Laplacian pyramid decomposition, and using the results published in [22], to determine the optimal number of decomposition levels. The algorithm [21] performed fairly well in a static scene and static camera scenarios however its performance rapidly degraded under dynamic conditions. A real time global motion compensation algorithm was developed in [23], to address the problem of non-static camera, performing spatial alignment of the input images. This algorithm was limited only to translational motion correction, so it was later replaced by the procedure proposed in [24], based on calculating only a reduced set of motion vectors in the input images, and determining the global motion in the scene. This procedure could correct both translational and rotational movements of the camera/scene. The error robustness of the procedure was further increased in [25], using a cross-validation of the resulting motion vectors to detect and isolate the incorrectly calculated ones.

In this paper, the contributions from [20]–[25], are joined in one high quality fully automatic EF procedure, along with one concept change and several new possibilities coming out of it. The proposed concept change is the idea to utilize the mobile device's viewfinder video stream, providing additional information to the existing two input images (UN and OV), in the form of low-resolution frame VF (viewfinder), captured without exposure bias. Although this is not equivalent as working with three full resolution input images, the information obtained from the VF image is enough to ensure highly adaptive selection criteria, leading to the robustness of the whole algorithm to the moving objects in the scene.

3. THE ALGORITHM OVERVIEW

The proposed procedure is designed to work using a Luminance/Chrominance based color space, which is the most common color representation method for the mobile devices with image capturing ability. The overview of the whole procedure is shown in Figure 1.

Normally, the end-user begins with scene pre-viewing on the viewfinder screen. At this time, the device measures the light in the scene and calculates the optimal exposure time, according to the selected metering mode. For the proposed EF pro-

cedure, the matrix or evaluative metering mode is preferred, since the spot metering mode could ob-

tain incorrect information about the overall light radiance in the scene.

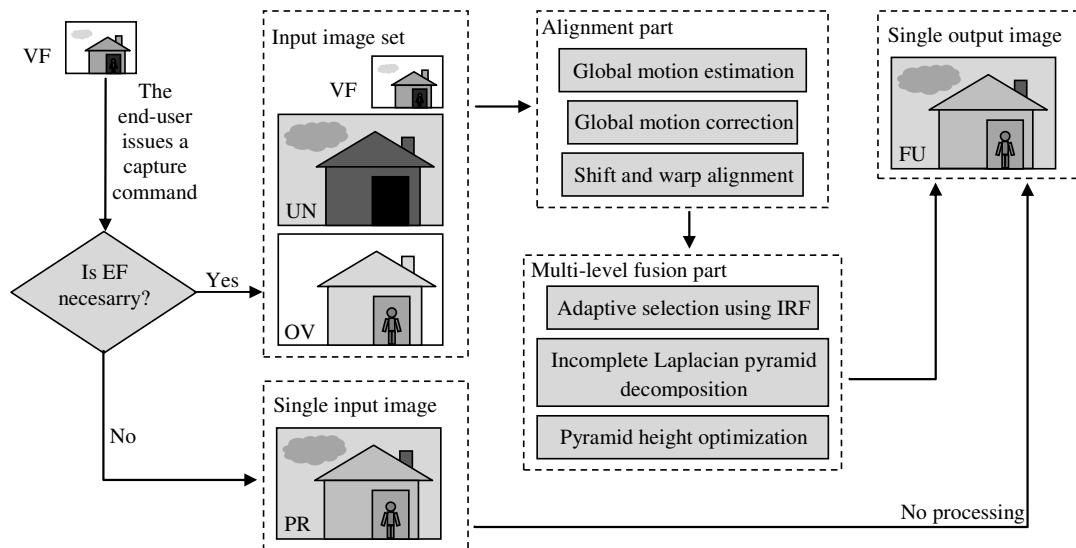


Fig. 1. The workflow of the fully automatic EF procedure. It starts with a single capture command issued by the end-user, results in a single output image stored on the mobile device. No user intervention is needed in the process, and there are no externally adjustable parameters.

When the scene is locked and the capture command is issued, the EF procedure first decides whether the scene contains high dynamic radiance range. If the dynamic range of the scene is normal or low, the device performs as normal digital camera, capturing a single, properly exposed input image (PR), using the calculated optimal exposure time. The PR image is without further processing stored in the device.

If the scene contains high dynamic radiance range, then instead of one, three input images are recorded. The first is a snapshot from the viewfinder screen in the moment of the capture command. It represents a low resolution, properly exposed image, VF. The second and third images are two exposure bracketed, full resolution images, UN and OV.

The captured input image set (VF, UN and OV) is subjected to the alignment part of the procedure (Section 4), where the VF and OV images are spatially aligned to match the background of the UN image. The aligned input image set is then pipelined to the multi-level fusion part of the procedure (Section 5), where the actual EF takes place. In the end, a single resulting output image is stored on the device, the fused image FU. The rest of this section addresses some of the global solutions implemented in the proposed procedure.

A) The necessity of the EF in the scene

A significant number of recorded scenes would not require the execution of the EF procedure, due to their low radiance dynamic range. Instead, a regularly captured single image, using the calculated optimal exposure time, would be able to represent correctly both the brightest and the darkest objects in the scene.

The decision part depicted in Figure 1 is based on the data present in the bit-stream used by the viewfinder screen of the device. Alternatively, the captured VF image could be used, or mostly recommendable, the live histogram option, if the device has such. The proposed procedure divides all the pixels from the preview image in two bins, according to their luminance levels. The first bin contains the so called non-saturated pixels, with luminance levels between two empirically obtained thresholds, valued at 20% and 80% of the maximum possible luminance level. The second bin contains the rest of the pixels, representing the saturated and the near saturated pixels, accounting both black and white saturation.

The experiments performed on a real world images show that in the scenes where the number of the non-saturated pixels is higher than the number of saturated and near saturated pixels in the

preview image, the EF procedure should be skipped.

Considering the small size of the preview image, and the simplicity of the decision process (luminance thresholding and pixel counting), this process can be performed in negligible time, or even preemptively, as a real time background process active whenever the camera is in use.

B) The optimal exposure bias

The exposure biases used in the bracketing when capturing the UN and OV images should be set high enough, so that maximum possible radiance range from the scene is covered. However, too high exposure biases can lead to a middle gap in the covered range, and incorrect representation of the medium bright objects. In [20], the mathematical analysis is performed to obtain the optimal value for the exposure bias.

