

Quaternionic keypoint description for multispectral imaging

Giorgos Sfikas

sfikas@cs.uoi.gr

Dept. of Computer Science &
Engineering

Ioannina, Greece

Dimosthenis Ioannidis

djoannid@iti.gr

Information Technologies Institute
Thessaloniki, Greece

Dimitrios Tzovaras

Dimitrios.Tzovaras@iti.gr

Information Technologies Institute
Thessaloniki, Greece

ABSTRACT

In this work, we propose a new keypoint descriptor that is suitable for multispectral inputs comprising up to 4 channels. Color images with or without an additional infrared or depth channel are some of the use-cases that can be handled by the proposed descriptor. Standard keypoint descriptors employ single-channel input gradients, thereby discarding potentially useful content. The proposed descriptor is based on a quaternionic representation of the input image thereby treating each pixel multispectral value holistically. Coupled with a suitable multispectral quaternionic detector, we show that the proposed detector leads to superior experimental results on a keypoint matching scenario.

KEYWORDS

quaternions, SIFT, Harris corners, polar representation

ACM Reference Format:

Giorgos Sfikas, Dimosthenis Ioannidis, and Dimitrios Tzovaras. 2020. Quaternionic keypoint description for multispectral imaging. In *11th Hellenic Conference on Artificial Intelligence (SETN 2020)*, September 2–4, 2020, Athens, Greece. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3411408.3411417>

1 INTRODUCTION

Keypoints, also known as features or interest points have been of interest to the computer vision community for many decades [15]. Some of the applications on which they are used include photogrammetry and 3D reconstruction, image retrieval and visual localization [17, 18]. Perhaps two of the most seminal works in the field are the Harris and SIFT algorithms, which have in turn led to numerous adaptations and extensions. ORB and SURF are also two other very popular algorithms that were proposed after SIFT [17]. Learning-based keypoint detectors and descriptors have been recently proposed [12, 16], following the trend of applying deep learners on most if not all vision problems. However, traditional, hand-crafted keypoint detection and description algorithms are still used in numerous practical problems (e.g. 3D reconstruction) [18, 19].

A feature that is almost ubiquitous in non-learning keypoint detectors and descriptors is that they operate over the first or second-order gradient of an image that is assumed to comprise only a single channel. If more than one channels exist, only the grayscale component is practically used, leading to effectively neglecting potentially useful information. The most well-known case of this paradigm are color images, which are made up of 3 channels. Furthermore, images that capture different modalities are possible and increasingly more easily available. Such modalities include Near Infrared images [9], thermography imaging [22] or depth sensors. These inputs can

be concatenated to, for example an optical sensor input, making up for a 4-channel input. Note that in most if not all of these cases, computing a single-channel component as in order to compute gradients is meaningless or obviously wrong (for example, running a keypoint descriptor on gradients computed on a weighted sum of optical and depth channels).

In this work, we propose a novel keypoint descriptor that can be used with multispectral images that are made up of up to 4 channels. This constraint is due to handling the input as a matrix of quaternionic values, which are intrinsically 4-dimensional objects. Combined with a recently proposed keypoint detector for quaternionic inputs, we show that the proposed detector extends and improves the well-known SIFT descriptor in a simple manner. We have run numerical trials on two datasets of color images; the first dataset has been captured aerially by a drone-mounted camera in an urban area, and the other dataset is comprised of images of rock carvings captured at the archaeological site of Vathy in Astypalaia, Greece. We have computed numerical results for keypoint matching precision for both datasets, where the proposed keypoint detector was shown to be capable of capturing well input channel cross-correlation, outperforming two more naïve descriptor variants.

With this work, we propose a keypoint descriptor that is suitable for multispectral inputs comprising up to 4 channels. We use a quaternion image representation, according to which each pixel value can be treated as a single entity. Quaternions are 4-dimensional generalizations of complex numbers, with one real and three imaginary, independent components. Quaternionic analysis has found numerous applications in multimodal signal processing, as well as lately in deep learning-based models (for example quaternionic convolutional networks [28], or a quaternion adaptation of capsule networks [27]).

We have structured the remainder of this paper as follows. In section 2 we have a brief review of related work, with respect to either quaternionic analysis or multispectral imaging. In sections 4 and 5 we describe the quaternionic interest point detector and descriptor that we employ in this work. We present numerical results in section 6 and close with a short discussion of the paper's contribution and future work in section 7.

2 RELATED WORK

Keypoint detection and description in multispectral imaging has been addressed by a number of previous works [1, 3, 9, 11]. In [9] and [23], the Harris corner detector is generalized by taking into account the sum of autocorrelation of matrices per band. A generalization of the SIFT descriptor is also proposed in [9], where description is invariant to a 180 degree change in gradient direction. This has been found to be useful in when combining certain types of

modalities like infrared, where gradient direction is not uncommon to be inverse to the one that corresponds to the RGB channels. Summing autocorrelations per band as done in [9] neglects to take into account channel cross-correlation; this has been addressed with the recently proposed Quaternion Harris detector, which treats the input as a quaternion-valued matrix. As quaternions are intrinsically 4-dimensional objects, in this manner an image with up to 4 channels may be processed. However, no new descriptor had been proposed in [11], and keypoint descriptions were handled with the SIFT descriptor, in practice describing only the luminance channel and discarding the rest of the information. In the current work, we extend the detector of [11] by proposing a suitable quaternionic detector that can effectively handle multispectral content.

