# Chapter 4
Proposed Framework

Abstract This chapter describes the full photo-mosaicing pipeline proposed in this monograph. This pipeline is intended to process datasets of thousands of images from large-scale underwater optical surveys. The first stages of the process involve the input sequence preprocessing, required to reduce artifacts such as the inhomogeneous lighting of the images, mainly due to the use of limited-power artificial light sources and the phenomenon of light attenuation and scattering. In this step, a context-dependent gradient based image enhancement is proposed, with allows equalizing the appearance of neighboring images when those have been acquired at different depths of with different exposure times. The pipeline follows with the selection of each image contribution to the final mosaic, based on different criteria, such as image quality and acquisition distance. Next, the optimal seam placement for all the images is found. A gradient blending, in a narrow region around the optimally found seam, is applied in order minimize the visibility of the joining regions, as well as to refine the appearance equalization along all the involved images. Finally, a novel strategy allowing to process giga-mosaics composed of tenths of thousands of images in conventional hardware is proposed. The technique divides the whole mosaic in tiles, processing them individually and seamlessly blending all of them again using a technique that requires low computational resources.

Keywords Image preprocessing · Inhomogeneous lighting compensation · Image enhancement · Gradient domain blending · Tone mapping · Giga-mosaicing

A full photo-mosaicing pipeline has been developed, conceived to address the most relevant specific problems of underwater imaging. Nevertheless, the application field of the proposed approach can be extended to the generation of conventional panoramas or maps from terrestrial or aerial images. Figure 4.1 shows the sequence of steps that are performed by our approach, which are intended to build high resolution blended photo-mosaics of the deep-seafloor.

Fig. 4.1 Full processing pipeline of the proposed underwater photo-mosaicing approach. Some of the processing steps can be executed using parallel computing techniques to increase the performance of the algorithm
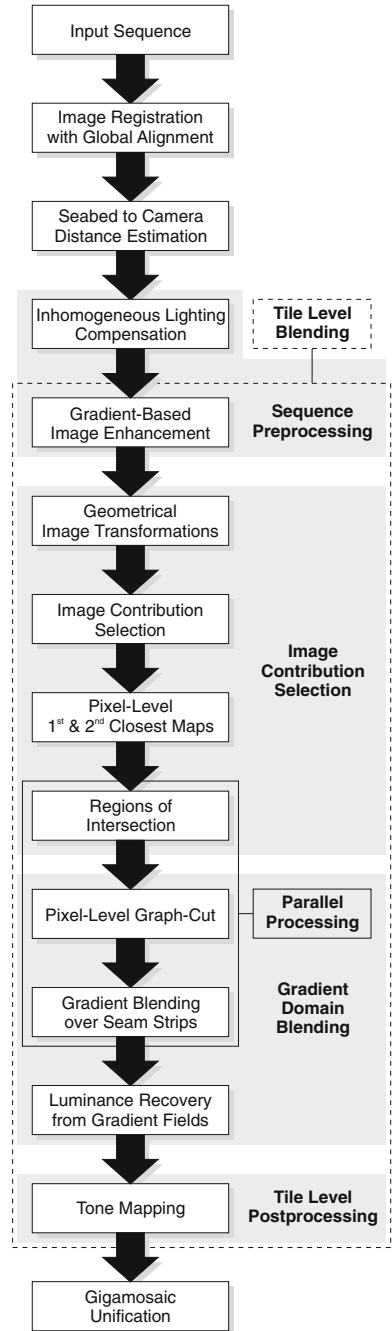
4.1 Input Sequence Preprocessing

Inherent underwater optical imaging problems have already been described in Sect. 1.2. Aside from *exposure variations*, which are a common issue in terrestrial images, other important problems are not directly addressed by conventional panorama generation software. To deal with these, image pre-processing is required, and is becoming a key step with a strong impact on the quality of the final photomosaic rendering.

### *4.1.1 Inhomogeneous Lighting Compensation*

The *lighting inhomogeneity* problem in deep waters is mainly due to the lack of natural global lighting, and to the necessary use of artificial light sources with limited power. Illumination systems are often rigidly attached to the AUV or ROV and light sources typically concentrate the rays into a given area where the camera is focused. The acquired image borders suffer from darkening due to *light attenuation*, principally induced by the *light absorption* of the water. The effect is similar to *vignetting*, although the phenomenon is not produced by the camera lens but by the medium itself. All images from a given sequence are affected, to some degree, by this factor. The illumination distribution from artificial light sources changes with the distance from the camera to the seafloor. Colors are also affected due to *light absorption*, resulting in *depth-dependant* color profiles of the images acquired.

Imaging conditions hinder the application of a single compensation function on all the images acquired in absence of precise information about the placement and nature of the light sources, the distance from the camera to the seabed, and the 3D structure of the scene. This circumstance results in the loss of a global terrain perception, which is a cognitive sensation factor highly dependant on lighting coherency [1].

A feasible correction of *lighting inhomogeneity* and *vignetting-like* artifacts in a single step consists of the application of a 2D "inverse illumination distribution" to the original input images [2–5]. The main aim of this operation is to enhance the luminance of the darkened image borders in order to obtain uniform illumination throughout the image. If a high sensitivity camera with a high pixel depth (>8 bpp) is available, not only the luminance but also the richness of detail can be enhanced in the region affected by the *light absorption*.

The illumination pattern describing the "inverse illumination distribution" function can be estimated from a subset of images showing low texture and reduced 3D structure (i.e. flat, sedimented terrain). As this function changes with the distance from the light source to the seabed, a three-step approach is proposed (Fig. 4.2) to correct the lighting artifacts. It is based on two main ideas: (1) the application of a depth dependant inverse illumination distribution, and (2) the automatic selection of the images to compute this pattern in a given depth-range based on the Total Variation (*TV*) metrics [6], as described below.

