Full high-dynamic range images for dynamic scenes

Ramirez Orozco R. [a], Martin I. [a], Loscos C. [c], and Vasquez P.-P. [d]

[a]Universitat de Girona, Girona, Spain
[b]University of Reims Champagne-Ardenne, Reims, France
[c]Universitat Politectica de Catalunya, Barcelona, Spain

ABSTRACT

The limited dynamic range of digital images can be extended by composing photographs of the same scene taken with the same camera at the same view point at different exposure times. This is a standard procedure for static scenes but a challenging task for dynamic ones. Several methods have been presented but few recover high dynamic range within moving areas. We present a method to recover full high dynamic range (HDR) images from dynamic scenes, even in moving regions. Our method has 3 steps. Firstly, areas affected by motion are detected to generate a ghost mask. Secondly, we register dynamic objects over a reference image (the best exposed image in the input sequence). Thirdly, we combine the registered input photographs to recover HDR values in a whole image using a weighted average function. Once matching is found, the assembling step guarantees that all aligned pixels will contribute to the final result, including dynamic content. Tests were made on more than 20 sets of sequences, with moving cars or pedestrians and different background. Our results show that Image Mapping Function approach detects best motion regions while Normalized Cross Correlation offers the best deal speed-accuracy for image registration. Results from our method offers better result when moving object are roughly rigid and their movement is mostly rigid. The final composition is an HDR image with no ghosting and all dynamic content present in HDR values.

Keywords: Computational Photography, High Dynamic Range Imaging, Movement Detection, Image Registration

1. INTRODUCTION

Computational photography is a relatively recent field of research. It is dedicated to overcome the limitations and to enhance conventional photography by mean of computational techniques.[1] High dynamic range imaging (HDRI) is a branch of computational photography aiming at enlarging the range of intensities represented in images between the smallest and the largest illumination value (also called dynamic range).

Auto exposure control algorithms present in most digital cameras determine the correct exposure for covering the light intensity range of a given scene. When the amount of energy that reaches the sensor exceeds the range of values allowed, bright areas appear over-exposed. On the contrary, if not enough energy reaches the sensor under-exposure takes place. Figure 1(b-f) shows an example of a sequence with over and under exposed areas.

HDR image generation aims to minimize over or under-exposure producing high quality irradiance maps. HDR formats can store at least 16 bits per color channel but most displays or printers allow to represent 8 bits per color channel. There are two choices to visualize an HDR images, either in a specifically built HDR display[2–4] or more commonly adapting them to LDR displays using tone mapping.[5] Figure 1a shows an example of a tone mapped HDR image. An HDR image can be captured using HDR sensors like Kodak KAC-9628, IMS Chips HDRC sensors, Silicon Vision Products, SMaL Camera or Pixim, but today these technologies are still fairly expensive[6] preventing them from being available to a large number. On the other hand, solutions for combining images with different exposures of the same scene into an HDR image[7] are already available in tools like Adobe Photoshop [*], HDRShop [†] or Photomatix [‡].

Further author information: (Send correspondence to)
Ramirez Orozco R.: rramirez@ima.udg.edu
Loscos C.: celine.loscos@univ-reims.fr
[*]http://www.adobe.com/es/products/photoshop.html
[†]http://www.hdrshop.com
[‡]http://www.hdrsoft.com

(a) Tone mapped HDR image

(b) t=1/400 EV=-2   (c) t=1/200 EV=-1   (d) t=1/100 EV=0   (e) t=1/50 EV=1   (f) t=1/25 EV=2

Figure 1: Set of low dynamic range images (LDR) with its relatives exposure times (b-f) and a tone mapped representation of the resultant HDR image (a). Captured on a NIKON D200 camera with auto bracketing function and aperture f/11.

Reconstructing HDR images from a set of LDR images works properly for static scenes like in Figure 1, but it is important to consider that capturing a set of LDR images takes at least the sum of their exposure times. This is enough to introduce differences in images if there are any dynamic objects in the scene. If a set of images from a dynamic scene are combined using conventional methods,[7–9] pixels from dynamic objects merge with the background, producing visible "ghosting" effects in the resulting image. Dealing with movement represents one the main challenges of HDR reconstruction.

There are several techniques that cope with this problem, but most of them propose either to exclude dynamic objects from the HDR image[10–12] or to replace the affected area with the best exposed LDR values.[13–18] In the first case, the result is not coherent with the scene because the dynamic objects are missing in the final image. In the second case there are over and under exposed areas in the HDR image because areas affected by movement are replaced with LDR content from the best exposed image only. There are also approaches for tracking dynamic objects through the sequence[19–21] but their results are still limited.

This work presents a technique to allow dynamic objects in HDR reconstruction by gathering HDR information of dynamic objects from a set of LDR images. Dynamic objects are not removed from the resulting image and areas affected by movement contain higher dynamic range than in previous works. Once regions in

movement are identified, each dynamic objects is registered to its position in a reference image. We therefore augment the dynamic range of values by combining registered images, allowing HDR values even in areas affected by movement. The best results are achieved in dynamic scenes where the objects in movement are roughly rigid as well as their movement in the image sequence. It means objects that don't change their shape considerably during the sequence (cars, motorbikes, planes) and that their motion trajectory can be averaged by a set of translations.

The remaining of this document is structured in four sections. Section 2 describes the state of the art of HDR generation. We describe advances in different steps of the HDR reconstruction process found in literature. In section 3 we make a detailed explanation of the different implemented techniques found in the literature and we present our solution for generating HDR images from dynamic scenes. In section 4 we discussion of the obtained results and finally, in section 5 we conclude and propose possible future improvements.

2. PREVIOUS WORK

In this section, we review each of the steps necessary to build a HDR image from differently exposed LDR images: acquisition, alignment, color adjustment, movement detection and management, final HDR generation.

2.1 LDR set acquisition

HDR images store wider range of color intensity than LDR images, thus providing more information from the scene and minimizing the risk of over or under-saturated areas.[5] If HDR images are acquired or built in this intention, every pixel has a physical meaning proportional to the scene irradiance. The mapping from color value to irradiance can be done with a non linear function, also called the inverse response curve of the camera (see section 2.3), or a linear function if data was acquired in RAW format. HDR images can be stored in RGB extended formats like RGBE, HDR or OpenEXR.[5] These formats store RGB color values using from 16 bits to 32 bits per color channel, enlarging the conventional RGB format used for LDR images.[5]

Complete reviews of existing techniques have been published.[5,22] They provide an excellent introduction to HDR imaging and analyze many important issues in this field.