The non-linear equation that maps the real luminance levels from the scene, Y , to the perceived lightness by a human observer, L^* , is given by the CIE (Commission Internationale de l'Eclairage) standard [28], in (1). The parameter Y_n is one of the three tristimuli (X_n , Y_n , Z_n) of a specific white color (referent white).

$$L^* = 116 \cdot f\left(\frac{Y}{Y_n}\right) - 16 \quad (1)$$

where

$$\begin{aligned} f\left(\frac{Y}{Y_n}\right) &= \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} & \text{for } \left(\frac{Y}{Y_n}\right) > \left(\frac{6}{29}\right)^3 \\ f\left(\frac{Y}{Y_n}\right) &= \left(\frac{841}{108}\right) \cdot \left(\frac{Y}{Y_n}\right) + \frac{4}{29} & \text{for } \left(\frac{Y}{Y_n}\right) \leq \left(\frac{6}{29}\right)^3 \end{aligned} \quad (2)$$

The unity value of luminance can be declared for the referent white ($Y_n = 1$). The lightness L^* can be linearly mapped to the digital luminance values (DLL : 0–255) in the 8-bit coded photography. Implementing these simplifications, (1) can be rewritten as

$$DLL = 2.55 \cdot \left(116 \cdot \sqrt[3]{Y} - 16\right) \quad \text{for } Y > 8.9 \cdot 10^{-3} \quad (3)$$

and its inverse

$$Y = \left(\frac{DLL + 40.8}{295.8}\right)^3 \quad \text{for } DLL > 19. \quad (4)$$

These equations give simplified transfer function for digital camera operation, accurate for relative luminance values above 0.89%, or digital lu-

minance levels above 19. Using the matrix metering mode, the automatic optimal exposure calculation should set the middle gray digital level ($DLL = 127$) to match the average luminance in the scene, leading to near symmetrical distribution of the white and black saturated areas in the properly exposed (PR) image. This allows exposure bracketing using symmetric exposure biases in the OV and UN images.

The optimal exposure biases of the exposure bracketed images OV and UN should be set in a manner that, the OV image would capture the border black saturated pixels ($DLL = 0$ in the PR image) as middle gray level, and the UN image capturing the border white saturated pixels ($DLL = 255$ in the PR image) as middle gray level. Because (4) is valid only for DLL above 19, the optimal exposure bias can be obtained for the underexposed image, and the bias for the overexposed image can be set symmetrically.

The increase or decrease of the exposure bias for one point corresponds to doubling or halving the relative luminance [28]. To capture the white saturated pixels in PR image as a middle gray level pixels in the UN image, the camera exposure bias EB should be decreased by ΔEB :

$$\Delta EB = \log_2 \left(\frac{Y_{WH.SAT}}{Y_{MID.GRAY}} \right) = \log_2 \left(\frac{255 + 40.8}{127 + 40.8} \right)^3 = 2.454 \text{ points.} \quad (5)$$

The closest standard values implemented in the digital cameras are 2.33 points or 2.66 points. However, in the low-end digital cameras built into the pocket mobile devices, the values of the possible exposure bias rarely exceed ± 2.00 points. Taking this into account, the ± 2.00 points exposure biases are adopted in the development of the proposed EF procedure, and all the experiments are performed using these values.

4. THE IMAGE ALIGNMENT PART

For the spatial alignment of the input images, the registration algorithm proposed in [24] is used, which estimates five representative motion vectors, named joint motion vectors, between each pair of input images, and performs the actual alignment using fast operations like shifts and slice-based warps. The motion estimation is performed on 8×8 pixel blocks, using the hierarchical technique [31]. The key points of the algorithm are presented here, however, for full comprehension of the process,

the reader is encouraged to study the work presented in [24].

A) Motion estimation between differently exposed images

All images in the input image set are captured with different exposure settings, rendering them ill-suited for the standard motion estimation techniques. The two major problems to overcome are the loss of contrast and details in the white/black saturated regions in the input images, and the difference in their overall brightness.

The solution of the first problem is to avoid the saturated regions. A simple activity threshold is implemented for the decision whether the 8×8 pixel block is suitable for motion estimation or should be avoided. The activity for block with elements x_i ($i = 1, 2, \dots, 64$) is calculated as in (6):

$$ACT = \frac{\sum_{i=1}^{63} |x_i - x_{i+1}|}{64}. \quad (6)$$

Calculation of this activity parameter, in addition to the saturated regions, also helps to avoid the flat regions of the image where the motion estimation is unreliable.

An example, showing typical values of the activity parameter, is shown in Fig. 2. Using a number of examples, the actual activity threshold is empirically set to 5. No block with lower activity is used for motion estimation. The reduced motion vector set scheme explained below is capable to compensate for missing motion vectors in the image.

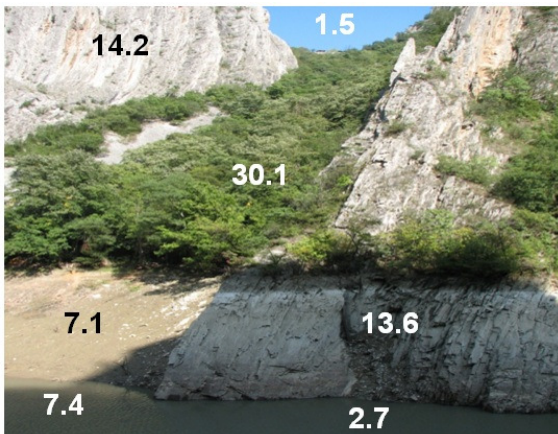


Fig. 2. Typical values of the activity parameter in common image areas.

The second problem, the different brightness and contrast in the input images is solved by adjusting the core of the motion estimation routine. Instead of searching for the most similar block in the terms of luminance, a different comparison tool is implemented, the content match, CM , defined in (7) and (8). The statistical background of the CM parameter is in some ways similar to the $SSIM$ [32], however, CM is more rudimentary, and less computationally demanding.

$$CM = \sum_{i=1}^{64} |C_A(i) - C_B(i)| \quad (7)$$

The parameters C_A and C_B are defined as content measures for the blocks **A** and **B**. The content measure for the block **A** with elements x_i ($i = 1, 2, \dots, 64$) is defined as in (8):

$$C_A(i) = \frac{(x_i - A_M) \cdot 64}{\sum_{p=1}^{64} |x_p - A_M|}, \quad (8)$$

where A_M is the mean value of the elements in **A**.

The content match parameter CM , calculated for two blocks, has lower values if the blocks have more similar content. This parameter is insensitive to changes in brightness or contrast, making it perfect for motion estimation in images captured with different exposure settings.

B) Reduced motion vector set

Five square sections are defined in each image, one in the image center and four in the corners, leaving margin to the image edges equal to 10% of the image diagonal (in pixels). The sizes of the sections should be set in such manner, that the displacements in all directions of up to 1% of image diagonal could be estimated. This represents the limit of the image alignment part, set by the number of experiments. The images displaced by a higher value often come with motion blur and are too degraded to be usable for EF. The actual size of the sections is recommended to be a power-of-two, in order to facilitate fast implementations of the down sampling and hierarchical motion estimation. For each section, one motion vector is calculated, called joint motion vector, obtained by averaging the best 4 representatives of the estimated 10 motion vectors at random locations (and complying to the activity rule) in the section. The best representatives among estimated vectors are decided by their (lowest) CM scores.

The prospect of using only 10 randomly chosen motion vectors in each section is derived from a statistical experiment performed on 338 freehand captured image pairs, and shown in Fig. 3. All the image pairs were manually aligned, and the ground truth motion vectors were extracted. The comparison between the ground truth motion vectors and the calculated joint vectors in terms of the Spearman's rank correlation coefficient, showed that a heavily reduced and randomly chosen motion vector set can be used in calculation of the joint vectors, without losing the accuracy. The result in Figure 3 states that introducing more than 10 vectors per section does not increase the accuracy further.