Fig. 4.2 Lighting pattern
compensation procedure. The
images of a sequence are
classified into depth subsets,
and a different lighting pat-
tern compensation function
is computed for each one.
The figure shows a set of n
images from which the $n/2$
images having the lowest
TV value have been selected.
Next, the images are averaged
and the result normalized and
smoothed using a Gaussian
filter with an adaptively
selected σ

Depth$_d$ Image Subset (n)

Selected Images (n / 2)

Average of
Selected Images

Gaussian
Smoothing
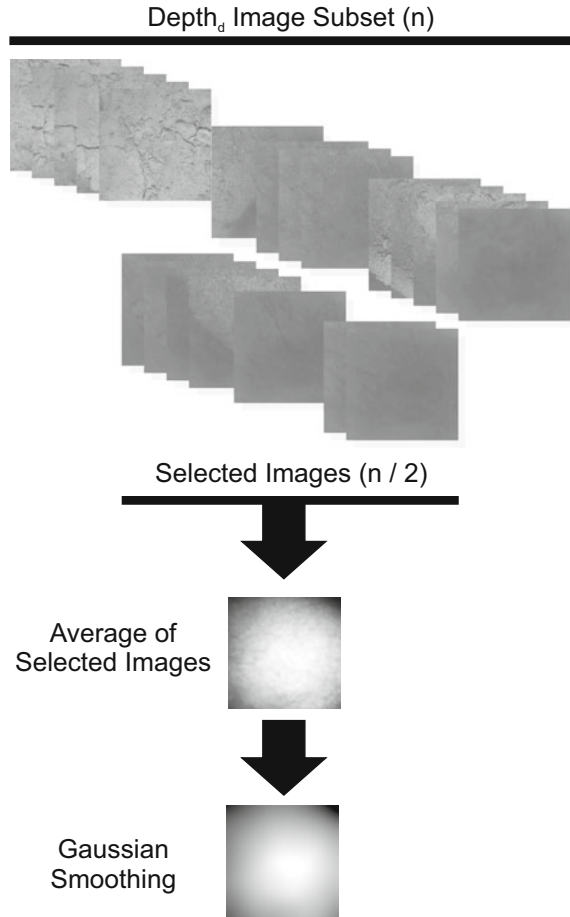
Quasi-Altitude Estimation

Underwater image acquisition platforms often record not only image sequences but
also other synchronized data like heading, acoustic positioning, surface Global Posi-
tioning System (GPS) positioning and altitude, among others. Unfortunately, camera
altitude is not always available for every data set. Consequently, as a first step, the
images of a given sequence should be classified according to altitude in order to
apply a different lighting correction function to each one, but assuming that precise
information about distance from the camera to the seafloor may not be available.
In order to solve that issue, a quasi-altitude estimation is now proposed to be used
instead.

   Given a sequence of images and its corresponding registration parameters onto
the photo-mosaic frame, it is possible to determine which ones were acquired closer

to the seabed and which ones further away by computing the size or scale of the image once registered to the 2D photo-mosaic coordinate system. Specifically, it is possible to consider only the diameter of the transformed pictures (i.e. the size of the longest diagonal) since this scale and the altitude are highly correlated when the focal length of the camera is assumed constant. Once an image list has been built and sorted according to their diagonal length, the images can be classified in subsets of similar altitudes.

Depth Sliding Window Strategy

The "inverse illumination distribution" changes with the distance from the camera to the seafloor, inasmuch as the light sources are rigidly attached to the UV. Consequently, this distribution should dynamically vary to compensate for depth fluctuations. In that sense, a depth sliding window strategy can be used. Given all the images of a given data set, the first step consists of sorting them by altitude, using sensor-acquired depth information or the quasi-depth estimation measure. The second step consists of opening a window centered on a given reference image in the sorted set with and arbitrary size depending on the frequency of the depth changes. The images in this window will be used to compute the "inverse illumination distribution" to be applied to the image on which the window is centered. With this strategy, a smooth variation of the function is ensured. Nevertheless, to avoid excessive computations, the step between reference images can be set to N instead of one image, and the function can be applied not only to the reference image but also to a small temporal neighbourhood determined by the value of N. In any case, this strategy will obtain an acceptably smooth variation of the function, in contrast with other strategies using a single function for all the images in the sequence, or those determining an arbitrary number of image depths.

Image Selection

For each image window, a distinct compensation function for the light distribution should be computed from images with a low texture content and homogeneous appearance. Low textured images are the best suited for this estimation due to their low average gradient length. An adequate ranking metric for the selection of these images is the *TV*.

$$TV = \frac{1}{W \cdot H} \sum_{x=1}^{W-1} \sum_{y=1}^{H-1} \|g(x, y)\| \tag{4.1}$$

Equation 4.1 shows the computation of the normalized *TV* for a given image, where W and H are the width and height sizes and $\|g\|$ notates the L_1 or L_2 norm of the g gradient vector. The *TV* values for the last row and column of a given image are set to 0.
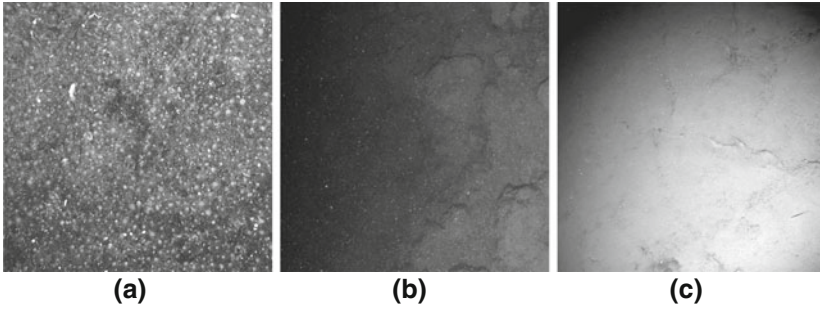
Fig. 4.3 a Example of *back-scattering* due to the reflection of rays from the light source on particles in suspension, hindering the identification of the seafloor texture. **b** Example of *forward scattering* caused by the local inter-reflection of the light suspended particles, hiding the terrain behind them. **c** Effects produced by *light absorption* of the water resulting in an evident loss of luminance in the regions farther from the focus of the artificial lighting