3 ELEMENTS OF QUATERNIONS

Quaternions have been introduced in 1843 by the Irish mathematician W.R. Hamilton. They extend the concept of complex numbers as sums of one real and one imaginary part, to numbers that have one real and three independent imaginary parts. In that sense, the space of quaternions is isomorphic to \mathbb{R}^4 . In particular, any quaternion q can be written in a unique way as

$$q = a + bi + cj + dk, \quad (1)$$

where $a, b, c, d \in \mathbb{R}$ and i, j, k are independent imaginary units. A useful, alternative representation of quaternions is the Cayley-Dickson form, according to which we write:

$$q = \zeta + \eta j, \quad (2)$$

where the difference with the analogous relation for complex numbers is that ζ and η are not necessarily real, so in general $\zeta = a + bi, \eta = c + di$. This scheme can easily be generalized to using any other couple of perpendicular imaginary units μ_1, μ_2 instead of i, j . In that, more general case, it is referred to as a symplectic decomposition [6], which is essentially a change of basis from $(1, i, j, k)$ to new imaginary units $(1, \mu_1, \mu_2, \mu_3)$. A third way to represent a quaternion is as a sum of its real and imaginary part:

$$q = S(q) + V(q), \quad (3)$$

where $S(q) = a$ and $V(q) = bi + cj + dk$.

Quaternions form a skew-field which we shall denote here as \mathbb{H} . This means that \mathbb{H} has all the properties of a field concerning quaternion addition and multiplication, except for quaternion multiplication commutativity: in general $pq \neq qp$ for $p, q \in \mathbb{H}$. Concerning quaternion imaginary units i, j, k , we have:

$$i^2 = j^2 = k^2 = ijk = -1,$$

and

$$ij = -ji = k, jk = -kj = i, ki = -ik = j. \quad (4)$$

. Conjugacy is defined as

$$\bar{q} = a - bi - cj - dk, \quad (5)$$

and, in a relation analogous to the one for complex numbers, we have

$$|q| = \sqrt{q\bar{q}} = \sqrt{\bar{q}q} = \sqrt{a^2 + b^2 + c^2 + d^2}. \quad (6)$$

Quaternions with a zero real part ($a = 0$) are called pure quaternions, and quaternions with $|q| = 1$ are called unit quaternions.

Matrix calculus can be extended to matrices with quaternionic elements, $\mathbb{H}^{m \times n}$. Quaternionic matrices can be written as tuples of complex matrices [14, 26] following a decomposition akin to the Cayley-Dickson form. For any $A \in \mathbb{H}^{m \times n}$, we can write $A = A_1 + A_2j$ where $A_1, A_2 \in \mathbb{C}^{m \times n}$ are unique complex matrices [26]. Furthermore, the mapping χ_A of any quaternionic matrix to a complex matrix can be defined:

$$\chi_A = \begin{bmatrix} A_1 & A_2 \\ -\bar{A}_2 & \bar{A}_1 \end{bmatrix} \quad (7)$$

Matrix χ_A is called the adjoint or complex adjoint of A . If for a complex matrix C there exists a quaternionic matrix Q such that $C = \chi_Q$, C is called symplectic [8], and equivalently $\chi_C^{-1} = Q$. The discussion of eigenvalues and eigenvectors of quaternionic matrices becomes more complicated than the real or complex matrix case from the very beginning. The first difficulty comes due to the noncommutativity of quaternion multiplication, which means that the following problems are distinct in the quaternionic case:

$$\begin{cases} Ax = \lambda x \\ Ax = x\lambda \end{cases}$$

The two relations are related to left and right eigenvalues respectively, and their respective eigenvectors. Another difficulty is that the well-known relation $\det(A - \lambda I) = 0$ is not directly applicable, as a quaternionic determinant is itself more complicated to use or even define (see for example Dyson 1972 [5]). It is known that any square quaternion-valued matrix has exactly n right eigenvalues that are complex with a non-negative part [26]. If these are real, we have $Ax = x\lambda = \lambda x$ and these will be also left eigenvalues. A fortunate case for quaternionic matrices is for Hermitian matrices, i.e. matrices that are equal to their conjugate transpose with conjugacy now defined in the quaternionic sense (eq. 5). Formally then $A = A^H$, where H denotes the conjugate transpose. For such matrices their right eigenvalues are real (however there may exist other left eigenvalues that are not right eigenvalues, and necessarily non-real, see for example [26]) as $x^H Ax = x^H x\lambda$ and λ must be real as the ratio of real numbers $x^H Ax, x^H x = \|x\|^2$ ($x^H Ax$ is real as it is equal to its conjugate due to $A = A^H$). Note that this does not follow necessarily if we take $Ax = \lambda x$, which is the formula for left eigenvalues.

