Robotica: (2020) volume 38, pp. 396–409. © Cambridge University Press 2019
doi:10.1017/S0263574719000699

Graph-Based Registration and Blending for Undersea Image Stitching

Xu Yang†, Zhi-Yong Liu†, Hong Qiao†, Jian-Hua Su†, Da-Xiong Ji‡, Ai-Yun Zang§ and Hai Huang¶*

† *State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, People's Republic of China,*
E-mails: xu.yang@ia.ac.cn, zhiyong.liu@ia.ac.cn, hong.qiao@ia.ac.cn, jianhua.su@ia.ac.cn
‡ *Ocean College, Zhejiang University, Zhoushan 316000, People's Republic of China.*
E-mail: 0015108@zju.edu.cn
§ *College of Engineering, Ocean University of China, Qingdao 266100, People's Republic of China.*
E-mail: zaysi@sina.com
¶ *National Key Laboratory of Science and Technology of Underwater Vehicle, Harbin Engineering University, Harbin 150001, People's Republic of China*

(Accepted April 10, 2019. First published online: May 31, 2019)

SUMMARY
Image stitching is important for the perception and manipulation of undersea robots. In spite of a well-developed technique, it is still challenging for undersea images because of their inevitable appearance ambiguity caused by the limited light in the undersea environment, and local disturbance caused by moving objects, ocean current, etc. To get a clean and stable background panorama in the undersea environment, this paper proposes an undersea image-stitching method by introducing graph-based registration and blending procedures. Specifically, in the registration procedure, matching the features in each undersea image pair is formulated and solved by graph matching, to incorporate the structural information between features. In the blending procedure, an energy function on the indirect graph Markov random field is proposed, which takes both image consistency and neighboring consistency into consideration. Coincidentally, both graph matching and energy minimization can be mathematically formulated by integer quadratic programming problems with different constraints; the recently proposed graduated nonconvexity and concavity procedure is used to optimize both problems. Experiments on both synthetic images and real-world undersea images witness the effectiveness of the proposed method.

KEYWORDS: Undersea image stitching; Feature correspondence; Graph matching; Energy minimization; Nonsubmodular function.

1. Introduction

Image stitching aims at combining two or more images with overlapped areas into a wide viewing composite, or a panorama. It offers help for robot tasks in inaccessible places for human beings, such as extraterrestrial celestial body, disaster, and deep sea environments, as it can provide a larger field of view over the operating areas by integrating images shot from different locations. In this paper, we focus on the stitching problem in the undersea environment.

Image stitching itself has long been an important topic in image processing and robot vision. It can be roughly divided into two main procedures, that is, registration and blending. The registration

* Corresponding author. E-mail: haihus@163.com

procedure searches for the image alignment based on the overlapped area, by selecting and estimating a proper transformation model, for example, rigid transformation, affine transformation, and projective transformation. The key component is the matching of the overlapped area, of which the methods can be further divided into two categories, that is, the area-based methods and the feature-based methods. The area-based methods search and match the overlapped area through the area appearance cues, for example, intensity information or related statistics, and typical methods include the correlation method and mutual information method. These methods find applications in tasks with rigid transformation such as the medical imaging, but they are usually less robust to occlusions and geometric distortions which are common in robot vision. Therefore, the feature-based methods are more popular in robot vision-related image stitching, which rely on local features dedicated to tackling the above changing factors. And the related works are discussed in Section 2. The blending procedure combines the registered images into a smooth panorama with minimization of visual differences across the overlapped area. Apparently the key component is blending the overlapped area, by synthesizing the redundant information in the overlapped area from associated images into a visually optimal part of the panorama. Early methods use the (weighted) average, also known as the feathering operation, to blend the overlapped area, which is useful in, for example, static landscape panorama. In robot vision, the blended area often contains exposure artifacts caused by the robot motion and ghosting caused by moving objects. In the last decade, researchers resort to the energy minimization-based pixel-labeling technique to tackle these problems. Specifically, searching the optimal value for each pixel is elegantly defined by the pixel labeling problem on Markov random field (MRF), which is solved by energy minimization methods such as graph cuts and loopy belief propagation. Their effectiveness has been shown in tackling the seam and ghosting in the overlapped area.

These advances make image stitching to some extent a solved problem in many daily tasks, which is even widely used on current smart phones. However, different from these daily usage, it is still a challenging problem to stitch undersea images. The main obstacles are the inevitable appearance ambiguity and local disturbance in undersea images. The appearance ambiguity is mainly caused by the limited light in the undersea environment, which is further scattered or absorbed by water molecules, plankton, or sands. The local disturbance of undersea images is mainly caused by moving objects (e.g. fish and other aquatic organisms), ocean current, or motion of the robot body. These obstacles would significantly deteriorate the performance of image-stitching methods, even the feature-based registration and energy minimization-based blending.

For the feature-based registration, the ambiguous appearance would lead to the poor discriminant ability of the feature descriptor. Therefore, it is intuitive to incorporate additional cues or constraints to make up the inadequate appearance information. The structural constraint, which requires that the corresponding features should keep structural consistency beyond maintaining the appearance similarity, could offer help to avoid abnormal feature assignments.[36] For the energy minimization-based blending, though the popular methods, for example, graph cuts, own the ability to tackle the seam and ghosting, they have problems in getting a clean and stable background panorama from undersea images with local disturbance, which in essence is due to their specific energy function requirement, as discussed in Sections 2 and 3. Therefore, a novel energy function is necessary for the undersea image-stitching problem, together with the corresponding optimization algorithm.

Based on these considerations, this paper proposes a novel undersea image-stitching method by introducing graph-based registration and blending procedures. Specifically, in the registration procedure, matching the features in each undersea image pair is formulated and solved by graph matching, to incorporate the structural information between features. In the blending procedure, an energy function on the indirect graph MRF is proposed, which takes both image consistency and neighboring consistency into consideration. The previous consistency is to remove the seam and ghosting and the latter one is to remove the local disturbance. Note the proposed energy function has a non-submodular form and is inappropriate to be optimized by, for example, the graph cuts algorithm. Coincidentally, both graph matching and energy minimization can be mathematically formulated by integer quadratic programming (IQP) problems with different constraints. Therefore, a recently proposed general-purpose combinatorial optimization framework, named by the graduated nonconvexity and concavity procedure (GNCCP), is used to optimize both problems, which has rare restriction on the energy form or graph structure.

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

The remaining paper is organized as follows: After the discussions of related works in Section 2, the proposed undersea image-stitching method is introduced in Sections 3–5, which is followed by the experimental evaluation in Section 6. Finally Section 7 concludes the paper.

2. Related Works

In this section, we first give some discussions on the image-stitching algorithms and then introduce their applications to undersea images.

2.1. Image-stitching algorithms

The related works can also be roughly divided into two categories according to the emphasis procedure, that is, the registration procedure and the blending procedure, which are, respectively, discussed in the following.