Equation 4.1 can be used with both L_1 or L_2 norms. In our experiments, we have selected the L_2 norm, i.e. Euclidean metrics, to evaluate the homogeneity of the images, because it allows characterizing the magnitude of the neighboring pixel variations (i.e. gradient vectors). Once the *TV* measure has been computed for all the images of a given altitude subset, an image subset of low *TV* is used to estimate the light distribution. The aim of the measure is to identify images containing structures rich in details. The presence of high frequency noise, mainly due to scattering on macroscopical particles in suspension of scattering (see Fig. 4.3), may skew the image quality evaluation. The *TV* magnitude of the image may inappropriately increase leading to scenarios where the dominant part of the metrics comes from high frequency noise. Nevertheless, the unwanted effects of the high frequency components can be avoided by building lower resolution images from the originals with $N \times N$ super-pixels. This simple approach significantly reduces the effects of the high frequency components in both the image and the *TV* measure. In practice, 8×8 linearly averaged super-pixels may produce good results for images of $1,024 \times 1,024$ pixels, which are reduced to 128×128 pixels. The images obtained save every important seabed feature but cancel the effects of the scattering phenomena, allowing the use of the *TV* as an image quality evaluation metrics. For each depth-range, the images with a *TV* value below the median can be used to compute the illumination correction function. To obtain this function, the selected images are averaged and the result is smoothed by a low-pass filter to reduce the remaining high frequency components, as explained below.

Compensation of Lighting Inhomogeneities

In order to compensate the light attenuation problems and obtain an image with a homogeneous illumination l_H, the acquired luminance values are divided by a given

compensation mask as shown in Eq. 4.2

$$l_H(x, y) = \frac{l(x, y)}{l_G(x, y)} \qquad (4.2)$$

where l is the image luminance values, l_G corresponds to the illumination pattern and l_C is the lighting compensation pattern before the Gaussian smoothing.

$$l_C(x, y) = \frac{1}{N} \sum_{k=1}^{N} l_k(x, y) \qquad (4.3)$$

Equation 4.3 computes the average value for every pixel position given a stack of N images. Finally, the compensation mask l_C obtained is smoothed with a low-pass Gaussian filter to obtain the illumination distribution l_G function. This distribution is then used for the *lighting inhomogeneity* compensation, as per Eq. 4.4, where $\langle \rangle$ denotes Gaussian smoothing.

$$l_G(x, y) = \langle l_C \rangle \qquad (4.4)$$

The value for σ used in the Gaussian convolution is selected adaptively for each altitude subset. Starting from the average image l_C in Eq. 4.3, a set of increasing values $\sigma_1, \sigma_2, \ldots, \sigma_k$ will be sequentially applied to it until the smoothed TV value is under a threshold $TV(l_{G(\sigma)}) < \varepsilon$. Values in the range of $\frac{d}{256}, \frac{d}{128}, \ldots, \frac{d}{32}$, where d is the shortest dimension of a given image, offer good results in practice. With this threshold condition the appropriate smoothness and uniformity of the blurred image are ensured.

4.1.2 Gradient-Based Image Enhancement

As the altitude of the robot increases, the effects of the previously mentioned *backscattering*, *forward scattering* and *light absorption* phenomena become more evident. The strategy proposed to enhance the high frequency details affected by these phenomena is a simple and global approach, selecting the highest quality image in a given surrounding region from the whole set, and using it as a contrast or gradient reference. To avoid unpredictable visual effects, the non-global approaches of homomorphic filtering [7, 8], Contrast Limited Adaptive Histogram Equalization (CLAHE) [9] (Fig. 4.4) and histogram specification [10] are not used, due to the following reasons. On the one hand, homomorphic filtering may lead to an excessively homogeneous appearance of the filtered image and to a loss of global consistency in the appearance of the photo-mosaic. The suppression of low frequencies performed by this kind of filter may provide some advantages in the visibility of local details, but in giga-mosaicing, depending on the zoom factor, every spatial frequency can be important to recognize and understand the nature and morphological attributes
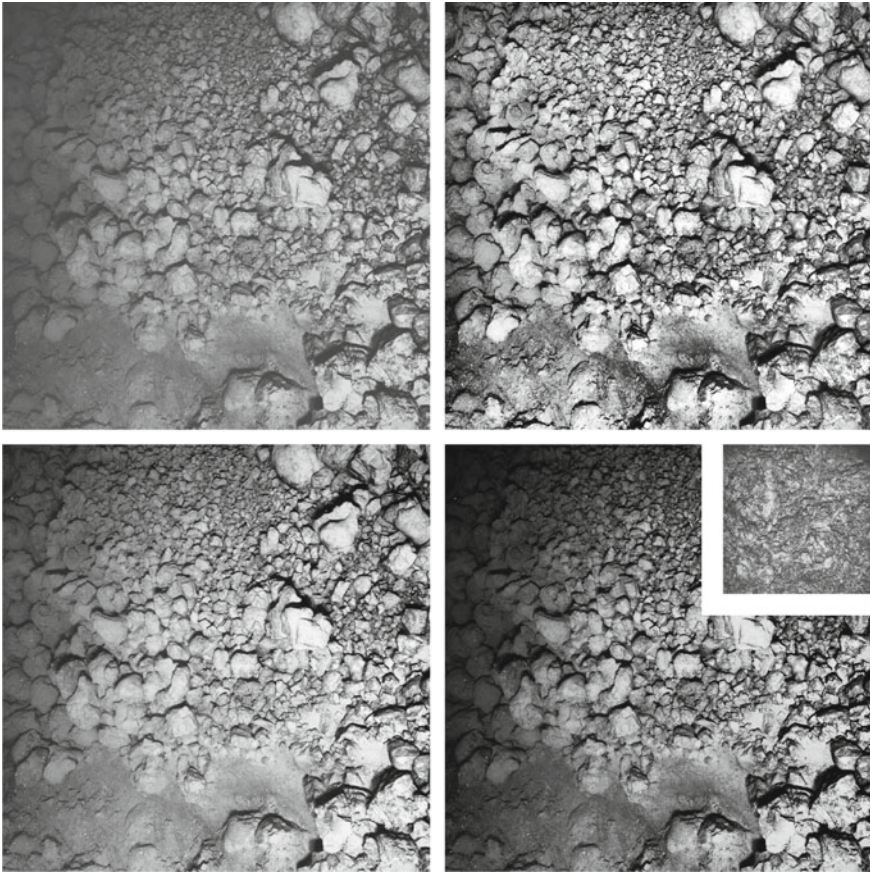
Fig. 4.4 (*Top-left*) Image lacking contrast on its *left side*. (*Top-right*) Image processed with a CLAHE algorithm, showing enhanced details in the originally lower-contrast regions. sssThe appearance of the processed image is less realistic than the original due to an aggressive level of local filtering. (*Bottom-left*) Image processed with a Butterworth homomorphic filter. The image evidences a generalized lack of contrast. (*Bottom-right*) Image resulting from the histogram specification of an apparently uniformly illuminated image into the test image. The image obtained has better contrast than the original, but still evidences problems in the *darkest* areas