4 QUATERNIONIC DETECTOR

In this section we briefly describe the quaternionic detector that was recently introduced in [11]. The well-known Harris keypoint detector was based on the premise of locating points that maximize an error function when displacement to any direction is taken into account. This error function measures the difference of values around the candidate point and a window displaced by some Δx . Formally we can write the error as:

$$E(x) = \sum_{x_n \in N(x)} g(x_n) |I(x_n) - I(x_n + \Delta x)|^2, \quad (8)$$

where $N(x)$ denotes the neighbourhood around candidate point x and a $g(x_n)$ function plays the role of a weight to penalize values that are calculated for points far from the candidate keypoint.

In the classical variant of the Harris keypoint detector, $I(x)$ is a grayscale image and takes values in some subset of \mathbb{R} . Assuming

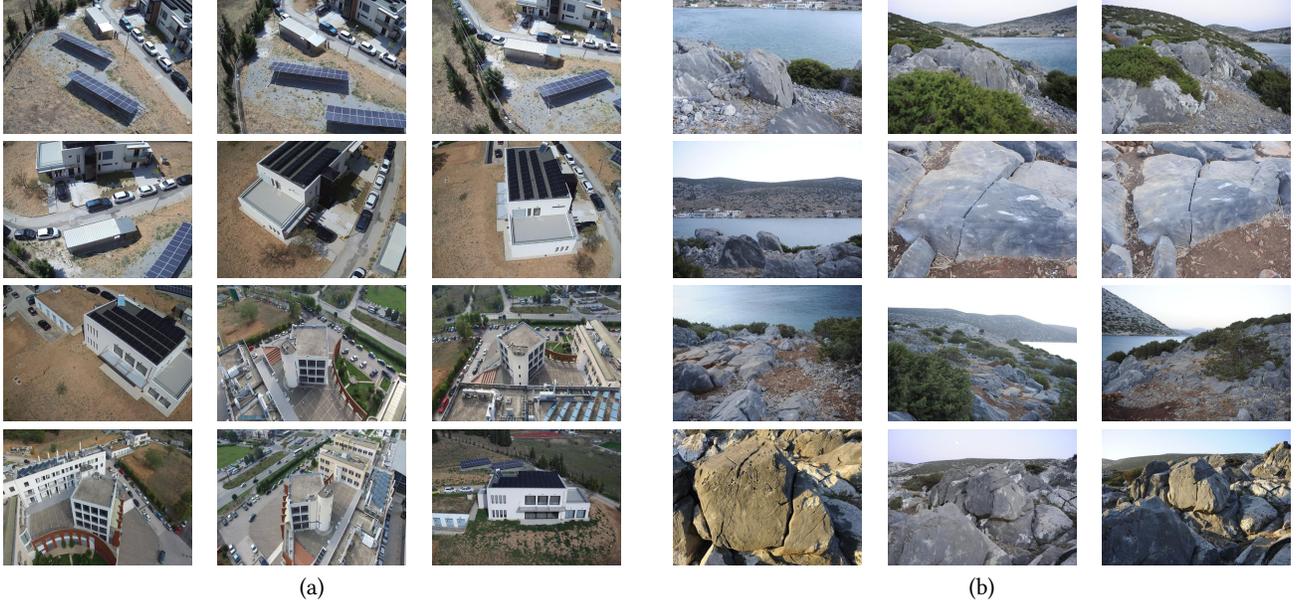


Figure 1: Color images used in our experiments. (a) The “CERTH” dataset. (b) The “Astypalaia” dataset. See text for details.

now that we have a multichannel input and each channel corresponds to the real part or one of the quaternionic imaginary parts, we can consider $I(x)$ as a quaternion-valued function, or equivalently consider I as a quaternionic matrix.

Following the idea that $I(x)$ is considered as a quaternion-valued function, we can use the quaternion-valued Taylor approximation of $I(x)$ [4] and continue to write the error of eq. 8 as:

$$\begin{aligned} E(x) &= \sum_{x_n \in N(x)} g(x_n) \Delta x^T \nabla I(x_n) \overline{\Delta x^T \nabla I(x_n)} \\ &= \sum_{x_n \in N(x)} g(x_n) \Delta x^T \nabla I(x_n) \overline{\nabla I(x_n)^T \Delta x} \\ &= \Delta x^T \left[\sum_{x_n \in N(x)} g(x_n) \nabla I(x_n) \nabla I(x_n)^H \right] \Delta x \\ &= \Delta x^T A_q \Delta x. \end{aligned}$$

In the above relation, matrix A_q is key to continue. We have considered

$$A_q = \sum_{x_n \in N(x)} g(x_n) \nabla I(x_n) \nabla I(x_n)^H, \quad (9)$$

to which we refer to as quaternionic autocorrelation or simply autocorrelation matrix. As in the non-quaternionic case, it is easy to see that this matrix is Hermitian. After computing this matrix, we can decide whether to characterize the candidate point as a keypoint by comparing the eigenvalues of A_q , or by taking into account a cornerness function:

$$c(A_q) = (\lambda_1 \lambda_2) - \kappa (\lambda_1 + \lambda_2)^2. \quad (10)$$

where parameter κ is empirically set to a small positive value.