Exposure value (EV) relates aperture and shutter speed, the two parameters conditioning the amount of light that reaches the sensor. LDR pictures for HDR generation are usually taken sequentially at regular EV intervals to ensure that each image contains useful information of different parts of the scene. Aperture is fixed in order to avoid optical effects due to aperture changes (like depth of field variations). White balance and the focus are also kept constant. An example of an LDR sequence taken for HDR generation is shown in Figure 1 (b-f) where t is the exposure time in seconds.

The auto-bracketing function available in many cameras is very useful to capture a set of LDR images at regular EV intervals. The camera automatically calculates the best aperture/shutter speed combination EV_0 for the given lighting conditions to minimize over or under-exposure. Once EV_0 is established, the camera captures darker and brighter pictures respectively increasing and decreasing the shutter speed.

2.2 LDRs Alignment

Every pixel in the same coordinate through the sequence should correspond to the same object in the scene to guarantee the coherence of the HDR image. During the capture process, misalignment happens often even using a tripod. This leads to unexpected artifacts like ghosting and blur in the reconstructed HDR image.

Ward[23] presented Median Bitmap Transform (MBT) to align a sequence of images. The algorithm transforms a set of 8 bit images to a bitmap, using the median intensity value as threshold. The median intensity value is insensitive to exposure variations. The difference of two bitmaps is defined as a logical XOR and it shows where the images are misaligned. Alignment is implemented iteratively to minimize differences respect to a reference image. MTB works properly for images with a rather bimodal brightness distribution. Different implementations and variations of this technique have been used in HDR.[13,14,17,24] This technique does not depend on the camera response function (see section 2.3) and its computational cost is very low.

Markowski[12] presented a GPU based approach to solve alignment using a bidirectional s-shape comparison function. The parameters of the function were tuned by trial and error. Another approach is based on scale

invariant feature transform (SIFT).[18, 25] A modified SIFT algorithm is used to extract key point descriptors that represent correspondences between key points in the reference image and the remaining images. After finding SIFT features, homographies are calculated using RANSAC algorithm to match all images and a reference image previously selected.

Oǧuz *et al.*[26] assume that misalignment between consecutive images is translational and the correlation between pixels tends to remain constant. The algorithm searches the most similar correlation map in consecutive images using Hamming distance as similarity measure. The main limitation of this method is that it can't solve rotation neither other more complex transformations.

2.3 Inverse Camera Response Function

A nonlinear mapping function converts irradiance values into pixel colors.[8] The most significant nonlinearities occur at the saturation point, where any pixel above or below a certain level is mapped to the same maximum or minimum image value. This is called camera response function (CRF) and its inverse is required to convert from pixel intensities I to irradiance E captured by the camera sensor.

Under the assumption that the CRF f is monotonically increasing, the existence of $f^{-1}(I)$ of Equation 1 is guaranteed. Several approaches[7–9] attempt to obtain $g(I)$.

$$E \cdot \Delta t = f^{-1}(I) = g(I) \tag{1}$$

Mann and Picard[7] proposed an automatic method to get $g(I)$. It consists of fitting the pixel values to an empirical function with gamma shape $g = \alpha + \beta I^\gamma$, where α is the minimal density obtained from a picture taken with lens covered, β is an arbitrary scale factor and γ is a contrast parameter estimated by regression. This method is highly restrictive so it does not lead to accurate results and does not support most real CRFs.[22]

Debevec and Malik[8] take the natural logarithm on both sides of the Equation 1 to approximate the CRF. They calculate both f^{-1} and E minimizing the error by mean of the least square method. This approach is less restrictive than Mann's and obtains good results for images that are not too noisy.[9]Mitsunaga and Nayar[9] presented an approach that does not require precise estimates of the exposure times. They improve the previous approach using a flexible polynomial model for representing a wide range of response functions.

Grossberg and Nayar[27] recovered the Image Mapping Function (IMF) that relates pixel intensities through the sequence. They demonstrated that having the histograms is enough to recover the CRF in sequences with both movement in camera and scene while the histograms remains roughly constant. Other works attempt to obtain HDR information using only one LDR image[28, 29] extrapolating information in under and over-exposed, but this is not enough for highly saturated or totally dark areas.

2.4 Ghosting detection and removal

Objects move in a scene, ghosting appears in the HDR image because information is merged from misaligned pixels. Loscos and Jacobs [30] classified in a recent overview approaches proposed to manage movement occurring during HDR acquisition. A combination between feature matching and optical flow is proposed[21] for HDR video sequence matching. This method is robust to changes in exposure and lighting, but if there are objects moving at high speed, artifacts may still appear. Bogoni[19] also used optical flow to perform per pixel registration applied after global registration. Kang *et al.*[20] proposed a procedure that handles both camera and objects movement. They captured a video sequence alternating long and short exposure times. Adjacent frames are warped and registered to finally make an HDR composition. This method is not suitable for big differences between frames and knowing the CRF in advance is required.

Some approaches focused on reconstructing the background omitting dynamic objects from the HDR image. Khan *et al.*[10] proposed a probabilistic method for weighting pixels without any explicit movement detection. This method is computationally expensive because it requires several iterations and for complex scenes, artifacts still persist. Pedone *et al.*[31] proposed a similar approach which requires less iterations. Granados *et al.*[11] presented a method for background estimation of a scene from a set of images that can be also applied to HDRI generation. It is based in two assumptions: background objects are static and they represent the major part

of the image. The output is a composition of background pixels obtained by minimizing a cost function of all possible compositions. A GPU-based application was presented by Markowski[12] which uses probabilities to automatically detect ghosting. All these methods have a different degree of success but all of them have the same drawback: dynamic objects are omitted in the final HDR image.