For the front one, the feature-based methods have generally overtaken the area-based methods, ever since the seminal work[5] based on scale-invariant feature transform (SIFT) feature.[25] And one main advantage of the feature-based methods is the robustness to changing factors,[12,32] such as the geometric distortions. Besides the SIFT feature, researchers have applied many types of local features to image stitching, of which some representative features include the speeded-up robust feature (SURF),[1,31] the binary robust independent elementary features (BRIEF),[6] the shape context feature,[2] etc. These methods are for general-purpose image-stitching tasks, which achieve superior performance on common natural images, especially images on land. A few works apply them in to the undersea images, which will be introduced in the next subsection.

For the latter one, early methods usually resort to the feathering technique [49] to blend the overlapped area, which is a pixel-wise operation using the weighted average. The multiband blending technique, which is widely used after the SIFT-based image-stitching work,[5] is essentially also a weighted average method. The feathering methods suffer from blurring and ghosting. Some other early methods utilize pixel labeling to model the image blending problem, such as the optimal seam method[8,26] and the method based on regions of difference.[35] These pixel labeling methods could well address the blurring and ghosting, but meanwhile they may further introduce incomplete objects in the panorama. Then the pixel labeling problem is elegantly formulated by Boykov et al.[4] as a pixel labeling problem on the MRF, and solved by graph cuts, a type of energy minimization methods. Later loopy belief propagation[29] and tree-reweighted message passing,[17] together with classic iterated conditional model,[30] are applied to the energy function. But these energy minimization methods are usually limited to specific graph structures or energy formulations (e.g. submodular function). In the last decade, many researchers turn to the continuous method, which is especially appropriate for robot vision due to its high efficiency. Various continuous optimization techniques have been applied to the relaxed continuous optimization problem, including those based on semi-definite programming,[33] convex relaxation QP (CQP),[28] L_2 norm constraint QP (L2QP),[21] spectral relaxation QP with affine constraints (SQP),[7] and our previous work.[22] Inspired by these works, the proposed method designs a stitching method specially for undersea images.

2.2. Undersea image applications

In the undersea image-stitching method proposed by Leone et al.,[19] the Harris corner point detector with certain specific improvements is used to extract the feature points, and the texture information is used to built the feature point descriptor. Then the correspondence between two feature point sets representing two undersea images is established by matching the feature point descriptors. The homography transformation, that is, the translations, rotations, and scaling effects, between two undersea images is estimated based on the correspondence, and then the stitched image is obtained by feathering-based blending. A similar scheme is used by Elibol et al.,[10] which is named by undersea optical mapping in their work. Differently, they adopted the SIFT feature point and descriptor extracted from the undersea images, of which the outlier assignments are refined by the famous random sample consensus (RANSAC) technique. In the real-time image-stitching method proposed by Ferreira et al.,[11] after the BRIEF-based motion estimation, the SURF is used in the feature correspondence step. Garcia-Fidalgo et al.[15] in their undersea image-stitching method used a feature which is a variant of BRIEF in the framework of bags of words, and the multiband blending is adopted to create the final undersea panorama.

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

Table I. Notations.

$\mathcal{G}, \mathcal{H}$	Weighted graphs representing the feature sets.
$\mathbf{G}, \mathbf{H}$	Weighted adjacency matrices associated with the weighted graphs.
$\mathbf{A}$	Cost matrix between vertex attributes in two graphs.
$\mathbf{X}, \mathbf{Y}$	Assignment matrix for registration and labeling matrix for blending.
$\mathcal{D}, \mathcal{S}$	Discrete domains, respectively, for registration and blending.
$\mathcal{C}, \mathcal{T}$	Relaxed continuous domains, respectively, for registration and blending.
$U(\mathbf{Y}_{ke})$	pixel-wise potential function.
$V(\mathbf{Y}_{ke}$	Neighbor-wise potential function.
$\mathbf{B}$	Energy matrix for blending.
$\|\cdot\|_F$	Matrix Frobenius norm.
$\mathrm{tr}(\cdot)$	Matrix trace.
$\mathrm{vec}(\cdot)$	Row-wise vectorization of a matrix.

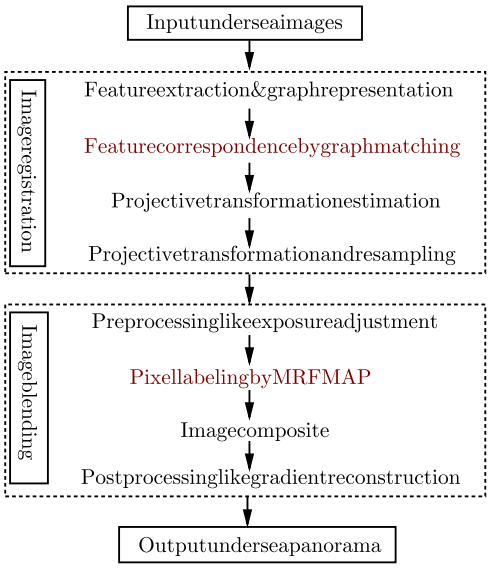

Fig. 1. Undersea image-stitching framework. The components in maroon are related to the main technical contributions of the paper.

Generally, most undersea image-stitching methods adopt the framework of the general-purpose SIFT-based image-stitching work,[5] which further design different local features for robustness or efficiency purposes. Few of them use the graph-based model to tackle the ambiguous appearance problem in the registration procedure, or the local disturbance problem in the blending procedure, which by contrast illustrates the novelty of the proposed method.

3. Undersea Image-Stitching Framework

First the general framework for undersea image stitching is briefly introduced, and then two main technical contributions of this paper are presented, as shown in Fig. 1. The main novelty of this paper is the introduction of graph algorithms into both the registration procedure and the blending procedure. Specifically, in Fig. 1 it can be observed that both the registration and blending procedures consist of a number of components, and the main technical contributions of this paper are, respectively, related to the component *feature correspondence by graph matching* in the registration procedure, and the component *pixel labeling by MRF maximum a posteriori (MAP)* in the blending procedure, which both take advantage of graph models in Fig. 1.

To avoid any confusion, a list of the main notations used in this paper is given in Table I.

4. Registration Based on Graph Matching

To integrate the appearance information and the structural relations, the feature points extracted from an undersea image are first represented by a weighted attributed graph $\mathcal{G}$, which is abbreviated as the

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

term graph below. Note in this case any type of local feature could be used as the feature extractor and descriptor, for example, Harris corner detector, SIFT extractor and descriptor, SURF extractor and descriptor, and BRIEF extractor and descriptor, which implies that the incorporation of structural cue lowers the demand of discriminant feature extractor and descriptor. It is straightforward to represent the feature set by a graph, by representing each feature point by a graph vertex, representing the link between a pair of feature points by a graph edge, describing the vertex by a so-called attribute using the feature descriptor, and describing the edge by a so-called weight using the spatial relation measures, for example, length and orientation of the link. Then feature correspondence can by defined by graph matching, that is, assigning the vertices in two graphs.

