Multiexposure and multifocus image fusion with multidimensional camera shake compensation

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- This article was published in the journal, Optical Engineering [Copyright (2013) Society of Photo-Optical Instrumentation Engineers]. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

Metadata Record: https://dspace.lboro.ac.uk/2134/13729

Version: Published

Publisher: © Society of Photo-Optical Instrumentation Engineers

Please cite the published version.
This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
Multiexposure and multifocus image fusion with multidimensional camera shake compensation

Alexis Lluis Gomez
Sara Saravi
Eran A. Edirisinghe
Multiexposure and multifocus image fusion with multidimensional camera shake compensation

Alexis Lluis Gomez
Sara Saravi
Eran A. Edirisinghe
Loughborough University
Digital Imaging Group
Loughborough LE11 3TU, United Kingdom
E-mail: s.saravi@lboro.ac.uk

Abstract. Multiexposure image fusion algorithms are used for enhancing the perceptual quality of an image captured by sensors of limited dynamic range. This is achieved by rendering a single scene based on multiple images captured at different exposure times. Similarly, multifocus image fusion is used when the limited depth of focus on a selected focus setting of a camera results in parts of an image being out of focus. The solution adopted is to fuse together a number of multifocus images to create an image that is focused throughout. A single algorithm that can perform both multifocus and multiexposure image fusion is proposed. This algorithm is a new approach in which a set of unregistered multiexposure/focus images is first registered before being fused to compensate for the possible presence of camera shake. The registration of images is done via identifying matching key-points in constituent images using scale invariant feature transforms. The random sample consensus algorithm is used to identify inliers of SIFT key-points removing outliers that can cause errors in the registration process. Finally, the coherent point drift algorithm is used to register the images, preparing them to be fused in the subsequent fusion stage. For the fusion of images, a new approach based on an improved version of a wavelet-based contourlet transform is used. The experimental results and the detailed analysis presented prove that the proposed algorithm is capable of producing high-dynamic range (HDR) or multifocus images by registering and fusing a set of multiexposure or multifocus images taken in the presence of camera shake. Further, comparison of the performance of the proposed algorithm with a number of state-of-the-art algorithms and commercial software packages is provided. In particular, our literature review has revealed that this is one of the first attempts where the compensation of camera shake, a very likely practical problem that can result in HDR image capture using handheld devices, has been addressed as a part of a multifocus and multiexposure image enhancement system. © 2013 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.OE.52.10.102007]

Subject terms: image registration; multiexposure multifocus image fusion; translational and rotational camera shake; wavelet-based contourlet transform; coherent point drift; scale invariant feature transform; random sample consensus; camera shake compensation.

Paper 130060SSP received Jan. 14, 2013; revised manuscript received Jun. 19, 2013; accepted for publication Aug. 16, 2013; published online Sep. 20, 2013.

1 Introduction

In the field of high-dynamic range (HDR) imaging technology, in the past decade there have been significant developments encouraged by the increased consumer demand for experiencing images that are perceptually closer in appearance to images perceived by the human psychovisual system. To this effect, HDR imaging sensors are at present replacing the so-called standard dynamic range (SDR) sensors in digital cameras. However, the lack of developments in image/video encoding algorithms and display technology capable of making practical use of HDR images makes it still important to find alternatives to rendering HDR scenes using SDR imagery. Therefore, a number of algorithms have been proposed in literature to fuse multiple-exposure SDR images that result in images that are perceptually similar to HDR images, i.e., images perceived by the human eye.

Multiexposure image fusion involves the fusion of multiple consecutive images of the same scene taken in quick succession by an SDR camera. However, a problem thus arises: camera shake can cause severe deregistration between multiple images that invalidate the direct applicability of many existing multiexposure image fusion algorithms. The camera shake can be translational (vertical, horizontal) or rotational (in-plane and out-of-plane), and it is thus important that compensation for both types of shake is carried out prior to image fusion.

Similarly, due to the limited depth of field of optical lenses, it is usually impossible to capture an image that contains all relevant objects in focus. A solution for this is a multifocus image fusion that fuses two or more images that are captured using different camera settings (i.e., different focuses) of the same scene in order to form a final image with uniform focus and sharp content. Due to the slightly different focus settings used in capturing the constituent images for fusion, in addition to the presence of camera
shake in translational (vertical, horizontal) and rotational forms, compensation for the effect of image zooming (hence multidimensional camera shake) is very important to consider.

