Matching of Large Images Through Coupled Decomposition

Panagiotis Sidiropoulos, *Member, IEEE*, Jan-Peter Muller, *Member, IEEE*

***Abstract*— In this paper, we address the problem of fast and accurate extraction of points that correspond to the same location (named tie-points) from pairs of large-sized images. First, we conduct a theoretical analysis of the performance of the full-image matching approach, demonstrating its limitations when applied to large images. Subsequently, we introduce a novel technique to impose spatial constraints on the matching process without employing subsampled versions of the reference and the target image, which we name coupled image decomposition. This technique splits images into corresponding subimages through a process that is theoretically invariant to geometric transformations, additive noise, and global radiometric differences, as well as being robust to local changes. After presenting it, we demonstrate how coupled image decomposition can be used both for image registration and for automatic estimation of epipolar geometry. Finally, coupled image decomposition is tested on a data set consisting of several planetary images of different size, varying from less than one megapixel to several hundreds of megapixels. The reported experimental results, which includes comparison with full-image matching and state-of-the-art techniques, demonstrate the substantial computational cost reduction that can be achieved when matching large images through coupled decomposition, without at the same time compromising the overall matching accuracy.**

***Index Terms*— Image matching, high-resolution imaging, image registration, image decomposition.**

I. Introduction

OVER the last few years, smartphones and digital cameras have made tens-of-megapixel images available to the general public, while new domains, such as high-resolution planetary mapping, have resulted in the release of images that reach up to several gigapixels (see [1]). As expected, the size increase raises scalability and computational complexity issues that make imperative the re-design of a number of image processing approaches.

Image matching is an essential step in multiple image processing applications, including image registration, image retrieval, stereo and 3D reconstruction. When dealing with large images, feature-based matching techniques are typically used. The latter employ extracted features (either in the form of corresponding tie-points [**?**], [**?**], [2], [3] or characteristic image curves [4], [5]) and perform image matching using characteristics of these features, thus limiting the computational cost required to accomplish this task.

The number of features is expected to increase with the image size, thus even the feature-based matching of large images becomes a computationally cumbersome task. State-of-the-art techniques designed to ameliorate image matching computational speed are mostly focused on the orthogonal issue, i.e. handling large datasets of "normal-sized" images [8]–[10]. All these techniques perform full-image matching, i.e. they don't impose any spatial constraints on the matching process. On the contrary, they ignore the actual image areas from where the tie-points are extracted, even though these may guide the matching process by defining the range and size of the tie-point neighbourhood, thus reducing the required time for matching large images.

When dealing with large images the most common way to circumvent the substantial computational cost is through pyramidal schemes. These are techniques that employ a number of sub-sampled versions of the reference and the target image to propagate matching results from the coarsest to the finest resolution. However, pyramidal schemes are fully dependent on the successful matching of all intermediate resolution versions, which is a non-trivial task for a large range of images (e.g. images of low contrast, images with a small number of distinctive features, etc.). If there are any mistakes made at any one level of the pyramid, these mistakes will be hugely amplified at the finest resolution level.

Additional challenges arise when images have partially changed. For example, when dealing with planetary images of the same area acquired at different times, it is straightforward that due to natural processes such as dust deposition or geological phenomena, the images may not be perfectly co-aligned. In such applications, it is desirable to have a sub-image ranking according to the matching quality that was achieved. In this way, a measure is attained that enables a discrimination between un-changed and potentially-changed areas. The latter may be subsequently either discarded from the matching output or become the input to a semantic analysis stage aiming to recognise the change (e.g. this may signify some unidentified dynamic natural processes [9], [10]).

In order to address the aforementioned issues, we present a novel adaptive image manipulation technique, which we name "coupled image decomposition". This is a two-step, iterative, global-to-local algorithm. At each iteration, the corresponding images (or sub-images) are initially coupled, before a concurrent image decomposition process generates multiple corresponding sub-images. The coupled decomposition

Manuscript received August 12, 2014; revised December 15, 2014; accepted February 24, 2015. Date of publication March 4, 2015; date of current version April 6, 2015. This work was supported in part by the Science and Technology Facilities Council through the Mullard Space Science Laboratory under Grant ST/K000977/1 and in part by the European Union's Seventh Framework Programme (FP7/2007-2013) through the IBM Multimedia Analysis and Retrieval System under Grant 607379. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Quanzheng Li.

The authors are with the Mullard Space Science Laboratory, University College London, London WC1E 6BT, U.K. (e-mail: p.sidiropoulos@ucl.ac.uk; j.muller@ucl.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2409978

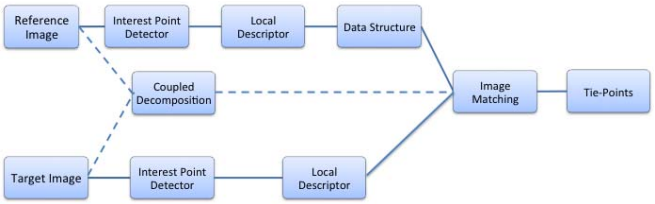


Fig. 1. Flowchart of the image matching pipeline from image input until tie-point extraction. The solid-lines in the flowchart corresponds to the typical processing used in image matching. The dashed lines show the newly introduced parts of the flowchart.

algorithm output defines an image neighbourhood for each feature in the first image and restricts matching into the corresponding neighbourhood in the other image. As a result, image matching is performed at the sub-image level, i.e. in pairs of sub-images that are matched independently from each other, thus significantly reducing the required computational time. Moreover, as it will be both theoretically and experimentally shown, this approach achieves an increase in the number of correctly identified tie-points, which can be used to enhance the overall matching accuracy. Finally, on top of the coupled decomposition algorithm, we introduce a technique that exploits the information redundancy of sub-image matching to improve the automatic estimation of the epipolar geometry as well as assess the sub-image matching quality.

The rest of the paper is organized as follows: related work is presented in Section II, followed by a theoretical analysis of the shortcomings of full-image matching of large images in Section III. Subsequently, the coupled decomposition method is presented in Section IV. In Section V, image matching and fundamental matrix estimation using coupled decomposition is introduced, while experimental validation is performed in Section VI. Section VII concludes the work.

II. Related Work

A. Tie-Point Matching

A standard tie-point matching pipeline for two input images, W, and Z [11] consists of two independently executed parts, i.e. the extraction of descriptive points p, q from W and Z respectively, and their subsequent matching (solid-line flowchart on Fig. 1). The latter can be considered as the definition of a function f from the set S_p ($p \in S_p$) to the set $S_q \cup \emptyset$ ($q \in S_q$) [6], based exclusively on local descriptors estimated at each point of S_p and S_q. Note that apart from the S_p subset that is mapped to $\emptyset$, which signify points of S_p that don't have a match to S_q, this function is bijective (i.e. one-to-one).

The most commonly used local descriptor in image matching tasks is the Scale-Invariant Feature Transform (SIFT) [11], which is a 128-dimensional local orientation histogram. Popular recent SIFT extensions and variations, such as SURF [12], GLOH [13] and DAISY [2] have been found to suffer from accuracy, compactness or speed loss respectively when compared with SIFT. As a matter of fact, there is evidence that SURF is not as accurate as SIFT in image matching applications [2], while GLOH and DAISY are slower than SIFT [13], [14], the latter being also less compact since it employs 200-dimension feature vectors.

No matter which descriptor is employed, when dealing with either large datasets of small images or very large images, the most time-consuming stage of the pipeline is point matching. Actually, the computational cost of a brute-force search (i.e. that examines all $(p, q) \in (S_p X S_q)$) may become prohibitive, while the use of efficient algorithms such as kd-tree search [15] and ball-tree search [16] are undermined by the high SIFT dimensionality [11], due to the curse of dimensionality [17].

The developed approaches for reducing the computational cost of point matching are designed either for applications in which the target image is compared to a large number of reference images or for applications in which the target image and the (single) reference image are exceptionally large. The former approaches typically perform matching through fast approximate nearest neighbour algorithms, such as priority search [18]. While recent variations limit the approximation error either by employing multiple kd-trees having between them a constant angle offset [7] or through a bi-directional image matching scheme [8], such techniques have non-trivial approximation errors, which may be tolerated or not depending on the application.

On the other hand, matching of large images usually involves some sort of pyramidal scheme in order to reduce the computational cost. Pyramidal schemes employ a set of sub-sampled versions of the reference and the target image. Matching is performed by propagating results from the coarsest to the finest level, in order to limit the computational time. In [19] the full images are matched only at the coarsest resolution, with the results being propagated and updated at progressively finer resolution. In a recent pyramidal technique, which was introduced in [20], coarse level registration results are used to decompose the images into 4 corresponding sub-images (using the center as the splitting point), which are subsequently independently matched in a finer level. This iterative algorithm starts from small sub-sampled images (128×128 or 256×256 pixels) and doubles the resolution at each step, until reaching the actual image dimensions. Finally, a similar approach was proposed in [21]. Matching is performed in a progressively finer resolution, while the matchings that are established at the coarsest levels are used to decompose the image planes into corresponding sub-images through Voronoi tessellation.

Pyramidal approaches follow the assumption that both the images and the image descriptors are robust to extreme scale differences (which in [20] case may reach up to 10^7). Unfortunately, this is not generally true, since theoretical scale invariance can be achieved only when the image resolution satisfies the Nyquist sampling criterion [22]. For a large range of images, including the majority of remote sensing products, this criterion is far from being satisfied, since it would require spatial frequency much more coarser than the actual ones. Additionally, this approach generates a more sparse set of tie-points than matching at full resolution, since the latter are typically extracted from extrema (e.g. corners) and the generation of coarse resolution versions is equivalent to image blurring. As a result, pyramidal schemes often fail to approximate the correct registration at the coarsest levels, an

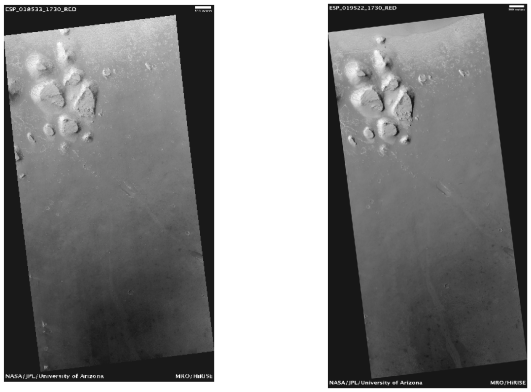
error that is propagated to the finer ones, thus contaminating the final matching outcome.

However, a technique that decomposes the reference and the target image into corresponding sub-images would be very useful not only for alleviating the overall computational cost but also for controlling the matching false acceptance and false rejection rates. This becomes apparent from the fact that typically a match is declared if the ratio of the distance of a query point p from its second nearest q_2 neighbour to the distance from its nearest neighbour q_1 ($p \in S_p$, $q_1, q_2 \in S_q$) is above a threshold. This approach raises scalability issues. since the second nearest neighbour distance over the nearest neighbour distance ratio depends on the size of S_q. When S_q is large enough and point p matches to q_1 then there is a high probability that by mistake q_2 is so similar to p that it will be erroneously omitted. On the contrary, when the set S_q is small and point p doesn't match to q_1 then it is possible by mistake that q_2 will be less similar than is required to correctly discard the match. Consequently, an increased false rejection rate is expected when the size of S_q is large enough, while an increased false acceptance rate occurs when the size of S_q is rather small. A global ratio threshold may bias the algorithm towards achieving a low false acceptance ratio even when a false rejection ratio is of major importance, e.g. when dealing with textureless or low-contrast images. This is especially the case when RANSAC [23] (or an analogous algorithm) can successfully prune erroneous matches in a post-processing step when the false rejection-false acceptance trade-off must be carefully examined so as to as to limit the number of correct matches that are missed (errors of omission).