The form of each value in A_q is perhaps notable: The diagonal terms of each rank-one component of A_q are of the form of $|I_x|^2$

and $|I_y|^2$, both of which are equal to a real sum of gradient squares for each image channel separately:

$$|I_x|^2 = I_{x0}^2 + I_{x1}^2 + I_{x2}^2 + I_{x3}^2, \quad (11)$$

and likewise for $|I_y|$. On the contrary, the off-diagonal terms $I_x \bar{I}_y$ and its conjugate $I_y \bar{I}_x$ are quaternionic, with each quaternion variate computed as a sum of products of quaternion variates that correspond to different channels. For example, the component of the term $I_x \bar{I}_y$ corresponding to the i imaginary unit is found as:

$$I_{x1} I_{y0} - I_{x0} I_{y1} + I_{x3} I_{y2} - I_{x2} I_{y3}, \quad (12)$$

The other imaginary components comprise terms with different combinations of channels. Hence, the off-diagonal terms of the autocorrelation matrix contribute information about image cross-channel correlation.

We can compute (right) eigenvalues and eigenvectors of autocorrelation A_q by considering its diagonalization and its complex adjoint form. We write:

$$U^H A_q U = \Lambda \quad (13)$$

from which follows that

$$\chi_U^H \chi_{A_q} \chi_U = \chi_\Lambda, \quad (14)$$

where we used $\chi_{AB} = \chi_A \chi_B$ and $\chi_{U^H} = \chi_U^H$ [26]. Note that the 4×4 matrix χ_{A_q} is also Hermitian (theorem 4.2.6 [26]). Eigenvalues will be obtained in two pairs of repeating values. Eigenvectors are found as $v_n = c_{n'} - \bar{d}_{n'} j$ for $n = 1, 2$ and $n' = 2n - 1$, where $w_k = [c_k \ d_k]^T$ is the k^{th} column of χ_U , considering reordered indices by eigenvalue magnitude. Note that one can still obtain a valid eigenvector if one uses $n' = 2n$ instead of $2n - 1$; this eigenvector will be a multiple of the eigenvector previously considered, by a factor of j multiplied from the right.

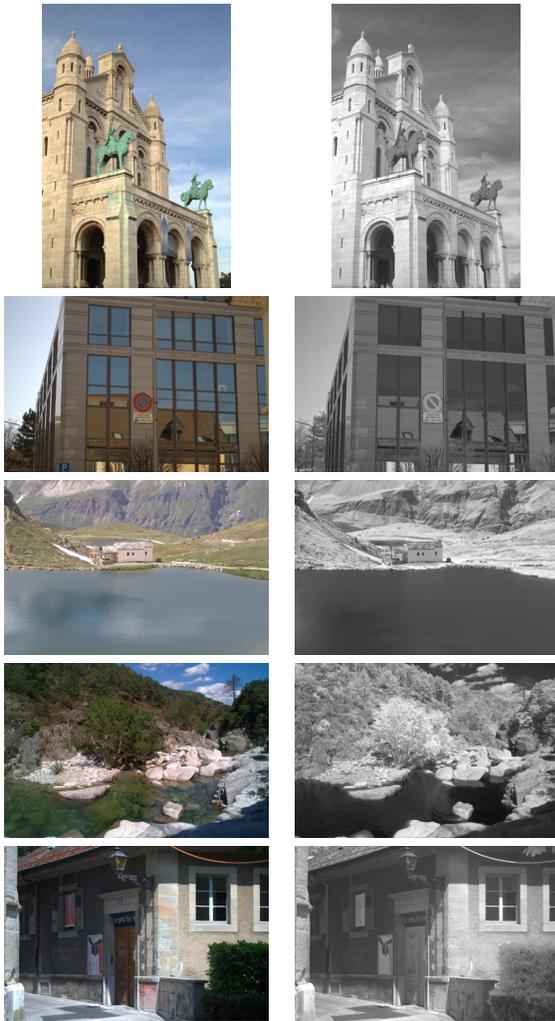


Figure 2: Images comprising four channels that were used in our experiments. The four channels are the Red-Green-Blue channels (left column) plus a near-infrared channel (right column).