Image fusion has been used in a wide variety of applications such as machine vision, automatic change detection, biometrics, medicine, remote sensing, etc. In general, the goal of image fusion is to integrate multiple images so that the resulting image is clearer and more intelligible.

Image fusion techniques can be classified into three different levels: pixel level, feature level, and decision level. Pixel-level fusion is the combination of pixels from multiple sources into a single resolution image. Feature-level fusion extracts features, e.g., edges, corners, lines, texture parameters, etc., from different image sources and then combines them into one final image. Decision-level fusion combines the results from multiple algorithms to yield a final fused decision. Methods of decision fusion include voting, statistical, and fuzzy logic based methods. The above categorization does not encompass all possible fusion methods, since input and output of data fusion may be different at different levels of processing. In practice, fusion procedures are often a combination of the three levels mentioned above. The advantage of pixel-level fusion is that images contain original data and therefore the pixel information is preserved. On the other hand, region- and decision-based methods are very useful because real-world objects usually consist of structures at different scales and human visual systems. In order to procure a high-quality fused image, all image fusion methods should be capable of performing a precise subpixel-level registration, accurately fusing all images when local/global changes of intensity happen, and/or, invariant to different levels of noise in the images, being able to compensate for moving objects in the scene.

The focus of the proposed research is the development of an end-to-end multiexposure and multifocus image fusion system that addresses the issues of camera shake, low dynamic range in SDR cameras, and the limited depth of field of optical lenses. A significant number of image fusion algorithms have been proposed in the literature. However, only a very few algorithms focus on the problem of camera shake, and these are severely limited in their ability to compensate the motion produced by the camera shake. The background of the proposed fusion technique comes from a combination of pyramidal contourlet and wavelet fusion techniques (feature-level fusion). The downsides of each of these fusion techniques by themselves are that the pyramidal and contourlet approaches are considered as computationally inefficient due to the redundancy presented in the pyramidal decomposition. The wavelet transform results are acceptable in natural images, but smooth edges cannot be detected powerfully because of its restricted three directions (horizontal, vertical, and diagonal) to detect features in the images. On the other hand, the contourlet transform is a two-dimensional transform that has the capability to effectively represent images containing curves and features. In the contourlet transform, multiscale and multidirection analyses are done separately. First, the Laplacian pyramid (LP) transform is used to perform a multiscale decomposition, and then a directional filter bank is used to filter the high-frequency components from each LP channel. Therefore a contourlet-based image fusion method produces improved outcomes as opposed to wavelet-based fusion. Thus, an approach proposed in Ref. provides a solution to the above shortcoming, i.e., wavelet-based contourlet transform (WBCT), which is nonredundant and has a multiresolution structure. The advantages of using WBCT are that it solves the problems of multiscale localization, directionality, and anisotropy. However, the WBCT fusion approach adds artifacts, and therefore the perception of the image fused, i.e., textures, edges, and intensity are highly affected. In Ref. we proposed an improved WBCT-based image fusion algorithm that compensates for camera shake. However, the evaluation and validation of the approach presented was limited and hence the justification of the performance improvements was not sufficiently strong.

In this paper, a new image fusion algorithm that allows compensating for camera shake (translational, rotational, and zoom) through a registration process and a new approach of the WBCT fusion that allows combining aspects of both pixel level and region level (thus minimizing artifacts) in order to fuse a set of images are proposed. Section 2 presents the operational and functional details of the proposed system. Section 3 provides experimental results and a detailed evaluation of the system’s performance. Finally, Sec. 4 concludes with an insight to further work.

2 Proposed System

The proposed system consists of a registration module which is used prior to image fusion. After the images have been registered and camera shake has been compensated for, a fusion module is used in order to produce a single HDR image from a set of multiexposure images or a single multifocus image from a set of images with different depths of field.

2.1 Image Registration

Figure 1 illustrates the block diagram of the proposed approach to registering a set of images with camera shake. The approach is based on the selection of a significant set of matching points, i.e., key-points, between a selected base image and an image to be registered and subsequently using them to calculate the transformation matrix for image registration. Sections 2.1.1, 2.1.2, and 2.1.3 describe the three key stages of the image-registration approach adopted.