B. Matching of Remote Sensing Spacecraft Images

Remotely sensed images of large areas of the earth and planetary surfaces constitute an image sub-class with several special characteristics, which render their matching a very challenging task. Firstly, matching accuracy is of fundamental importance. Actually, any matching error is likely to result in distortions being reprojected in geocoded images and/or maps, which are going to be used in tasks that require high accuracy. For example, images taken by Mars orbiters, with a resolution on the order of metres per pixel, have been extensively used as part of the selection of landing sites for Mars lander and rover missions [24], [25]. It is likely that any error in the employed maps would jeopardize the mission outcome. As a result, approaches that might compromise accuracy are avoided as far as one is able to.

On the other hand, a typical remote sensing image is on the order of tens-to-hundreds of megapixels, and can reach up to 10 gigapixels (e.g. such as the 25cm high-resolution Mars images from the HiRISE camera [1]). Thus, brute-force tie-point matching are very likely to result in a prohibitive computational cost. The combined requirements for high accuracy and reliability with a realistic computational time leads to the manual decomposition of the input images into corresponding, non-overlapping sub-images. However, such a task is cumbersome, un-scalable and subject to human errors. Alternatively, a pyramidal scheme may be employed (see [19]). However, pyramidal schemes induce severe performance degradations when dealing with texureless areas, which are very common in planetary data, since the scenery often involves extended flat or dusty areas such as valleys, with a small number of distinctive features (Fig. 2).



(a) (b)

Fig. 2. An example of corresponding image pairs from Martian imagery.

Moreover, remote sensing images arise from sampling of a 2D projection of the 3D land surface. As a result, the remote sensing image is independent from the sampling frequency of the input image (i.e. the pixel resolution) only if the latter satisfy the Nyquist sampling criterion [22]. In a typical remote sensing scenario, this implies very fine pixel resolution, i.e. on the order of millimetres. Such a resolution is not currently available. Consequently, in this type of image, a scaling variance due to aliasing is to be expected.

Image matching of remote sensing images, especially planetary images of bodies with atmospheres, also has to deal with the fact that the planet's surface is exposed to unpredictable local changes, due to weather or dust/aerosol/pollution phenomena [26], seasonal processes [9], clouds, possible impacts by large meteorites which do not burn up in the thin atmosphere, etc. These are often impossible to model due to the currently limited understanding of the natural mechanisms that trigger them. Apparently, it would be desirable to identify such local changes, not only to discard them from the image matching process but additionally because they refer to an unknown natural process whose extent is unknown.

Finally, remote sensing images are commonly acquired using pushbroom cameras, i.e. line cameras that acquire images perpendicular to the spacecraft flight direction, which scan distinct surface areas as the spacecraft flies forward [27]. It has been shown that matching images from pushbroom cameras only partially satisfy typical linear epipolar constraints, i.e. the epipolarity curves may not be considered linear but only partially linear [27]. This property provides additional merit for an algorithm that would automatically decompose image pairs into corresponding sub-images, as the one that is described analytically in the next section.

III. Full-Image Matching of Large Images

Full-image matching of large images is associated with an increased computational cost, since image matching by default

has quadratic computational complexity. In this section we are going to show that full-image matching of large images is also associated with degraded accuracy, which can not be improved through parameter tuning.

This analysis derives from the fact that matches are typically identified through the nearest neighbour distance ratio (NNDR), i.e. the ratio of the distance to the nearest neighbour over the distance to the second nearest neighbour. For example, in the original SIFT publication [11] a point in the reference image was declared matched if NNDR was less than a threshold $H = 0.8$, while it was further claimed that the exact threshold value is not critical on the matching performance.

When matching two sets of feature points coming respectively from the reference and the target image, it is fair to assume that the distances between a feature point in the reference image that does not have any match in the target image and all N_W feature points in the target image are random, i.e. coming from a distribution $f_D(d)$ (the corresponding cdf is $F_D(d)$). Then the joint distribution of the distance to the nearest neighbour d_1 and the second nearest neighbour d_2 is [28]:

$$f_{D_1,D_2}(d_1,d_2) = N_W(N_W-1)F_{D_1}^{N_W-2}(d_1)f_{D_1}(d_1)f_{D_2}(d_2) \tag{1}$$

where $d_1 \le d_2$. The probability P of correct rejection is the probability of $d_1 \le Hd_2$, i.e.,

$$\begin{aligned} P &= N_W(N_W-1)\int_{-\infty}^{\infty} F_{D_1}^{N_W-2}(d_1)f_{D_1}(d_1) \\ &\left(\int_{d_1}^{Hd_1} f_{D_2}(d_2)dd_2\right)dd_1 \\ &= \int_{-\infty}^{\infty} F_{D_1}^{N_W-2}(d_1)(F_{D_1}(Hd_1)-F_{D_1}(d_1))f_{D_1}(d_1)dd_1 \end{aligned} \tag{2}$$

By expanding Eq. (2) we get:

$$\begin{aligned} P = N_W(N_W-1)[&\int_{-\infty}^{\infty} F_{D_1}(Hd_1)F_{D_1}^{N_W-2}(d_1)f_{D_1}(d_1)dd_1 \\ &-\int_{-\infty}^{\infty} F_{D_1}^{N_W-1}(d_1)f_{D_1}(d_1)dd_1] \end{aligned} \tag{3}$$

Since $\int_{-\infty}^{\infty} N_W F_{D_1}^{N_W-1}(d_1)f_{D_1}(d_1)dd_1 = 1$ the above equation is re-written as:

$$\begin{aligned} P = N_W(N_W-1)&\int_{-\infty}^{\infty} F_{D_1}(Hd_1)F_{D_1}^{N_W-2}(d_1)f_{D_1}(d_1)dd_1 \\ &-(N_W-1) \end{aligned} \tag{4}$$

Finally, $[F_{D_1}(Hd_1)F_{D_1}^{N_W-1}(d_1)]' = Hf_{D_1}(d_1)F_{D_1}^{N_W-1}(d_1) + (N_W-1)F_{D_1}(Hd_1)F_{D_1}^{N_W-2}(d_1)f_{D_1}(d_1)$, so

$$P = N_W(1-H\int_{-\infty}^{\infty} f_{D_1}(d_1)F_{D_1}^{N_W-1}(d_1)dd_1) - (N_W-1) \tag{5}$$

or

$$P = 1 - N_W H\int_{-\infty}^{\infty} f_{D_1}(d_1)F_{D_1}^{N_W-1}(d_1)dd_1 \tag{6}$$

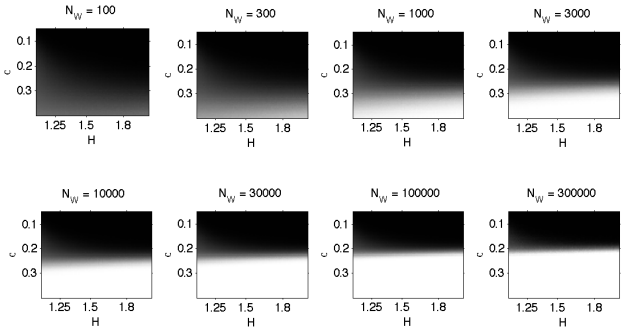


Fig. 3. The F-score dependence from the tie-point number N_W and the NNDR threshold H as a function of the non-matching distance standard deviation σ. Black-coloured pixels correspond to F-score=1, white-coloured pixel to F-score=0 and grey-coloured pixels to all intermediate values, using linear colour stretch.

Consequently, the mis-detection ratio E_{md} is

$$E_{md} = N_W H\int_{-\infty}^{\infty} f_{D_1}(d_1)F_{D_1}^{N_W-1}(d_1)dd_1 \tag{7}$$

Eq. (7) implies that the mis-detection ratio is determined by the distribution $f_D(d)$, the threshold H and the number of tie-points in the target image N_W. Following a similar rational we can conclude that the false rejection ratio and the mis-classification ratio is also determined by the above parameters, along with the distribution of the matching feature point distances $g_D(d)$. Consequently, this parameterisation can be employed to examine matching performance dependence from the number of points N_W and the NNDR threshold H.

In this implementation, it is assumed that correct matches show consistently low distance values, thus $g_D(d)$ was selected as a uniform distribution with minimum value 0.01 and maximum value 0.1. On the other hand, $f_D(d)$ was supposed to be a normal distribution, with mean value 0.5 (since distance takes a value within the range [0, 1]) and variable standard deviation σ. The above selections reduce the number of degrees of freedom to 3: the standard deviation σ, the threshold H and the number of tie-points in the target image N_W.

The matching quality measure in this case is the F-score, i.e. the harmonic mean of the information retrieval measures, Recall and Precision [29]. F-scores were estimated using the above model by sampling N_W, H and σ. More specifically, F-score was estimated for N_W equal to 100, 300, 1000, 3000, 10000, 30000, 100000 and 300000. On the other hand, σ samples were taken in the interval [0.05, 0.4] using a step 0.005 and H in the interval [1.1, 2] using a step 0.01.

As a result, for each N_W value, a matrix M of dimension 71×91 was retrieved. $M(i, j)$ value is the estimated F-score when $\sigma = 0.05 + 0.005(i-1)$ and $H = 1.1 + 0.01(j-1)$. These tables were quantised and visualised in Figure 3. Black-coloured pixels correspond to $F = 1$, while white-coloured pixels to $F = 0$. All intermediate F-score values are drawn using a linear grey-scale stretch, i.e. F-score x corresponds to grey level $255x$. We selected to demonstrate these results as grey-level images because the purpose of this synthetic experiment is not to give quantitative F-score estimates but rather to examine the qualitative dependence of F from N_W and H.

As expected, Fig. 3 confirms that higher standard deviation of the non-matching distance scores results in lower performance. As a matter of fact, σ expresses the inherent matching ambiguity that varies according to the content of the image, the camera parameters, the imaging conditions, etc as well as the capability of the employed descriptor to index the local image content. A rough discrimination can be generated by labeling matchings with low σ values as "inherently easy" while matchings with high σ values as "inherently hard".

Fig. 3 shows that while the former class seems rather robust to the number of tie-points N_W, the latter's performance is severely degraded when the number of tie-points increase. Consequently, a full-image matching of large images is exposed to unpredictable parameters (such as those modeled through σ) and for a significant range of images (e.g. remote sensing spacecraft images) is expected to lead to poor matching performance. Moreover, the matching performance can not be tuned through NNDR threshold H, since H values do not seem to play a critical role in the matching performance, especially for large N_W values. As a matter of fact, Fig. 3 implies that avoiding the selection of a very low H value seems to be the only necessary H tuning.

This inability of H to effect accurate matching of large images suggests the need to limit the number of candidate matches. The latter can be achieved by imposing spatial constraints on the matching process. In the next section we introduce such a technique, named coupled decomposition.

IV. Coupled Image Decomposition

The image matching pipeline is illustrated in Fig. 1. The solid lines denote the previously established process while the dashed ones represent the ones introduced in this work for the coupled decomposition step. As already discussed in Section II-A, in the state-of-the-art approach, initially interest point detection is conducted on the reference and the target image (e.g. through a Harris-Laplace detector [30]), before processing the extracted informative image points to generate local descriptors (usually SIFT [11]). Thus, a point is projected from image space to descriptor space. During matching, each target image descriptor is matched with a reference image descriptor or discarded if no match is found (possibly using a tree structure to facilitate efficient queries), and the matches are reprojected from the descriptor space to the image space. This approach does not impose any spatial constraints on the matching process, which is fully conducted in the descriptor space, while pyramidal schemes impose spatial constraints based on sub-sampled versions of the images, a strategy that is problematic when dealing with large images. Instead, in this section we introduce a novel technique, named coupled decomposition, imposing spatial constraints on pairs of images using only the full resolution images, which subsequently are matched.

A. Line and Point Invariants

We assume that Z and W are the reference and the target input images, respectively. As a matter of fact Z, (as well as W) is used to represent both the reference (target) image per se as well as its luminance at a point p_i (q_j), in which case this is denoted as $Z(p_i)$ ($W(q_j)$), or $Z(x_i, y_i)$ ($W(x'_j, y'_j)$) if the point coordinates are given. Moreover, for a pair of points p_i and p_j, θ_{ij} is the angle between the vector $p_i p_j$ with the horizontal line that pass through p_i.