Mathematically, the collection of the weights in a graph $\mathcal{G}$ can be represented by weighted adjacency matrices $\mathbf{G}^i$, $i = 1 \cdots d$. The number of weighted adjacency matrices d depends on the weight dimension. For instance, when using the distance between feature points, that is, the link length, as the edge weight, only one adjacency matrix $\mathbf{G}^1$ is enough for a graph, where each non-diagonal entry $\mathbf{G}^1_{ij}$ denotes the distance between the ith and jth vertices in $\mathcal{G}$. Given two graphs $\mathcal{G}$ and $\mathcal{H}$ of sizes M and N, respectively, their matching can be represented by an assignment matrix $\mathbf{X} \in \{0, 1\}^{M \times N}$, where $\mathbf{X}_{ia} = 1$ means that the ith vertex in $\mathcal{G}$ is assigned to the ath vertex in $\mathcal{H}$. If the one-to-one matching assumption is adopted, then the assignment matrix becomes a partial permutation matrix, and the domain can be defined by

$$\mathcal{D} := \left\{ \mathbf{X} | \sum_i \mathbf{X}_{ia} \leq 1, \sum_a \mathbf{X}_{ia} = 1, \mathbf{X}_{ia} = \{0, 1\} \right\}. \tag{1}$$

Without loss of generality, it is assumed that $M \leq N$ hereafter. The pre-calculated differences between vertex attributes are stored in an attribute cost matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, where $\mathbf{A}_{ia}$ denotes the distance between the attribute of the ith vertex in $\mathcal{G}$ and that of the ath vertex in $\mathcal{H}$. Based on the above mathematical representations, the correspondence between features can be found by minimizing the following graph matching objective function:

$$\begin{aligned} &\mathbf{X}^* = \min_{\mathbf{X}} \ \alpha \sum_{i=1}^{d} \|\mathbf{G}^i - \mathbf{X}\mathbf{H}^i\mathbf{X}^T\|_F^2 + (1-\alpha)\mathrm{tr}(\mathbf{A}^T\mathbf{X}), \\ &\text{s.t } \mathbf{X} \in \mathcal{D}. \end{aligned} \tag{2}$$

The weight parameter $\alpha \in [0, 1]$ is used to balance the pairwise terms $\sum_{i=1}^{d} \|\mathbf{G}^i - \mathbf{X}\mathbf{H}^i\mathbf{X}^T\|_F$ and the unary term $\mathrm{tr}(\mathbf{A}^T\mathbf{X})$. The above optimization problem is a non-deterministic polynomial (NP) hard high-order combinatorial optimization problem with factorial computational complexity, for which the approximate method is necessary. We use the GNCCP,[23] a combinatorial optimization framework based on the continuous method, to approximately solve the problem. To use the GNCCP, first the discrete domain $\mathcal{D}$ needs to be relaxed to its convex hull $\mathcal{C}$, which is defined by

$$\mathcal{C} := \left\{ \mathbf{X} | \sum_i \mathbf{X}_{ia} \leq 1, \sum_a \mathbf{X}_{ia} = 1, \mathbf{X}_{ia} \in [0, 1] \right\}. \tag{3}$$

And the GNCCP also makes use of the property that $\mathcal{D}$ is exactly the extreme point set of its convex hull $\mathcal{C}$. An extreme point in a convex set is the point which does not locate in any open line segment between two points in the set; for more introduction, please refer to ref. [3]. First the original optimization problem (2) is approximated by a relatively simple convex optimization problem over the continuous domain $\mathcal{C}$, and step by step it is implicitly transformed to be a concave optimization problem over $\mathcal{C}$. Note by a clever design, both the above convex optimization problem and concave optimization problem have exactly the same global optimum as (2) over the discrete domain $\mathcal{D}$. And the optimum point of the concave optimization problem over $\mathcal{C}$ lies in its extreme point set, that is, $\mathcal{D}$ by the property mentioned above. Therefore, a discrete assignment matrix could be automatically obtained when the GNCCP terminates at the concave optimization problem.

In each step of the GNCCP process, the subproblem is optimized by the conditional gradient descent method,[14,16] also known as the Frank–Wolfe algorithm. The gradient of the original function (2) can be deduced following ref. [27], which takes the following form:

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

$$\nabla = \alpha \sum_{i=1}^{d} (2\mathbf{X}(\mathbf{H}^{iT}\mathbf{X}^T\mathbf{X}\mathbf{H}^i + \mathbf{H}^i\mathbf{X}^T\mathbf{X}\mathbf{H}^{iT}) - 2(\mathbf{G}^i\mathbf{X}\mathbf{H}^{iT} + \mathbf{G}^{iT}\mathbf{X}\mathbf{H}^i)) + (1-\alpha)\mathbf{A}. \tag{4}$$

The solution $\mathbf{X}^*$ indicates the assignments between feature points in two undersea images. As mentioned in Sections 1 and 2, once these assignments are obtained, the projective transformation between the images can be estimated. Before the projective transformation estimation, we first employ the maximum-likelihood estimation sample consensus (MLESAC)[34] to refine the assignments, or say to remove the outlier assignments. The MLESAC is a generalization of the famous RANSAC.[13] Different from RANSAC, it aims at the solution which maximizes the likelihood instead of the number of inliers, and is particularly appropriate for the estimation of complex surfaces or more general manifolds from points.[34] Then the projective transformation matrix between two images is estimated based on the refined inlier assignments. If a frame sequence sampled from, for example, a video clip is provided, the projective transformation matrices are estimated sequentially following a similar way in our previous work.[22]

5. Blending Based on MRF MAP

After the projective transformation estimation, the undersea images are then warped and blended together to get a clean and stable background panorama, which at the same time contains no seam and ghosting. The blending is defined on the MRF by a pixel labeling problem, which means assigning each pixel in the panorama a label from a limited discrete set. All the pixels together with their neighboring relations are represented by an MRF, and two pixels $k, l \in \{1, 2, \ldots, P\}$ with a neighborhood relationship are denoted by $kl \in \mathcal{N}$, where P is the pixel number and $\mathcal{N}$ denotes the neighborhood set. The label used in this paper is the undersea image sequence number, and it is denoted by $e \in \{1, 2, \ldots, Q\}$, where Q is the label number. The assignments between the pixels and the labels are represented by a labeling matrix $\mathbf{Y} \in \{0, 1\}^{P \times Q}$, where $\mathbf{Y}_{ke} = 1$ denotes labeling the pixel k by the label e. Under certain assumptions on the panorama, the labeling matrix can be obtained by minimizing the following energy function, which is widely used in energy minimization problems[4,7]

$$E(\mathbf{Y}) = \sum_k U(\mathbf{Y}_{ke})\mathbf{Y}_{ke} + \sum_{kl \in \mathcal{N}} V(\mathbf{Y}_{ke}, \mathbf{Y}_{lf})\mathbf{Y}_{ke}\mathbf{Y}_{lf}. \tag{5}$$