Variance of irradiance across different exposed images is considered an indicator of movement.[5,14] A variance map is created storing the weighted variance of pixels over different exposures. Movement clusters are detected applying a threshold over the variance map. Morphological erosion and dilation[14] or flood filling[5] have been also applied to obtain more defined cluster areas. This technique has been used in HDR frameworks and softwares like Photosphere.[17] Another solution is to measure uncertainty using entropy as an indicator of potential movements. Local entropy at each pixel location in a neighborhood of radius five is computed, which generates a map denoted as the uncertainty image UI. Entropy or variance are not affected by intensity values but both fail when dynamic range is quite big.[17]

Pece and Kauts[17] proposed an algorithm to detect and isolate clusters of dynamic pixels in a sequence. Their method applies MTB[23] for each LDR and calculates differences between them. Grosch[13] also used MTB but includes also rotations. Gallo *et al.*[15] detect region patches which do not cause artifacts when combined with a reference image. The HDR image is generated using only these patches. This method assumes that pixels from consecutive images have a linear relation and applies a threshold over the deviation of pixels from the linear model. The solution proposed by Raman *et al.*[32,33] detects ghosting using block based comparison between exposures. Their method also applies a threshold to differences respect to a predicted value. The CRF to generate predicted values is estimated through a sixth order polynomial. If the inverse camera response function is monotonic increasing, then the pixels that break this relation pertain to regions of movement or another unexpected variation of intensity over the sequence. Sidive[16] select such pixels and classify them as movement which is truth under certain conditions, but it is not robust. Grossberg and Nayar[27] proposed the intensity mapping function (IMF) which can be used to predict pixels. The main contribution of this method is that alignment can be avoided because small scene movement does not change the histogram significantly. This is also valid for scene with moving objects if histograms remain approximately constant in the sequence. Zhengguo *et al.*[34] used the IMF to detect moving objects forward and backward in the sequence using a threshold. Heo *et al.*[18] calculated a joint probability density function between the reference image and the rest of LDRs to estimate the global intensity transfer functions.

2.5 HDR Generation

The last step in HDR generation corresponds to get HDR values for every pixel. Results are more precise if LDR images intensity values are transformed to irradiance values using the CRF. Irradiance values are combined using a weighted average function:[7–9]

$$E(i,j) = \frac{\sum_{n=1}^{N} w(I_n(i,j))(\frac{f^{-1}(I_n(i,j))}{\Delta t_n})}{\sum_{n=1}^{N} w(I_n(i,j))} \qquad (2)$$

where $E(i,j)$ is the irradiance at location (i,j) in the HDR image, N is the total number of exposures, $I_n(i,j)$ is the value of the pixel at location (i,j) in the nth exposure and Δt_n is the exposure time of the nth exposure. f^{-1} is the CRF associated to the used camera. w is the weighting function used to minimize the contribution of under and over exposed pixels. Typical weighting functions have Gaussian or Hat shape. Mann and Picard[7] used the derivative of the CRF for each color channel as weighting function. Devebec and Malik[8] and Khan *et al.*[10] used simple hat functions.

2.6 Discussion

There is a fair amount of methods that were proposed to build HDR images from non HDR camera. Movement is an important issue for which solutions were proposed but none exist that build full HDR information regardless of the type of movement that may have occur. In the following, we describe different combinations of methods we tried to create full HDR content even in areas in movement.

3. RECOVERING FULL HDR IMAGES

The aim of this paper is to provide HDR values even in areas in movement during acquisition of the multiple exposure LDR images. No assumption is made on the movement. It can appear at different regions of the image. However, we do assume that objects in motion are mostly rigid. Moreover movement should concern well defined areas of the image rather than the full image. In other words, we assume that pixels belonging to static objects and background were aligned for all input images.

We studied how previous approaches could be adapted and paired in order to reach our goal. We consider three different steps like shown in Figure 2.

1. Ghost Detection: Four of the most used methods for ghost detection have been implemented and compared achieving different degrees of success. The result of this step is a mask that contains pixels affected by movement.

2. Registration: The second step is registering regions affected by movement to a reference image. Four similarity measures for image registration have been implemented and compared. Except for MTB, it is the first time that normalized cross correlation, mutual information and sum of squared difference was applied for HDR image acquisition. An image pyramid is implemented to speed up the registration.

3. HDR compositing and assembling: Finally, the HDR image is composed using a weighted average function. Pixels from dynamic areas are combined after registration which provide HDR information for such areas. Ghosting areas are replaced with the obtained HDR values.

We studied different combinations of approaches chosen for each step. We detail in the following the implementation of chosen methods that were found appropriate.
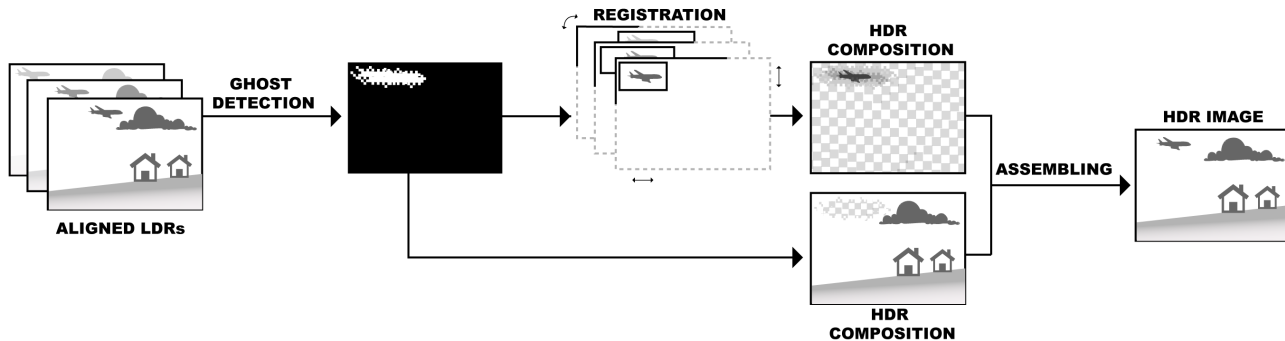

Figure 2:   Sequence of steps to achieve a full HDR image in dynamic scenes.

We implemented four well-known methods : Median Threshold Bitmap-based, variance-based, CRF-based, and IMF-based, and compare them when generating a motion mask. This mask represents parts of the image where movement occurred.

