Multispectral image registration based on keypoint matching and homography estimation for cultural heritage artifacts

Polykarpos Karamaouyas\textsuperscript{1}, Athanasios Zacharopoulos\textsuperscript{2}, Kostas Hatzigiannakis\textsuperscript{2}, Michalis Andrianakis\textsuperscript{2}, Kristalia Melessanaki\textsuperscript{2}, Xenophon Zabulis\textsuperscript{1}

The registration of Multispectral Imaging (MSI) datasets in the field of cultural heritage poses unique challenges. When the fast, portable, and non-invasive systems that employ digital cameras along with band pass filters are used, the different optical properties of the filters and vibrations of the systems reveal images that are misaligned. Image registration is necessary in order to generate a useful, aligned, MSI sequence, or otherwise, a spectral cube. This work regards the registration of multispectral images of planar, or approximately planar, artifacts, such as paintings. An approach that capitalizes upon the planarity of the imaged surfaces is proposed which, in turn, employs homography estimation to achieve the registration of these images. In this context, state-of-the-art homography estimation methods are comparatively evaluated and a non-linear, robust homography method is selected as the most accurate. The selected method utilizes correspondences across consecutive images of the sequence, which are established from matched keypoint features. For the above comparative assessment, a quantitative evaluation method is proposed that is based on markers. Using this method, a benchmark dataset for multispectral image registration, annotated with ground truth, has been compiled and is publicly availed.
Multispectral image registration based on keypoint matching and homography estimation for cultural heritage artifacts

Polykarpos Karamaoynas¹, Athanasios Zacharopoulos², Kostas Hatzigiannakis², Michalis Andrianakis², Kristalia Melessanaki², Xenophon Zabulis¹

1. Institute of Computer Science
2. Institute for Electronic Structure and Laser

Foundation for Research and Technology — Hellas (FORTH)
N. Plastira 100, Vassilika Vouton
Heraklion, Crete, 700 13 Greece
Web: http://www.ics.forth.gr/
Tel: +30 2810 391600, Fax: +30 2810 391601

E-mail: karam@ics.forth.gr, azacharo@iesl.forth.gr, kostas@iesl.forth.gr, alina@iesl.forth.gr, zabulis@ics.forth.gr


© Copyright 2016 by FORTH
1 Introduction

Multispectral Imaging (MSI) is a diagnostic technique which combines digital imaging with spectroscopic analysis to recover spatial and spectral information about an artifact. MSI is implemented by producing a sequence of images, each one acquired at a narrow spectral band. In general, imaging techniques require an illumination system to illuminate the object to be depicted and an imaging sensor to capture the light backscattered from the object. For the purposes of MSI, a monochromator, most often a series of bandpass filters, is interfered in the light path either in front of the illumination system or in front of the imaging sensor. The outcome of this technique is a set of successive images, one spectral image for each spectral band, called spectral cube. Further processing of the spectral cube results to the extraction of useful information on the materials comprising the object under study.

MSI can be applied in several fields e.g. in medicine, in agriculture, in remote sensing, in food industry. One of the important applications of MSI is in cultural heritage (art conservation, archaeology, art history) [1]. Cultural heritage objects are extremely delicate, in the sense that any intervention or even contact may harm them. There are numerous diagnostic techniques for cultural heritage objects which require contact with the artwork (i.e. placement of markers) or, even more, sample removal. The advantage of MSI is that as an imaging technique performs remotely, that is contactlessly and non-invasively.

MSI has been proved to be a useful tool for restorers, archeologists and art historians. Stratigraphic analysis [2], monitoring of conservation interventions [3] and enhancement of fainted patterns [4] are among the potentials of MSI in cultural heritage. Additionally, study of the materials comprising the artwork is possible. This process is called spectrometry. For this purpose a pixel or a few pixels - corresponding to the area of the physical object comprised of the material to be studied - are selected and the intensity along the spectral cube, that is the reflectance spectrum, is extracted. The reflectance spectrum gives information about the composition of the materials.