Using the above nomenclature, the profile $\bar{S}(p_i, \theta)$ is defined as the set of points p_j of Z for which $\theta_{ij} = \theta$ or $\theta_{ij} = \theta + \pi$, i.e. the set of points of a line passing from p_i. We further define the mean profile $\bar{Z}(p_i, \theta, \Xi)$ as the mean luminance of the profile points:

$$\bar{Z}(p_i, \theta, \Xi) = E[p_j], \quad p_j \in Z, \ \theta_{ij} = \theta \ or \ \theta_{ij} = \theta + \pi \quad (8)$$

In the above equation $E[.]$ is the expected value operator. Moreover, since the only available information is the image luminance in integer coordinates, the mean profile depends on the employed interpolation method, Ξ. Nevertheless, since the luminance variation caused by different interpolation is usually small and is further reduced through averaging, we assume that $\bar{Z}(p_i, \theta, \Xi_1) \approx \bar{Z}(p_i, \theta, \Xi_2) = \bar{Z}(p_i, \theta)$ for two different interpolation methods Ξ_1, Ξ_2.

Through the mean profile, we define for each point p a function $f_{\bar{Z},p}(\theta)$ that maps each angle θ to the mean profile $\bar{Z}(p, \theta)$. If two images, Z and W, differ only in terms of geometric transformations and additive noise (i.e. Z can be reproduced from W after a set of geometric transformations and noise addition) and if p and q are two corresponding points in Z and W (i.e. p is where point q is mapped through the geometric transformations) then the following equation stands:

$$f_{\bar{Z},p}(\theta) = f_{\bar{W},q}(\theta + \phi) \quad (9)$$

In the above equation ϕ signifies the orientation angle difference of W and Z.

Finally, in order to find two corresponding points in Z and W, two different approaches are proposed, depending on the image matching scenario. More specifically, when Z and W completely overlap Z and W centroids are selected as corresponding points, since geometric transformations and additive noise leaves the image centroid invariant, i.e.:

$$\begin{aligned}(\sum x_i Z(x_i, y_i), \sum y_i Z(x_i, y_i)) \\ = (\sum x'_j W(x'_j, y'_j), \sum y'_j W(x'_j, y_j))\end{aligned} \quad (10)$$

However, when the reference and the target image overlap only partially then the reference and the target image centroid are not expected to correspond. When dealing with partially overlapping image pairs, a 1-element sample of the set of matches is proposed to be used as the corresponding point. More specifically, this is estimated by matching one by one the feature points extracted from Z with the complete set of feature points extracted from W, until an unique match is identified. The matching point coordinates in Z and W determine the pair of corresponding points. In the rest of this paper, the two variations are going to be named mean-based coupled decomposition and match-based coupled decomposition, respectively.

B. Coupled Image Decomposition Algorithm

The coupled image decomposition algorithm is presented in Alg. 1. The input of the algorithm consists of the reference

Algorithm 1 Coupled Image Decomposition Algorithm

Input: A reference image Z, a target image W and two corresponding (root) points p_1, q_1.

1: (Coupling) Estimate $f_{\bar{Z},p_1}(\theta)$ and $f_{\bar{W},q_1}(\theta)$ and the angle ϕ that maximizes their correlation.

2: (Decomposition) Decompose images Z and W in M radial sections Z_1, Z_2, ..., Z_M and W_1, W_2, ..., W_M. More specifically, the pixel p_i that belongs to Z is assigned to Z_j, $j = \lfloor M\theta_{1i}/2\pi \rfloor + 1$ while pixel q_i that belongs to W is assigned to W_j, $j = \lfloor M(\phi + \theta_{1i})/2\pi \rfloor + 1$.

3: If the algorithm has reached the maximum number of iterations, K, return the decomposition of both images into corresponding sub-images. If not, for each new pair of sub-images repeat steps 1 and 2.

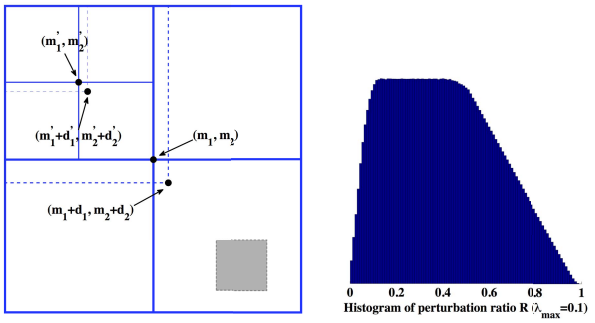
image, Z and the target image, W. The coupled decomposition algorithm starts by estimating the corresponding points p and q respectively and by estimating the profile functions $f_{\bar{Z},p}(\theta)$ and $f_{\bar{W},q}(\theta)$ on them. As a matter of fact, since the profile functions are continuous, we employ a quantised version of them, which aggregates the mean profiles for angle ranges of $\Delta\theta$. As $\Delta\theta$ gets smaller the quantised profile function becomes the profile functions introduced in the previous subsection.

From Eq. 9 it is assumed that $f_{\bar{W},q}(\theta)$ is a (circularly) translated version of $f_{\bar{Z},p}(\theta)$. The angle translation parameter ϕ is the one that maximizes the profile vector correlation. Consequently, a polar coordinate system fit in p with the axis at angle θ (relative to a horizontal line passing through p) corresponds to a polar coordinate system fit in q with the axis at angle $\theta + \phi$. This property is used to decompose images into corresponding pairs of sub-images, Z_i and W_i.

The sub-images Z_i (as well as W_i) are actually radial sections of the polar co-ordinate system that is defined on the point p (q). For simplicity reasons, we use a constant angular range for each radial section. For example, if the corresponding point is the point p_1 and 4 radial sections are used then the range of each radial section is $\pi/2$ and the first, second, third and fourth sub-image would include the image points p_i for which $0 \leq \theta_{1i} \leq \pi/2$, $\pi/2 \leq \theta_{1i} \leq \pi$, $\pi \leq \theta_{1i} \leq 3\pi/2$ and $3\pi/2 \leq \theta_{1i} \leq 2\pi$, respectively.

For each of the corresponding sub-images, the coupled decomposition process is iterated, until the maximum number of iterations is reached. The algorithm output is a set of flags that determine for each W and Z pixel the sub-image that they belong to. If W and Z differ only in terms of geometric transformations and additive noise, then points in Z_i would correspond to points in W_i and vice versa. Consequently, instead of matching the complete images Z and W, image matching can be performed on a set of corresponding sub-images Z_i and W_i.

The parameters that control coupled decomposition output are the quantization step $\Delta\theta$, the number of iterations K and the number of radial sections M. In order to tackle cases in which the differences between Z and W are not limited to geometric transformations and additive noise, the sub-images Z_i, W_i are enlarged by a factor $(1 + a)$ before matching. The overlap ratio a is the fourth parameter of this algorithm.



(a) (b)

Fig. 4. (a) Z image plane and the centroid perturbations when decomposing Z coupled to W' instead of W. The grey rectangle in the bottom-right corner represents the image region that $W - W'$ differ, while the solid and the dashed lines the sub-image boundaries when decomposing Z coupled to W and W', respectively. (b) $|R|$ mean value histogram when using image pair 1 (Table I) and 10^5 runs. At each run λ_1 was randomly selected ($\lambda_{max} = 0.1$) and pixels were randomly changed in the bottom-right corner until it was reached. Then λ_2, λ_3 was estimated from the image, and finally R from Eq. 14.

C. Algorithm Properties

1) Robustness and Invariance: Mean-based coupled decomposition is by default invariant to image rotation, translation and scaling. Moreover, the estimation of both centroid points and profile vectors in Algorithm 1 includes exclusively mean pixel values. As a result, the mean-based coupled decomposition outcome is expected to be invariant to both additive white noise and global radiometric differences due to different lighting conditions or transparency of the atmosphere. Similar conclusions can be drawn for match-based coupled decomposition, except from the non-theoretical invariance to geometric transforms. However, since the corresponding points are established through a single matching pair of points, elaborate features and matching can be employed to confirm that the corresponding points would be robust to geometric transforms.

Match-based coupled decomposition is additionally robust to local changes, i.e. changes that happen only to a limited image area, since such changes are expected to have limited effects both on the determination of the corresponding points and on the points' mean profiles. On the other hand, robustness to local changes is not straightforward for the mean-based coupled decomposition variation.

Actually, in order to examine this property, we assume that two target images, W and W', are to be matched with a reference image, Z. We further assume that Z and W differ only in terms of geometric transformations, additive noise and global radiometric differences, i.e. that by using Z and W we can achieve perfectly accurate mean-based coupled decomposition (meaning that no points in Z_i, W_j, $i \neq j$ refer to the same point in world coordinates). Additionally, W' is considered identical to W, except an area in the bottom-right corner. In Fig. 4a, the reference image plane is demonstrated ($M = 4$), along with the Z sub-image boundaries for both input pairs, i.e. Z-W and Z-W' (solid and dashed lines, respectively). We focus this analysis on the top-left sub-image Z_1, examining

whether the boundary "error" triggered by the modification in the bottom-right sub-image Z_2 is propagated to the sub-images in which Z_1 is decomposed in further coupled decomposition iterations.

Apparently, local change robustness can be examined via the ratio of the distance of corresponding sub-image centroids. More specifically, if $p_0(m_1, m_2)$ ($p'_0(m_1+d_1, m_2+d_2)$) is the image centroids of the reference image estimated through coupled decomposition of Z and W (Z and W') and $p_1(m'_1, m'_2)$ ($p'_1(m'_1+d'_1, m'_2+d'_2)$) is the corresponding top-left sub-image centroid, then a local change robustness measure R is

$$R = \sqrt{d_1'^2 + d_2'^2}/\sqrt{d_1^2 + d_2^2} \tag{11}$$

R depends on several parameters. In order to simplify the analysis, it is assumed that i) due to the modification in the bottom-right area, the centroid is displaced evenly in both coordinates, i.e. that $d_1/m_1 = d_2/m_2 = \lambda_1$ and ii) the top-left centroid p_1 coordinates are analogous to the image centroid p_0, i.e. that $m'_1/m_1 = m'_2/m_2 = \lambda_2$.

The number of pixels that lie in the top-left sub-image is $N_1 = m_1 m_2$ when decomposing Z coupled with W while $N'_1 = (1+\lambda_1)^2 m_1 m_2$ when decomposing Z coupled with W', i.e. in the second case $(\lambda_1^2 + 2\lambda_1) m_1 m_2$ additional pixels are assigned to the top-left sub-image. Consequently,

$$\begin{aligned} m'_1 + d'_1 = {} & \lambda_2 m_1 (1/(1+\lambda_1)^2) \\ & + \lambda_{31}(m_1 + d_1)((\lambda_1^2 + 2\lambda_1)/(1+\lambda_1)^2) \end{aligned} \tag{12}$$

where $\lambda_2 m_1 = m'_1$ is the top-left sub-image centroid longitude of the N_1 pixels that are common to the top-left sub-image centroid corresponding to W, and $\lambda_{31}(m_1+d_1)$ is the centroid longitude of the pixels that are assigned to the top-left sector only when Z is decomposed coupled with W'. $\lambda_{32}(m_2 + d_2)$ is the corresponding latitude. By further assuming that $\lambda_{31} = \lambda_{32} = \lambda_3$,

$$\begin{aligned} \sqrt{d_1'^2 + d_2'^2} = {} & \sqrt{m_1^2 + m_2^2}\ \lambda_1(\lambda_1 + 2) \\ & \times (\lambda_2 - \lambda_3(\lambda_1 + 1))/(1+\lambda_1)^2 \end{aligned} \tag{13}$$

$$R = (\lambda_1 + 2)(\lambda_2 - \lambda_3\lambda_1 - \lambda_3)/(1+\lambda_1)^2 \tag{14}$$

In Eq. 14, λ_1 represents the relative image centroid perturbation due to a local change, λ_2 the normalized original top-left sub-image centroid position (relative to the image centroid position) and λ_3 the centroid position of the changed pixels relative to the perturbed image centroid position.