3.1 Ghosting mask generation

3.1.1 Median Threshold Bitmap

The MTB algorithm[23] computes the median value of pixels intensity for each image i and uses it as a threshold to produce bitmaps B_i. Pece[17] proposed a simple Bitmap Movement Difference(BMD) to get the dynamic pixels from a sequence. If there is not movement in the scene, each pixel is expected to have the same value in all B_i bitmaps. The difference between all bitmaps can be calculated like a logical XOR or summing all values like in Equation 3 and selecting pixels that are ($M \neq 0$ & $M \neq N$) where N is the number of images. The result is a bitmap containing dynamic objects but also some noise that could lead to artifacts in the ghost mask. Morphological operations of erosion and dilation help to eliminate such noise.

$$M = \sum_{n=1}^{N} B_i \qquad (3)$$

3.1.2 Variance-based methods

Weighted variance methods[5, 14] assume that pixels from areas affected by movement have a higher variance across the sequence. This method computes the variance of pixel intensities over the different exposed images using Equation 4 and store them in a variance image (VI). A weighting function W_k is used to minimize the influence of under- and over-exposed values, we have implemented the hat function proposed by Khan $et\ al.$[10] which is the same used in the final HDR composition. High variance values are selected from the VI applying a threshold to obtain a binary image of dynamic pixels. Some high variant pixels remain in the binary image that corresponds to noise or very small movements. Morphological erosion and dilation help to define the final mask.

$$VI(i,j) = \frac{\sum\limits_{k=0}^{N} W_k(i,j)E_k(i,j)^2 / \sum\limits_{k=0}^{N} W_k(i,j)}{(\sum\limits_{k=0}^{N} W_k(i,j)E_k(i,j))^2 / (\sum\limits_{k=0}^{N} W_k(i,j))^2} \qquad (4)$$

3.1.3 Methods based on the CRF

Sidibe $et\ al.$[16] proposed an approach based on the assumption that the CRF is monotonically increasing. Pixels from consecutive images must be related by the same relation as their exposure times. In a sequence of N exposures $\forall k \in [1...N]$ if the exposure times $\Delta t_k < \Delta t'_k$, it is possible to assume that the following relation (5) is valid for any pixel in (i,j)

$$I_{(i,j),k} \leq I_{(i,j),k'} \qquad (5)$$

The LDR images are organized increasingly by their exposure times and pixels that break the relation of Equation 5 are selected. This method not only detects movements but also the unexpected variation of a pixel color.[16] Gallo $et\ al.$[15] improve this result assuming a linear relation $y = x + ln(EV)$ between images and using a threshold to select pixels far from this line.

3.1.4 Method based on the IMF

Instead of a linear relation, we tried the IMF proposed by Grossberg $et\ al.$,[27] which does not require to recover the CRF and is very accurate. The IMF is deduced from properties of the cumulative histogram $H()$. The area below the curve for an image with intensities in the range $[0, X]$ is given by $H(X)$, being h the continuous histogram:

$$H(X) = \int_0^X h(u)du$$

The mapping function $I_2 = \tau(I_1)$ between two images is given by:

$$\tau(I_1) = H_2^{-1}(H_1(I_1)) \qquad (6)$$

Differences between consecutive images are calculated and combined with a logical XOR in a ghosting mask that contains pixels affected by motion through the sequence. We also apply morphological dilation and erosion to isolate motion clusters from noise introduced by under or over-exposure.

3.2 Image Registration

Once the ghosting mask is selected, we get the area of interest that contains dynamic pixels in each image. The best exposed image is used as reference for the registration step like described in the next section.

Image registration is an intense research field. Registration is the process of matching two or more images of the same scene taken under different conditions (sensor, time, viewpoint or optical settings). During registration one image is taken as reference and the rest (target images) are transformed until a matching is found.[35] Surveys classifying and analyzing several techniques were presented in.[35-38] Registration techniques have been traditionally used in remote sensing, medical images, cartography or computer vision. Most registration methods can be classified in two main groups:

- **Feature-based methods** use salient structures (borders, lines or points) spread all over the image, recognizable in both images and stable in time. Some of these methods have been used for alignment in HDR reconstruction and were mentioned in section 2.2. However, dynamic objects in a sequence may not be well defined due to large exposure times and movement. This makes the feature detection difficult through the sequence.

- **Intensity based methods** attempt to match images without any feature detection, matching directly pixels intensities without feature selection. They define a measure of similarity between the source and the target and adjust transformations until the similarity measure is optimal.

We implemented intensity based registration using four similarity measures in order to analyze the results. From the ghost detection step we get a sequence of sub images containing dynamic objects and a ghost mask. These sub images can still be reduced. The partial differences between pairs of them are smaller than the whole affected area. Differences between consecutive images are calculated generating partial ghost masks and sub images are cropped according to the partial masks like shown in Figure 3.
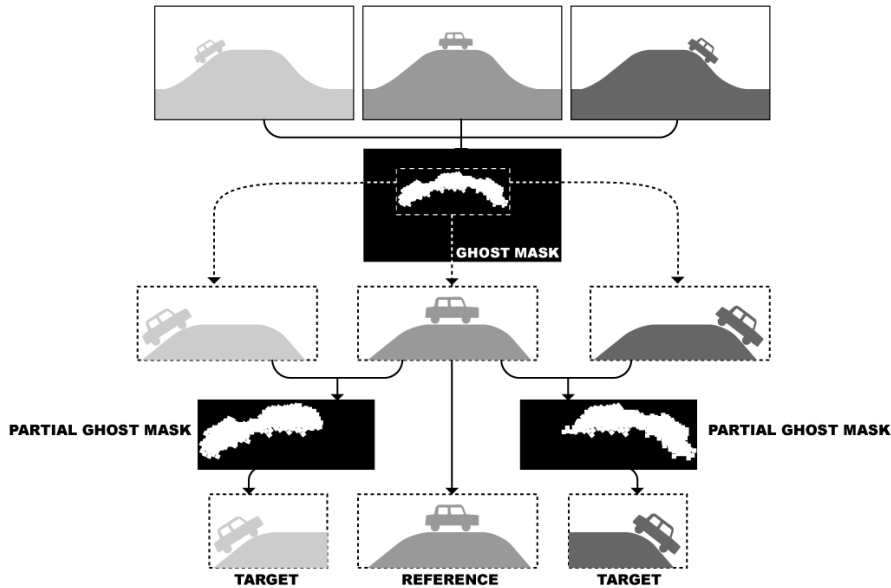

Figure 3: Defining target images from the ghost mask.