The pixel-wise energy $U(\mathbf{Y}_{ke})\mathbf{Y}_{ke}$ penalizes labeling the pixel k by the label e, where

$$U(\mathbf{Y}_{ke}) = \begin{cases} 0 & \text{if the } e\text{th image is available for pixel } k, \\ \infty & \text{otherwise.} \end{cases} \tag{6}$$

The term implies that the labels for the pixels in the non-overlapped area are directly determined by the pixel-wise term, and for the overlapped area the labels only need to be chosen from the sequence numbers of associated undersea images, which are further determined by $V(\mathbf{Y}_{ke}, \mathbf{Y}_{lf})\mathbf{Y}_{ke}\mathbf{Y}_{lf}$. The neighbor-wise potential $V(\mathbf{Y}_{ke}, \mathbf{Y}_{lf})$ is designed based on the abovementioned assumptions, and one most popular way for its construction is as follows:

$$V'(\mathbf{Y}_{ke}, \mathbf{Y}_{lf}) = \begin{cases} \text{Dist}(I(k, e), I(k, f)) + \text{Dist}(I(l, e), I(l, f)) & \text{if } kl \in \mathcal{N}, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $I(k, e)$ means the appearance descriptor, such as the gray value or RGB value, of pixel k in the eth image, and $\text{Dist}(\cdot, \cdot)$ denotes the distance measure between two pixel appearance descriptors. Note if using proper distance measure, for example, the Euclidean distance, $V'(\mathbf{Y}_{ke}, \mathbf{Y}_{lf})$ is known to be a submodular function, which is appropriate to be minimized by the state-of-the-art graph cuts algorithm. Generally, this potential prefers the neighborhood from the same image and seam area from the most similar images, and thus its main advantages are removal of ghosting and least visible seam. An interesting characteristic is that the potential considers the discontinuity between neighboring pixels to be reasonable, which helps to include more foreground objects in the panorama. However, as mentioned in Section 1, it can hardly deal with the local disturbance, for example, moving object and ocean current, in the undersea image. Therefore, in this paper, the energy function is given by

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

$$V(\mathbf{Y}_{ke}, \mathbf{Y}_{lf}) = V'(\mathbf{Y}_{ke}, \mathbf{Y}_{lf}) + V''(\mathbf{Y}_{ke}, \mathbf{Y}_{lf}), \tag{8}$$

with the additional term $V''(\mathbf{Y}_{ke}, \mathbf{Y}_{lf})$ defined by

$$V''(\mathbf{Y}_{ke}, \mathbf{Y}_{lf}) = \begin{cases} \text{Dist}(I(k, e), I(l, f)) & \text{if } kl \in \mathcal{N}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

By the way, it is set $V''(\mathbf{Y}_{ke}, \mathbf{Y}_{lf}) = 0$ when $kl \notin \mathcal{N}$, but actually no matter 0 or any other constant all imply that the labels of k and l are independent. An advantage of using 0 is the highly sparse form of the matrix $\mathbf{B}$ in (11). Minimizing only the potential $V''(\mathbf{Y}_{ke}, \mathbf{Y}_{lf})$ means that the neighboring pixels in the panorama should have coherent appearance descriptors, which helps to suppress the local disturbance and result in a clean and stable background panorama. However, with $V''(\mathbf{Y}_{ke}, \mathbf{Y}_{lf})$, the energy function $V(\mathbf{Y}_{ke}, \mathbf{Y}_{lf})$ is no longer a submodular function, even when $\text{Dist}(\cdot, \cdot)$ adopts the Euclidean distance measure. Though graph cuts can be applied to the non-submodular energy function,[18,30] by truncating nonsubmodular terms, that is, replacing them by submodular approximations and minimizing the latter terms. But this technique in practice only works when there are few violating submodular terms, and with more nonsubmodular terms, the performance of the graph cuts would significantly deteriorate. Therefore, similar to the above graph matching problem, we also resort to the GNCCP because of its generality. Before introducing the optimization process, first $E(\mathbf{Y})$ is transformed to the following IQP problem in the matrix form:

$$\begin{aligned} &\mathbf{Y}^* = \arg\min E(\mathbf{Y}) = \arg\min \text{vec}(\mathbf{Y})^T \mathbf{B} \text{vec}(\mathbf{Y}), \\ &\text{s.t. } \mathbf{Y} \in \mathcal{S},\ \mathcal{S} := \left\{ \sum_e \mathbf{Y}_{ke} = 1, \mathbf{Y}_{ke} = \{0, 1\} \right\}, \end{aligned} \tag{10}$$

where $\text{vec}(\mathbf{Y})$ denotes the row-wise vectorized $\mathbf{Y}$ and the matrix $\mathbf{B} \in \mathbb{R}^{PQ \times PQ}$ is defined by

$$\mathbf{B}_{(k-1)Q+e,(l-1)Q+f} = \begin{cases} U(\mathbf{Y}_{ke}) & \text{if } k = l \text{ and } e = f, \\ \dfrac{1}{2} V(\mathbf{Y}_{ke}, \mathbf{Y}_{lf}) & \text{if } k \neq l \text{ or } e \neq f. \end{cases} \tag{11}$$

Note the pixel-wise energy $\sum_k U(\mathbf{Y}_{ke})\mathbf{Y}_{ke}$ is actually $\text{vec}(\mathbf{Y})^T \text{diag}(\mathbf{B})$, where the operation $\text{diag}(\mathbf{B})$ means transforming the diagonal entries in $\mathbf{B}$ to a vector. When $\mathbf{Y}_{ke}$ takes value from $\{0, 1\}$, there is $\text{vec}(\mathbf{Y})^T \text{diag}(\mathbf{B}) = \text{vec}(\mathbf{Y})^T \text{diag}(\mathbf{B}) \text{vec}(\mathbf{Y})$[1], and it is the reason why the unary terms are located in the diagonal entries of $\mathbf{B}$. In the definition of $\mathcal{S}$, it means by $\sum_e \mathbf{Y}_{ke} = 1$ that each pixel is assigned to only one label.

To apply the GNCCP, the IQP problem (10) is first relaxed to a continuous optimization problem as follows:

$$\begin{aligned} &\mathbf{Y}^* = \arg\min E(\mathbf{Y}) = \arg\min \text{vec}(\mathbf{Y})^T \mathbf{B} \text{vec}(\mathbf{Y}), \\ &\text{s.t. } \mathbf{Y} \in \mathcal{T} := \left\{ \sum_e \mathbf{Y}_{ke} = 1, \mathbf{Y}_{ke} = [0, 1] \right\}, \end{aligned} \tag{12}$$

where the continuous domain $\mathcal{T}$ is the convex hull of $\mathcal{S}$. Then similar to graph matching, the GNCCP starts from a convex optimization problem, which is a convex approximation of (12), and gradually transforms it to a concave optimization problem, of which the optimal point over $\mathcal{T}$ is exactly the optimal point of (12). Consequently, a final discrete solution is automatically obtained when the GNCCP terminates. Each subproblem in the GNCCP process is also optimized by the conditional gradient descent method, with the gradient function given by

$$\nabla E(\mathbf{Y}) = 2\mathbf{B}\text{vec}(\mathbf{Y}). \tag{13}$$

6. Experimental Evaluation

The proposed method is first evaluated on the synthetic images by comparing it with some popular methods, and then applied to real-world undersea images.