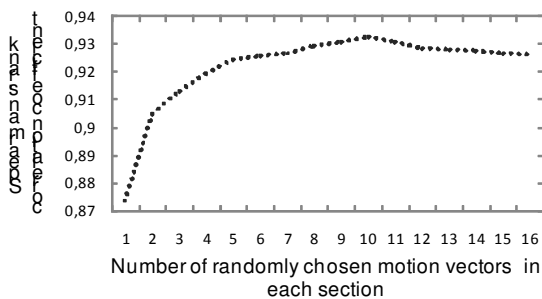


Fig. 3. Correlation between the ground truth motion vectors and joint motion vectors calculated from 338 image pairs captured with free hand. In the cases where the number of randomly chosen motion vectors is lower than 4, the whole set is used in calculation of the joint vectors. In the other cases, the 4 lowest CM scoring representatives are used

C) Validation of the motion estimation results

Although a high correlation is calculated between the ground truth motion vectors and the estimated joint vectors, the actual error check shows that total of 150 out of 1690 joint vectors were incorrectly estimated, contained in 69 out of 338 image pairs. 21 of those 69 image pairs have displacements that are out of the reach of the proposed algorithm, too complex or higher than 1% of the image diagonal. That leaves 48 image pairs where some of the joint vectors are erroneously calculated. Thus, a post-processing procedure is implemented to test and validate the results of the motion estimation.

The obtained five joint vectors per image pair are subjected to k -fold cross-validation of their values, testing each of them for consistency with the globally calculated motion, and correcting up to 2 out of 5 inconsistent joint vectors. Using a modeling of the global 3D motion as proposed in

[30], an analysis performed in [25] shows that only two joint vectors are enough to calculate the global motion of the whole scene, allowing to the test consistency of the remaining joint vectors against the calculated global motion in their respective locations. The testing is performed using the "leave-one-out" strategy, where each joint vector is tested against the global motion calculated and averaged from all possible 2 out of 4 combinations from the other 4 joint vectors. The correctness of the tested joint vector is decided by comparing the estimated error to the empirically obtained threshold value. To remove the false positives, a second iteration of testing is performed, in which all the joint vectors are tested again, however, this time the global motion is calculated using only the joint vectors that were cleared as correct in the first iteration.

Considering the low amount of processed data, the validation procedure is computationally irrelevant. Performance wise, as shown in [25], it offers an error detection rate of 92%, and error correction rate of 68%. Naturally, these rates would be heavily dependent on the statistical properties of the input images, and the camera operating behavior of the end-user.

D) Fast shift and warp alignment

After obtaining the five joint vectors, a simple and fast image alignment procedure takes place. All the operations are performed only on one of the images in the processed pair (usually the OV image is aligned respective to the UN image). The VF image, being much smaller in size than OV and UN, is not aligned because the global motion is unnoticeable when the input images are downsampled to the size of the VF image.

At first, the global displacement vector is calculated as the mean value from the Cartesian coordinates of the five joint vectors. This vector is used for the translational shift of the image. Then, the global vector coordinates are subtracted from the coordinates of the four corner joint vectors, obtaining vector differentials. At the end, the slice-based warping is applied to the image, to match the motion expressed by the vector differentials. For the warping, the image is divided in horizontal and vertical slices which movements are extrapolated using the vector differentials and image geometry. The slices are shifted only by integer number of positions, and following the rule that the shift of each slice cannot be higher than 1 pixel relative to its neighboring slices.

E) Computational complexity

The Table 1 contains the information about the computational complexity of the alignment part, without the computationally irrelevant parts, the cross-validation and shift/warp procedures. The whole alignment procedure works in under one operation per output pixel, which is negligible compared to the rest of the EF procedure. Such efficiency makes the alignment part a perfect pre-processing tool for global motion compensation of the images captured with hand held imaging device. The analysis in the Table 1 is performed for a 5 megapixel image pair. However, input images with higher resolutions would actually lower the per-pixel computational load, as the total number of pixels increases faster than the total number of operations.

The actual down sampling method used in the hierarchical motion estimation is explained later in the text, because the same method is used for the pyramid decomposition in the EF procedure.

Table 1

*The alignment part: computational complexity
(Number of operations)*

Op. type	Add.	Shift	Sign	Mul.	Σ
Per comparison ¹	507	4	192	2	705
Per vector ²	22k	180	8k6	90	32k
Per section ³	228k	1800	86k	900	317k
Only motion estimation ⁴	1140k	9000	432k	4500	1586k
Down sampling of the sections	762k	0	109k	0	870k
Total per one output image	1902k	9000	541k	4500	2456k
Normalized ⁵ per output pixel	0.38	≈ 0	0.11	≈ 0	0.49

¹⁾ Calculating the CM factor between two 8×8 blocks

²⁾ 45 comparisons = 5 resolution levels · 9 comparisons per level

³⁾ Reduced set with 10 motion vectors per section

⁴⁾ Five sections in image (without down sampling of the sections)

⁵⁾ Five MP image = 2592×1944 pixels

5. THE MULTI-LEVEL FUSION

The overview of the main part of the EF procedure, the Laplacian-pyramid based multi level fusion, is shown in Fig. 4. General luminance/chrominance color space is used, the actual one (e.g. YUV, YIQ, YPbPr) being irrelevant, as

the chrominance components are simply copied from the input images to the output, according to the weight maps obtained using the luminance. The multi level fusion is performed only on the luminance.

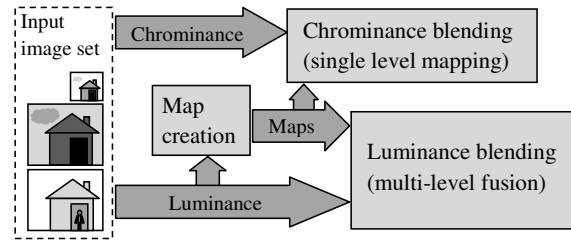


Fig. 4. The main part of the EF algorithm. All processing is performed on the luminance, whereas the chrominance is simply copied from the input images.

A) Adaptive selection and weight map creation

This module locates the optimally exposed parts from the input images, and creates weight maps where the selection information is stored. The selection criteria are threshold based, namely, all the pixels from the OV image with luminance lower than threshold $T_{H,OV}$, are selected as a content source for the fusion process. The rest of the pixels in the OV image are declared as white-saturated, or near white-saturated. In the saturated areas, the matching pixels from the UN image are tested against threshold $T_{H,UN}$, and all having luminance higher than threshold, are also selected as a content source for the fusion process. In some cases, the same locations can be white-saturated in the OV image, and black-saturated (underexposed) in the UN image. Such double saturation (DS), usually occurs when object movement is present in the scene, because the saturated moving object occupying certain scene location in one of the images, can reveal the saturated background at the same location in the other image. Because neither of the full resolution OV and UN input images does not contain usable data in the DS regions the pixels from the low resolution VF image are selected as content source for those regions. Considering that the DS regions are usually relatively narrow regions present at the revealing edge of the moving object, the loss of resolution is often visually indistinguishable, and the benefit of the chrominance copying is noticeable, coloring the otherwise white or black DS regions.