of the seabed structures. On the other hand, histogram specification is highly dependent on the reference image, and therefore the modified image may often loose its realistic appearance. Therefore a simple but robust local contrast stretching can be applied to equalize a given sequence of images.

Image Quality Estimation

There is not a single and objective criterion to identify the image with the highest visual quality from a given set because the concept of "quality" involves different cognitive aspects. However, phenomena affecting image detail richness and sharpness, such as *scattering* and *light absorbtion*, are known to grow with the distance from the camera to the seabed.

This simple and fast approach may lead to poor results when the selected image presents an over-exposed region, for example, due to being acquired too close to the seabed under strong illumination. A more robust selection of the reference image is to use *TV* to rank image quality also. Thus, the image with the highest *TV* may be selected as the reference image while ensuring that over-exposed regions do not affect this selection. According to our experimental validation, the image with the highest *TV* coincides in most cases with the closest one to the seabed on a given survey, and with the second or the third closest images in the few remaining cases.

Global Contrast Stretching

The *TV* value of the reference image selected is used to compute the stretching factors that will be applied for a global contrast (or gamma amplification) on all the other images. This stretching factor should be selected below a given threshold T_s to avoid overamplification of areas of poor contrast, e.g. textureless sediment-covered regions. T_s depends on the Signal-to-Noise Ratio (SNR) of the image, which can vary highly according to water quality, lighting intensity, and/or the camera sensor. Despite the application of these gradient corrections, the merging of images from highly different depth categories will unavoidably produce noticeable seams due to their distinct blurring levels. The stretching factor $\frac{TV_{reference}}{TV(k)}$ is applied to enhance the x and y gradient components of the k-th image.

4.2 Image Registration with Global Alignment

While image registration is not directly related to the blending procedure and, therefore, is not at the core of the work presented here, the accuracy of image registration will significantly affect the final quality of the photo-mosaic rendered.

Even when navigation data (such as USBL positioning, heading, depth, etc.) are available, pair-wise image registration is still required to ensure a precise camera motion estimation. Pair-wise registration can be performed using a feature-based approach, involving the well known image feature detectors and descriptors of Harris [11], SIFT [12] and SURF [13], among others. When building a 2D photo-mosaic from a set of images acquired by a camera close to the seabed, the *planar assumption* of the scene can be violated due to the microbathymetry of the seafloor.

As already stated in Sect. 2.3.2. The 3D geometry of the scene, in addition to the short camera distance, results in *parallax*. This problem increases the difficulty of estimating the 2D planar transformation between consecutive images, often leading to *misregistrations*, resulting in double contour effects during blending.

A global alignment strategy [14, 15] is required to reduce the inaccuracies of a simple sequential pair-wise registration, as explained in Sect. 2.4. The strength of the global alignment arises from closing-loops because they allow a significant improvement of the camera trajectory estimate when re-visiting an already mapped area. In absence of loop-closings, and considering input sequences of thousands of images, the drift accumulated by the pair-wise transformations leads to significantly inconsistent (missaligned) photo-mosaics.

4.3 Image Contribution Selection

The *parallax* effect will influence both image registration and image blending procedures. On the one hand, image panorama software often fails to register sequences with strong *parallax* since they assume camera rotation only. On the other hand, and even using the best possible registration, the double contouring problem will appear when merging two or more images if the vehicle (and the camera) translates and the scene is not perfectly planar.

The solution to avoid ghosting artifacts is the use of information from a single image for each pixel of the final photo-mosaic whenever possible. Blending is performed in a narrow region around the optimally computed seams, and consequently information from more than one image is fused only in a small fraction of the final photo-mosaic. Ghosting may occur in those regions, but its noticeability is significantly localized and dependent on the width of the transition region.