All target images will be registered over the reference image. Registration could be highly time consuming, the use of a pyramid of images helps to speed it up. In each level images are sampled down by a factor of two, like shown in Figure 4. Using the images from last level, the target images are translated and rotated iteratively

over the whole reference image evaluating the similarity measure in each iteration. Once we have found the best match for the lowest level of the pyramid, we climb to previous levels. Then we check only for transformations in an offset around the match point obtained in the previous step. The size of the offset depends on the size of target images but usually searching in a vicinity of 10 pixels around the previous matching position gives good results.
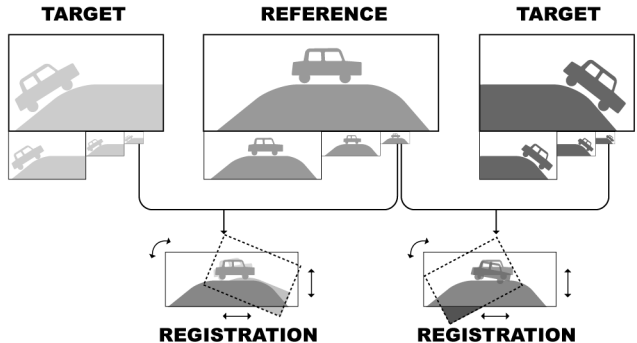

Figure 4: Image pyramid registration

The following similarity measures were implemented and the results are discussed in the next section.

Sum of Squared Difference (SSD): is the simplest and more intuitive way of measuring similarity.[39–41] The minimum value of SSD corresponds to the transformation that better match the images.

$$SSD = \sum_{n=1}^{N}(A_n - T(B_n))^2$$

However there are some problems using this criterion as a measure for HDR registration because it is not invariant to changes in lighting conditions across the image sequence[42]

Normalized Cross Correlation (NCC): assumes that corresponding intensities in the images have a linear relationship.[43] This metric has been used in images taken with the same device at different times.[44] We implemented it using an approach presented by Lewis.[42] The best matching corresponds to the transformation that maximizes the NCC value.

$$NCC = \frac{\sum\limits_{n=1}^{N} A_n \cdot T(B_n)}{\sqrt{\sum\limits_{n=1}^{N} A_n^2 \cdot \sum\limits_{n=1}^{N} T(B_n)^2}}$$

Mutual Information: is a measure of statistical dependence between two variables.[44] Since Viola *et al.*,[45] several papers have been presented mainly for image registration either minimizing joint entropy or maximizing mutual information. Most of them use a pyramidal approach to speed up the registration process.[35] We use Normalized Mutual Information as similarity measure.

$$H(A, B) = \frac{H(A) + H(B)}{H(A, B)}$$

Median Bitmap Difference: allows to find the correct transformation for registration if we find the transformation that minimizes such difference. This approach was presented by Grosch.[13]

3.3 HDR composition

The ghost mask allows to classify pixels in static or dynamic ones. The static pixels can be composed in an HDR image using a weighted average[7–9] defined in Equation 7 and the weighting function of Equation 8 proposed by Khan et al.[10]

$$E(i,j) = \frac{\sum_{n=1}^{N} w(I_n(i,j))(\frac{f^{-1}(I_n(i,j))}{\Delta t_n})}{\sum_{n=1}^{N} w(I_n(i,j))} \quad (7) \qquad\qquad w(I) = 1 - (2\frac{I}{255} - 1)^{12} \qquad (8)$$

In each step we exclude the pixels affected by movement from the compositing function. Pece and Kautz[17] suggested to replace all pixels in the ghost mask by the best exposed LDR. The ghost mask contains pixels from dynamic objects all over the sequence, but usually movement does not affect more than two or three consecutive images. Gallo et al.[15] calculated partial ghost mask for each pair of consecutive images. This ensure that only pixels from dynamic objects are excluded in each LDR image. The result of combining static pixels from all LDRs is an HDR image with LDR values only in pixels affected by movement. Such pixels will be replaced with HDR values recovered for dynamic regions.

The dynamic pixels after the registration step are aligned and ready to be composed using the same Equation 7. To prevent from artifacts produced by small misalignment, in the HDR composition we only consider pixels that are correctly aligned. To distinguish such pixels, we calculate the difference between aligned images using IMF threshold and exclude the differences. The result is an HDR sub-image of the area affected by movement. Recovered HDR values from dynamic objects are replaced in the HDR image producing a coherent HDR final result.