[1]Note the equation does not hold when $\mathbf{Y}$ is relaxed to the continuous domain introduced in (12).

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

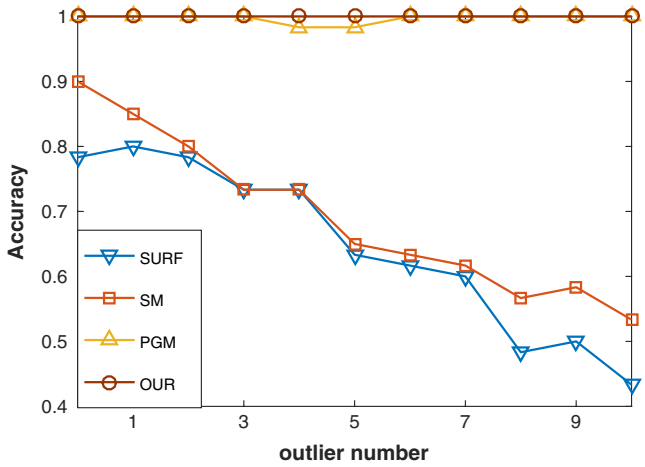

Fig. 2. Feature correspondence accuracy on manually labeled undersea images.

6.1. On synthetic images

Before the application to real-world images, it is necessary to assess some claimed properties of the proposed method.

The first simulation on the synthetic images is to assess the feature correspondence performance based on graph matching in the registration procedure. The methods for comparison include the unary method using only the SURF descriptor, the spectral matching (SM) method,[20] the probabilistic spectral graph matching (PGM) method,[9] and the proposed method denoted by OUR. The latter three methods are all graph matching-based methods, for which the SURF descriptor is used as the vertex attribute, the distance between feature points is used as edge weight, and the graph structure is built by the Delaunay triangulation technique. The weight parameter in Eq. (2) is set to be 0.5. Note there are no ground truth assignments for feature points automatically extracted from the undersea images by instance the Laplacian of Gaussian (LOG) operator. Therefore, for a quantitative comparison, an undersea image data set with manually labeled points[36] is used. This data set consists 9 undersea image pairs, which are labeled with 20 ground truth inlier points and 10 outliers as shown in Fig. 3.

The matching accuracies of different methods are compared with respect to the outlier number, which is increased from 0 to 10 by a step size of 1. The quantitative comparison result is illustrated in Fig. 2, from which it can be observed that the proposed method outperforms the other ones. By incorporating the structure cues into the SURF unary descriptor, the three graph-matching algorithms all improve the performance compared with the unary method. Some matching instances by the four methods are given in Fig. 3.

The second simulation is to assess whether the proposed method could result in a clean and stable background panorama, which is mainly related to the blending procedure. Therefore, the methods included for comparison include the feathering method, the multiband blending method, the graph cuts method, and the proposed one denoted by OUR. The feathering method is implemented by ourselves, while the multiband blending method and the graph cuts method are implemented by publicly available codes. For all these methods, the synthetic images are registered by the graph-matching-based procedure. For the synthetic images, as shown in Fig. 4, they are generated by placing an additional portunid crab image patch, respectively, in two undersea images fetched from an undersea image sequence data set[15] which is also used in the following experiment.

The stitching results are illustrated in Fig. 5. The ghosting and artifact phenomena are observed by the feathering method in the two portunid crab image patches related areas, which is consistent with our prediction. Despite also taking a weighted average way, the multiband blending method overcomes the ghosting and artifact problems, but leaves one portunid crab image patch in the panorama. Though a different one, the panorama by the graph cuts method also contains a portunid crab image patch. The latter observation is reasonable because the submodular energy for the graph cuts method, that is (7), prefers neighboring pixels from the same images and allows the discontinuity, but this idea is inconsistent with our goal of a clean and stable background panorama. Last, the proposed method avoids the influence of the "moving portunid crab" and obtains a relatively clean background

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

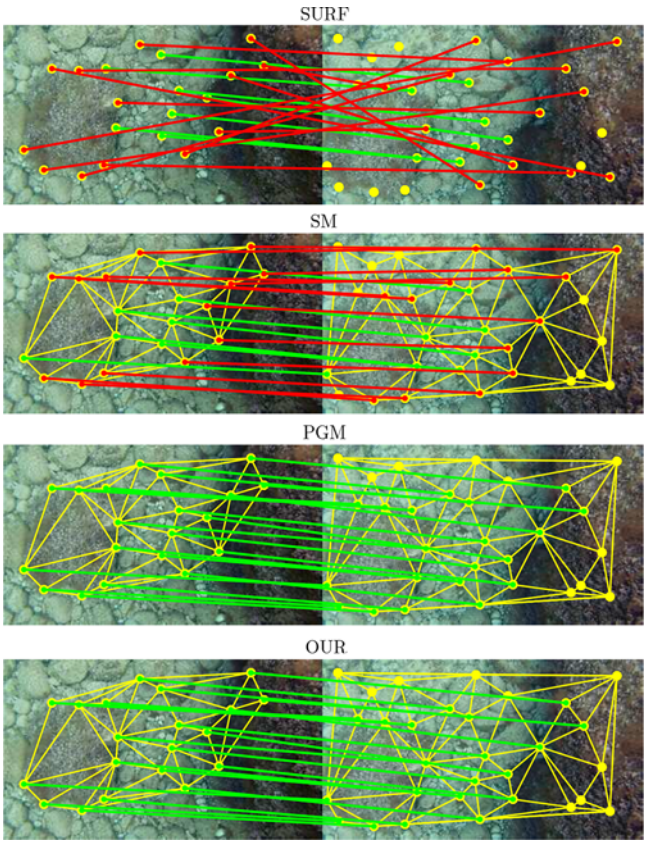

Fig. 3. Feature correspondence instances on manually labeled undersea images. The manually labeled points are shown in yellow, the correct assignments are shown in green, and the incorrect assignments are shown in red.

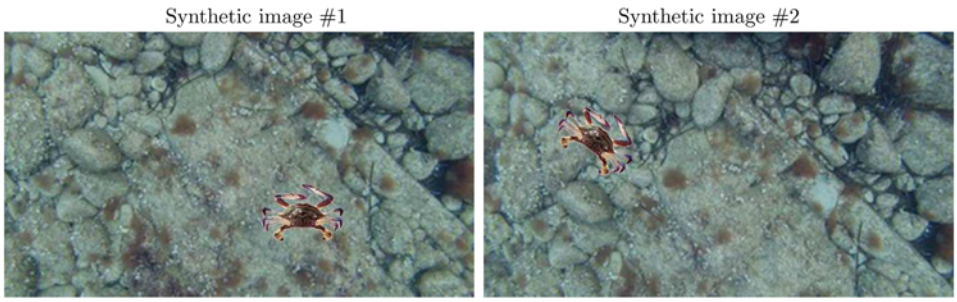

Fig. 4. Synthetic images. An additional portunid crab image patch is placed in two undersea images, with different locations and orientations.

panorama. And it should be noted that its cost is some residual pixels of the portunid crab image patch, which are similar to the background, together with some blurring pixels in the seam areas.