We have experimentally confirmed the validity of the assumptions of Eq. 11, finding that in the testset employed in this work, that 98% of the values of $(d_1 m_2)/(m_1 d_2)$ vary between 0.93 and 1.07 and the 98% of the values of $(m'_1 m_2)/(' m_2 m_1)$ vary from 0.96 and 1.04. Moreover, the experimentally determined value ranges of λ_1, λ_2, λ_3 are $[0, 0.1]$, $[0.3, 0.7]$ and $[\lambda_2, 2\lambda_2]$. It can be straightforwardly computed that for this range $|R| < 1$. For example, by selecting λ_1 to be randomly selected in a range from 0 to λ_{max}, where $\lambda_{max} = 0.01, 0.02, 0.05, 0.1$ and averaging over 10^5 runs, the estimated mean $|R|$ value was 0.368, 0.372, 0.383 and 0.4, respectively. The results of a small experiment on a planetary image (Fig. 4b) are in line with the reported mean $|R|$ values. Such values signify that a local change disturbing the image centroid position is minimized by > 2.5 at each iteration. Consequently, even when the number of iterations is small, the mean-based coupled decomposition outcome is robust to moderate local changes.

Additional evidence for the method robustness is given by the fact that image matching is typically robust to small displacements of the image boundaries. As a matter of fact, if a decomposition in sector centroid p generates a sub-image Z_i with N_i pixels, C_i of which are correctly matched with pixels of W_i, and p moves by d pixels, then the number of pixels N'_i in the new sub-image W'_i would be $(N_i - d^2) \leq N'_i \leq (N_i + d^2)$. Consequently, the expected value of correctly matched pixels that will be missed is less than $C_i d^2/N_i$. If Z_i is assumed square with size d', then the expected percentage of missed pixels is less than $(d/d')^2$. In mean-based coupled decomposition, d' depends on the image size as well as the total number of iterations, K. Thus, in the case of large-sized images, the generated sub-image sizes are on the order of hundreds or thousands of pixels, which implies that even a misplacement of tens of pixels is not expected to significantly degrade the algorithm output.

However, when severe local changes are expected, match-based coupled decomposition is suggested to be employed instead.

2) Computational Concerns: The most computationally demanding part of Algorithm 1 is the interpolation required for the mean profile estimation. However, when dealing with large images, as in this work, the digital image pixel grid provides a rather dense image representation, which can be used to circumvent interpolation in the mean profile estimation step. More specifically, in order to generate the mean profile, all image pixels are parsed sequentially and the angle θ_{1i} with the corresponding image or sub-image centroid p_1 is estimated and quantized, before the pixel illumination is adjusted to the corresponding mean profile.

In this way each pixel is visited only twice at each iteration in mean-based coupled decomposition (once for the profile vector estimation and once for sector centroid estimation) and only once at each iteration in match-based coupled decomposition, since the corresponding points are estimated through matching. Actually, the overall computational complexity of mean-based coupled decomposition and match-based coupled decomposition is $O(2K(N_Z + N_W) + T_{ang})$ and $O(K(N_Z + N_W) + T_{ang} + T_{mat})$, respectively. In the above equations K is the number of iterations, N_Z and N_W the number of pixels at the reference and target image, respectively, T_{ang} the computational time related to the estimation of the angle that maximizes the correlation of profile vectors and T_{mat} the computational time needed to identify the match-based corresponding points in Z and W. In large images T_{ang} and T_{mat} is negligible, i.e. $O(2K(N_Z + N_W) + T_{mat}) \approx O(2K(N_Z + N_W))$ and $O(K(N_Z + N_W) + T_{ang} + T_{mat}) \approx O(K(N_Z + N_W))$. As is subsequently shown, even for very large-sized images, only a small number of iterations is required (typically $K \leq 5$), while both N_Z and N_W are much larger than 10^6. Consequently, the computational complexity

Algorithm 2 Tie-Point Extraction Using Coupled Decomposition

Input: A reference image Z, a target image W, an empty tie-point repository S and a sub-image index $i = 1$.
1: Extract the characteristic points of Z and W using some state-of-the-art technique (e.g. SIFT [13]).
2: Perform coupled decomposition using Algorithm 1.
3: Estimate the target set K as the intersection of M_r and M_t target sets (i.e. the sets from which M_r and M_t take values).
4: Estimate Z_i and W_i and enlarge them by $(1 + a)$.
5: Match points of Z_i and W_i and append the identified matches to S.
6: $i = i + 1$. Go to step 4.

of the coupled decomposition algorithm is linear to the sum of the total number of pixels in the reference and the target image, independently whether their size is almost equal or not. This small increase in the computational cost, caused by the coupled image decomposition, results on a significant decrease in the overall computational cost, as will be demonstrated in the experimental results section.

V. Image Matching Using Coupled Decomposition

The set of corresponding sub-images, which is the output of coupled decomposition, can be used to develop several image matching solutions that focus either on the decrease of the computational complexity required for image matching or on exploiting the information redundancy associated with the fact that all sub-images belong to the same image. In this section, we introduce two such techniques.

A. Tie-Point Extraction Using Coupled Decomposition

In this rather straightforward technique, the coupled decomposition outcome is used to impose spatial constraints on the tie-points of the target image that can be matched to the tie-points of the reference image. When no overlap is used (i.e. $a = 0\%$) this is equivalent to decomposing both reference and target images in corresponding sub-images and performing tie-point extraction independently on each pair of sub-images, before aggregating the identified tie-points. A non-zero overlap rate a can be used to determine the tradeoff between computational complexity and the number of identified tie-points. The actual algorithm is presented in Algorithm 2.

The above algorithm achieves a significant reduction in the required computational time. Supposing that in the reference image (the target image) the total number of extracted interest points are C_Z (C_W), the candidate matches that need to be checked are reduced from $C_Z C_W$ to $\sum C_{Z_i} C_{W_i}$, where i is the sub-image index. In large images, in which C_Z and C_W may be on the order of millions, this constitutes a substantial computational gain. In this section, we demonstrate that the computational gain increases with K (i.e. the total number of iterations of Algorithm 1) and decreases with overlap rate a.

In order to establish this property for mean-based coupled decomposition, it is initially assumed that $a = 0$, M is the number of radial sections per iteration, K the total number of iterations, and C_Z and C_W the total number of tie-points in Z and W, respectively. Then the average number of tie-points in each sub-image would be C_Z/M^K and C_W/M^K. Moreover, the fact that the image centroid usually lies near the image center implies that, generally speaking, C_{Z_i} and C_{W_i} are more probable to take values near their average than take extremely low or high values, i.e. that the probability density function maximum and the average value coincide. It is also fair to assume that the distribution is symmetric around its mean value since there is no generally applicable reason for $P(C_{Z_i} \leq (C_Z/M^K)) \neq P(C_{Z_i} \geq (C_Z/M^K))$ or $P(C_{W_i} \leq (C_W/M^K)) \neq P(C_{W_i} \geq (C_W/M^K))$. Finally, C_{Z_i} and C_{W_i} are not expected to be mutually independent, since coupled decomposition generates pairs of matching sub-images.

All of the above specifications are satisfied if the distribution of C_Z and C_W is a bivariate normal distribution with mean vector $(C_Z/M^K, C_W/M^K)$. In this case the mean value of the product $C_{Z_i} C_{W_i}$ is

$$E[C_{Z_i} C_{W_i}] = \varrho_{ZW} \sigma_Z \sigma_W + (C_Z C_W)/M^{2K} \tag{15}$$

where σ_Z (σ_W) is the marginal standard deviation of C_{Z_i} (C_{W_i}) and ϱ_{ZW} is the correlation coefficient. If the reference and the target image match $\varrho_{ZW} \approx 1$ and the achieved computational gain G can be expressed as

$$G = \frac{C_Z C_W}{\sum C_{Z_i} C_{W_i}} \approx M^{2K} \frac{C_Z C_W}{C_Z C_W + M^{2K} \sigma_Z \sigma_W} \tag{16}$$

If $a > 0$ and assuming that the number of extracted interest points is proportional to image size, G becomes

$$G \approx \frac{M^{2K}}{(1+a)^2} \frac{C_Z C_W}{C_Z C_W + M^{2K} \sigma_Z \sigma_W} \tag{17}$$

Eq. (17) can be refined by considering that when an already sampled population is sub-sampled, resulting in B output bins for each input bin, the output standard deviation is $\sqrt{B}$ times smaller than the input one [31]. If $\sigma_{Z_0}(\sigma_{W_0})$ is the standard deviation of the reference image (target image) the feature points distribution at the first iteration, Eq. 17 is rewritten as:

$$G \approx \frac{M^{2K}}{(1+a)^2} \frac{C_Z C_W}{C_Z C_W + M^{K+2} \sigma_{Z_0} \sigma_{W_0}} \tag{18}$$

Since the mean number of points assigned to each sub-image in the first iteration is C_Z/M and C_W/M, the final expression of computational gain G is

$$G \approx \frac{M^{2K}}{(1+a)^2 (1 + M^K C V_Z C V_W)} \approx \frac{M^K}{(1+a)^2 C V_Z C V_W} \tag{19}$$

where $CV_Z = \sigma_{Z_0} M / C_Z$, $CV_W = \sigma_{W_0} M / C_W$ are the coefficients of variation [31] for the feature point distribution of the reference and target image, respectively.

While the above analysis stands for the mean-based coupled decomposition variation it may be extended for the match-based coupled decomposition variation if the single corresponding point needed for decomposition is identified near the center of Z and W. This can be achieved if Z feature

Algorithm 3 Coupled Decomposition Enhanced Fundamental Matrix Computation

Input: A set of matched tie-points S and the coupled decomposition outcome.
1: For each pair of sub-images Z_i and W_i estimate the set $S_i \in S$ of matched tie-points that lie in Z_i and W_i respectively. Set the current number of iterations ($K' = 1$) and the minimum distance vector ($d_k = \infty$).
2: For each $i, S_i \neq \emptyset$ split randomly S_i in two parts S_{it} and S_{iv}.
3: Use S_{it} to estimate the fundamental matrix $F_i^{K'}$ through a state-of-the-art algorithm (e.g. the gold standard method of [34]).
4: Using $F_i^{K'}$ and S_{iv} as the validation set, estimate the mean reprojection error $d_i^{K'}$.
5: If $d_i^{K'} < d_i$, $d_i = d_i^{K'}$, $F_i = F_i^{K'}$.
6: If $K' < K_{max}$, $K' = K' + 1$ and go to step 2, else go to step 7.
7: Estimate the median d_m of all d_i and the value nearest the 84.1% quantile d_{m+s}. The outlier threshold T_0 is $T_0 = d_m + 2(d_{m+s} - d_m)$.
8: For each pair of sub-images Z_i and W_i, if $d_i < T_0$ return F_i, else return a flag that W_i area may be changed and the reprojection error d_i.

points are not parsed randomly but according to their distance from the image center, favouring feature points near the center of Z.

Note that the computational gain expressed in Eq. (19) refers to the number of candidate matches that are examined in a brute-force approach that doesn't employ any data structure to organize interest points. However, it is apparent that coupled decomposition is orthogonal to data structuring, thus it can be used along with data structuring to further optimize the computational gain if this is gauged necessary.

B. Partial Fundamental Matrix Estimation and Sub-Image Matching Assessment

The previous sub-section examined the use of coupled decomposition for matching of coupled decomposition. However, coupled decomposition modifies the tie-point extraction part of image matching. Consequently, coupled decomposition may be encompassed in multiple matching schemes, each with a different architecture, objectives, etc.

In this sub-section, we introduce an algorithm that deals with one of the most common image matching objectives, i.e. the estimation of the fundamental matrix [32] that describes the epipolar geometry between two images. Actually, being a sub-image matching scheme, coupled decomposition may comply with the partial epipolarity of image pairs acquired by pushbroom cameras [27], since a distinct fundamental matrix can be estimated for each sub-image. The algorithm output also includes a rank of the corresponding sub-images according to the reprojection error. Coupled decomposition output is robust to local changes (Section IV-C1), thus, these are assumed to correspond not just to sub-image misalignments but also to grey-level image changes.