4. RESULTS AND DISCUSSION

Several tests were made using more than 20 sets of LDR images with different exposures. All images were captured using a tripod and auto bracketing, and keeping constant all parameters except the shutter speed. Results vary depending on the ghost detection technique, the registration similarity measure and the threshold used in each case. The following is dedicated to analyze results from each step and each method.
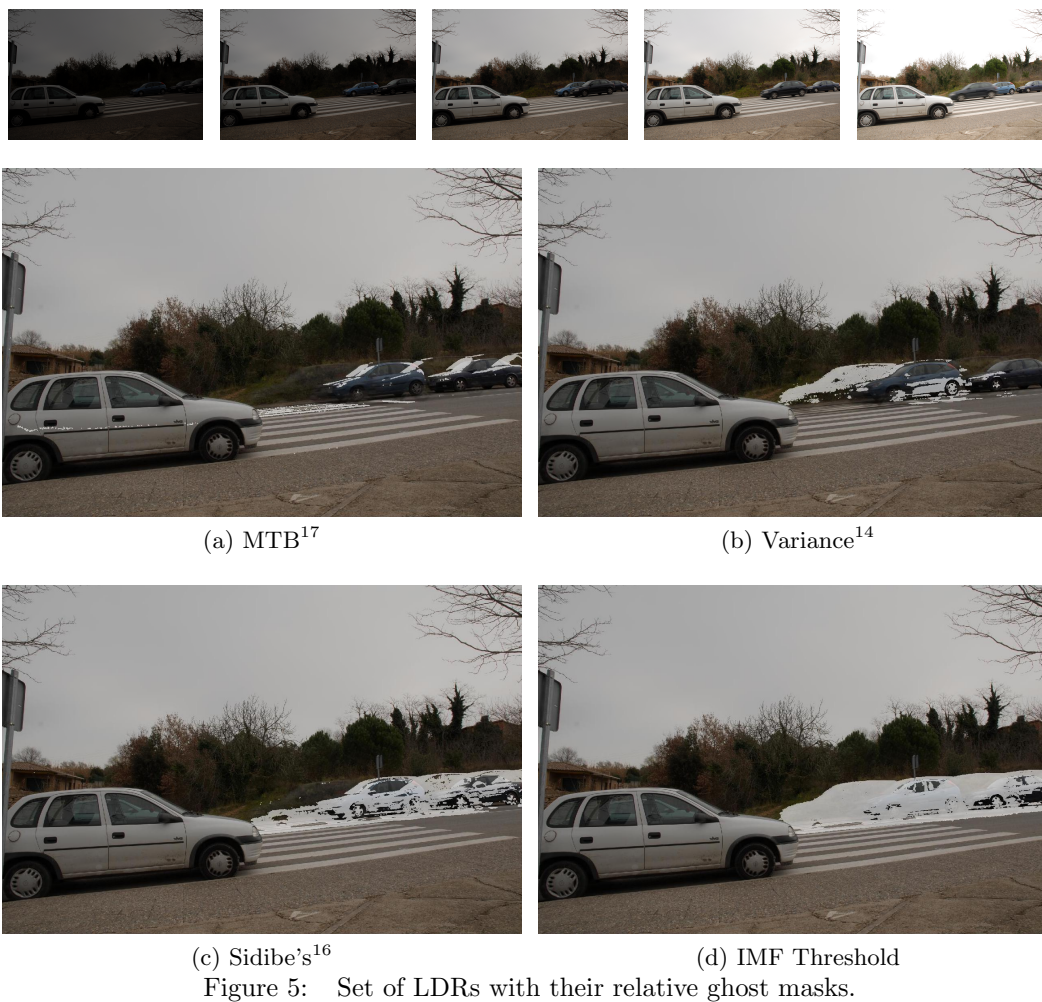
4.1 Ghost Detection

In Figure 5, five LDR images from a scene of a black car moving in a street with trees and stationed cars (also of dark colors) in the background. It shows the result of ghost detection using the four implemented methods.

The MBD method works properly for scenes with a rather bimodal brightness distribution. It is also very fast and easy to implement. If the dynamic object and the background are both smaller or bigger than the median value of intensities, it fails. In Figure 5a, only the front glass are detected because the rest is very similar to the background.

Variance method may fail in zones where the dynamic range is too big (highlights, shining objects, direct sun light) or when the movement is too slow and produces overlapping.[17] The success of using variance in ghost detection also depends on the relation between colors of the dynamic object and the background. If the color of the dynamic object is similar to the background the method fails.[16] Figure 5b shows an example where the variance method fails because dynamic pixels are very similar to the background. The threshold value is also an issue to take into account since the results directly depend on it.

The method proposed by Sidibe[16] is very simple and does not depend on threshold values. It is based on the assumption that pixel intensities increase through the sequence, which is not always true in dynamic scenes. Any situation that breaks such assumption is detected as movement, like when the black car is moving in front of a dark background in Figure 5c.

The most robust method in our tests is the one based on the IMF. Getting the IMF allows to predict values between pair of images and compare predicted values with the actual ones. The results of this method are the best in all tested cases. Nevertheless, selecting well the threshold value is very important to succeed using this method.

(a) MTB[17]

(b) Variance[14]

(c) Sidibe's[16]

(d) IMF Threshold

Figure 5:   Set of LDRs with their relative ghost masks.

4.2 Image Registration

The results of image registration highly depend on the previous step. The target images are selected from the bounding box of pixels affected by movement, like shown in Figure 3. If the ghost detection is not correct, bounding boxes do not correspond to dynamic objects and the registration step may fail too. The goal of the registration algorithm is to find the optimal value (minimum or maximum depending on the measure) of the similarity measure.

Figures 6 shows an example of registration of two images from a sequence and the results obtained with the four implemented measures presented in section 3.2.

(a) Reference

(b) Target

(c) SSD

(d) MBD

(e) NCC

(f) MI

Figure 6:   Reference and target of image 5 from a scene of 5 LDR images with a car moving through

Even when SSD finds a correct matching, in some cases it is not appropriate for registering images with very different exposures. The distance between images is directly affected by over- and under-exposed values. The MBD method is suitable for most of cases. Its implementation is very simple and it is the fastest of the tested methods. However, similarity to variance for ghost detection, if the background and the object have similar colors, MBD can fail.

MI is an statistical measure, hence it is not affected by intensity changes. Most of results obtained using it are good but it depends directly on the entropy of images which can be perturbed by over or under saturated pixels. It is also the slower method among the implemented. We obtained the best relation cost/performance when using NCC. It is insensitive to intensity changes which makes it an strong measure for images with different exposures.

Our registration method only evaluates rotations and translations in the image space. It is possible to have other transformations in real scenes like scaling, shearing or more complex deformations due to perspective or rotations in object space.

4.3 HDR composition

In most previous work, dynamic areas were either removed reconstructing the background or replaced by LDR content, producing incoherence within the scene or LDR areas in the final result. Our results are consistent in terms of the dynamic range recovered in the final HDR image.