Concerning the occluding edge of a moving object, where a double non-saturated region would occur, the pixels from the OV image are given pri-

ority in the selection. This decision is made based on the most common real world usage of the EF, where the darker object (properly exposed in the OV image) is in front of the brighter background (properly exposed in the UN image), e.g. photographing people against bright windows, door-frames, or located in shadows. In the vice-versa scenario (bright object against dark background), usually the object employs some sort of added lightning (light emanating from the object itself, spotlight, or camera flash), making the difference in the lightning of the object and the background so high, that it is highly probable at least one of them would be (black or white) saturated in each of the input images.

The thresholds $T_{H,OV}$ and $T_{H,UN}$ are adaptively calculated using the reconstruction of the camera response function (CRF), calculated as proposed in [2], by solving a system of linear equations, with a set of sample pixel values from matching locations in input images with known exposure times. To obtain the samples, a grid layered over the input images is used, leaving border margins of 1/8 of the image dimensions.

This yields samples evenly distributed across the images, representing the majority of the image content. Concerning the total number of samples, the higher number should represent the image statistics better, however, the experiments show that any scene would be well represented by a total number of 529 (23×23 grid) locations for pixels samples. The opportunity for calculation of the CRF is another benefit from introducing the VF image into the input image set. As can be seen in Fig. 5, the obtained CRFs are much more consistent if the pixels from VF image are added in the mix.

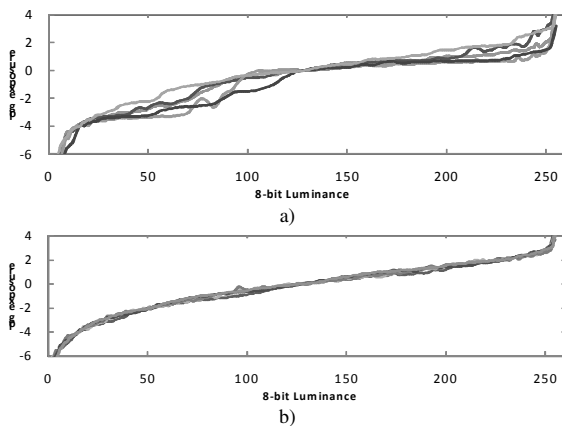


Fig. 5. Calculated camera response functions for 5 image pairs or triplets captured with the same camera settings, (a) using only OV and UN images, (b) using OV, UN, and VF images.

After obtaining the CRF for given input image set, the threshold values are calculated as the maxima of the absolute second derivative of CRF, as in (9):

$$T_H = DLL \left| \max \left(\frac{d^2 CRF}{dDLL^2} \right) \right| \quad (9)$$

$T_{H,OV}$ must be in the interval $200 < DLL \leq 255$

$T_{H,UN}$ must be in the interval $0 \leq DLL < 50$

where DLL is the 8-bit digital luminance level 0–255.

B) Forward (in place) pyramid implementation

In the proposed EF procedure, a multi-level fusion process is adopted, decomposing the input image set into a Laplacian pyramid [26]. If the decomposition is performed in N pyramid levels, an input image \mathbf{G}_0 is decomposed into the members of the Gaussian part of the pyramid ($\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_N$), representing the down-scaled versions of the input image, and into the members of the Laplacian part of the pyramid ($\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_N$), representing the spatial activity and image details across the different scales. The weight maps, generated with the adaptive selection criteria, are decomposed only into the Gaussian parts of the pyramids. At each pyramid level, the Laplacian parts of the input pyramids are fused into one pyramid, according to weight maps. The reconstruction of the obtained single pyramid generates the output image. The reconstruction of the Gaussian member at level $L-1$ can be performed using the members from the upper level L according to (10).

$$\mathbf{G}_{L-1} = \uparrow(\mathbf{G}_L) + \mathbf{D}_L \quad (10)$$

where the operator \uparrow represents the two step processing of the \mathbf{G}_L , including up-scaling and low pass filtering. Inductively replacing the equations from upper pyramid levels (10) can be rewritten in such way, which states that for the reconstruction of the original image \mathbf{G}_0 only the last member of the Gaussian pyramid part, and all the members of the Laplacian pyramid part are needed. In general case with N levels of decomposition, the reconstruction equation obtains the following form:

$$\mathbf{G}_0 = \left(\sum_{i=1}^N \uparrow^{i-1}(\mathbf{D}_i) \right) + \uparrow^N(\mathbf{G}_N) \quad (11)$$

where the operator \uparrow^X designates X times consecutive application of the operator \uparrow .

The expression in (11) states that the output image can be built in forward manner, incorporating into it each level of the fused pyramids, as it is calculated, discarding the necessity for keeping the results from all the decomposition levels for the reconstruction, which saves up a considerable memory space.

C) Fast image scaling for the pyramid decomposition

The qualitative and quantitative properties of the image decomposition into a Laplacian pyramid are highly dependent on the implemented low pass filtering. Also, the low pass filtering is the main bottleneck regarding the computing efficiency of the whole procedure. In the proposed work, two different low pass filters are implemented in the decomposition process, separate for the down-scaling and the up-scaling operations. Both filters are designed to work using only additions and binary shifts on integer numbers. This design significantly reduces the computational load of the processing, at the cost of a small amount of rounding error, visually unnoticeable in the output image.

The first filter, integrated within the down-scaling operation, performs averaging of all the 8-connected neighbors of the processed element. For example, the element $g_L(m/2, n/2)$ in the decimated image \mathbf{G}_L on the L -th level of the Gaussian part of the pyramid, is calculated using 8 elements from the lower level image \mathbf{G}_{L-1} , as shown in (12).

$$g_L\left(\frac{m}{2}, \frac{n}{2}\right) = \frac{\sum_{i=m-1}^{i=m+1} \sum_{j=n-1}^{j=n+1} g_{L-1}(i, j)}{8} \quad (12)$$

The element $g_{L-1}(m, n)$ itself is excluded from the calculation, in order to obtain division by 8 instead of division by 9, increasing the speed of the algorithm. The explained down-scaling is applied on every position with even numbered coordinates m and n in the image \mathbf{G}_{L-1} .

The second filter is included into the up-scaling operation, used for the creation of the image $\uparrow(\mathbf{G}_L)$, which is low pass filtered representation of the image \mathbf{G}_{L-1} . The image \mathbf{G}_L is up-scaled keeping all the original columns, and introducing new ones between each two of them, valued as average of their closest two original neighbors. Then the process is repeated for the image rows.

D) Incomplete pyramid and static scenes edge enhancement

Considering the lower resolution VF image, compared to UN and OV, an incomplete pyramid blending is employed. The first few levels of the pyramid fusion are performed only on OV and UN images, without the data from the VF image. Exception to this is the chrominance copying in the base decomposition level, where up-scaled chrominance from the VF image is used. The low sensitivity of the human vision to the chrominance sharpness, allows the most rudimentary up-scaling methods to be used, without visible loss of detail.