4.3.1 Image Discarding

Each pixel of the photo-mosaic is obtained from a single image pixel whenever possible. To maximize the quality of the final photo-mosaic, the contribution from sharper and informative images should be prioritized. Image blending algorithms take into account the information of all the available images. Unfortunately, this may lead to unnecessary contributions of low quality images even when higher quality information is available in a given area. Therefore, discarding low quality images will ensure that their information is not taken into account in any sense. Furthermore, ignoring these images will also impact the optimal seam finding step, reducing the number of paths to be computed, and consequently speeding up the process. The developed discarding procedure is described below.

First, the frames of the original images are mapped into the global photo-mosaic frame using the image registration parameters in order to know their shape and

area coverage in the final photo-mosaic coordinate system. The depth estimation is computed, assuming that depth information is not available in the navigation data. It is possible to discard low quality images covering a region of the scene if higher quality ones are available for that area. The discarding procedure is performed using logical operations on the polygons describing the images, which is an efficient approach requiring few resources.

Each image is defined as a trapezoid described by four vertices corresponding to the four image corners once registered to the photo-mosaic frame. Additionally, the polygons are sorted decreasingly according to their corresponding image TV value. At each step of the iterative process, a new image trapezoid of the sorted list is added to the final photo-mosaic polygon using simple binary operators. If the area covered by the new trapezoid has already been fully covered by the photo-mosaic polygon (i.e. the trapezoid does not intersect the photo-mosaic polygon and lies inside this one), the image is discarded because this same region is supposed to have already been covered by higher quality images. Otherwise, if the image to be added contains information from a non-covered area, the photo-mosaic polygon is updated and the image is accepted.

4.3.2 Pixel-Level First-Closest and Second-Closest Maps

The proposed blending methodology determines the first and second closest maps at pixel level. The *first closest map* contains, for each pixel coordinate of the photo-mosaic, the index of the image whose center is closest (see Fig. 4.5). The *second closest map* does the same, but with the second closest image indices. Similar to [16], the overlap of these two maps will use a graph-cut algorithm to compute the seam-strips for blending. For every *seam pixel* two image indices are selected. Therefore, every pixel outside the seams (most of the photo-mosaic) is associated to a single image.

The Euclidean distance between a pixel $I^M(x, y)$ in the photo-mosaic frame and the center of a given n-th image $I^n(x, y)$ is weighted by a factor $w_n(s)$, as shown in Eq. 4.5:

$$d_M^n(x, y) = w_n(s) \cdot \sqrt{(x_M - x_n)^2 + (y_M - y_n)^2} \qquad (4.5)$$

where the scalar factor $w_n(s)$ is a size-ratio between the n-th image and the image having the smallest area once registered. For time efficiency reasons, the ratio is not computed based on the area of the warped images, but on the length of their diameters, as explained in Sect. 4.1.1, to obtain a rough but fast approximation, as shown in Eq. 4.6:

$$w_n(s) = s_{\min}/s_n \qquad (4.6)$$

where $s_{\min}$ is the diameter of the smallest image for a given set and s_n is the diameter of a given n-th image.

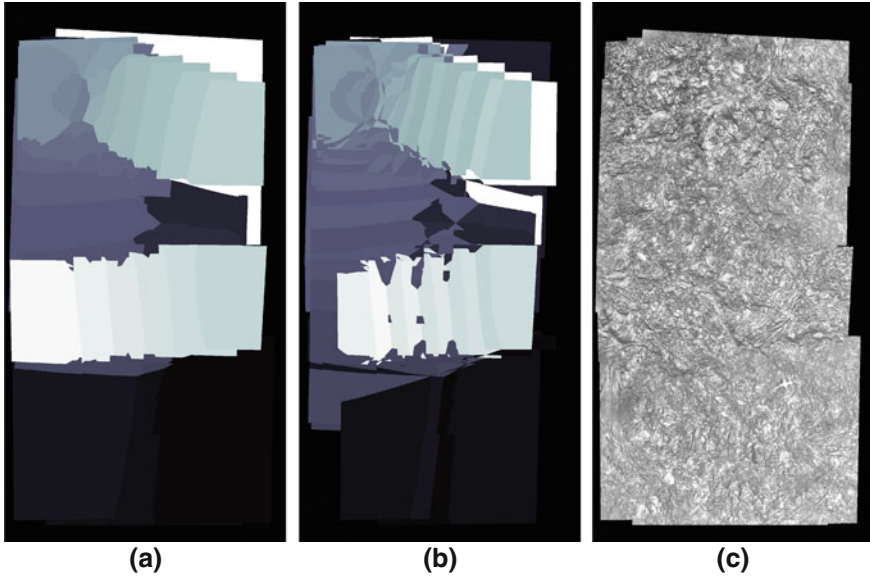<div align="center">(a)   (b)   (c)</div>

Fig. 4.5 **a** First closest map and **b** second closest map corresponding to the registered images finally blended into the **c** photo-mosaic. The *blue* level of every pixel in the closest maps represents the index of the image having the closest and second closest image centers. The distance measure gives more priority to pixels belonging to images which have been acquired at a lower altitude, consequently showing a higher level of detail

This weighting prioritizes pixels from images acquired at low altitudes, close to the seabed, and consequently less affected by underwater imagery artifacts. This weighting also maximizes the contribution of "higher-quality" images to the final photo-mosaic image. Therefore, in cases like the one shown in Fig. 4.6, only a small percentage of the pixels from the smaller overlapping image are lost while computing the smooth transition, while the most significant percentage of the original image is preserved.