The algorithm is presented in Algorithm 3. In the first 6 algorithmic steps the matching quality of each corresponding sub-image is estimated through a common state-of-the-art practice, used for validation when no ground-truth is available. More specifically, the tie-points' set is split into training and validation sub-sets. The latter is assumed to include actual matches that are going to validate the accuracy of the fundamental matrix that is computed through the former. A fixed number of iterations ensures a decent approximation of the minimum reprojection error.

The intermediate output of the first 6 steps is a reprojection error vector, d_i. If d_i is a normal distribution sample, then the mean and the median reprojection error would concur, while the 84.1% quantile error value would correspond to the sum of the mean (i.e. the median) and the standard deviation. Taking into account that 95.5% of a normal distribution $N(\mu, \sigma)$ values are smaller than $T_0 = \mu + 2\sigma$, T_0 is employed as an outlier detection threshold (μ in the above equation is the median value).

Summarising, for each sub-image pair with reprojection error higher than T_0 a region is identified as with grey-scale change (and may be further examined in applications that require the characterisation of the change), while the fundamental matrices for the rest of the sub-images are retrieved.

VI. Experimental Results

A. Datasets and Experimental Setup

The experimental dataset consists of 30 pairs of remote sensing images from orbiter missions around Mars and Phobos (Table I). The image pairs differ in relation to the camera and data type, the resolution, the image size, etc. More specifically, the 30 pairs can be classified into 6 sub-sets, named "HiRISE JPG", "CTX", "Phobos SRC", "HRSC", "Mixed" and "HiRISE".

The first and the last sub-sets come from the NASA Mars Reconnaissance Orbiter (MRO) mission camera HiRISE [1]. The former (pairs $1-6$) includes 6 pairs of subsampled images of $7-12$ megapixels, which constitute a dataset with a size similar to the one that is currently available from the latest smartphones, digital cameras, etc. The latter (pairs $28-30$) includes 3 pairs of raw HiRISE images, which depict large areas of the Martian surface with $25cm$ per pixel resolution [1], reaching up to 2.4 gigapixels. The second sub-set (pairs $7-12$) are 6 image pairs taken by the NASA MRO mission Context Camera (CTX) [33]. The resolution in this dataset is on the order of few metres per pixel (typically $6-12$ metres [33]), while their size vary from 18 to 87 megapixels. The third subset (pairs $13-18$) consists of 6 pairs of small planetary images, which are mainly used to examine the size limits under which coupled decomposition becomes redundant. These are less-than-one-megapixel Phobos images taken at a maximum resolution of 2.3 metre per pixel resolution from the ESA Super Resolution Channel (SRC) of High Resolution Stereo Camera (HRSC) which is on board Mars Express mission [34]. The latter mission is also the source of the fourth subset (pairs $19-24$). This consists of 6 pairs of HRSC [34] imagery, with a resolution of $12.5-25$ metres per pixel and a size of 27 to 220 megapixels. Finally, the fifth dataset (pairs $25-27$) includes 3 mixed pairs, i.e. an image acquired from CTX image and another from HRSC.

In the current experimental setup, four measures of comparison are employed: the computational time T, the

TABLE I
IMAGE PAIRS DATASET (SEE TEXT FOR FURTHER DETAILS)

Id	Reference Image	Target Image	Ref. Image Size	Tar. Image Size	Sub-set	Bit-depth
1	ESP_012435_2015_RED	ESP_012725_2015_RED	(4096, 2048)	(4096, 2048)	HiRISE JPG	8
2	PSP_003125_1665_RED	PSP_003191_1665_RED	(5107, 2048)	(5107, 2048)	HiRISE JPG	8
3	ESP_018533_1730_RED	ESP_019522_1730_RED	(3671, 2048)	(3671, 2048)	HiRISE JPG	8
4	ESP_017465_1730_RED	ESP_017966_1730_RED	(5921, 2048)	(5629, 2048)	HiRISE JPG	8
5	PSP_007736_2345_RED	PSP_008303_2345_RED	(6462, 2048)	(6331, 2048)	HiRISE JPG	8
6	PSP_003492_1995_RED	PSP_003558_1995_RED	(4769, 2048)	(4211, 2048)	HiRISE JPG	8
7	B01_009906_2226_XN_42N016W	B01_010196_2227_XN_42N0 16W	(7168, 5000)	(7168, 5000)	CTX	16
8	B01_009894_1665_XI_13S042W	B02_010606_1666_XN_13S042W	(8192, 5000)	(8192, 5000)	CTX	16
9	B08_012934_1073_XN_72S225W	B09_013066_1073_XN_72S225W	(17408, 5000)	(17408, 5000)	CTX	16
10	B16_015999_1534_XN_26S034W	B17_016276_1534_XN_26S034W	(7168, 5000)	(8192, 5000)	CTX	16
11	G14_023809_1416_XN_38S136W	G14_023875_1416_XN_38S136W	(5120, 3760)	(5120, 3760)	CTX	16
12	G14_023809_1416_XN_38S136W	G15_024099_1416_XN_38S136W	(5120, 3760)	(12288, 5000)	CTX	16
13	h1574_0005_sr2	h1607_0005_sr2	(750, 650)	(719, 600)	Phobos SRC	16
14	h4913_0005_sr2	h4946_0005_sr2	(1018, 1008)	(900, 900)	Phobos SRC	16
15	h5362_0005_sr2	h5381_0005_sr2	(820, 700)	(750, 600)	Phobos SRC	16
16	h4414_0004_sr2	h4447_0005_sr2	(1018, 1008)	(1018, 1008)	Phobos SRC	16
17	h3245_0005_sr2	h5277_0005_sr2	(550, 560)	(800, 500)	Phobos SRC	16
18	h6896_0042_sr2	h6896_0045_sr2	(1018, 1008)	(1018, 1008)	Phobos SRC	16
19	h2895_0000_nd3	h2895_0000_s23	(15252, 7264)	(7626, 3632)	HRSC	16
20	h2917_0000_nd3	h2895_0000_nd3	(9700, 4024)	(15252, 7264)	HRSC	16
21	ha866_0000_nd3	ha866_0000_s13	(28087, 7846)	(14044, 3923)	HRSC	16
22	h2895_0000_nd3	h2895_0000_s13	(15252, 7264)	(7626, 3632)	HRSC	16
23	h2917_0000_nd3	h2917_0000_s13	(16690, 5923)	(8345, 2962)	HRSC	16
24	h2917_0000_nd3	h2895_0000_s13	(9700, 4024)	(4802, 3632)	HRSC	16
25	HRSC-MER-B	CTX-MERB-DRG	(4981, 2901)	(9901, 4501)	Mixed	16
26	B18_016617_1562_XI_23S345W	h6488_0000_nd3	(4000, 5056)	(700, 2000)	Mixed	16
27	h1927_0000_nd3	P01_001422_1747_XN_05S222	(5000, 2700)	(9571, 4240)	Mixed	16
28	ESP_011310_1395_RED	ESP_011811_1395_RED	(62722, 27201)	(80392, 33187)	HiRISE	16
29	ESP_011476_1600_RED	ESP_014100_1600_RED	(64481, 29067)	(65567, 29523)	HiRISE	16
30	ESP_011767_1640_RED	ESP_011833_1640_RED	(47762, 15693)	(53265, 15733)	HiRISE	16

number of extracted tie-points C, the rigid mean reprojection error Err_R and the warping mean reprojection error Err_W. Rigid mean reprojection errors are used to contrast fundamental matrix estimation versus partial fundamental matrix estimation (as discussed in Section V-B), while warping mean reprojection error is used to compare the tie-point quality that is achieved from the introduced techniques with the quality achieved from full-image matching and two techniques taken from the literature, namely, [22], [23].

The evaluation of the fundamental matrix estimation is based on the fact that for the employed dataset (as well as most planetary image datasets) there is no unambiguous ground-truth to estimate the achieved accuracy. In order to circumvent this inherent impediment, we estimate the accuracy of the fundamental matrix estimated from a tie-point set, S_1, relative to another set of tie-points, S_2. This is achieved by initially splitting the sets of tie-points into a training and a testing set, S_{itr} and S_{ite}, $i \in \{1, 2\}$. The training and testing sub-sets are randomly selected so as $S_{itr} \cup S_{ite} = S_i$, $S_{itr} \cap S_{ite} = \emptyset$, $\#(S_{itr}) = \#(S_{ite})$. Subsequently, the fundamental matrix computed from S_{1tr} is used to project $S_{1te} \cup S_2$ and the fundamental matrix computed from S_{2tr} is used to project $S_{2te} \cup S_1$. The two reprojection errors are linearly combined using weights proportional to the cardinality (i.e. the number of matched tie-points) of sets S_1, S_2.

However, when comparing more than 2 methods the above measure is not practicable since it requires pairwise comparison of all methods. Consequently, tie-point quality that is achieved through some image matching method was evaluated using warping reprojection error. More specifically, for each image pair of Table I, full-image matching was performed. The total number of estimated tie-points N is a constant parameter for all methods. Since sub-image matching consistently estimates more tie-points than full-image matching, RANSAC was employed to prune these sets to N tie-points. Subsequently, the tie-point sets were randomly split into a training and a testing sub-set, the former used to estimate a 3D warping function and the latter to evaluate the mean reprojection error caused by the warping function. Finally, the reported error value is the average of the mean reprojection errors, estimated over 1000 iterations.

Apart from the parameters and algorithms that are examined in this section, the rest of the implementation is based on state-of-the-art techniques. Interest point detection was conducted through Harris-Laplace corner detection [30], SIFT [11] was selected as the local descriptor, the "gold standard" method of [32] was used for fundamental matrix estimation and least squares for warping function parameter estimation.

B. Parameter Tuning

In this subsection, we test the dependence of the tie-point matching outcome on the coupled decomposition algorithm parameters. As explained in Section IV-B, there are 4 parameters involved in coupled decomposition, K, M, a and $\Delta\theta$. The following analysis is conducted on mean-based coupled decomposition, however, the examination of match-based coupled decomposition (which is not presented here due to length limitations) led to similar conclusions, the only difference being that overlap seems redundant (i.e. $a = 0\%$ is the default value for match-based coupled decomposition).

The number of sub-images that coupled decomposition generates is M^K. Consequently, both of these two parameters determine the decomposition coarseness. M is also associated with the grid shape, while K with the robustness to local changes (Section IV-C1). When dealing with planetary images, the robustness to local changes is very important, while no specific symmetric pattern is expected. Consequently, in the following experiments, we selected a fixed, small value for M ($M = 4$), and used K to examine the dependence of tie-point performance from the number of generated sub-images.

In the experiments, we examined five alternatives for K ($1 \leq K \leq 5$), two profile vector quantization angles

TABLE II

$\Delta\theta$ IMPACT ON THE TIE-POINT MATCHING ALGORITHM. EACH SCORE IS THE RATIO OF THE MEAN SCORE WHEN $\Delta\theta = \pi/720$ DIVIDED BY THE MEAN SCORE WHEN $\Delta\theta = \pi/180$

$E[\hat{T}]$	$Max(\hat{T})$	$Max(\hat{T}*)$	$E[\hat{Err}_R]$
1.0833	1.5484	1.1277	1.0268
$E[\hat{C}]$	$Max(\hat{C})$	$Min(\hat{C})$	$\#(\hat{Err}_R < 1)$
0.9989	1.0392	0.9225	16

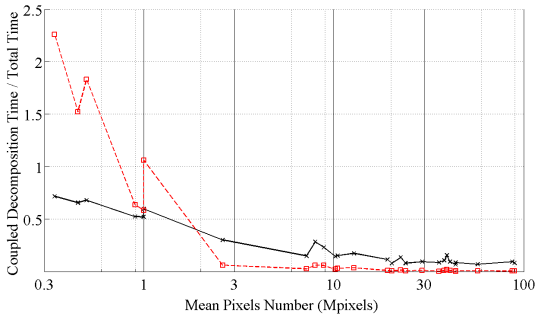


Fig. 5. A log-linear plot of the coupled decomposition runtime ratio versus the image size. Red dashed line: T_{CD}/T_0. Black solid line: T_{CD}/T_1.