When the pyramid blending reaches the level where the decomposed input images are down-scaled to resolutions near to the resolution of the VF image, then this third image is introduced in the further process. Regarding the weight maps, since they are available in full resolution, they can be present in the decomposition at all the pyramid levels. If the memory load is more prioritized than the computational load, optionally instead of three, only two weight maps could be stored at the same time, since all three must be normalized to unity at each point, and at each level of decomposition.

In the static scenes, a noticeable edge enhancement could be achieved simply changing the fusion criterion in the base pyramid level. Instead of map blending, the Laplacian member with higher absolute magnitude is selected for participation in the fused Laplacian member. This leads to increased sharpness of the final image. This scheme is performed under the condition that the number of DS pixels does not surpass 0.1% of the total number of image pixels. Such low number of DS pixels indicates a static scene and perfect spatial alignment of the input images. If the number of DS pixels is higher than this threshold, the procedure falls back to the standard map blending of the Laplacian members in the base pyramid level.

E) The optimal number of pyramid decomposition levels

Although this parameter is often neglected in the EF algorithms based on multi-level fusion, the height of the pyramid significantly affects the visual quality of the fused results. A decomposition in too few levels cannot build the interdependency between separate image parts, leading to visible luminance transitions among the image regions, known as halo artifacts. Contrary a decomposition

in too many levels causes a global averaging of image features at a higher pyramid levels. The effect could be an over-compensation of the luminance dynamic range, decreasing the contrast of the final output image and introducing de-saturation of the colors.

The optimal pyramid height for artifact free output image is different for various input images, and depends on the image contents and statistics. Hence, designing an autonomous algorithm for decision on the optimal number of decomposition levels is a challenging task. An empirically derived such algorithm is proposed in [22].

In this work, the actual number of decomposition levels is not used as a representative of the pyramid height, since it varies with the initial size of the image. A more convenient form of representation is the resolution (number of rows: R_G) of the last (highest) member in the Gaussian part of the pyramid. The experiments performed in [22] showed that for the best visual quality, the decomposition should be executed at least up to the resolution of $R_G = 8$.

When that level is reached, two parameters are considered as stopping criteria for the further decomposition. The first parameter is the image activity. High enough activity could mask the halo effect, and the pyramid decomposition could be stopped at lower level without concern about visible halo artifacts. Taking into the account that the luminance transitions of the image down-scaled at the level L , are well represented by the Laplacian member \mathbf{D}_L , the image activity at that level, ACT_L , is defined as the absolute average of the \mathbf{D}_L .

The second parameter to be considered is the smoothness of the level L Gaussian member of the weight map decomposition, \mathbf{GM}_L . Smooth and uniform \mathbf{GM}_L leads to uniform participation shares of the input images in the fused images at the observed decomposition level. Non-smooth \mathbf{GM}_L with high difference elements leads to a fused image at that decomposition level, in which some parts are built mainly or exclusively from the features of only one of the input images, causing relatively large luminance transitions among the regions in the output image, or halo artifacts. In such case, the pyramid decomposition should continue to the next higher level, until a relatively smooth \mathbf{GM}_L is obtained.

The smoothness of \mathbf{GM}_L is calculated testing its extreme values (minimum and maximum) against the adaptive interval $[I_L, I_H]$. If both extremes have values within the tested interval, then

\mathbf{GM}_L is **smooth-order 2**. If only one of the extremes is within the interval, then \mathbf{GM}_L is **smooth-order 1**. Otherwise, \mathbf{GM}_L is **non-smooth**. The endpoints of the interval for pyramid level L depend on the image activity, as in (13):

$$[I_L, I_H]_{\text{Level}=L} = [M_V - (ACT_L + T_{H1}), ACT_L + T_{H1}], \quad (13)$$

where T_{H1} is empirical threshold and M_V is the maximum possible value of the elements in the weight map.

Two criteria are tested for stopping the pyramid decomposition: too high activity, or too smooth map. If the level $R_G = 2$ is reached, the pyramid decomposition is stopped unconditionally. The $R_G = 1$ resolution level is never achieved. For $R_G = 8$ and $R_G = 4$ levels the decomposition is stopped if ANY of the following three conditions is met:

- 1) The map is smooth-order 2.
- 2) The map is smooth-order 1
AND $ACT_L > T_{H3}$.
- 3) $ACT_L > T_H(L)$

The used threshold values are experimentally adjusted. For 8-bit luminance image their values are: $T_{H1}=225$; $T_{H3}=11$; $T_H(L)=15$ for level $R_G = 8$; and $T_H(L)=10$ for level $R_G = 4$.

6. EXPERIMENTAL RESULTS

A) Experimental setup

The experimental validation of the performance of the proposed EF procedure, **PROP**, was executed in two ways, the quality test and the efficiency test. The quality test is focused on grading the fused image appearance, by both visual and objective means. The efficiency test includes implementation of the algorithm core on different mobile devices and measurement of the execution times.

For the quality test, an EF image test collection available online [39], is used, which contains 128 triplets of exposure bracketed images, captured with exposure biases of 0, -2 and +2 points. The image sets cover various usage scenarios, including static or handheld camera, static or moving objects, indoor or outdoor scenes, objects of interest near to, or far from the camera, images captured with magnification (zoom), etc. Also, different statistical properties of the recorded scenes are included in the database: large flat surfaces, objects

with many details or prominent textures, surfaces with sky or water, regular geometric shapes, different image resolutions, etc.

The usual time delay between capturing the consecutive images in one exposure bracketed set is around 400 milliseconds, making the movements of the camera or the movements of the objects in the scene easily noticeable.

The triplets were provided as input sets (VF, UN and OV) for the proposed EF procedure. In order to simulate the view screen image VF, the zero-exposure-biased image from the triplet was down-scaled to a 1/4 resolution in each dimension.

The obtained fused results were evaluated visually, and compared against 5 other algorithms generating 9 other fused results. The algorithm in [10] (LI), was the first choice, which is state-of-the-art algorithm with broader usage in multi-focus and multispectral fusion in addition to the EF. This algorithm is inherently robust to moving objects in the scene. However, the proposed default values for the free parameters $r_1 = 45$ and $r_2 = 7$, (working well on image resolution of 512×512 pixels), are not suitable for higher image resolutions. Hence, two fused results are generated with this algorithm, **LI45/7**, using the default values for the free parameters, and **LI100/16**, using the values $r_1 = 100$ and $r_2 = 16$. It should be noted that the second set of parameters (100/16), dramatically increase the computational load of the algorithm [10], due to filtering with a very large support sized filters. Of course, the time complexity of the guided filtering can be kept linear and independent on the filter size by using a table of summed areas [27]. However, this method rapidly increases the memory consumption because it includes creation of one table per each source image and per each weight map, equally sized with their image or map counterparts, and with much higher bit-depth. For the method [27] to be used on five megapixel 8-bit input images (2592×1944) all the tables would have to use at least 31 bits per element, or four times more memory than the source images and weight maps themselves.