Figure 7 shows in false color the amount of pixels that finally contribute to an HDR value. Notice that for most pixels in the dynamic areas HDR values result of combining at least 3 LDR values.
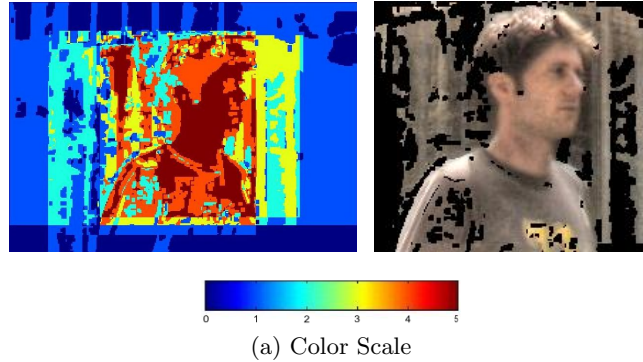
(a) Color Scale

Figure 7:   HDR reconstruction of dynamic objects.

Some artifacts may persist in the border of regions obtained from the combination of different number of LDRs. Even when the ghost detection works reasonably well, the ghost mask has no defined borders. Small holes or discontinuous borders make it difficult to get well defined masks of dynamic objects to clearly separate dynamic from static pixels.
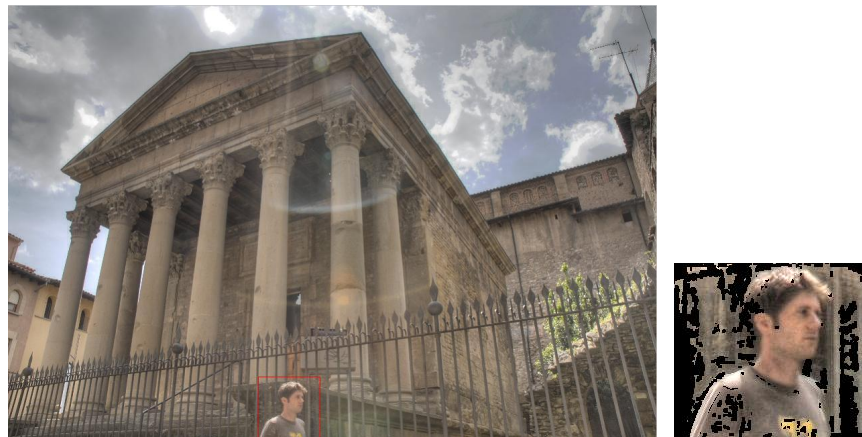

Figure 8:   Whole HDR image from a dynamic scene.

5. CONCLUSIONS AND FUTURE WORK

We presented a method for compositing HDR images of dynamic scenes. Our method detects areas affected by movement, matches them in a reference image and recovers HDR values from such areas in a sequence of LDR images. Promising results were obtained for scenes where dynamic objects were roughly rigid. We implemented some state-of-the-art algorithms according to descriptions given by their authors. The selected algorithms were developed in Matlab and a GUI was implemented for supporting the tests. The algorithms were tested using several set of images from different scenes.

Regarding the ghost detection, we implemented four approaches obtaining different degrees of success. Even when all implemented techniques produce good results for some kind of scenes, the best results in general are obtained using a threshold over differences of pixels predicted with the IMF and the actual values. It is very important step since all the process lies on its results. Any improvement of this step will have a positive impact in the final result.

We applied registration techniques to the HDR reconstruction problem. When the clusters of pixels affected by movement in the ghost mask are closed and well-defined areas, the selection of the target images works properly. The registration step shows the best results when using MI or NCC being the last one faster than

MI. The most important improvement that could be done to our work is during the registration step. The obtained results are good mostly for roughly rigid dynamic objects. For deformable objects it may fail, this can be addressed in future work subdividing the dynamic objects and match them by parts.

The values from LDR images are combined into a full HDR image. After the registration step, HDR values are recovered for areas affected by movement. These values are replaced in the HDR image. The result is a consistent HDR image with less LDR values than in any previous work. Some artifacts may be still introduced in this step because of artifacts in the ghost mask. Neighbor areas resulting of combining different mount of LDRs may contain visible borders. To avoid such problems, the blending technique presented by Gallo *et al.*[15] could be tested.

ACKNOWLEDGMENTS

REFERENCES

[1] Lukac, R., [*Computational Photography: Methods and Applications*], Digital Imaging and Computer Vision, Taylor & Francis Group (2010).

[2] McLaughlin, C., "High dynamic range displays." Online (June 2007).

[3] Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., Ghosh, A., and Vorozcovs, A., "High dynamic range display systems," in [*Proc. of SIGGRAPH '04 (Special issue of ACM Transactions on Graphics)*], (Aug. 2004).

[4] Dolby., "Dolbys high-dynamic-range technologies: Breakthrough tv viewing," (February 2012).

[5] Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P., [*High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*], The Morgan Kaufmann Series in Computer Graphics, Morgan Kaufmann Publishers Inc (August 2005).

[6] Cerman, L., *High Dynamic Range Images from Multiple Exposures*, Master's thesis, Czech Technical University in Prague (2006).

[7] Mann, S. and Picard, R. W., "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in [*Proceedings of ISandT*], 442–448 (1995).

[8] Debevec, P. and Malik, J., "Recovering high dynamic range radiance maps from photographs," in [*In proceedings of ACM Siggraph (Computer Graphics)*], **31**, 369–378 (1997).

[9] Mitsunaga, T. and Nayar, S., "Radiometric self calibration," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], **1**, 2 vol. (xxiii+637+663) (1999).

[10] Khan, E., Akyuz, A., and Reinhard, E., "Ghost removal in high dynamic range images," in [*IEEE International Conference on Image Processing*], 2005 –2008 (oct. 2006).

[11] Granados, M., Seidel, H.-P., and Lensch, H. P. A., "Background estimation from non-time sequence images," in [*Graphics Interface*], 33–40 (2008).

[12] Markowski, M., "Ghost removal in hdri acquisition," in [*Central European Seminar on Computer Graphics*], (2009).

[13] Grosch, T., "Fast and robust high dynamic range image generation with camera and object movement," in [*Vision, Modeling and Visualization, RWTH Aachen*], 277–284 (2006).

[14] Jacobs, K., Loscos, C., , and Ward, G., "Automatic high-dynamic range generation for dynamic scenes," *IEEE Computer Graphics and Applications* **28**, 24–33 (March-April 2008).

[15] Gallo, O., Gelfand, N., Chen, W.-C., Tico, M., and Pulli, K., "Artifact-free high dynamic range imaging," *IEEE International Conference on Computational Photography (ICCP)* (April 2009).