5 PROPOSED QUATERNIONIC DESCRIPTOR

The basis of the proposed keypoint descriptor is the SIFT descriptor [17]. We will first briefly reiterate the basic points for computing SIFT, then discuss the proposed quaternionic descriptor. In order to construct the standard SIFT descriptor (here we shall also refer to it as *Vanilla SIFT* to distinguish from the other variants we shall discuss in what follows), first gradient magnitude and orientation maps are computed over 16×16 neighbourhood around the keypoint. This region is then divided into a grid of 4×4 non-overlapping cells. For each of these cells, a 8-bin gradient orientation histogram is computed. After weighting each contribution to the histogram according to its distance from the keypoint, the cell histograms (a total of $4 \times 4 = 16$ histograms) are concatenated to a single vector of $16 \times 8 = 128$ values, which constitutes the SIFT descriptor for the point of interest.

While SIFT has proven to be a remarkably robust descriptor, our problem with applying it to a multispectral image input is that now gradients are quaternionic, as are pixel values. As gradients are computed to the horizontal and vertical direction, the quaternionic gradient would effectively be a $4 + 4 = 8$ -dimensional number. We can then safely hypothesize that the 8 bin-histograms prescribed by vanilla SIFT would be inadequate for partitioning the space of quaternionic orientations, due to their high dimensionality. While we could work in the direction of simply adding more bins to the orientation histogram, we would be faced with a *curse-of-dimensionality* problem: As the dimension and bins increase, the number of data remain the same, which are the keypoint neighbourhood gradients in our case. Consequently, any orientation histogram constructed in this manner would be of little practical value.

In order to create a descriptor that would combine the robustness of SIFT, while avoiding to compute a descriptor only on a single-channel function of the original input, we consider using a polar representation of image pixel values, and hence by extension of the whole quaternionic image I . The polar representation of a quaternion $q \in \mathbb{H}$ is:

$$q = |q|e^{\mu\theta}, \quad (15)$$

with $\theta \in \mathbb{R}$ and $\mu \in \mathbb{H}$ a pure unit quaternion. Values μ and θ are called the eigenaxis and eigenangle (or eigenphase) of the quaternion [2, 6, 13]. The eigenaxis and eigenangle can be computed as [2]:

$$\mu = V(q)/|V(q)| \quad (16)$$

and

$$\theta = \tan^{-1}(|V(q)|/|S(q)|) \quad (17)$$

We can proceed by considering instead of the four quaternionic channels a, b, c, d , the magnitude, eigenaxis and eigenangle as our new image per-pixel representation. Out of these values, we keep the magnitude and eigenangle as more convenient due to their being in \mathbb{R} , and compute gradients $\nabla|q|$ and $\nabla\theta$. Over these gradients we then separately compute SIFT descriptions, that is as if each one were the gradient information of a grayscale input. We concatenate the two SIFT descriptors into a single vector of $128 \times 2 = 256$ values. This constitutes the proposed keypoint descriptor, to which we shall refer to as *Quaternion SIFT* in the remainder of this paper.

Adding phase information to the keypoint is crucial because it effectively encodes information about cross-channel correlation. Considering each channel separately is also suboptimal for the same reason, as we shall see in the experiments section.

6 EXPERIMENTS

We have run tests on three image collections, to which we shall refer with the acronyms CERTH, Astypalaia and RGB+NIR respectively. The first collection (“CERTH”) is comprised of images captured with an unmanned aerial vehicle (UAV), flown above the premises of the Centre for Research and Technology, located in Thessaloniki, Greece. A total of 12 images were shot with a DJI M200 UAV in Thessaloniki, at a resolution of 500×375 pixels. The second collection (“Astypalaia”) is comprised of 12 color images captured at the archaeological site located close to the village Vathy at the island of Astypalaia, Greece. The rock carvings that have been found on the site are practically invisible to the naked eye and/or without

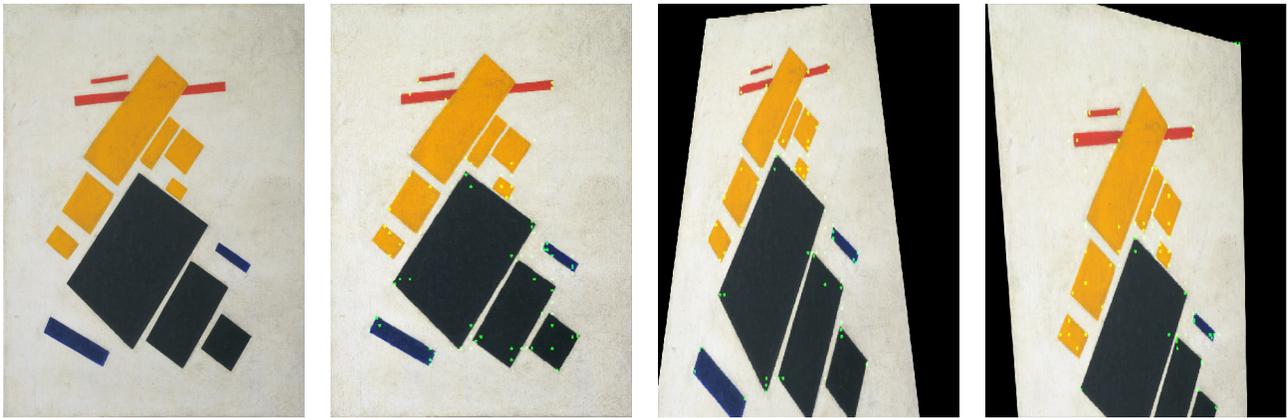


Figure 3: Illustration of our experimental setup. From left to right: Original image, Same image with overlaid Quaternion Harris keypoints (method described in sec. 4, Original image distorted by a random perspective transform (the two last images correspond to two different random transforms) with keypoints computed and overlaid on the new images. In the experiments section (results in tables 2, 1), precision is measured as percentage of correctly identified matches between keypoints in the original and the transformed image, and over 50 different random transformations.