($\Delta\theta \in \{\pi/180, \pi/720\}$) and three overlap ratios ($a \in \{0\%, 20\%, 50\%\}$). The parameter tests were mutually independent through the neutralization of the effect of the other two parameters. For example, when the computational time T relatively to $\Delta\theta$ was examined, then (for each image pair of Table I) $T(K, a, \Delta\theta)$ was estimated for all valid parameter combinations, and the ratio $\hat{T}$ of the mean T value of combinations with $\Delta\theta = \pi/720$ over the mean T value of combinations with $\Delta\theta = \pi/180$ was the evaluation output.

Table II summarizes the algorithm parameters as a function of the profile vector quantization angle $\Delta\theta$. The reported scores are the mean and the maximum computational time ($E[\hat{T}]$, $Max(\hat{T})$), the maximum computational time if the Phobos image pairs are ignored ($Max(\hat{T}*)$), the mean, maximum and minimum number of tie-points ($E[\hat{C}]$, $Max(\hat{C})$, $Min(\hat{C})$), the rigid mean reprojection error ($E[\hat{Err}_R]$) and the number of image pairs (out of 30 image pairs) for which the rigid reprojection error of $\Delta\theta = \pi/720$ is smaller than the reprojection error of $\Delta\theta = \pi/180$ ($\#(\hat{Err}_R < 1$). Each score is the ratio of the mean score when $\Delta\theta = \pi/720$ divided by the mean score when $\Delta\theta = \pi/180$.

Table II provides evidence that $\Delta\theta$ does not play a critical role in the algorithm performance. As a matter of fact, all three metrics (i.e. computational cost, tie-point quantity and reprojection error) seems to fluctuate randomly around 1 for all large-image pairs. Following these clues, we have decided in our implementation to use a fixed value of $\Delta\theta = \pi/720$.

It should be noted that in the small-image pair sub-set (Phobos sub-set) the relative computational time seems to increase with $\Delta\theta$. As a matter of fact, the mean computational time for this sub-set is 1.364, while all 6 pairs exhibit computational time higher than 1.199. The reason for this difference between small-image and large-image pairs is that quantization density has an effect on the required coupled decomposition time, which constitutes a significant part of the computational workload only in small-image pairs.

TABLE III

THE IMPACT OF THE OVERLAP RATIO a ON THE IMAGE MATCHING ALGORITHM. THE RESULTS ARE NORMALIZED BY DIVIDING THE ACHIEVED SCORES WITH THOSE CORRESPONDING TO 0% OVERLAP

$E[\hat{T}_{20}]$	$E[\hat{T}_{50}]$	$E[\hat{C}_{20}]$	$E[\hat{C}_{50}]$	$E[\hat{Err}_{20}]$	$E[\hat{Err}_{50}]$
1.3371	2.2602	1.5879	1.8993	0.9734	0.9147

TABLE IV

INCREASE OF COMPUTATIONAL TIME AND NUMBER OF TIE-POINTS VERSUS OVERLAP RATIO a

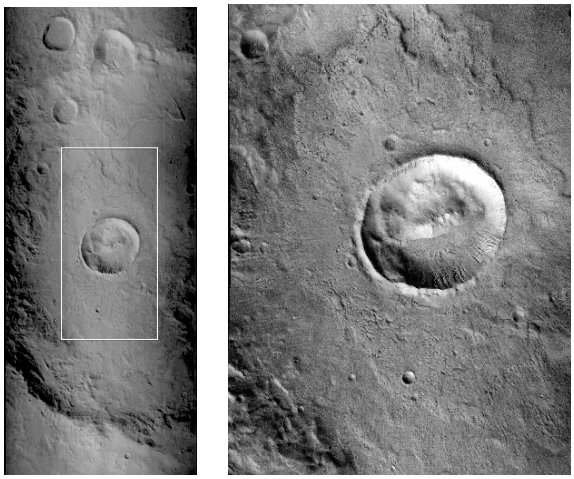
Dataset	$E[\hat{T}_{50}]$ ($E[\hat{T}_{20}]$)	$E[\hat{C}_{50}]$ ($E[\hat{C}_{20}]$
HiRISE JPG	2.502 (1.3743)	1.9471 (1.6278)
CTX	2.9938 (1.5291)	1.748 (1.522)
Phobos SRC	1.0674 (1.0141)	1.9017 (1.5744)
HRSC	2.5158 (1.4196)	1.9 (1.5834)
Mixed	2.1827 (1.3596)	2.1005 (1.6757)
HiRISE	3.1169 (1.7008)	1.8143 (1.5712)

This is further shown in Fig. 5, in which the ratio of the coupled decomposition runtime T_{CD} over the total required image matching time is plotted. When no coupled decomposition is used the total time is denoted with T_0 (T_{CD}/T_0 is shown in the red line) while when coupled decomposition is used with T_1 (T_{CD}/T_1 is shown in the black line). Runtime is plotted against "mean pair size" (i.e. the harmonic mean of the pixel number of the two images) for the first five datasets. The one excluded is the HiRISE dataset, for which T_0 is prohibitively large. As a hint, the "largest" included pair is No. 21, for which image matching without coupled decomposition requires more than 10 days in a 2.7 GHz Intel Core i7, 16Gb RAM workstation.

The size scale represented in Fig. 5 covers images from less than 1 megapixel to 100 megapixels. It can be deduced that as the size increases, coupled decomposition stage constitutes a progressively smaller part of the image matching workload. Finally, red line shows that $T_{CD} > T_0$ for images smaller than $1-2$ megapixels. Consequently, image matching computational time can benefit from coupled decomposition only when dealing with images larger than $1-2$ megapixels.

As already noted, three overlap levels were tested, 0%, 20% and 50%. Table III summarizes the overlap ratio impact in the employed experimental dataset. Actually, the reported scores in Table III are the mean computational times, the mean number of estimated tie-points and the mean reprojection error, both for 20% and 50% overlap. These scores use the same convention as in Table II, i.e. they are normalised by the corresponding scores when no overlap is used. The first conclusion that can be derived from this Table is that a high overlap ratio enhances both the quality (i.e. the reprojection error) and the quantity of the extracted tie-points, but with a large consumption of the available computational resources.

In order to get a more complete insight, the computational time and number of tie-points scores are reported separately for each dataset in Table IV. This table provides evidence that



(a) (b)

Fig. 6. Image pair 12 (Table I). The right image is a close-up of the encircled area of the left image. The image sizes are asymmetrically adjusted for printing reasons. The actual size of the right image is 1/3 of the left image.

the overlap triggers a quadratic increase in the required computational time when dealing with images of size on the order of tens of megapixels. As the image size exceeds 1 gigapixel the computational time increase becomes approximately cubic. On the other hand, the increase in the quantity of tie-points doesn't depend on the image size but on whether the specific pair setup favours a correct coupled decomposition. For example, if the local changes on the target image are so extensive that the assumptions of Algorithm 1 don't stand, then the resulting coupled decomposition output discrepancy may be alleviated through the image enlargement implied by a non-zero overlap rate. Such a case is demonstrated in Fig. 6, in which image pair No. 12 is demonstrated. The two images of pair 12 only partially match, thus the number of points increase that is achieved via an overlap of 50% is more than 120%. On the contrary, when the reference and the target image differ mostly in terms of geometric transformations, additive noise and global radiometric difference, coupled decomposition would benefit less from a non-zero degree of overlap. Consequently, selecting the employed overlap ratio a depends on an a priori estimate of the extent of the non-invariant transformations that are required to match the reference and the target image. In our implementation we selected $a = 20\%$ as a tradeoff between accuracy and speed.

The final parameter that is analysed is the number of iterations, K. As already mentioned, this is related to the required local change robustness and the false rejection ratio. Additionally, as shown in Fig. 7, computational complexity gain is controlled mainly through the tuning of K. Fig. 7 plots the logarithm of the computational time required for tie-point matching (counted in seconds) as a function of the image pair size, for different K values.

Fig. 7 implies that there is a certain "size zone", in which coupled decomposition achieves optimized speed when using a fixed number for K. So, for images on the order of megapixels $1 - 2$ iterations achieve the optimal speed, while $K = 3$ is proposed for images of size around 10 megapixels, $K = 4$ for images between 30 to 100 megapixels and $K = 5$ for images on the order of hundreds of megapixels. Finally, $K = 6$ gives the fastest image matching pipelines for image pairs with more than several gigapixels. In this work, K is tuned to the maximum score that allows each sub-image to have on average 1,000 SIFT points. This score, which is consistent with the analysis of Section III, implies that $2 \leq K \leq 6$.

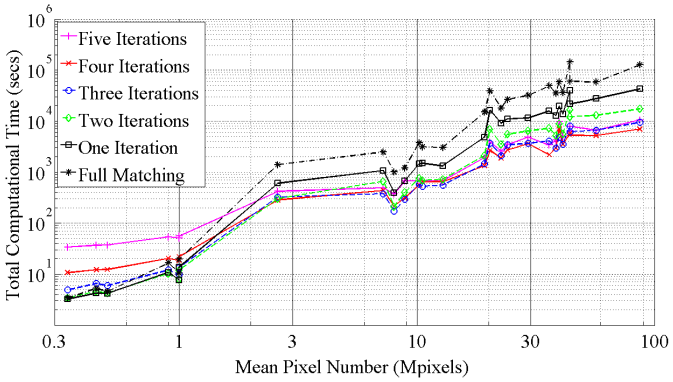


Fig. 7. The logarithm of the required computational time, counted in seconds, as a function of the pair size.

C. Method Comparison

After tuning coupled decomposition algorithms we continue with the evaluation of their performance, when comparing both will full-image matching and with other state-of-the-art techniques. More specifically, the included techniques are:

- The full-image matching algorithm (named F) that performs straightforward tie-point extraction without involving coupled decomposition or any other sub-image matching approach. This is the baseline technique.
- Mean-based coupled decomposition algorithm (named CD_c), using the parameters implied by the previous section.
- Match-based coupled decomposition algorithm (named CD_m), which also use the parameters implied by the previous section.
- The pyramidal scheme of [21] (named P_v), which establish matchings in the coarsest levels and use them to decompose the image planes into corresponding sub-images through Voronoi tessellation, and, finally,
- The pyramidal scheme of [20] (named P_r), which register the images in coarse level and decompose the registered images into 4 corresponding sub-images, using the center as the splitting point.

As already said, for all methods the number of retrieved tie-points was the same and determined by F. More specifically, full-image matching was followed by RANSAC [23] to prune any erroneous matches. The number of tie-points retrieved by this method for each image pair of the dataset is demonstrated in the second column of Table V. Exception to this is the mixed pairs (i.e., pairs 25–27), for which full image matching (as well as the state-of-the-art methods) fails. The number of tie-points in this case is the number of tie-points estimated from CD_m, which is the only non-failing technique.