4.3.3 Regions of Intersection

The overlap between the first and second closest maps determines the regions where the pixel level graph cut should be performed. Therefore, for each overlapping patch, the texture from the two best-quality images is available, and the graph cut is used to find the optimal boundary seam, determining the contribution of each one in the final photo-mosaic. Each region of intersection $ROI_{i,j}$ between the two images i and j, where i is the closest image, j is the second closest image, and $R_{i,j}$ denotes the photo-mosaic region where i and j coincide, is defined as $ROI_{i,j} = R_{i,j} \cup R_{j,i}$.
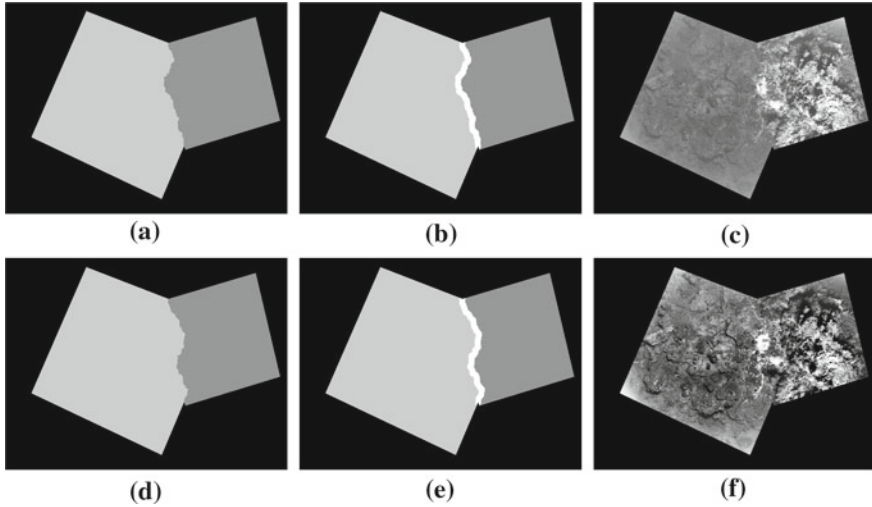
Fig. 4.6 Example of a pixel level graph-cut performed between two overlapping images acquired at different altitudes, and consequently evidencing differences in appearance. **a** Result of the graph cut performed on the images without enhancement, **b** depicts, in *white*, the narrow strip (20 pixels on each side of the cut) where the gradient domain blending is performed and **c** shows the blended image pair. **d** is the result of the graph cut performed on the images after being enhanced according to the proposed neighboring based enhancement approach, **e** depicts, in *white*, the narrow strip where the gradient domain is performed and **f** shows the blended image pair. Notice that the results of the pixel-level graph-cuts are different before and after the application of the image enhancements

4.4 Gradient Domain Blending

4.4.1 Pixel-Level Graph-Cut

The proposed blending strategy uses an optimal seam finding algorithm to compute the best boundaries in the overlapping image areas. A pixel level graph cut is performed on the regions of intersection determined by the first and second closest maps. In contrast to [16], the graph-cut is performed at the pixel level in order to guarantee maximum accuracy of the cut, given that the main aim of the algorithm is to achieve a high image quality. The algorithm searches for the boundary that minimizes the cost of the transition from one side to the other of the border line for every pair of pixels. The function has three weighted terms controlling the behavior of the cut:

$$C = \mu_1 \cdot f(I_1, I_2) + \mu_2 \cdot s(g_1, g_2) + \mu_3 \cdot L \qquad (4.7)$$

The first term $\mu_1 \cdot f(I_1, I_2)$ measures the intensity differences between overlapping pixels. The second term $\mu_2 \cdot s(g_1, g_2)$ measures the gradient vector differences along the boundary B seam. Finally, the third term $\mu_3 \cdot L$ measures the length L of the

seam. The three weighting factors μ_1, μ_2 and μ_3 control the behavior of the cut. The gradient term, which is not been used in such a way in the literature [16], allows us to deal with differently exposed overlapping regions. Here an intensity-based graph cut will consider that the differences between neighboring pixels are large even if the registration is accurate, and thereby avoid those regions where the cut should be performed. Instead, if the difference between the gradient vectors along the seam path is used, the optimal seam will be found independently of the differences of image exposure. In the case of misregistration of moving elements in the scene, the term $\mu_2 \cdot s(g_1, g_2)$ avoids bisecting those elements by having the seam line by-pass them. This is due to the fact that even a large value of L in the by-pass has less cost than crossing a double contour with large gradients of a given structure. The gradients are also less sensitive to other illumination issues, such as those caused by artificial lighting and non-uniform lighting. Furthermore, working in the gradient domain compensates the exposures when recovering the luminance images from the gradient vectors. Despite the benefits of the gradient term, the intensity term is kept in order to favor low photometric differences when registration is highly accurate. Therefore, a weighted addition between both intensity and gradient domain terms is proposed.

The effects of *parallax* and registration inaccuracies are minimized since the graph cut tends to place the seam in textureless regions where morphological differences are low. For the same reason, cuts over moving objects tend to be avoided, thus benefiting the visual consistency of the blended results.

Performing a graph cut, especially at pixel level, is usually a computationally expensive operation when the size of the region to process is significantly large. Nevertheless, the regions on which the graph cut is working, determined by the intersection between the first and second closest maps, are rarely large. Furthermore, this process can be parallelized, taking advantage of recent multi-core processors, to speed up the execution in one of the main bottlenecks of the processing pipeline.

4.4.2 Gradient Blending Over Seam Strips

Once an optimal seam has been estimated, a smooth transition between neighboring regions needs to be performed. Even for sequences where the images have been preprocessed to solve non-uniform illumination problems such exposure artifacts and contrast level equalization, the graph cut result may lead to an image with noticeable seams. Therefore, smoothing the transition between the image patches is required. The image fusion around the computed seams should be performed in a limited region, being both wide enough to ensure a smooth transition and narrow enough to reduce the noticeability of ghosting and double contouring. According to our experience, a transition strip of 10 pixels at each side of the seam (i.e. a 20 pixels transition region) has been demonstrated to be appropriate for sequences of 1-Mpixel images.