guidance from an expert, making for a challenging detection problem [24]. All 12 images were captured with a NIKON D700 camera, at a resolution of 500×332 pixels. The images from the first two sets can be seen in Fig. 1. Regarding the third set (“RGB+NIR”), it is comprised of 5 images that are each made up of a total of 4 channels. The first 3 channels correspond to the color, i.e. Red-Green-Blue channels, while the 4th channel corresponds to the Near Infrared band. We can examine these images in Fig. 2. One can observe that the Near Infrared band is not strictly correlated to the visible-spectrum channels (see for example the intensity of lake pixels in the NIR band, which in the visible spectrum appears transparent; also [3] for a discussion of RGB-NIR correlation).

In all tests we have used the quaternionic detector described in sec. 4 with cornerness function hyperparameter set to $k = 0.04$, and a multiscale-affine implementation was used following the implementation in [18]. The experimental pipeline was set up as follows. For each dataset image we calculated 50 transformed versions of the original image. Each transformation was a random perspective transformation computed by moving 4 control points by random displacements. This results each time to a transformation T which we later use to compute ground truth displacements for keypoints. Next we compute keypoints for the original image as well as on all transformed images. Subsequently, we compute keypoint descriptors for all instances, using the descriptor proposed in sec. 5. Ideally, each keypoint x in the original image should have a matching keypoint at position $T(x)$ in the randomly transformed image; if that is the case (within a distance threshold of 2 pixels) then we consider a hit for the given keypoint. Matches were considered by computing the Euclidean distance between descriptors in \mathbb{R}^{256} . An example showing detected keypoints on the original image and transformed versions of the original can be observed in fig. 3¹.

Numerical precision results for each image of the two sets can be examined in tables 1, 2 and 3 for the *Astypalaia*, *CERTH* and *RGB-NIR* datasets respectively. Each row in these tables corresponds to an image in Fig. 1, where we traverse images from left to right and from top to bottom (for example, the 5th row on the table corresponds to the image on the 2nd row, 2nd column). In the last row we see mean and standard deviation over all images in the dataset. The proposed detector (Quaternion SIFT) was compared against two other descriptor variants: Vanilla SIFT and Multiband SIFT. As described in section 5, in Vanilla SIFT we map the image to grayscale then compute the classical SIFT descriptor. In Multiband SIFT, we compute SIFT for each channel separately, then concatenate the result to a single vector. Hence Quaternion SIFT, Vanilla SIFT and Multiband SIFT result to descriptor vectors in \mathbb{R}^{256} , \mathbb{R}^{128} , \mathbb{R}^{512} ² respectively. In all cases, the proposed Quaternion SIFT outperforms the other SIFT variants. Interestingly, this is also the case versus Multiband SIFT, despite this latter being larger than Quaternion SIFT. We can justify this result due to Multiband SIFT not capturing channel cross-correlation as effectively as Quaternion SIFT, since all channels are considered independently. Furthermore, the similar results of Multiband SIFT to Vanilla SIFT can be explained due to R-G-B channels being closely correlated to one another, hence each new channel SIFT does not contribute significantly to the end result. In the results for the RGB+NIR dataset (Fig. 3), where there is a fourth component that is only weakly correlated to the visible spectrum channels [3], we can note that the figures for Multiband SIFT are on average more than 1.5% points better than Vanilla SIFT. While this is a small difference in absolute terms, the same difference on the sets that do not comprise a NIR channel is much smaller still (0.3% in both *Astypalaia* and *CERTH*). This difference can be attributed to the existence of the extra channel in RGB+NIR. More importantly, the proposed Quaternion SIFT fares still better

¹The depicted original image is a digitized version of the painting “Airplane flying” by K.Malevich (1915).

²Or \mathbb{R}^{384} if the image does not comprise a fourth component.

than the other variants. Compared to Multiband SIFT in particular, it outperforms it while being much more compact in size.

Table 1: Precision values keypoint matching in the Astypalaia dataset. Columns correspond to compared descriptors, rows correspond to different images in the dataset. The last row shows mean +- st.deviation over all image results in the set. See text for correspondence of rows with images of Fig. 1. Higher values are better.