The second algorithm for comparison is [6] (MER), which represents a milestone in the EF algorithms based on Laplacian pyramid decomposition. This algorithm can accept both input sets containing 2 images OV and UN, (**MER2**), or containing the whole triplet OV, UN and VF, (**MER3**). The third comparison was made using the commercial software Photomatix from HDR-Soft [38], with all the user adjustable parameters

set at their default values. This software was also tested with two (**PMX2**) or three (**PMX3**) input images.

The fourth comparison is performed against the algorithm proposed in [8], which although computationally very demanding, includes advanced methods for compensating both the camera and the object motion in the process of image capturing. Again, two or three input images are optional, and the obtained results are denoted as **SEN2** and **SEN3**. Finally, the last comparison is with the predecessor of the PROP, the algorithm [21] (**KAR**), which uses the same Laplacian pyramid core as PROP, however, does not include the adaptive selection based on CRF, does not include the reduced motion vector set global motion compensation procedure, and works without VF image data.

For all of the results that are generated using three input images (MER3, PMX3, SEN3), the full resolution VF image was provided. For LI and MER algorithms, that do not embed image alignment part, the reduced vector set procedure explained above was used as a preprocessing tool. For KAR algorithm, the image alignment part that works only with translational displacements disengaged and replaced with the reduced MV set GMC procedure for the experiment.

B) Quality test: visual comparison

Few representative examples of the exposure bracketed triplets of images are shown in Fig. 6. The explanations of their statistical significances for various usage scenarios are placed in Table 2. All examples are 5 MP images.

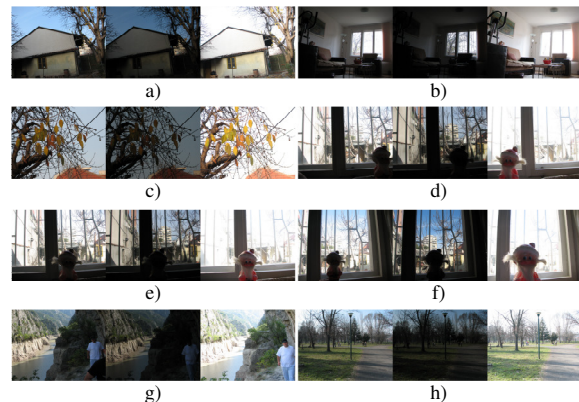


Fig. 6. Eight representative cases of exposure bracketed input image sets.

Table 2
Descriptions of the presented examples

Ex.	Camera mobility	Object motion	Comment
(a)	Static	No	Outdoor, large flat surfaces
(b)	Static	No	Indoor, very high radiance dynamic range
(c)	Handheld	No	Captured using magnification (zoom)
(d)	Static	Yes	Total displacement of a rigid object
(e)	Static	Yes	Fractional displacement of a rigid object
(f)	Static	Yes	Z-axis displacement of a rigid object
(g)	Handheld	Yes	Natural, non-rigid motion, near to camera
(h)	Handheld	Yes	Natural, non-rigid motion, far from camera

The fused results using a total of ten different methods are shown in Fig. 7. The results obtained with the proposed EF procedure are placed in the bottom row. For better preview, the reader is encouraged to examine the electronic form of the examples, available for download at [39]. In visual terms, the proposed EF procedure obviously yields the best comparative results. The LI scheme, whereas working well with lateral and fractional object displacements, suffers from total displacements or z-axis displacements. Also a prominent halo artifacts are present in LI45/7, due to the relatively small filter support compared to the image size. As can be seen in the LI100/16 result, the halo artifacts can be decreased with increment of the r_1 and r_2 parameters, however, a procedure for adjustment of the parameter values according to image content or size is not provided in [10]. The MER method produces fused results with high visual quality in the static scenes, however, its quality degrades significantly under dynamic conditions, and also, it does not include its own global alignment part. The commercial software requires at least three full resolution input images to produce reasonable output image (the PMX2 result is below any acceptable quality), yet it struggles with moving objects, too. Ghosting artifacts are noticeable in all three results featuring rigid-object motion. SEN2 and SEN3 are compensating the global and local motion very well (the different object position is a result of different input image being chosen as referent image), however, in addition to its immense complexity, this algorithm outputs true HDR images, which have to be tone mapped af-

terwards, process that requires user intervention, in order to adjust the proper luminance on the regions of interest (the face of the toy Santa Clause in the example, which is too dark). Finally, the KAR algorithm suffers from ghosting artifacts and discolorations in the moving object scenarios. Although it retains more vivid colors than PROP in some of the cases (the sky, the green flora...), that is result of a generally darker fused images, consequence of the non-adaptive selection thresholds.

C) Quality test: Numerical comparison

Along with visual evaluation, several numerical objective fusion metrics are employed to test the proposed EF procedure. The used metrics are the gradient based Q_{pABF} [33], the structure based Q_Y [35], the information theory based Q_{MI} [34], the phase congruency based Q_P [36] and visual information fidelity based Q_V [37].

The achieved numerical scores averaged across all of the tested images are shown in Figure 8, each metric values normalized to the highest scoring result. Among 10 tested methods, the proposed EF procedure is top 4 according to the gradient based metric Q_{pABF} , and top 3 according to the rest four metrics. In terms of the metric Q_V , the proposed EF procedure is considerably better than the competition. Overall, if scores from all the metrics are averaged, the proposed EF procedure is tied for the best score with the LI45/7 method, Table 3. However, the LI45/7 method is visually inferior to the proposed procedure. The high objective grades it receives are mainly the consequence of its robustness to moving objects in the scene, and preservation of the salient features in the input images. The presence of the severe halo artifacts in the LI45/7 results, represented by relatively slow luminance transitions in the image, is not considered by the quality metrics, which are based on salient features, such as edges, corners, structure or contours, generally more rapid luminance transitions.

Table 3
Overall Averaged Scores from all Quality Metrics (%)