2.1.1 Scale invariant feature transform based key-point selection

The scale invariant feature transform (SIFT) is an algorithm that is capable of detecting and describing local features of an image. Its invariance to rotation, scale, and translation has made it a popular algorithm in many areas of computer vision and pattern recognition. SIFT is also partially invariant to illumination changes and robust to local geometric distortion.

In the proposed approach to image registration, a base image is selected from among the set of images in which another image from the set is registered to the base image. The algorithm first uses SIFT to find significant feature/key-points in both images. In the case of multiexposure registration, it is noted that the base image is considered to be the image with medium level of exposure among the multiexposure image set when comparing the exposure...
changes that are represented within the image set being fused.

Due to the relatively large number of feature points that may be selected by SIFT in carrying out the matching of key-points between the base image and the image to be registered, it is likely that two geometrically noncorresponding points of the two images may match as they result in the minimum distance. Therefore, reducing key-point outliers prior to the matching key-points will improve the reliability of matching and hence the outcome of the eventual image registration task. This requires the use of the stage that follows.

2.1.2 Using random sample consensus to remove matching point outliers

Random sample consensus (RANSAC)\(^2\) algorithm is an iterative method to approximate factors of a mathematical form from a set of experimental data which include outliers. RANSAC is able to do robust estimations of the model parameters; it can estimate the parameters with a high degree of accuracy even when a significant amount of outliers are present in the data set.

The key-point data sets generated from the SIFT stage consist of inliers and outliers. Within the purpose of using the RANSAC algorithm for the task at hand, outliers are defined as key-points that are found to be present in the image being registered but not found in the base image in a specially closer location. The outliers may result from poor illumination conditions, noise, etc. In our approach all SIFT key-points resulting from the stage described in Sec. 2.1.1 from the base image and the image being registered are first fed to RANSAC algorithm. The base image matching points are assumed theoretically to be the inliers. Then RANSAC fits a model to these inlier points and tests the points from the image being registered against the fitted model. If a point fits to the model, it will be regarded as an inlier. The model is recalculated from each and every inlier, followed by an error estimation of the inliers relative to the model. The outlier key-points are finally removed from the set of key-points of the image being registered. The set of inlier key-points of the image being registered are subsequently used for corresponding point matching with the set of key-points of the base image.

As stated above, the removal of outliers using RANSAC results in increasing the reliability of subsequent SIFT key-point matching. Thus, the result of the stage described above is two point sets from base image and the image being registered that can now enter the final stage of a typical SIFT algorithm, i.e., key-point matching. A Euclidean distance between feature vectors of key-points is used to find corresponding point sets, by selecting all point sets whose distance is below a specified, experimentally determined threshold.

2.1.3 CPD algorithm for registration

In this section, the use of the coherent point drift (CPD)\(^2\) algorithm is described to register the images. CPD allows preparing the set of images to be fused in the subsequent stage of the proposed approach. CPD is based on point set registration and aims to form links between two given sets of points to find the corresponding features and the necessary transformation of these features that will allow the images to be registered.

In the CPD approach, there are two methods for registering an image: rigid and nonrigid point set approaches. These two are based on a transformation model principle. The key characteristic of a rigid transformation is that “distance between points are preserved,” which means it can only be used in the presence of translation rotation, but not under scaling and skewing distortions as nonrigid transformation allows (affine transformations). The method used in the proposed approach is a nonrigid transformation, which provides the flexibility of registering images under rotational, translational, and nonuniform scaling and skewing conditions.

A probabilistic method has been introduced in Ref. 27 for point set registration, which is named the CPD method. Given two point sets, a Gaussian mixture model (GMM) is applied to the first point set by maximizing the likelihood and subsequent fit to the second point set. The GMM results in a set of centroids and variances for a reference set of points, which can be used to calculate the likelihood of an input data point set with reference point set. Subsequently, the point sets become aligned and the correspondence is achieved using the subsequent likelihood of the GMM.