6.2. On real-world undersea images

In this experiment, the proposed method is evaluated on two real-world undersea image data sets. The widely used stitching system, that is, SURF-based registration and multiband blending, is used as a comparison method, which is denoted by CM standing for *classical method*. Before the final computational performance comparison, the visual performance of the two methods are first compared.

The first real-world data set is an undersea image sequence shot at the Valldemossa harbor seabed (Mallorca, Spain).[15] This total data set contains 201 images with 320×180 pixels, which forms a loop around a central point. In each image, the key points are extracted by SURF[1] together with the

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

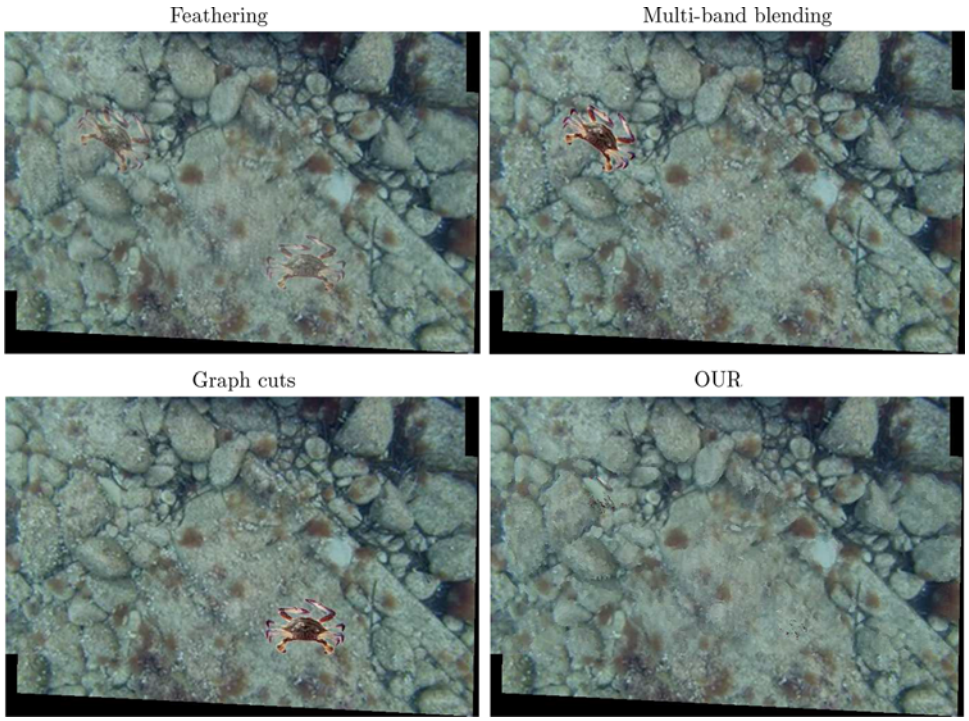

Fig. 5. Synthetic image-stitching results.

descriptors. The graphs are constructed as described in Section 4. Specifically, the key points are represented by the graph vertices, with the SURF descriptor as the vertex labels. The graph structure, that is, the set of the graph edges, is built by the Delaunay triangulation technique.[24] The length and orientation of each link in the graph structure are used as the two-dimensional edge weight. After the registration procedure, all the images are blended in a batch manner for convenience, which would lead to high storage expense. Therefore, as illustrated in Figs. 6, eight image samples in the sequence are selected and stitched in this simulation, with sequence number interval to be five.

The stitching result is shown in Fig. 7. Obvious seam can be observed for the CM method, together with slight displacements of certain component images in the corresponding panorama. The displacement is mainly due the less robust feature correspondence using only the unary information, while the visible seam is related to the multiband blending which is essentially based on the weighted average idea. On the other hand, despite a large sequence number interval, the proposed method illustrates smooth transition across the images, which is attributed to the two graph-based procedures. Note some distortions can be observed in some parts of the panorama, which are because of the projection transformation model in the image registration procedure.

Then the two methods are applied to another undersea data set[2] sampled from the video released by the French Research Institute for Exploitation of the Sea (IFREMER), which is shot along the Mid-Atlantic Ridge in the North Atlantic Ocean. The samples to be stitched are illustrated in Fig. 8. The experimental setting are the same with the above simulation. The stitching results are given in Fig. 9, which validate the effectiveness of the proposed method.

Finally the computational performance of the proposed method is evaluated by comparing the running time with respect to the number of images to be stitched. The comparison platform is a desktop computer with Intel i5 CPU (dual cores and 2.2 GHz for one core) and 8 GB RAM. The classical method, that is, the abovementioned SURF-based registration and multiband blending, is implemented by publicly available C++ codes, while the proposed method is implemented by Matlab. The comparison result is shown in Table II, from which it can be observed that the proposed method spends more time, and neither of the two methods is real time, especially the proposed method. As the two methods are implemented by different programming languages, the running time is further

[2] The data set is named by ODEMAR, which is available at https://github.com/emiliofidalgo/bimos

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

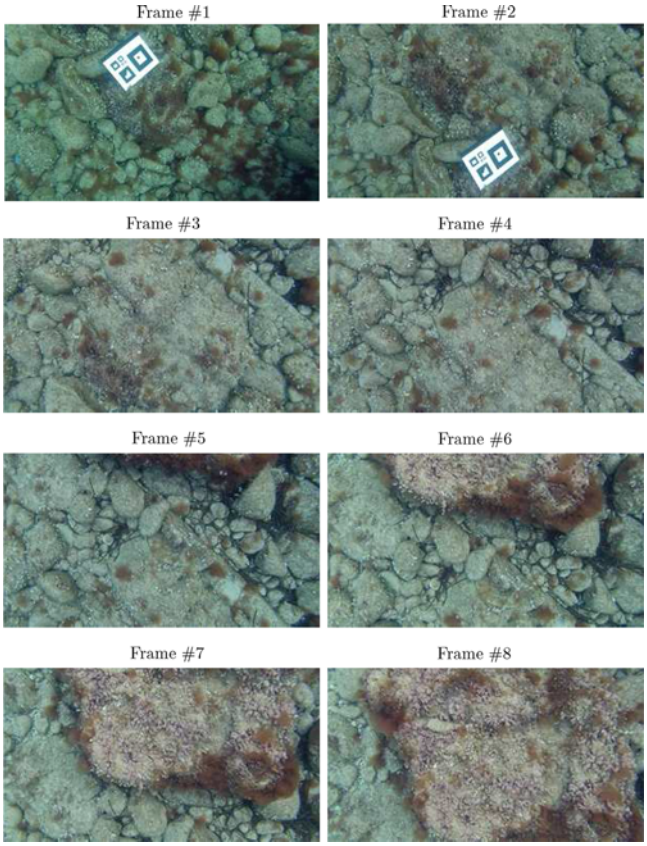

Fig. 6. Eight samples selected from undersea image sequence shot at Valldemossa harbor seabed.