Vanilla SIFT	Multiband SIFT	Quaternion SIFT (proposed)
28.0 ± 5	28.0 ± 5	34.1 ± 5
31.9 ± 4	32.5 ± 4	38.2 ± 4
32.2 ± 4	32.6 ± 4	39.1 ± 3
38.7 ± 5	38.9 ± 5	43.8 ± 5
31.5 ± 5	31.9 ± 5	35.4 ± 5
28.8 ± 6	29.0 ± 6	32.8 ± 6
30.0 ± 4	30.4 ± 4	36.4 ± 4
28.0 ± 5	28.0 ± 5	34.1 ± 5
31.9 ± 5	32.4 ± 5	37.3 ± 5
30.8 ± 6	31.2 ± 5	33.6 ± 5
33.9 ± 3	34.7 ± 3	38.8 ± 3
33.7 ± 5	34.2 ± 5	37.8 ± 5
31.6 ± 2	31.9 ± 2	36.7 ± 2

Table 2: Precision values keypoint matching in the CERTH dataset. Columns correspond to compared descriptors, rows correspond to different images in the dataset. The last row shows mean +- st.deviation over all image results in the set. See text for correspondence of rows with images of Fig. 1. Higher values are better.

Vanilla SIFT	Multiband SIFT	Quaternion SIFT (proposed)
37.2 ± 4	37.5 ± 4	41.1 ± 4
37.7 ± 5	38.1 ± 5	41.6 ± 4
35.3 ± 4	35.5 ± 4	38.9 ± 5
40.7 ± 4	40.9 ± 4	44.6 ± 4
38.2 ± 5	38.5 ± 5	41.8 ± 5
38.5 ± 4	38.9 ± 4	42.4 ± 4
35.2 ± 5	35.7 ± 5	40.1 ± 5
37.0 ± 4	37.2 ± 4	41.2 ± 4
34.5 ± 5	34.8 ± 5	39.3 ± 5
31.4 ± 3	31.7 ± 3	37.4 ± 3
35.2 ± 4	35.5 ± 4	40.7 ± 3
33.5 ± 3	33.7 ± 3	36.8 ± 3
36.2 ± 2	36.5 ± 2	40.5 ± 2

7 CONCLUSION AND FUTURE WORK

We have proposed a keypoint descriptor that is suitable for color and in general multispectral images that comprise up to four channels. The descriptor is based on a polar decomposition of each image quaternionic value, and we have shown that it manages to capture

channel correlation in a more efficient manner than two other descriptor variants. This new quaternionic keypoint descriptor was proposed as a complement to the recently proposed quaternionic keypoint detector [11].

As future work, we plan to perform more extensive results on multispectral images, including optical+infrared pairs [10, 22] and Stokes images [7, 20], multi-modal medical imaging [21] or combining with video processing [25]. Also we plan to compare the proposed Quaternion SIFT descriptor to learning-based detectors such as Hardnet [16] or the earlier learning-free multispectral SIFT of [3]. Concerning the use of the proposed descriptor itself, another idea would be to explore potential uses as a pixel-level cue based on local autocorrelation, or combine with other relatively unused information such as eigenvector direction of the quaternionic autocorrelation. Finally, we could experiment on relaxing the constraint of having at most four channels, by either extending our analysis to algebras of higher dimensionality or combining with dimensionality reduction techniques.

ACKNOWLEDGMENTS

The work presented was carried out as part of the eDREAM project which is co-funded by the EU’s Horizon 2020 framework programme for research and innovation under grant agreement No 774478. The authors would also like to thank the anonymous reviewers for their useful suggestions.

Table 3: Precision values keypoint matching in the RGB+NIR dataset. Columns correspond to compared descriptors, rows correspond to different images in the dataset. The last row shows mean +- st.deviation over all image results in the set. Rows correspond with image rows in Fig. 2. Higher values are better.

Vanilla SIFT	Multiband SIFT	Quaternion SIFT (proposed)
37.7 ± 3	38.5 ± 4	40.3 ± 4
23.9 ± 5	26.1 ± 5	26.6 ± 5
41.8 ± 4	44.6 ± 4	44.7 ± 4
37.6 ± 5	39.1 ± 5	41.4 ± 4
35.2 ± 4	35.9 ± 4	37.5 ± 4
35.2 ± 6	36.8 ± 6	38.1 ± 6

REFERENCES

- [1] Catherine Achard, Erwan Bigorgne, and Jean Devars. 2000. A sub-pixel and multispectral corner detector. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Vol. 3. IEEE, 959–962.
- [2] Dimitrios S Alexiadis and Petros Daras. 2014. Quaternionic signal processing techniques for automatic evaluation of dance performances from MoCap data. *IEEE Transactions on Multimedia* 16, 5 (2014), 1391–1406.
- [3] Matthew Brown and Sabine Süsstrunk. 2011. Multi-spectral SIFT for scene category recognition. In *CVPR 2011*. IEEE, 177–184.
- [4] Dong Cheng and Kit Ian Kou. 2016. Properties of Quaternion Fourier transforms. *Preprint* (2016).
- [5] Freeman J Dyson. 1972. Quaternion determinants. *Helvetica Physica Acta* 45, 2 (1972), 289.
- [6] Todd A Ell and Stephen J Sangwine. 2007. Hypercomplex Fourier transforms of color images. *IEEE Transactions on image processing* 16, 1 (2007), 22–35.
- [7] Sylvain Faisan, Christian Heinrich, Giorgos Sfikas, and Jihad Zallat. 2013. Estimation of Mueller matrices using non-local means filtering. *Optics express* 21, 4 (2013), 4424–4438.