In Sec. 2.1.1 it was mentioned that the SIFT-based approach to identifying key-points, which can subsequently be used in matching, allows the selection of key-points invariant to translation, rotation, and scale. It was further mentioned that SIFT is also partially invariant to illumination changes and robust to local geometric distortion. Therefore, the use of SIFT will allow the robust selection of potentially matchable key-points in the presence of camera shake (due to invariance to translation and rotation) and zooming (due to
invariance to scale). Further, the process will be partially invariant to illumination changes and effects due to the presence of geometric distortion. Given statement above in a typical HDR image capture and creation process of a handheld device/camera which involves fusing together a number of SDR images captured under different exposure and focus settings, in quick time succession, may introduce relative vertical, horizontal, rotational movement, zooming, illumination changes and geometric distortion, the use of SIFT with above-mentioned invariance are well justified. Further, the use of nonrigid point set CPD algorithm (see Sec. 2.1.3) allows the proposed image registration approach to handle the presence of camera zoom between the constituent images.

Figure 2 illustrates images captured for the unit testing of the performance of the image registration module, which is applied prior to the multiexposure image fusion. The images were taken without the aid of a tripod, and the fusion was performed with two different image exposures in RGB domain. Figure 2(a) shows an overexposed image considered as the base image for the registration process, and Fig. 2(b) is the relevant underexposed image, which is to subsequently be registered to the base image. Figures 2(c) and 2(d), respectively, illustrate the distribution of all SIFT key-points of the over- [Fig. 2(a)] and underexposed [Fig. 2(b)] images. A close inspection of the two point sets indicates that the number of key-points found on the two images is not equal and the point sets are not uniformly distributed. A number of key-points in both images do not appear to find matching key-points from the corresponding image. This is better illustrated in Figs. 2(e) and 2(f) by the black and red circles, respectively, i.e., feature points that do not have matching counterparts in their opposite exposure image. In order for the images to be registered via point-set matching, the key-points that do not find matches from the corresponding image should be removed as outliers. Note that CPD is able to address point-set matching where the distributions are nonuniform once the above outliers are removed. Therefore,

![Figure 2](image-url)

Fig. 2 Image registration process. (a) and (b) Original set of multiexposure images. (c) and (d) Plotted scale invariant feature transforms (SIFT) key-points of overexposed and underexposed images, respectively. (e) and (f) Selected by random sample consensus (RANSAC) application in overexposed and underexposed images, respectively. (g) and (h) Plotted key-points after random sample consensus (RANSAC) application in overexposed and underexposed images, respectively. (i) and (j) Nonmatching points are eliminated using RANSAC in overexposed and underexposed images, respectively. (k) Key-points points to be registered. (l) Key-points registered using coherent point drift. (m) Fused image without registration. (n) Enhanced fused image with registration.
The registration process has been illustrated and explained for the case of multiexposure images. It is noted that the approach works exactly the same for multifocus images registration.

2.2 Multiexposure and Multifocus Image Fusion

Once all images are registered, a WBCT is used for identifying regions of maximum energy from within the multiexposure or multifocus images to generate a fusion decision mask, which is later used for fusing the multiexposure or multifocus images.

2.2.1 Wavelet-based contourlet decomposition

The WBCT decomposition approach can determine the high-frequency contourlet sub-bands, which contain details such as texture and edges, and the low-frequency contourlet sub-bands that contain the fuzzy, spread-out information such as background information of an image. This property of WBCT allows determining a fusing criterion where the best areas of a set of images are selected according to the level of energy of the sub-bands.

A previous attempt to fuse images with the WBCT approach applied a three-level decomposition in an LP fashion using the wavelet transform, which allowed decomposing an image into a nine high-pass and one low-pass sub-bands. Then, a directional contourlet transform was applied on each high-pass sub-band of the wavelet decomposition. In this WBCT approach, the image is only decomposed into one level of the wavelet transform, which allows decomposing the image into four sub-bands, and on each high-pass wavelet sub-band, a directional contourlet transform is applied. The levels of decomposition are reduced, as a difference to the three-level decomposition of the WBCT approach in Ref. 21, to reduce the inconsistency of areas fused and loss of details at each level of decomposition. Artifacts produced by Ref. 21 and improvements of the proposed approach are shown in the Experimental Results section.

Figure 3 illustrates the frequency partitioning of a given image when WBCT is used.

Figure 4 illustrates the block diagram of the fusion module. Note that for clarity of illustration, Fig. 4 does not illustrate the low-frequency contourlet transform. The mentioned fusion process is explained in the following section.

2.2.2 Proposed WBCT image fusion approach

The basic idea of the fusion algorithm is to generate a generic fusion mask that can be used to fuse all high- and low-frequency contourlet sub-bands, and the low-pass wavelet sub-band.