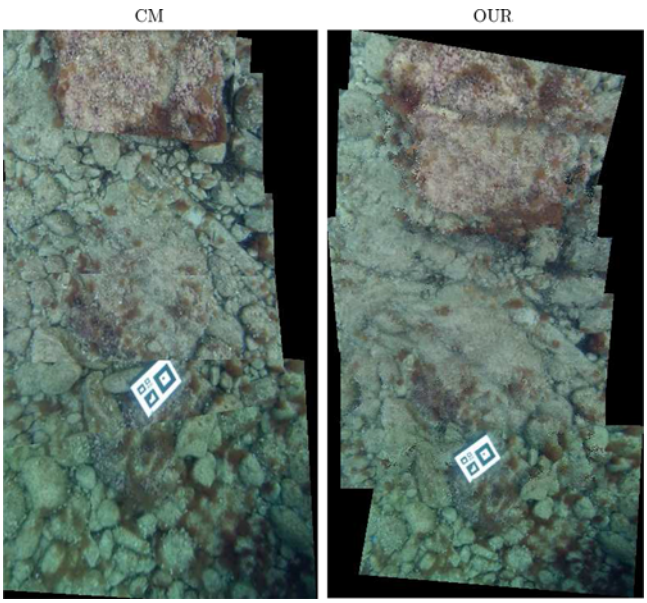

Fig. 7. Stitching result on undersea image sequence shot at Valldemossa harbor seabed.

compared in a logarithmic manner and the result is shown in Fig. 10. The slope rates indicate the changing trend and speed with respect to the number of images. It can be observed that the proposed method involves a higher computation complexity than the competitor, which is mainly due the adoption of the graph-based models.

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

Table II. Running time comparison result. CM1 and CM2 denote the results obtained by the classical method, respectively, on the two real-world undersea image data sets, while OUR1 and OUR2 denote those by the proposed method. The running time is measured in seconds.

Number of images	2	3	4	5	6	7	8
CM1	15.87	44.31	72.74	167.65	258.03	273.70	381.06
OUR1	2.18	3.10	3.17	3.97	4.89	6.39	7.10
CM2	2.32	6.90	7.66	9.06	11.80		
OUR2	33.24	90.01	165.67	278.26	542.96		

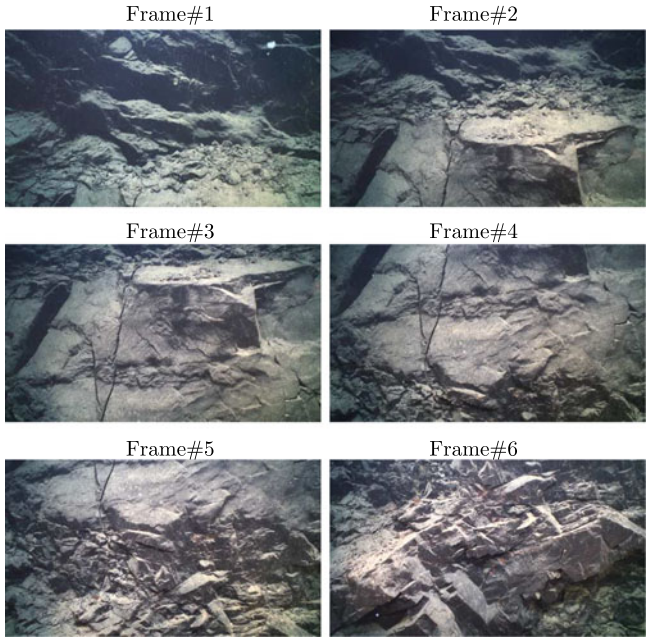

Fig. 8. Six samples selected from undersea images shot by IFREMER.

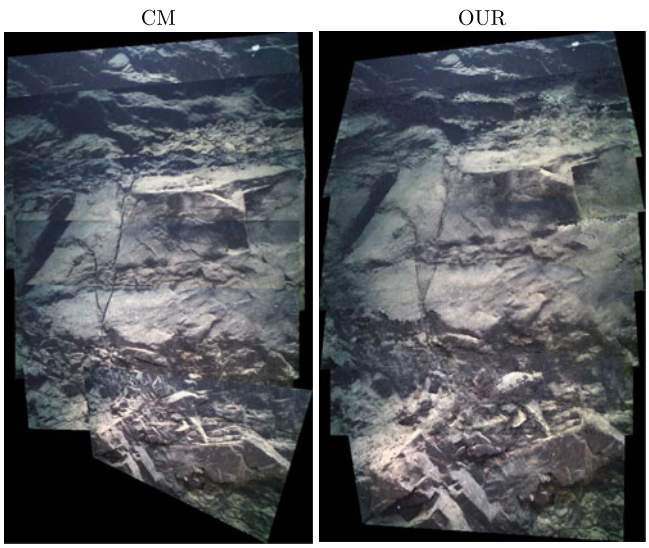

Fig. 9. Stitching result on the undersea images shot by IFREMER.

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

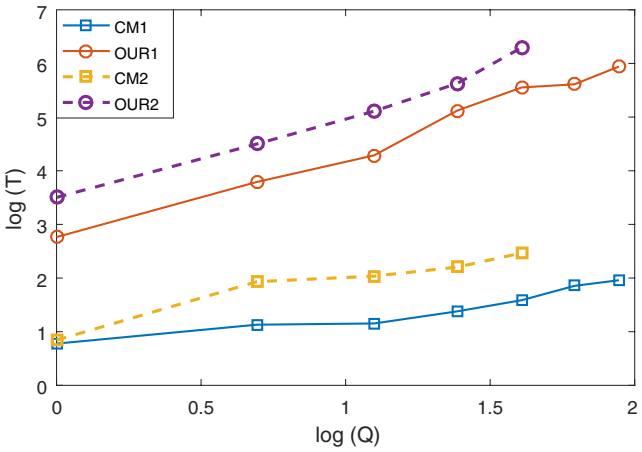

Fig. 10. Running time comparison in logarithmic manner. The running time is measured in seconds, and Q denotes the number of images to be stitched. CM1 and CM2 denote the results obtained by the classical method, respectively, on the two real-world undersea image data sets, while OUR1 and OUR2 denote those by the proposed method.

7. Conclusion and Future Works

This paper aims at a clean and stable background panorama in the undersea environment, and proposes to introduce graph-based registration and blending procedures to, respectively, address the appearance ambiguity and local disturbance problems. Experiments are performed on both synthetic images and real-world undersea images. It has been shown that the goal of a clean and stable background panorama is generally realized by the proposed method. The future works may focus on two aspects, with the first one to deal with the blurring phenomenon in the seam areas and the second one to reduce the computational complexity and storage expense of the proposed method. Besides, the proposed method could be generalized to other environments with less discriminative appearance descriptors.