[16] Sidibe, D., Puech, W., and Strauss, O., "Ghost detection and removal in high dynamic range images," in [*17th European Signal Processing Conference. EUSIPCO 2009*], HAL - CCSD, Glasgow, Scotland (August 2009).

[17] Pece, F. and Kautz, J., "Bitmap movement detection: HDR for dynamic scenes," in [*Conference for Visual Media Production 2010 (CVMP 2010)*], (2010).

[18] Heo, Y. S., Lee, K. M., Lee, S. U., Moon, Y., and Cha, J., "Ghost-free high dynamic range imaging," in [*Proceedings of the 10th Asian conference on Computer vision - Volume Part IV*], ACCV'10, 486–500, Springer-Verlag, Queenstown, New Zealand (2011).

[19] Bogoni, L., "Extending dynamic range of monochrome and color images through fusion," in [*Pattern Recognition, 2000. Proceedings. 15th International Conference on*], **3**, 7 –12 vol.3 (2000).

[20] Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R., "High dynamic range video," *ACM Trans. Graph.* **22**(3), 319–325 (2003).

[21] Sand, P. and Teller, S. J., "Video matching," *ACM Trans. Graph.* **23**(3), 592–599 (2004).

[22] Banterle, F., Debattista, K., Artusi, A., Pattanaik, S., Myszkowski, K., Ledda, P., and Chalmers, A., "High dynamic range imaging and low dynamic range expansion for generating hdr content," *Computer Graphics Forum* **28**(8), 2343–2367 (2009).

[23] Ward, G., "Fast, robust image registration for compositing high dynamic range photographs from handheld exposures," *JOURNAL OF GRAPHICS TOOLS* **8**, 17–30 (2003).

[24] Guthier, B., Kopf, S., and Effelsberg, W., "Histogram-based image registration for real-time high dynamic range videos," in [*Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*], 145 – 148 (September 2010).

[25] Tomaszewska, A. and Mantiuk, R., "Image registration for multi-exposure high dynamic range image acquisition," in [*Proc. Int'l Conf. Central Europe on Computer Graphics, Visualization, and Computer Vision (WSCG)*], (2007).

[26] Akyüz, A. O., "Photographically guided alignment for hdr images," in [*Proceedings of EUROGRAPHICS 2011*], (April 2011).

[27] Grossberg, M. D. and Nayar, S. K., "Determining the camera response from images: what is knowable?," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**, 1455 – 1467 (nov. 2003).

[28] Wang, L., Wei, L.-Y., Zhou, K., Guo, B., and Shum, H.-Y., "High dynamic range image hallucination," in [*SIGGRAPH '07: ACM SIGGRAPH 2007 sketches*], 72, ACM, New York, NY, USA (2007).

[29] Rempel, A. G., Trentacoste, M., Seetzen, H., Young, H. D., Heidrich, W., Whitehead, L., and Ward, G., "Ldr2hdr: on-the-fly reverse tone mapping of legacy video and photographs," in [*ACM SIGGRAPH 2007 papers*], SIGGRAPH '07, ACM, New York, NY, USA (2007).

[30] Loscos, C. and Jacobs., K., "High-dynamic range imaging for dynamic scenes," in [*Computational Photography, Methods and Applications*], Lukac, R., ed., 259–281, CRC Press (October 2010).

[31] Pedone, M. and Heikkil, J., "Constrain propagation for ghost removal in high dynamic range images," in [*VISAPP (1)*], 36–41 (2008).

[32] Raman, S., Kumar, V., and Chaudhuri, S., "Blind de-ghosting for automatic multi-exposure compositing," in [*ACM SIGGRAPH ASIA 2009 Posters*], SIGGRAPH ASIA '09, 44:1–44:1, ACM, New York, NY, USA (2009).

[33] Raman, S. and Chaudhuri, S., "Reconstruction of high contrast images for dynamic scenes," *The Visual Computer* **27**, 1099–1114 (2011). 10.1007/s00371-011-0653-0.

[34] Li, Z., Rahardja, S., Zhu, Z., Xie, S., and Wu, S., "Movement detection for the synthesis of high dynamic range images," in [*Image Processing (ICIP), 2010 17th IEEE International Conference on*], 3133 –3136 (sept. 2010).

[35] Zitova, B., Flusser, J., and Sroubek, F., "Image registration methods: a survey," in [*Proceedings of the International Conference on Image Processing*], IEEE (September 2005).

[36] Brown, L. G., "A survey of image registration techniques," *ACM Comput. Surv.* **24**, 325–376 (December 1992).

[37] Maintz, J. B. and Viergever, M. A., "A survey of medical image registration.," *Medical Image Analysis* **2**(1), 1–36 (1998).

[38] Wyawahare, M. V., Patil, P. M., and Abhyankar, H. K., "Image registration techniques. an overview," *Int. Journal Signal Proccesing, Image Processing and Pattern Recognition* **2** (2009).

[39] Anuta, P., "Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques," *Geoscience Electronics, IEEE Transactions on* **8**, 353 –368 (oct. 1970).

[40] Svedlow, M., Mcgillem, C., and Anuta, P., "Image registration: Similarity measure and preprocessing method comparisons," *Aerospace and Electronic Systems, IEEE Transactions on* **AES-14**, 141 –150 (jan. 1978).

[41] Rezaie, B. and Srinath, M., "Algorithms for fast image registration," *Aerospace and Electronic Systems, IEEE Transactions on* **AES-20**, 716 –728 (nov. 1984).

[42] Lewis, J. P., "Fast normalized cross-correlation," in [*Vision Interface (1995)*], 120–123 (1995).

[43] Crum, W. R., Hartkens, T., and Hill, D. L. G., "Non-rigid image registration: theory and practice," *British Journal of Radiology* **77 Spec No 2**, S140–53 (2004).

[44] Roshni, V. and Revathy, K., "Using mutual information and cross correlation as metrics for registration of images," *Journal of Theoretical and Applied Information Technology* **4** (June 2008).

[45] Viola, P. and Wells, W.M., I., "Alignment by maximization of mutual information," in [*Computer Vision, 1995. Proceedings., Fifth International Conference on*], 16 –23 (jun 1995).