2.2.3 High- and low-frequency contourlet sub-band fusion

In order to generate the fusion mask, the absolute energy of high-frequency coefficients is calculated in a block-based manner, comparing and selecting the block with the higher energy (by comparing the two corresponding blocks in the under- and overexposed images). The fusion mask records in a binary fashion the selected block position in the image according to the correspondence of the block selected. The energy of a region can be calculated as follows:

Region energy $E$ of a high-frequency sub-band $E_H$ (where $H = (l, m, n)$, l-level of wavelet decomposition, $m$-LH, $H$, and $HH$ bands of wavelet decomposition, $n$-directional contourlet sub-band) of an image $X$ can be calculated using Eq. (1).

$$E_H(x) = \sum_{(x,y) \in H} f_H(x,y)^2,$$

where $f_H(x,y)$ is the coefficient at location $(x,y)$ of the high-frequency sub-band $H = (l, m, n)$ Figure 5 shows an example of a fusion mask obtained from the calculation of the energy of the high-frequency contourlet sub-bands. Note that the fusion mask is a binary image in which a value of 1 represents the fact that the higher-energy block comes from the underexposed image and a value of 0 represents the fact that the higher-energy block comes from the overexposed image. Further, for illustration purposes, only

![Image](image.png)

**Fig. 3** Frequency partitions obtained with wavelet-based contourlet transform (WBCT).
the fusion mask is drawn to be of similar in dimensions to the original images in Fig. 5. It is noted that a single element of the fusion mask represents an entire block area of the original images.

After the fusion mask has been created, all the high- (i.e., LH, HL, and HH) and low-frequency contourlet sub-bands can be combined following a block-by-block approach, using an alpha blending approach, as follows:

\[
SB_{\text{fused}} = SB_u \times \alpha + SB_o \times (1 - \alpha),
\]

where \( \alpha \) is the value of fusion mask for the block, \( SB_u \) is the sub-band of the underexposed image, and \( SB_o \) corresponds to the sub-band of the overexposed image. \( SB_{\text{fused}} \) is the sub-band of the fused block. Equation (2) works the same way when fusing multifocus images, with the multixposure images replaced by the multifocus images.

Once all contourlet sub-bands are alpha blended, the low-pass wavelet sub-band is fused as described in the next section.

### 2.2.4 Low-pass wavelet sub-band fusion

The low-pass wavelet sub-band is the sub-band that holds the luminance information of an image. Thus, in order to generate an HDR image, it is important to keep the best luminance areas of all images; i.e., clipped dark or clipped bright areas should not be considered in the fusion of the low-pass wavelet sub-band. WBCT previous fusion attempts fused the low-pass wavelet sub-band by calculating the average of the low-pass wavelet sub-bands of the multiple exposure images. However, our investigations revealed that this approach produced a hazy fused image.

In order to solve this shortcoming, the fusion mask generated by the calculation of the energy of the high-frequency contourlet sub-bands (Sec. 2.2.2) is used again to fuse the low-pass wavelet sub-band. However, in this case, the fusion mask is blurred with a Gaussian kernel with a standard deviation of 1.3 prior to its use in fusion. The standard deviation was determined experimentally so that the Gaussian kernel could blur the mask in a more efficient way. After blurring the fusion mask, the low-pass wavelet sub-bands are obtained using Eq. (2). Figure 6(b) shows the fusion mask obtained from Sec. 2.2.2 after being blurred with the Gaussian kernel.

### 2.3 Reconstructing the Fused Image

After obtaining the best low- and high-frequency sub-bands of contourlet transform using the approach presented in Sec. 2.2.2 and the best low-frequency sub-band of wavelet
transform as described in Sec. 2.2.3, the complete fused image is reconstructed using inverse WBCT approach.

3 Experimental Results

Our investigations revealed that when adopting the WBCT approach from Ref. 21, the output images added undesirable artifacts such as loss of resolution, ghosting of moving objects, blockiness, chromatic aberrations (decoloring around the edges), a hazy appearance of the image due to LP decomposition/reconstruction, and average blending of wavelet transform band. This is because at each level of decomposition, the sub-bands were fused and reconstructed with a nonconsistent selection of areas within the sub-bands. That is, areas selected at different bands and at different levels of the decomposition were not the same as the fusion rule did not guarantee consistency throughout all bands. This fact makes the inverse of WBCT be inaccurate when reconstructing textures, edges, and colors, which results in visible artifacts in the final image.