- [8] Douglas R Farenick and Barbara AF Pidkowich. 2003. The spectral theorem in quaternions. *Linear algebra and its applications* 371 (2003), 75–102.
- [9] D. Firmenichy, M. Brown, and S. Süsstrunk. 2011. Multispectral interest points for RGB-NIR image registration. In *2011 18th IEEE International Conference on Image Processing*. IEEE, 181–184.
- [10] G.Sfikas, A.Noula, J.Patacas, D.Ioannidis, and D.Tzovaras. 2019. Building thermal output determination using visible spectrum and infrared inputs. In *International Conference on Energy and Sustainable Futures (ICESF)*.
- [11] G.Sfikas, D.Ioannidis, and D.Tzovaras. 2020. Quaternion Harris for Multispectral Keypoint Detection. IEEE.
- [12] Axel Barroso Laguna, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. 2019. Key: Net: Keypoint Detection by Handcrafted and Learned CNN Filters. *arXiv preprint arXiv:1904.00889* (2019).
- [13] Nicolas Le Bihan. 2017. The geometry of proper quaternion random variables. *Signal Processing* 138 (2017), 106–116.
- [14] Nicolas Le Bihan and Jérôme Mars. 2004. Singular value decomposition of quaternion matrices: a new tool for vector-sensor signal processing. *Signal processing* 84, 7 (2004), 1177–1199.
- [15] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. 2005. A comparison of affine region detectors. *International journal of computer vision* 65, 1-2 (2005), 43–72.
- [16] Anastasiya Mishchuk, Dmytro Mishkin, Filip Radenovic, and Jiri Matas. 2017. Working hard to know your neighbor’s margins: Local descriptor learning loss. (Dec. 2017).
- [17] S.J.D. Prince. 2012. *Computer vision: models, learning, and inference*. Cambridge University Press.
- [18] E. Riba, D. Mishkin, D. Ponsa, E. Rublee, and G. Bradski. 2019. Kornia: an Open Source Differentiable Computer Vision Library for PyTorch. <https://arxiv.org/pdf/1910.02190.pdf>
- [19] Johannes L Schonberger, Hans Hardmeier, Torsten Sattler, and Marc Pollefeys. 2017. Comparative evaluation of hand-crafted and learned local features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1482–1491.
- [20] Giorgos Sfikas, Christian Heinrich, Jihad Zallat, Christophoros Nikou, and Nikos Galatsanos. 2011. Recovery of polarimetric Stokes images by spatial mixture models. *JOSA A* 28, 3 (2011), 465–474.
- [21] Giorgos Sfikas, Christophoros Nikou, Nikolaos Galatsanos, and Christian Heinrich. 2008. MR brain tissue classification using an edge-preserving spatially variant Bayesian mixture model. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 43–50.
- [22] Giorgos Sfikas, Joao Patacas, Charalampos Psarros, Antigoni Noula, Dimosthenis Ioannidis, and Dimitrios Tzovaras. 2019. A Deep Neural Network-Based Method For The Detection And Accurate Thermography Statistics Estimation Of Aerially Surveyed Structures. In *19th International Conference on Construction Applications of Virtual Reality*.
- [23] Julian Stottinger, Allan Hanbury, Theo Gevers, and Nicu Sebe. 2009. Lonely but attractive: Sparse color salient points for object retrieval and categorization. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 1–8.
- [24] Giorgos Tsigkas, Giorgos Sfikas, Anastasios Pasialis, Andreas Vlachopoulos, and Christophoros Nikou. 2020. Markerless detection of ancient rock carvings in the wild. *Pattern Recognition Letters* (2020).
- [25] Dimitrios Tzovaras and Michael G Strintzis. 1998. Motion and disparity field estimation using rate-distortion optimization. *IEEE Transactions on Circuits and Systems for Video Technology* 8, 2 (1998), 171–180.
- [26] Fuzhen Zhang. 1997. Quaternions and matrices of quaternions. *Linear algebra and its applications* 251 (1997), 21–57.
- [27] Yongheng Zhao, Tolga Birdal, Jan Eric Lenssen, Emanuele Menegatti, Leonidas Guibas, and Federico Tombari. 2020. Quaternion Equivariant Capsule Networks for 3D Point Clouds. In *IEEE European Conference in Computer Vision (ECCV)*.
- [28] Xuanyu Zhu, Yi Xu, Hongteng Xu, and Changjian Chen. 2018. Quaternion convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 631–647.