Acknowledgment

The authors thank the associate editor and anonymous reviewers whose comments greatly improved the paper. This work is supported partly by the National Natural Science Foundation (NSFC) of China (grants 61633009, 61503383, U1613213, 61627808, 91648205, and U1509212), partly by the National Key R&D Program of China (grants 2016YFC0300801 and 2017YFB1300202), and partly supported by the Strategic Priority Research Program of Chinese Academy of Sciences (grant XDB32000000).

References

1. H. Bay, T. Tuytelaars and L. Van Gool, "SURF: Speeded Up Robust Features," *Proceedings of the 7th European Conference on Computer Vision*, Austria (2006) pp. 404–417.
2. S. Belongie and J. Malik, "Matching with Shape Contexts," *Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries*, Hilton Head Island, SC, USA (2000) pp. 20–26.
3. S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, UK, 2004).
4. Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222–1239 (2001).
5. M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int'l. J. Comput. Vision.* **74**(1), 59–73 (2007).
6. M. Calonder, V. Lepetit, C. Strecha and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," *European Conference on Computer Vision* (Springer, Berlin, Heidelberg, 2010) pp. 778–792.
7. T. Cour and J. B. Shi, "Solving Markov Random Fields with Spectral Relaxation," *Proceedings of the International Conference on Artificial Intelligence and Statistics*, San Juan, Puerto Rico (2007)
8. J. Davis, "Mosaics of Scenes with Moving Objects," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, USA (1998) pp. 354–360.
9. A. Egozi, Y. Keller and H. Guterman, "A probabilistic approach to spectral graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(1), 18–27 (2013).

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press

10. A. Elibol, R. Garcia and N. Gracias, "A new global alignment approach for underwater optical mapping," *Ocean Eng*. **38**(10), 1207–1219 (2011).
11. F. Ferreira, G. Veruggio, M. Caccia, E. Zereik and G. Bruzzone, "A Real-Time Mosaicking Algorithm Using Binary Features for ROVs," *Proceedings of the Mediterranean Conference on Control and Automation*, Chania (2013) pp. 1267–1273.
12. G. Finlayson, S. Hordley, G. Schaefer and G. Y. Tian, "Illuminant and device invariant colour using histogram equalisation," *Pattern Recogn.* **38**(2), 179–190 (2005).
13. M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*. **24**(6), 381–395 (1981).
14. M. Frank and P. Wolfe, "An algorithm for quadratic programming," *Naval Res. Logis. Q.* **3**(1–2), 95–110 (1956).
15. E. Garcia-Fidalgo, A. Ortiz, F. Bonnin-Pascual and J. P. Company, "Fast Image Mosaicing Using Incremental Bags of Binary Words," *Proceedings of the IEEE International Conference on Robotics and Automation*, Stockholm, Sweden (2016) pp. 1174–1180.
16. M. Jaggi, "Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization," *Proceedings of the International Conference on Machine Learning*, Atlanta, GA, USA (2013) pp. 427–435.
17. V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell*. **28**(10), 1568–1583 (2006).
18. V. Kolmogorov and C. Rother, "Minimizing nonsubmodular functions with graph cuts a review," *IEEE Trans. Pattern Anal. Mach. Intell*. **29**(7), 1274–1279 (2007).
19. A. Leone, C. Distante, A. Mastrolia and G. Indiveri," A Fully Automated Approach for Underwater Mosaicking," *OCEANS* (2006) pp. 1–6.
20. M. Leordeanu and M. Hebert, "A Spectral Technique for Correspondence Problems Using Pairwise Constraints," *Proceedings of the IEEE International Conference on Computer Vision*, Beijing, China (2005) pp. 1482–1489.
21. M. Leordeanu and M. Hebert, "Efficient MAP Approximation for Dense Energy Functions," *Proceedings of the International Conference on Machine Learning*, Pittsburgh, Pennsylvania, USA (2006).
22. C. Li, Z. Y. Liu, X. Yang, H. Qiao and J. H. Su, "Stitching contaminated images," *Neurocomputing* **214**, 829–836 (2016).
23. Z. Y. Liu and H. Qiao, "GNCCP - graduated nonconvexity and concavity procedure," *IEEE Trans. Pattern Anal. Mach. Intell*. **36**(6), 1258–1267 (2014).
24. J. D. Loera, J. Rambau and F. Santos, *Triangulations: Structures for Algorithms and Applications* (Springer-Verlag, Berlin, Heidelberg, 2010).
25. D. G. Lowe, "Object Recognition from Local Scale-Invariant Features," *Proceedings of IEEE International Conference on Computer Vision*, Kerkyra, Greece, vol. 2 (1999) pp. 1150–1157.
26. A. Mills and G. Dudek, "Image stitching with dynamic elements," *Image Vision Comput*. **27**(10), 1593–1602 (2009).
27. T. P. Minka, "Old and New Matrix Algebra Useful for Statistics," Technical report (2001).
28. P. Ravikumar and J. Lafferty, "Quadratic Programming Relaxations for Metric Labeling and Markov Random Field Map Estimation," *Proceedings of the International Conference on Machine Learning*, Pittsburgh, Pennsylvania, USA (2006) pp. 737–744.
29. J. Sun, N. N. Zheng and H. Y. Shum, "Stereo matching using belief propagation," *IEEE Trans. Pattern Anal. Mach. Intell*. **25**(7), 787–800 (2003).
30. R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen and C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell*. **30**(6), 1068–1080 (2008).
31. C. Q. Tang, G. Y. Tian, X. T. Chen, J. B. Wu, K. Li and H. Meng, "Infrared and visible images registration with adaptable local-global feature integration for rail inspection," *Infrared Phys. Technol*. **87**, 31–39 (2017).
32. G. Y. Tian, D. Gledhill and D. Taylor, "Comprehensive interest points based imaging mosaic," *Pattern Recogn. Lett*. **24**(9), 1171–1179 (2003).
33. P. H. S. Torr, "Solving Markov Random Fields Using Semi Definite Programming," *Proceedings of the International Conference on Artificial Intelligence and Statistics*, Key West, Florida, USA (2003).
34. P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vision Image Understand*. **78**, 138–156 (2000).
35. M. Uyttendaele, A. Eden and R. Skeliski, "Eliminating Ghosting and Exposure Artifacts in Image Mosaics," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, HI, USA (2001) pp. II509–II516.
36. X. Yang, Z. Y. Liu, H. Qiao, Y. B. Song, S. N. Ren and S. W. Zheng, "Underwater image matching by incorporating structural constraints," *Int'l. J. Adv. Robot. Syst*. to appear, https://doi.org/10.1177/1729881417738100.

https://doi.org/10.1017/S0263574719000699 Published online by Cambridge University Press