Figure 7(b) shows a cropped section of a fused image obtained using the approach presented in Ref. 21 if it is compared with Fig. 7(a) (original underexposed image). Artifacts such as blockiness, chromatic aberrations, and loss of resolution artifacts can be observed near the edges and around the plants in the middle of the image. These artifacts are created due to the reasons explained above.

Due to the multiple images that are needed to create HDR or multifocus images, if objects move while capturing the images, a ghosting or duplicate effect will appear in the image. Figure 7(d) shows a fused image with ghosting effects mentioned. The arrows in Fig. 7(d) point to the (people walking, cars, and the lorry) ghosting effects.

The use of the fusion mask allowed us to reduce the inconsistency of areas fused and therefore minimize the appearance of artifacts, such as haziness and ghosting effects. Further, by reducing the decomposition levels to one, we have kept the spectral degradation to a minimum.

Figure 8 represents experimental results on a set of five multieposure image pairs. For each set in a top-to-bottom and left-to-right order, the five images illustrated represent the underexposed image (a), the overexposed image (b), the output obtained with the WBCT approach of Ref. 21 with no compensation for camera shake (c), the same with compensation for camera shake using the approach proposed in Ref. 28 (d), and finally the fused image obtained when using the camera shake compensation approach and the novel WBCT approach presented in this paper (e). The results clearly demonstrate the fact that the removal of camera shake is vital for a good-quality fused image and can be successfully addressed by the novel approach presented in Sec. 2.1. The results also demonstrate that the use of the proposed novel WBCT approach to fusion results in images that are sharper and more visually pleasing and natural appearing.

Figure 9 presents results of a more detailed investigation into the impact of the proposed fusion algorithm. The WBCT-based fusion approach of Ref. 21 is used in order to compare the performance of the proposed approach. The comparison of Fig. 9(a) with Fig. 9(b) shows that the proposed approach is capable of enhancing the visible
Fig. 8 Experimental results on a set of five multiexposure image pairs. For each set in a top-to-bottom and left-to-right order, the five images represent (a) the underexposed image, (b) the overexposed image, (c) the output obtained with the WBCT approach of Ref. 21 with no compensation for camera shake, (d) the same with compensation for camera shake using the approach proposed Ref. 28, and (e) the fused image obtained when using the camera shake compensation approach and the novel WBCT approach presented in this paper.
dynamic range of the image, making it look more visually pleasing. In Fig. 9(a), the image haziness and colors are better preserved than in Fig. 9(b). In order to illustrate the proposed approach’s ability to enhance the sharpness of edges, color aberrations, smudging, and the reduction of blocking artifacts, Figs. 9(c) and 9(d) illustrate a cropped and significantly zoomed area of the images that were illustrated in Figs. 9(a) and 9(b). The proposed approach does not add unwanted artifacts [see Fig. 9(c)]. It is noted that in the results being presented in Fig. 9, all experiments were conducted on a pair of images that underwent prior registration to compensate for camera shake, using the novel approach that was presented in Sec. 2.1, i.e., any possibility of camera shake has been removed and the comparison of quality is being done after this initial stage of the proposed system, the basic idea being to prove the added value of the novel approach to WBCT-based fusion algorithm presented in this paper.

When objects in the image move at the time of capture, ghosting effects may appear in the fused image. Figure 10(a)
shows how our approach is less sensible to ghosting artifacts in comparison to the previous approach [Fig. 10(b)] based on WBCT presented in Ref. 21.

Due to the nature of the WBCT approach adopted, the proposed algorithm could also be used to fuse a set of multifocus images. In the experiments carried out it could be

![Multifocus image fusion comparison](image1)

**Fig. 11** Multifocus image fusion comparison: (a) and (b) show the original set of multifocus images. (c) is the result of the image fused with the algorithm proposed. In (d), a spatial frequency fusion approach is used. (e) is the resulting fused images using a Laplacian pyramid fusion approach.