LI45	LI100	MER2	MER3	PMX2	PMX3	SEN2	SEN3	KAR	PROP
96	95	93	85	83	82	95	93	90	96

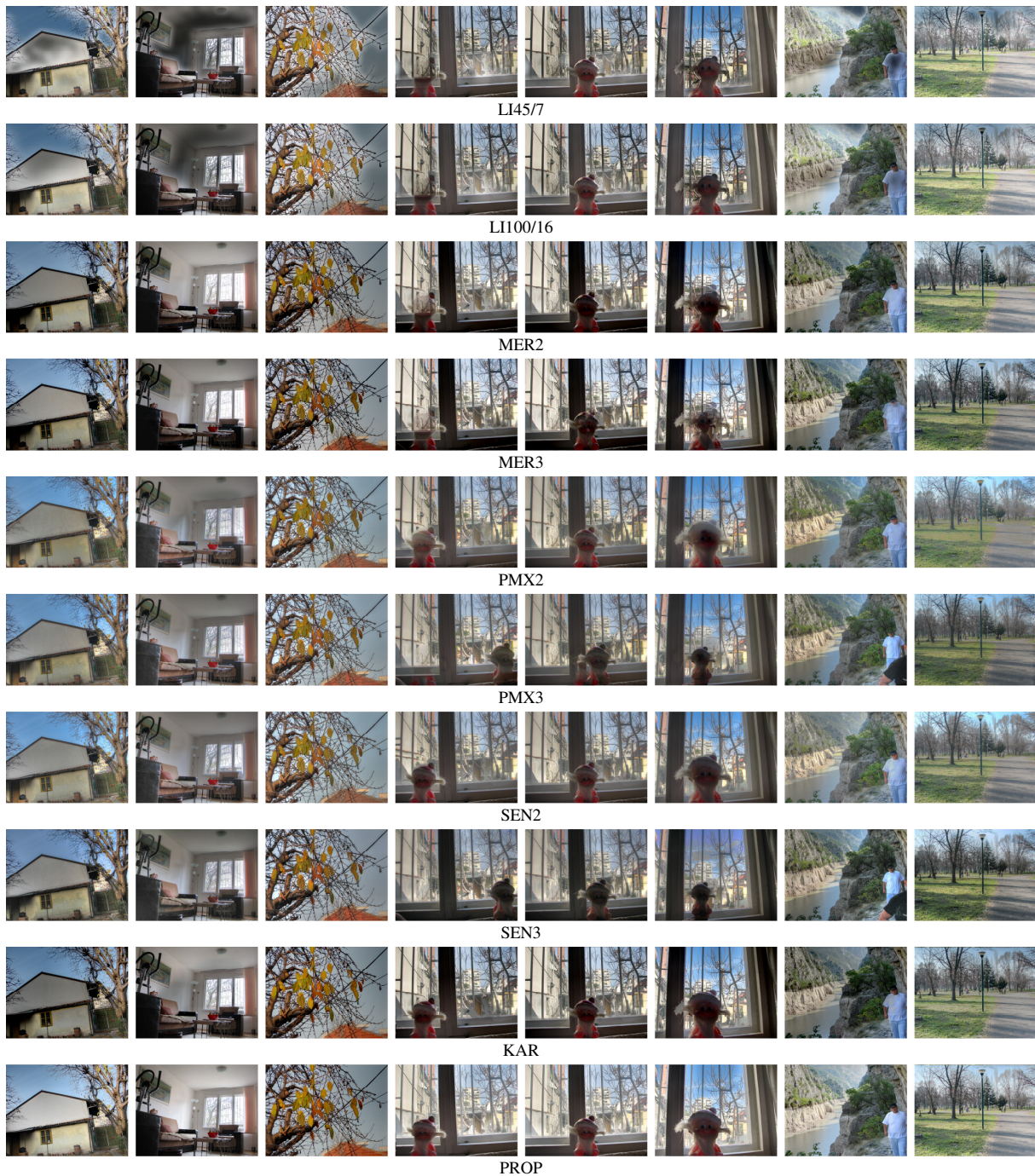


Fig. 7. Visual comparison of the performances of different algorithms. Examples by columns, algorithms by rows

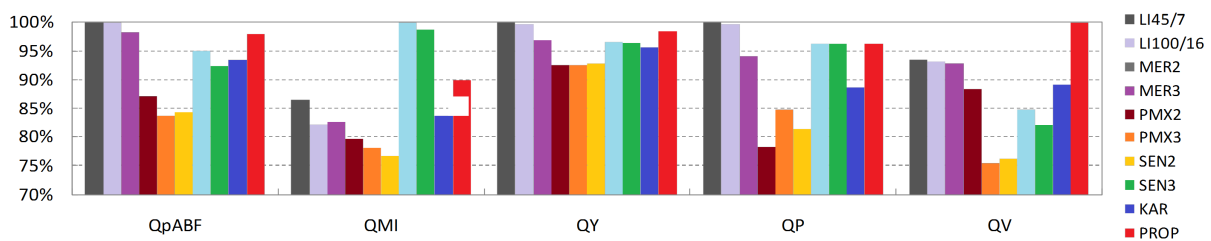


Fig. 8. Quality scores of the compared results, normalized to the highest

D) Efficiency test: Execution Times

In order to test the computational efficiency of the proposed EF procedure, the core of the algorithm (only the multi-level fusion part) was implemented in JAVA/C and compiled as an Android application. The application was tested on several mobile devices of different quality class, computational performances, and price-range. The execution times obtained for input images of VGA sizes are shown in Table 4. As it can be seen from the Table 4, all the execution times are shorter than 0.3 seconds, and as short as 60 milliseconds. Hence, the computationally most demanding part of the algorithm, its multi-level fusion core, works in acceptable wait time implemented on any modern mobile device.

Table 4

Algorithm Core Performance on Various mobile Devices

Model	CPU (GHz)	RAM (Bytes)	Processing times (ms)		
			min	max	avg
Sony Xperia SP	1.7 (dual)	1 G	57	64	60.5
Samsung Galaxy S3	1.4 (quad)	1 G	101	115	105.5
Lenovo A850	1.3 (quad)	1 G	135	156	142.5
Sony Ericsson Live	1	512 M	181	210	194.2
Samsung P3113(tablet)	1 (dual)	1 G	187	220	201.6
HTC Desire 300 1	(dual)	512 M	191	226	208.3
HTC Desire X	1 (dual)	768 M	212	262	229.9
HTC Desire 200 1		512 M	254	309	277.7

7. CONCLUSIONS

A complete and high quality exposure fusion procedure was presented in this paper, designed specifically for usage on a mobile platform devices. Several novel solutions are introduced, among them, accelerated and efficient computing, automatic calculation of the optimal fusion parameters without any user intervention, viewfinder datastream utilization, reduced motion vector set for input image alignment, and overall decision for the necessary of the fusion process.

The experimental results rate the presented procedure very high among the state-of-the art competition, both in the quality and in the efficiency terms.

Further research can be made towards a hardware-related procedure optimization, and device-specific implementation, e.g., a dedicated DSP processors available in the modern mobile devices can be utilized in the down-scaling, up-scaling and the filtering operations present in the procedure, which can render the whole procedure even more efficient.