A new transition smoothing approach is proposed in this book. The applied method is a weighted average around the seams in the gradient domain, as shown in Eq. 4.8, where g_x^1, g_y^1, g_x^2 and g_y^2 are the x and y gradient fields for the two involved images, $\hat{g}_x$ and $\hat{g}_y$ are the x and y gradient fields after the blending and μ is the smoothing transition function. Concretely, a 3rd order Hermite function is applied. The advantage of performing the weighted average in the gradient domain is the automatic compensation for different exposures between neighboring images when the luminance image is integrated from the gradients as a final step.

$$g_x(x, y) = \mu \cdot g_x^1(x, y) + (1 - \mu) \cdot g_x^2(x, y)$$
$$g_y(x, y) = \mu \cdot g_y^1(x, y) + (1 - \mu) \cdot g_y^2(x, y)$$

(4.8)

4.5 Luminance Recovery from Gradient Fields

After independently processing each overlapping strip region around the seams, the resulting patches need to be unified into a single, larger image. Each patch processed should be updated on the final photo-mosaic image, while information which belongs to regions without overlap should be recovered from the corresponding original images.

Once the final gradient domain photo-mosaic has been composed after the "strip-blending", a non-integrable or inconsistent gradient field is obtained. In order to recover the luminance values from the gradient fields, a multigrid Poisson solver [17] is used.

4.6 Tone Mapping

The solution provided by the gradient solver is defined up to a free additive term on the recovered intensity value. Consequently, a mapping algorithm such as Minimum Information Loss [18] should be applied to determine this factor. The main goal of the mapping algorithm is to appropriately manipulate the dynamic range of the computed image in order to make it fit into the limited range of a display device while keeping the maximum amount of detail information.

4.7 Giga-Mosaic Unification

The photo-mosaicing pipeline described is currently implemented in Matlab™, using Matlab EXecutable (MEX) files and parallel computing when possible. This allows the efficient blending of photo-mosaics up to 60 Mpixels in a standard personal computer with 4 GB of RAM in less than 5 min. Nevertheless, this mosaic size (i.e. <0.1 Gpixels) is small at the gigapixel scale in which this work is interested, and

a solution should be used to reach the desired 5–15 Gpixels required to process the currently available data sets.

The amount of RAM may become a limitation when dealing with gigapixel images, especially if the images have more than 8 bpp (e.g. 16-bpp grayscale images or 24/48-bpp color images). The strategy proposed to reduce the computer requirements consists of decomposing the problem into sub-problems (i.e. rectangular *tiles*) in order to sequentially solve them and finally unify them into the final mosaic image.

The price of this decomposition is the need of a second level of blending of the tiles. This one is similar to the "strip-blending" presented in Sect. 4.4.2 applied to the optimal seams, but is performed in the intensity domain. This second level of blending is performed only in the intensity domain for computational reasons. When compared with gradient domain operations, intensity blending is inexpensive and can deal with large amounts of data. Furthermore, this method does not lead to loss of quality due to the particular conditions in which it is applied. There are two reasons for the need of a blending step between neighboring tiles. The first is the different free factor of every tile after the luminance recovery using the Poisson solver since this factor is multiplicative when working with *log I* values. The second is the nature of the Poisson solver which spreads the inconsistency of the gradient fields along the whole area recovered. After multiplying the pixel intensities of every tile with the corresponding constant factor, a *tile-overlap* intensity blending has to be performed. This kind of blending will compensate the gradient differences of overlapping tiles coming from different Poisson solutions. The decomposition necessarily differs from the theoretically exact Poisson solution, given that the errors due to gradient inconsistencies will be differently spread by the solver in both cases. Nevertheless, these differences are negligible in practice.

Although the *tile-level* pipeline described above is straightforward, its technical implementation deserves further clarifications owing to the need to manage available computational resources with such large problems (i.e. gigapixel photo-mosaics).

The rectangular "canvas" of the full photo-mosaic is divided into a regular grid of overlapping *tiles* in order to process it using an *out-of-core* algorithm [19]. The size of the *tiles* depends on the available RAM. For time efficiency, the space required to store a single tile and a full *global-strip* (i.e. a row of tiles) is allocated to memory, avoiding an excessive amount of slow hard drive sequential accesses.

A weighted average smoothing in the intensity domain is used to join neighboring *tiles* in a given rectangular overlapping region. In our experiments, the size of the overlapping regions varied between 15 and 25 % of the *tile* size depending on the initial spatial image arrangement. Once a *tile* has been processed, it is stored in the current *global-strip*, performing a blending with the previously processed one (when available). When a single *global-strip* has been processed, it is stored in the hard drive to save RAM space and the same procedure is repeated on the next one. The strategy used to blend two neighboring *tiles* is also used to blend two neighboring *global-strips*. Performing the blending in this structured way avoids the problem of simultaneously fusing more than two images of a given region, which may make the computation of a transition function of the overlapping areas more complex. Figure 4.7 shows the giga-mosaic unification strategy described above.