![Perceptual quality comparison](image2)

**Fig. 12** Comparison of the performance of the proposed algorithm with three state-of-the-art algorithms. (a) and (b) are underexposed and overexposed images, respectively. The outputs produced by (c) the proposed approach, (d) our previously published approach of Ref. 28, (e) Photomatix, and (f) Photoshop.
verified that the quality of the multifocus images fused were preserved and, in most cases, enhanced in comparison to other fusion algorithms such as fusion using LP pyramid decomposition/reconstruction and spatial frequency fusion. Figure 11 shows the multifocus images fused and a comparison of the outputs of the algorithms mentioned above. Figure 11 shows two sets of images each marked from (a) through (d) that were used to evaluate the performance of the proposed approach. The results clearly show that the proposed approach performs better than the spatial frequency algorithm based and the LP fusion based approaches. It is noted that although the LP fusion approach did not add artifacts, the resulting fused image is perceptually blurred. The spatial frequency algorithm added stitching artifacts such as oversharpening, and further luminance wise flat backgrounds made the algorithm fail to output a visually pleasing image [see Fig. 11(d)].

In the experiments performed with multifocus images, the depths of fields had to be restricted to a certain focus range. This is because objects can be bigger in appearance from one image to another. If the range is not restricted, occlusion problems will arise and it will make the fusion algorithms fail (add occlusion artifacts) when fusing the images.

The results of the images fused with the proposed algorithm clearly illustrated how and why the proposed approach can compensate for camera shake and also showed the improvement of subjective quality performance when fusing multiexposure and multifocus fusion in comparison to other fusion algorithms.

Figure 12 compares the performance of the proposed algorithm with three state-of-the-art algorithms, namely our previous approach to multifocus and multiexposure image fusion presented in Ref. 28 and the outputs generated by two commercial products, namely Photomatix and Photoshop. A visual comparison of results presented in Fig. 12 demonstrate the proposed algorithm’s ability to produce HDR images that are visually pleasing, free of artifacts, and appear to have a wide dynamic range and contrast.

4 Conclusions

In this paper a new algorithm capable of compensating for camera shake of a number of degrees of freedom and capable of producing HDR images and effectively fused multifocus images was proposed. This algorithm enables users to create HDR and multifocused images with an SDR camera without the aid of stabilizing devices such as tripods, due to the compensation in place for the removal of camera shake. The proposed approach registers a set of multiexposure or multifocus images with an algorithm that is based on SIFT feature point selection, followed by the use of RANSAC algorithm for removing outliers in matching, and finally a fast CPD algorithm that uses a fast Gaussian transform and low-rank matrix estimation to register displaced images. After the set of multiexposure images are registered, a novel wavelet-based Contourlet transform approach for image fusion is used. The experiments conducted with the proposed approach enabled one to separately demonstrate the positive impact of both the camera shake compensation algorithm and the image fusion algorithm. In particular, their performance was compared with a number of state-of-the art approaches, especially the most popular commercial products. Experiments revealed the proposed approach’s ability to significantly improve the visual quality of fused images, minimizing or completely removing unwanted artifacts often created by other state-of-the-art approaches. Results were demonstrated using a large set of test images captured specifically for the experiments conducted.

References


**Alexis Lluis Gomez** obtained his BS in computer science from The State University of New York at Buffalo in 2009. He completed his MSc in Internet computing and network security at Loughborough University in 2010. He joined Loughborough University as a research associate in December 2010 and has been working in a partnership with Apical Limited on a KTP Project “QPipe,” where the research carried out aims to generate next generation of image sensor processors. He started his PhD studies in December 2011.

**Sara Saravi** completed her high school education at Talented School of Farzanegan, Iran, and obtained her BSc degree in software engineering from Azad University of Tabriz, Iran. She completed her MSc degree in multimedia and Internet computing from Loughborough University in September 2009. She started her PhD studies in December 2009. Recently she completed her doctoral thesis on “Use of coherent point drift in computer vision applications.”

**Eran A. Edirisinghe** obtained his BScEng (hons) degree from University of Moratuwa, Sri Lanka, in 1994. He completed his MSc and PhD degrees at Loughborough University in 1996 and 1999, respectively. He joined Loughborough University as a lecturer in computer science in July 2000 and was promoted to a senior lecturer in 2004. He was awarded the title of Reader in Digital Imaging, Loughborough University in April 2008. He was appointed as the head of the Department of Computer Science, Loughborough University in August 2011 and was promoted to a chair in July 2012. His research has been funded by the Engineering and Physical Science Research Council (EPSRC), Technology Strategy Board (TSB) and industry. He has also acted as a reviewer for many esteemed journals, international grant awarding bodies, and on various advisory roles to industry and public sector organizations.