REFERENCES

- [1] Mann, S. and Picard, R. W.: On being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures, In: *Proc. IS&T*, Washington D.C., USA, 1995, pp. 422–428.
- [2] Debevec, P. E. and Malik, J.: Recovering high dynamic range radiance maps from photographs, In: *Proc. SIGGRAPH*, Los Angeles, USA, 1997, pp. 369–378.
- [3] Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J.: Photographic tone reproduction for digital images, *ACM Trans. Graph.*, vol. **21**, no. 3, pp. 267–276 (Jul. 2002).
- [4] Rubinstein, R.: *Fusion of Differently Exposed Images*, Technion, Israel IT, Final Project Report, Oct. 2004.
- [5] Goshtasby, A.: Fusion of multi-exposure images, *Image and Vision Computing*, vol. **23** (6), pp. 611–618 (Jun. 2005).
- [6] Mertens, T., Kautz, J., and Van Reeth, F.: Exposure Fusion, In: *Proc. Pacific Graphics*, Maui, USA, 2007, pp. 382–390.
- [7] Shen, R., Cheng, I., Shi, J., and Basu, A.: Generalized random walks for fusion of multi-exposure images, *IEEE Trans. Image Processing*, vol. **20**, no. 12, pp. 3634–3646 (Dec. 2011).
- [8] Sen, P., Kalantari, N. K., Yaesoubi, M., Darabi, S., Goldman, D., and Shechtman, E.: Robust patch-based HDR reconstruction of dynamic scenes, *ACM Trans. Graph.*, vol. **31**, no. 6, art. 203, pp. 1–111 (Dec. 2012).
- [9] Li, X., Li, F., Zhuo, L., and Feng, D.: A layered-based exposure fusion algorithm, *IET Image Processing*, Vol. **7** (7), pp 701–711 (Oct. 2013).
- [10] Li, S., Kang, X., and Hu, J.: Image fusion with guided filtering, *IEEE Trans. Image Processing*, Vol. **22**, No. 7, pp. 2864–2875 (Jul. 2013).
- [11] Ward, G.: Fast, robust image registration for compositing high dynamic range photographs from handheld exposures, *J. of Graphics Tools*, vol. **8** (2), pp. 17–30 (2003).
- [12] Im, J., Jang, S., Lee, S. and Paik, J.: Geometrical transformation-based ghost artifacts removing for high dynamic range image, In: *Proc. ICIP*, Brussels, Belgium, 2011, pp. 365–368.
- [13] Jacobs, K., Loscos, C. and Ward, G.: Automatic high-dynamic range image generation for dynamic scenes, *IEEE Comp. Graphics and Applications*, vol. **28** (2), pp. 84–93 (Feb. 2008).

- [14] Zimmer, H., Bruhn, A. and Weickert, J.: Freehand HDR imaging of moving scenes with simultaneous resolution enhancement, *Comp. Graph. Forum*, vol. **30**, no. 2, pp. 405–414 (Apr. 2011).
- [15] Hossain, I. and Gunturk, B. K.: High dynamic range imaging of non-static scenes, In: *Proc. SPIE*, San Francisco, USA, 2011, vol. **7876**, pp. 78760P–78760P–8.
- [16] Park, S., Oh, H., Kwon, J. and Choe, W.: Motion artifact-free HDR imaging under dynamic environments, In: *Proc. ICIP*, Brussels, Belgium, 2011, pp. 361–364.
- [17] Hu, J., Gallo, O., Pulli, K.: Exposure Stacks for Live Scenes with Hand-held Cameras, In: *Proc. ECCV*, Florence, Italy, 2012, pp. 499–512.
- [18] Jinno, T. and Okuda, M.: Multiple exposure fusion for high dynamic range image acquisition, *IEEE Trans. Image Processing*, vol. **21**, no. 1, pp. 358–365 (Jan. 2012).
- [19] Zheng, J., Li, Z., Zhu, Z., Wu, S. and Rahardja, S.: Hybrid patching for a sequence of differently exposed images with moving objects, *IEEE Trans. Image Processing*, vol. **22**, no. 12, pp. 5190–5201 (Dec. 2013).
- [20] Kartalov, T., Ivanovski, Z. and Panovski, Lj.: Optimal exposure value shift in acquisition of hdr images, In: *Proc. CIIT*, Bitola, Macedonia, 2010, Available online: <http://www.dipteam.info>.
- [21] Kartalov, T., Ivanovski, Z. and Panovski, Lj.: Fully automated exposure fusion algorithm for mobile platforms, In: *Proc. ICIP*, Brussels, Belgium, 2011, pp. 369–372.
- [22] Kartalov, T., Ivanovski, Z. and Panovski, Lj.: Optimization of the pyramid height in the pyramid-based exposure fusion algorithms, In: *Proc. INFOTEH*, Jahorina, Bosnia and Herzegovina, 2011, Vol. **10**, Ref. B-III-7, pp. 219–223, Available online: <http://www.dipteam.info>.
- [23] Kartalov, T., Ivanovski, Z. and Panovski, Lj.: A real time global motion compensation for multi-exposure imaging algorithms, In: *EUROCON*, Lisbon, Portugal, 2011, [IEE-EXplore link](#).
- [24] Kartalov, T., Ivanovski, Z. and Panovski, Lj.: Using reduced motion vector set for real time motion registration in HDR imaging, In: *Proc. MMSP*, Pula, Italy, 2013, pp. 344–349.
- [25] Kartalov, T., Ivanovski, Z. and Panovski, Lj.: Error resilient global motion registration for image fusion, In: *Proc. ETAI*, Ohrid, Macedonia, 2013, Available online: <http://www.dipteam.info>.
- [26] Burt, P. J. and Adelson, E. H.: The laplacian pyramid as a compact image code, *IEEE Trans. Communications*, vol. **com-31**, no. 4, pp. 532–540 (Apr. 1983).
- [27] Crow, F. C.: Summed-area tables for texture mapping, In: *Proc. SIGGRAPH 84*, vol. **18**, no. 3, pp. 207–212 (Jan. 1984).
- [28] *Colorimetry – Part 4: CIE 1976 L*a*b* Colour Space*, CIE Draft Standard, CIE DS 014-4.3/E, 2007.
- [29] *Photography – Digital still cameras – Determination of exposure index, ISO speed ratings, standard output sensitivity, and recommended exposure index*, ISO International Standard 12232, 2006.
- [30] Tsai, R. Y. and Huang, T. S.: Uniqueness and estimation of 3-D motion parameters of rigid objects with curved surfaces, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. **6**, no. 1, pp. 13–27 (Jan. 1984).
- [31] Nam, K. M., Kim, J. S., Park, R. and Shim, Y. S.: A fast hierarchical motion vector estimation algorithm using mean pyramid, *IEEE Trans. Cir. Sys. for Video Technology*, vol. **5**, pp. 344–351 (Apr. 1993).
- [32] Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Processing*, Vol. **13**, No. 4, pp. 600–612 (Apr. 2004).
- [33] Xydeas, C. and Petrović, V.: Objective image fusion performance measure, *Electronics Letters*, vol. **36** (4), pp. 308–309 (Feb. 2000).
- [34] Hossny, M., Nahavandi, S. and Creighton, D.: Comments on information measure for performance of image fusion, *Electronics Letters*, vol. **44**, no. 18, pp. 1066–1067 (Aug. 2008).
- [35] Yang, C., Zhang, J., Wang, X., Liu, X.: A novel similarity based quality metric for image fusion, *Inf. Fusion*, vol. **9**, 2, pp. 156–160 (Apr. 2008).
- [36] Zhao, J., Laganieri, R. and Liu, Z.: Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement, *Int. J. Innovative Computing, Information and Control*, vol. **3**, no. 6, pp. 1433–1447 (Dec. 2007).
- [37] Han, Y., Cai, Y., Cao, Y. and Xu, X.: A new image fusion performance metric based on visual information fidelity, *Inf. Fusion*, vol. **14** (2), pp. 127–135 (Apr. 2013).
- [38] HDRsoft Photomatix Pro 5.0. (<http://www.hdrsoft.com/>)
- [39] Digital image processing team (<http://dipteam.feit.ukim.edu.mk>).