TABLE V

TIE-POINT ACCURACY AND COMPUTATIONAL COST OF COUPLED DECOMPOSITION METHODS (CD_c AND CD_m) IN COMPARISON TO FULL-IMAGE MATCHING (F) AND PYRAMIDAL SCHEMES (P_v AND P_r)

		F		CD_c		CD_m		P_v		P_r	
Id	#(Points)	Time (sec)	Err_W	Time (sec)	Err_W	Time (sec)	Err_W	Time (sec)	Err_W	Time (sec)	Err_W
1	9681	1002	0.4603	353	0.4238	171	0.4204	273	**0.419**	186	0.4463
2	1706	3728	**0.5711**	880	0.7919	426	0.9395	820	0.877	388	0.8066
3	3730	2466	**0.261**	651	0.3179	388	0.3334	546	0.3245	Fail	Fail
4	8671	3102	0.4049	915	**0.3986**	488	0.4856	807	0.4547	417	0.4732
5	12456	3037	**0.3848**	1191	0.4156	621	0.3969	500	0.4431	338	1.6312
6	1528	1210	0.5607	1068	**0.5347**	1038	0.7663	865	0.6839	Fail	Fail
7	30704	49167	0.5384	8666	0.4465	6032	0.4583	5281	0.4444	5119	**0.438**
8	2401	74473	0.6285	9151	0.4921	5078	0.5213	8113	**0.4584**	Fail	Fail
9	16030	285441	2.2742	13067	0.9717	10760	**0.8797**	20364	1.3237	Fail	Fail
10	12471	64417	0.5626	5353	**0.5414**	4637	2.6318	6471	0.5496	Fail	Fail
11	28837	39699	0.585	5788	**0.5277**	4006	0.5316	4068	0.5546	Fail	Fail
12	19293	125590	17.4898	13639	**2.1919**	6127	17.6742	7397	16.4187	Fail	Fail
13	117	43	0.4058	47	**0.3876**	50	0.4572	110	0.4474	85	0.4362
14	209	67	0.6739	81	0.6961	64	**0.5894**	63	0.6469	58	0.7111
15	35	41	1.1783	52	1.0693	33	**1.0185**	74	1.1978	63	1.139
16	89	74	0.8242	53	1.0907	91	**0.6761**	86	0.7282	75	0.6844
17	11	39	70.12	44	**26.97**	44	82.6	Fail	Fail	Fail	Fail
18	660	43	0.4236	31	0.4274	27	**0.4118**	49	0.4247	44	0.4186
19	17141	23294	0.9489	2947	0.8262	2262	**0.7808**	3477	0.8185	Fail	Fail
20	10563	233756	14.8769	Fail	Fail	12170	**7.3282**	10831	8.6331	Fail	Fail
21	7590	169180	**1.007**	11217	1.2589	9635	1.0767	11281	1.0543	Fail	Fail
22	9443	29578	1.1492	2136	1.0756	1730	0.9581	2263	**0.9451**	Fail	Fail
23	23471	25446	1.1192	2803	0.9781	3416	**0.9667**	4324	0.9729	Fail	Fail
24	2845	29318	4.1544	3597	0.9654	3188	**0.7466**	2735	5.0241	Fail	Fail
25	1777	Fail	Fail	Fail	Fail	3735	**4.6791**	Fail	Fail	Fail	Fail
26	159	Fail	Fail	Fail	Fail	1625	**11.32**	Fail	Fail	Fail	Fail
27	360	Fail	Fail	Fail	Fail	11053	**8.4027**	Fail	Fail	Fail	Fail

Apart from the two first columns, Table V summarises the performance of the 5 examined techniques regarding both computational time and warping mean reprojection error. A first conclusion that can be learned from Table V is the extreme computational cost of full-image matching when dealing with large images, which is not justified by its performance. As a matter of fact, full image matching may be tens of times slower than all of the included sub-image matching techniques, while not being more accurate than them. For example, full-image matching of image pair 9 requires more than 3 days and its mean reprojection error is 2.2742 pixels, while CD_m requires almost 3 hours and its mean reprojection eror is 0.8797 pixels.

In summary, F is less accurate than CD_c in $16/27$ image pairs, than CD_m in $17/27$ image pairs, than P_v in $13/27$ image pairs and than P_r in $5/27$ image pairs. When focusing on large images the full-image accuracy is even worse, as expected from the analysis of Section 3. From the 4 largest image pairs (pairs 9, 12, 20 and 21) only for pair 21 full-image matching performs better than sub-image matching. On the other 3 pairs its performance is far worse than the maximum achieved from sub-image matching. More specifically, for pair 9 its mean reprojection error is 2.59 times larger than CD_m, for pair 12, 7.98 times larger than CD_r and for pair 20, 2.03 times worse than CD_m. The most characteristic example is pair 9, which consists of large and noisy images (due to dust in the atmosphere on the time that they were acquired). The low-quality of the images along with their large size cause the full-image matching to be much worse both in terms of accuracy and of computational cost from all examined sub-image matching techniques.

Passing on the comparison of individual techniques, from the 27 image pairs the maximum accuracy was achieved from CD_m on 12 pairs, from CD_r on 7 pairs, from F on 4 pairs, from P_v on 3 pairs and from P_r on 1 pair, i.e. the coupled decomposition technique outperformed pyramidal schemes as well as full-image matching on $19/27$ images. Additionally, while coupled decomposition is consistently much faster than full-image matching (reaching up to 26.5 times for image pair 9 and match-based coupled decomposition) for large images, it is faster than pyramidal schemes in half of the examined pairs. More specifically, a coupled decomposition variation required the least computational time of all 5 techniques in $13/27$ image pairs, usually CD_m since the non-zero overlap ratio of mean-based coupled decomposition increased the corresponding computational cost. Even when CD_m didn't require the least time, it's cost was near the optimal value. Actually, CD_m required less than $1.25T_{opt}$ (where T_{opt} is the optimal time) for all but 2 image pairs (pairs 5 and 16). On the other hand, while CD_c computational cost was usually larger than the rest of the sub-image matching methods, it never reached prohibitive levels. For the 23 image pairs that CD_c didn't fail, 16 required less than 1 hour to match the input images, while in all cases the computational time was

less than 4 hours, i.e. more than a level of magnitude less than the time required for full-image matching.

Note that a "Fail" was declared either when the mean reprojection error was more than 100 pixels. From the 5 examined techniques only CD_m does not fail in any of the image pairs. The fact that it requires only a single matching (from the full resolution images) allows us to employ more detailed (and time consuming) matching techniques, which are robust in most of the cases. On the contrary, all other techniques fail on the 3 mixed dataset pairs. CD_c fails additionally on pair 20, for which the reference and the target image only partially overlap, thus their centroids are very far from each other. P_v additionally fails on pair 17, which depicts Phobos from a great distance and with very different viewing angles, thus resulting in the performance of all methods in this pair to be poor.

P_r on the other hand fails on $18/27$ image pairs. The reason is that it requires a series of successful sub-image matchings, of resolution that is much coarser than the initial resolution. For example, in image pair 9 the initial matching takes place at a resolution in which each (sub-sampled) pixel corresponds to a 68×68 patch in the original image. So extreme sub-sampling undermines the possibility of correct matching, especially in images with a small number of distinctive features. The dependence of P_r from the number of features is also implied by the fact that the largest image pair that P_r succeed (pair 7) is also the pair with the largest number of matched (tie-points) features.

The pyramidal scheme of [21] seems to be much more robust. Its accuracy is typically worse than coupled decomposition, while its computational complexity is typically on the same level of CD_c and worse than CD_m. Actually, the computational complexity of P_v degrades (in comparison to coupled decomposition) with the image size, since the initialization of Voronoi decomposition depends on the correct identification of 3 tie-points. In substantially large images extreme sub-sampling inhibits the establishment of matching points, thus Voronoi tessellation initiliases from finer resolutions. The latter reduce the available number of iterations, which cause larger sub-images to be matched, i.e. an increased computational cost. For example, while on all 6 image pairs 1-6 (with less than 10Mpixels) P_v is faster than CD_c, on 4 out of 6 image pairs 19-24 (with size 15-200Mpixels) CD_c is faster.

Overall, contrary to both full-image matching and pyramidal schemes, matching based on coupled decomposition achieves to combine speed, accuracy and robustness. From the two introduced variations, CD_m seem to be the most suitable for a generic selection, being faster in $23/27$ pairs and more accurate in $16/27$ pairs. However, CD_c seem to outperform CD_m when the reference and the target image have the same resolution and the images fully overlap (pairs 1-12).

D. Fundamental Matrix Estimation Through Coupled Decomposition

In this subsection, we evaluate the coupled decomposition potential to enhance the fundamental matrix estimation for a pair of matching images. Actually, three different levels of coupled decomposition engagement are included in this setup. More specifically, the three approaches are the following i) the full-image matching algorithm that performs straightforward tie-point extraction without involving coupled decomposition (named Ξ_{BF}), ii) an algorithm that employs (mean-based) coupled decomposition as described in Algorithm 2 but estimates a single, global, fundamental matrix for the pair of images (named Ξ_{TP}) and iii) an algorithm that use both (mean-based) coupled decomposition and Algorithm 3, as described in Section 3 to estimate a separate fundamental matrix for each pair of corresponding sub-images (named Ξ_{CD}).

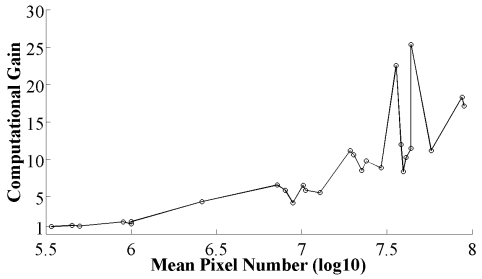


Fig. 8. Computational time of Ξ_{CD} (and Ξ_{TP}) over the Ξ_{BF} required time.

The computational gain of the algorithm, with the parameter selection strategy of sub-section VI-B is demonstrated in Fig. 8. Apart from the small-image dataset, in all the other datasets the computational gain is substantial. More specifically, the mean overall gain is 8.583, while the gain that was achieved for each dataset varies from 5.74 (HiRISE JPG dataset) to 13.86 (HRSC dataset). Note that in this comparison Ξ_{CD} and Ξ_{TP} are not distinguished, since the fundamental matrix computational time is negligible in comparison to the one associated with tie-point extraction.

The achieved computational gain does not trigger less accurate epipolar geometry estimation. This becomes apparent through the estimation of the rigid mean reprojection error when using epipolar geometry to transform the target image to the reference image. In both Ξ_{BF} and Ξ_{TP} variations, a single fundamental matrix is estimated for the complete image while in Ξ_{CD} a distinct fundamental matrix for each (determined through coupled decomposition) sub-image. The results are summarized in Table VI. Pairs 28-30 are omitted in this experiment because the computational time related to Ξ_{BF} was prohibitive.

The scores reported in Table VI show that the abundance of tie-points that is achieved through coupled decomposition may be used to enhance the matching accuracy of large-images. While the matching methodology is rather simple, compared to sophisticated matching pipelines such as [19], sub-pixel accuracy is achieved in 12 image pairs through Ξ_{CD} and in 9 pairs through Ξ_{TP} (compared to 5 pairs through Ξ_{BF}). On the contrary, if using Ξ_{CD} (Ξ_{TP}) in 6 pairs (9 pairs) the mean reprojection error is above 10 pixels. These are mostly cases in which the terrain elevation varies significantly in Z and W or cases in which images have a large scale difference, such as the mixed sub-set (pairs $25-27$), thus violating the assumptions that epipolarity requires. Additionally, this

TABLE VI
MEAN REPROJECTION ERROR RESULTS

Id	1	2	3	4	5
Ξ_{BF}	2.3959	4.5798	0.3821	1.0423	1.1311
Ξ_{TP}	1.968	7.0054	0.3636	0.1578	0.2739
Ξ_{CD}	0.5415	1.9506	0.5142	0.244	0.4536
Id	6	7	8	9	10
Ξ_{BF}	1.2767	11.8657	20.6592	11.8391	2.0698
Ξ_{TP}	0.7435	8.1208	15.1576	2.9186	3.3754
Ξ_{CD}	0.9543	0.7033	3.3651	1.7206	0.8697
Id	11	12	13	14	15
Ξ_{BF}	27.2144	28.9135	0.1927	1.1382	0.4625
Ξ_{TP}	23.482	22.2673	0.2385	1.0803	0.1973
Ξ_{CD}	2.2248	5.0963	0.2384	0.7517	16.417
Id	16	17	18	19	20
Ξ_{BF}	1.9004	221.5	0.0482	0.2369	165.36
Ξ_{TP}	5.0035	162.95	0.0659	0.4547	332.33
Ξ_{CD}	1.0079	Fail	0.4141	0.2303	111.68
Id	21	22	23	24	25
Ξ_{BF}	24.7373	6.7609	1.0298	402.45	99.5811
Ξ_{TP}	8.8041	4.4007	0.7004	195.54	51.2317
Ξ_{CD}	6.4956	1.7341	0.1979	95.483	11.7967
Id	26	27			
Ξ_{BF}	262.64	956.07			
Ξ_{TP}	219.99	201.91			
Ξ_{CD}	374.31	64.2668			

group includes image pairs in which the mean-based coupled decomposition algorithm assumption that the reference and the target image centroid refer to the same world coordinates diverge significantly from reality (e.g. image pairs 20 and 24).