Fig. 4.7 Tiling scheme for the gga-photo-mosaic blending. Each *tile* is processed as an independent photo-mosaic and blended with previously processed neighboring ones in a given *global-strip* (i.e. a row of blended tiles), using a weighted average in the luminance domain. Next, each two neighboring *rows* are blended using the same approach. The Giga-photo-mosaic is the result of joining all the *global-strips*
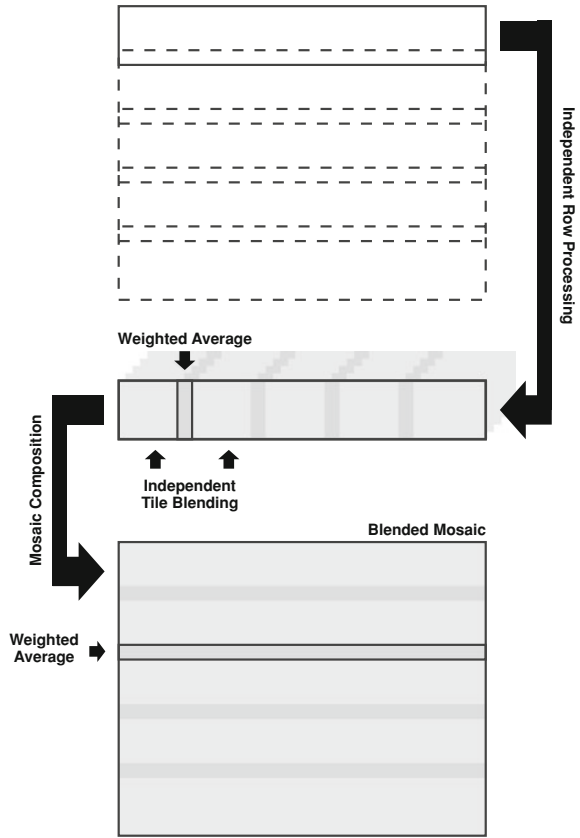
4.8 Conclusions

The main underwater imaging issues affecting underwater photo-mosaicing have been treated by the approach presented. For each one of the specific underwater imagery problems, a working solution has been presented and a new processing pipeline has been defined. In the preprocessing stage, an adaptive non-uniform illumination compensation based on a sliding window on the depth sorted image sequences has been proposed. This function allows not only giving an homogeneous appearance to a sequence of images, but also enhances hidden details in the case of high dynamic range images. Concerning exposure variations, the blending strategy based on the image gradients allows the avoidance of dealing with this problem, inasmuch as gradient methods are not sensitive to exposure variations. In the context of gradient domain methods, a novel hybrid luminance and gradient based graph-cut strategy has been presented, allowing the avoidance of problems concerning exposure variations and moving objects in the scene. Light attenuation and forward scattering lead to loss of contrast and poor detail in the images. In order to solve this issue, an adaptive

image enhancement, based on the selection of the highest quality image in a given surrounding as the image sharpness reference, has been presented. The approach allows giving an homogeneous appearance to the images involved, and to enhance, up to a reasonable level, the sharpness of the original images. Finally, and aiming to efficiently generate high-resolution large-scale mosaics, a method to subdivide the mosaic into smaller and easily processable tiles has been presented.

References

1. Goldstein, E.B.: Sensation and perception. In: PSY 385 Perception Series. Cengage Learning, Stamford (2010)
2. Garcia, R., Nicosevici, T., Cufi, X.: On the way to solve lighting problems in underwater imaging. In: Proceedings of the MTS/IEEE OCEANS Conference, vol. 2, pp. 1018–1024, Oct 2002
3. Capel, D.: Image Mosaicing and Super-Resolution. Springer, Berlin (2004)
4. Rzhanov, Y., Gu, F.: Enhancement of underwater videomosaics for post-processing. In: Proceedings of the MTS/IEEE OCEANS Conference, pp. 1–6, Oct 2007
5. Gracias, N., Negahdaripour, S., Neumann, L., Prados, R., Garcia, R.: A motion compensated filtering approach to remove sunlight flicker in shallow water images. In: Proceedings of the IEEE OCEANS Conference, pp. 1–7, Sept 2008
6. Chan, T., Jianhong, S.: Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods. Society for Industrial and Applied Mathematics, Philadelphia (2005)
7. Guillemaud, R.: Uniformity correction with homomorphic filtering on region of interest. In: Proceedings of the International Conference on Image Processing (ICIP), pp. 872–875 (1998)
8. Ebner, M.: Color constancy. In: The Wiley-IS&T Series in Imaging Science and Technology. Wiley, New Jeresy (2007)
9. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., Romeny, B.T.H., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Comput. Vis. Graph. Image Process. **39**(3), 355–368 (1987)
10. Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Comput. Graph. Appl. **21**(5), 34–41 (2001)
11. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of the Alvey Vision Conference, Manchester, UK, pp. 189–192, Aug 1988
12. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
13. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: European Conference on Computer Vision, pp. 404–417 (2006)
14. Ferrer, J., Elibol, A., Delaunoy, O., Gracias, N., Garcia, R.: Large-area photo-mosaics using global alignment and navigation data. In: Proceedings of the IEEE OCEANS Conference, pp. 1–9, Oct 2007
15. Elibol, A., Garcia, R., Delaunoy, O., Gracias, N.: A new global alignment method for feature based image mosaicing. In: Proceedings of the International Symposium on Advances in Visual Computing (ISVC), Part II, pp. 257–266. Springer, Berlin, Heidelberg (2008)
16. Gracias, N., Mahoor, M., Negahdaripour, S., Gleason, A.: Fast image blending using watersheds and graph cuts. Image Vis. Comput. **27**, 597–607 (2009)
17. Kazhdan, M., Hoppe, H.: Streaming multigrid for gradient-domain operations on large images. ACM Trans. Graph. (SIGGRAPH) **27**(3), 1–10 (2008)
18. Neumann, L., Matkovic, K., Purgathofer, W.: Automatic exposure in computer graphics based on the minimum information loss principle. In: IEEE Computer Society, Hannover, Germany, vol. 0, pp. 666–667 (1998)
19. Vitter, J.S.: External memory algorithms and data structures: dealing with massive data. ACM Comput. Surv. (CSUR) **33**(2), 209–271 (2001)