Finally, two special cases are image pairs 15 and 17, in which only algorithm Ξ_{CD} fails. These are small-sized image pairs, in which the tie-point sets are so sparse that a tie-point split into distinct sub-images does not provide the necessary number of points for a fundamental matrix to be accurately estimated (pair 15) or even to be estimated at all (pair 17). This feature implies a lower image size limit, under which Algorithm 3 cannot be successfully executed.

Note that in 20 out of 27 pairs Ξ_{TP} outperforms Ξ_{BF} in terms of mean reprojection error. Since, Ξ_{TP} employs coupled decomposition only to split images into corresponding sub-images, the above result is consistent with the theoretical analysis of Section III.

Further evidence for the validity of Section V-A conclusions are given through the analysis of the results of the Phobos image sub-set, in contrast to the results of all other sub-sets. Due to the small size of the Phobos sub-set imagery, the false acceptance ratio when employing Ξ_{TP} algorithm increases, thus undermining the accuracy of the subsequent matching. As a matter of fact, in 3 out of 6 Phobos pairs Ξ_{TP} exhibit worse mean reprojection error than Ξ_{BF}. On the contrary, in the 5 large-sized sub-sets, in only 4 out of 21 images, matching accuracy deteriorates when using coupled decomposition. These are either images that matching was already very accurate without performing coupled decomposition (pairs 2 and 19) or pairs that the reference and the target image centroid diverge significantly (pair 20).

Finally, Ξ_{CD} outperforms Ξ_{BF} in 21 out of 27 and Ξ_{TP} in 19 out of 27 image pairs. Ξ_{CD} performs worse than

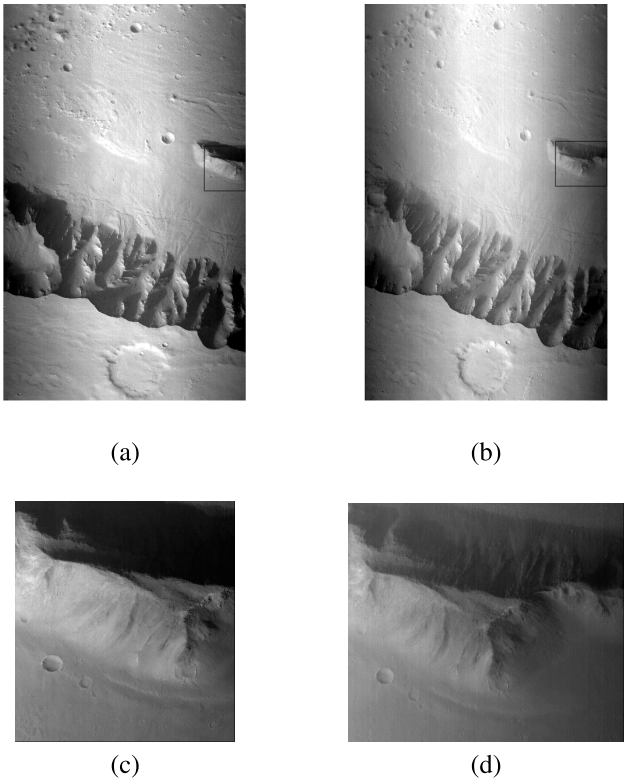


Fig. 9. Image pair 8 (Table I) and a grey-level changed area. (a) CTX image B01_009894_1665_XI_13S042W. (b) CTX image B02_010606_1666_XN_13S042W. (c), (d) The corresponding sub-images of (a) and (b) respectively with the maximum reprojection error among the set of 64 corresponding sub-images.

Ξ_{TP} in the Phobos sub-set, since in 4 out of 6 cases the output is less accurate than the one achieved through the brute-force approach. On the other hand, image pairs that are acquired from pushbroom cameras benefit significantly from Ξ_{CD}, as the results of the CTX sub-set demonstrate. Actually, the partial epipolarity constraints that are imposed by Ξ_{CD} achieve to reduce the mean reprojection error up to more than 10 times compared to both Ξ_{BF} and Ξ_{TP} (pairs 7, 11).

E. Identification of Sub-Image With Grey-Level Changes

An output of Algorithm 3 is a list of sub-images with grey-level changes. In this section we present an example of such sub-images, illustrated in Fig. 9. This comes from image pair 8. The mean-based coupled decomposition method was followed from Algorithm 3 as described in the previous section. The focused sub-images are these with the maximum (rigid) reprojection error. In this example the hill that is in the middle-right part of images 9(a) and 9(b) is identified as a pair of sub-images with a grey-level difference. A close-up demonstrated in Figs. 9(c) and 9(d) reveals that the north slope of the hill can not be distinguished in the former image, most probably due to the illumination conditions at the time that the image was acquired, while it is not so severely shaded in the latter image.

VII. CONCLUSIONS AND FUTURE WORK

In this work, we have presented a novel technique called coupled image decomposition. This technique has demon-

strated that it can be used to circumvent the computational cost boost that is usually associated with the matching of massively sized images, whilst generating information that may both enhance the achieved accuracy and estimate local regions that can not be successfully matched. In the future we are going to work towards the optimization of image matching accuracy, as well as the characterization of the grey-scale changed areas so as to achieve our goal of ending up with an efficient automatic local change detection approach.

REFERENCES

[1] A. S. McEwen *et al.*, "Mars reconnaissance orbiter's high resolution imaging science experiment (HiRISE)," *J. Geophys. Res., Planets*, vol. 112, no. E5, 2007.

[2] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.

[3] O. Chum and J. Matas, "Homography estimation from correspondences of local elliptical features," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 3236–3239.

[4] M. Xia and B. Liu, "Image registration by 'super-curves,"' *IEEE Trans. Image Process.*, vol. 13, no. 5, pp. 720–732, May 2004.

[5] L. Baboulaz and P. L. Dragotti, "Exact feature extraction using finite rate of innovation principles with an application to image super-resolution," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 281–298, Feb. 2009.

[6] K. Mikolajczyk and J. Matas, "Improving descriptors for fast tree matching by optimal linear projection," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–8.

[7] C. Silpa-Anan and R. Hartley, "Optimised KD-trees for fast image descriptor matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[8] K. Li and S. Zhou, "A fast SIFT feature matching algorithm for image registration," in *Proc. Int. Conf. Multimedia Signal Process. (CMSP)*, May 2011, pp. 89–93.

[9] A. S. McEwen *et al.*, "Seasonal flows on warm Martian slopes," *Science*, vol. 333, no. 6043, pp. 740–743, 2011.

[10] S. Byrne *et al.*, "Distribution of mid-latitude ground ice on Mars from new impact craters," *Science*, vol. 325, no. 5948, pp. 1674–1676, 2009.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[12] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.

[13] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[14] J. R. R. Uiljings, A. W. M. Smeulders, and R. J. H. Scha, "Real-time visual concept classification," *IEEE Trans. Multimedia*, vol. 12, no. 7, pp. 665–681, Nov. 2010.

[15] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[16] J. K. Uhlmann, "Satisfying general proximity/similarity queries with metric trees," *Inf. Process. Lett.*, vol. 40, no. 4, pp. 175–179, 1991.

[17] A. M. Kibriya and E. Frank, "An empirical comparison of exact nearest neighbour algorithms," in *Knowledge Discovery in Databases: PKDD* (Lecture Notes in Computer Science), vol. 4702. Berlin, Germany: Springer-Verlag, 2007, pp. 140–151.

[18] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1997, pp. 1000–1006.

[19] J. R. Kim and J.-P. Muller, "Multi-resolution topographic data extraction from Martian stereo imagery," *Planetary Space Sci.*, vol. 57, nos. 14–15, pp. 2095–2112, 2009.

[20] B. P. Jackson and A. A. Goshtasby, "Adaptive registration of very large images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2014, pp. 351–356.

[21] Z. Wu and A. Goshtasby, "Adaptive image registration via hierarchical Voronoi subdivision," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2464–2473, May 2012.

[22] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[23] 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*, vol. 24, no. 6, pp. 381–395, 1981.

[24] R. L. Kirk *et al.*, "High-resolution topomapping of candidate MER landing sites with Mars orbiter camera narrow-angle images," *J. Geophys. Res., Planets*, vol. 108, no. E12, 2003.

[25] R. L. Kirk *et al.*, "Ultrahigh resolution topographic mapping of Mars with MRO HiRISE stereo images: Meter-scale slopes of candidate Phoenix landing sites," *J. Geophys. Res., Planets*, vol. 113, no. E3, 2008.

[26] N. T. Bridges, F. Ayoub, J.-P. Avouac, S. Leprince, A. Lucas, and S. Mattson, "Earth-like sand fluxes on Mars," *Nature*, vol. 485, pp. 339–342, May 2012.

[27] T. Kim, "A study on the epipolarity of linear pushbroom images," *Photogrammetric Eng. Remote Sens.*, vol. 66, no. 8, pp. 961–966, 2000.

[28] H. A. David and H. N. Nagaraja, *Order Statistics*. New York, NY, USA: Wiley, 2003.

[29] D. L. Olson and D. Delen, *Advanced Data Mining Techniques*. Berlin, Germany: Springer-Verlag, 2008.

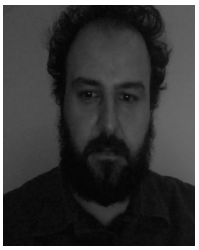
[30] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vis. Conf.*, 1988, pp. 147–151.

[31] L. Kish, *Survey Sampling*. New York, NY, USA: Wiley, 1965.

[32] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[33] M. C. Malin *et al.*, "Context camera investigation on board the Mars reconnaissance orbiter," *J. Geophys. Res., Planets*, vol. 112, no. E5, 2007.

[34] R. Jaumann *et al.*, "The high-resolution stereo camera (HRSC) experiment on Mars express: Instrument aspects and experiment conduct from interplanetary cruise through the nominal mission," *Planetary Space Sci.*, vol. 55, nos. 7–8, pp. 928–952, 2007.



Panagiotis Sidiropoulos (M'12) received the B.Sc. degree in electrical and computer engineering and the M.Sc. degree in informatics from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003 and 2007, respectively, and the Ph.D. degree in machine learning from the Centre for Vision, Speech and Signal Processing, University of Surrey, Surrey, U.K., in 2012. He is currently a Post-Doctoral Research Associate with the Mullard Space Science Laboratory, University College London, U.K. His main research interests are related to the use of machine learning and multimedia processing techniques to planetary science.

Jan-Peter Muller received the B.S. (Hons.) degree in physics from the University of Sheffield, U.K., in 1976, the M.Sc. and D.I.C. degrees in atmospheric physics and dynamics from Imperial College London, London, U.K., in 1977, and the Ph.D. degree in planetary meteorology from University College London (UCL), London, in 1982. He is currently a Professor of Image Understanding and Remote Sensing with UCL, where he has been a Faculty Member since 1984. He has been a member of the Moderate Resolution Imaging Spectrometer and the Multi-Angle Imaging SpectroRadiometer science teams since 1990. He is currently a Deputy Coordinator of the EU-FP7 QA4ECV project, generating a unique 35-year record of Earth's land surface albedo from European and U.S. sensors, and a Coordinator of the EU-FP7 iMars project, generating a 40-year long record of co-registered and orthorectified orbital images of Mars to map change detection using data mining and crowd sourcing methods. His research interests include practical applications of machine vision to remote sensing of the Earth and planets, with a special focus on operational applications of stereo photogrammetry and other 3D imaging technologies for applications in climate modeling, weather forecasting, and planetary exploration. He was the Chair of the U.K. Aurora Advisory Committee, and is the Chair of the Committee on Earth Observation Satellites WG on Cal/Val Terrain Mapping Subgroup, and the International Society for Photogrammetry and Remote Sensing Working Group on Global DEM Interoperability.