A typical example of the above mentioned methodology is the study of pigments in [5]. Pigments have been used since the ancient ages to decorate objects. Depending on the era, the geographical location, and the artist, different pigments have been used to achieve almost the same color. For instance, Egyptian blue, ultramarine, Prussian blue etc. are typical pigments used for the color blue. Revealing of the exact pigments used in an artwork can help conservators in a conservation process and art historians in their studies. In some cases only traces of the material of interest is present in the artwork, and the spatial accuracy of the MSI is of paramount importance.

Since our focus is in the portable, fast, non-invasive systems that are based on digital cameras combined with a series of band pass filters, there are certain challenges that are to be faced. The portability of the system, that allows for in-situ acquisition, for example introduces possible camera pose changes and vibrations during the acquisition which in turn are evident as translation shifts and even rotations between...
the different spectral images. In addition, the versatility of the system and the properties of the optical elements comprising the MSI sensor, e.g., the objective lens, could introduce the need for refocusing at certain wavelength bands and thereby slightly changes in the focal distance, as well as, the Field of View (FOV). This change of FOV will be manifested as a scaling of the image. Furthermore, the camera pose may slightly change due to the required physical interaction (refocusing) with the camera. As a result of the above the acquired spectral images are not perfectly aligned and, consequently, a specific pixel of the spectral cube does not correspond to a specific object point.

A registration of spectral images is required before the spectrum extraction, so that a physical point is imaged at the same coordinates on all images of the spectral cube. The more accurate this registration is, the more spatially precise spectral measurements can be achieved. As it will be discussed in Section 2, many MSI registration methods have been proposed for the art domain as well as others, such as remote sensing and medicine, which offer a valuable algorithmic background.

In this work, an accurate multispectral image registration method is proposed, for the domain of cultural heritage and in particular that of planar paintings. This method is based on keypoint matching and a robust estimation of the underlying homography. This method is comparatively evaluated against the state-of-the-art methods, more accurate results than commonly utilized approaches to this problem. For the evaluation, landmarks are utilized to obtain ground truth in datasets that were compiled for this purpose and publicly availed as benchmarks.

2 Related work

A comprehensive, generic review of image registration methods can be found in [6]. Registration methods are there abstracted as performing four tasks, in the following order: feature detection, feature matching, transform model estimation, and image resampling and transformation. Depending on if the matched features are localized points or areas, registration methods are characterized as feature-based or area-based, respectively.

A number of works that employ image registration in the medical domain (i.e., see [7] for a survey) utilize area-based registration methods, mainly because tissue usually appears textureless and does not avail distinct keypoint features. Another difficulty in medical images is that tissue is usually deformable and of unknown shape. The work in [8] utilizes a stereo pair to reconstruct the 3D surface of the imaged tissue and register MSIs at the locations of detected keypoints, however it is able to register only very small regions of interest rather than entire images. The most relevant domain to this work is that of remote sensing which is reviewed next, in Section 2.1 followed by a review of MSI registration in the domain of cultural heritage in Section 2.2.
2.1 Remote sensing

The MSI registration problem has been extensively studied in the context of remote sensing. Despite differences, the existing work provides useful insight for the registration of MSI in the context of cultural studies too. Similarly to the field of artworks, a central problem in the MSI registration in remote sensing is the matching of the image content between different spectral images. A factor that reinforces this problem is that remote sensing often employs only a few spectral images [9, 10, 11, 12, 13, 14] making, thus, the matching of images more demanding.

Remote sensing methods have utilized both feature [12, 13, 15, 16] and area [10, 17] based registration approaches, and combinations of the two [9]. Despite the fact that area based methods suffer from illumination difference and are usually less accurate than feature based methods, the main reason for their use is that in remote sensing often image content is homogeneous and, due to lack of features, performance of feature based approaches is hindered. The employed keypoint features are SIFT [12, 13, 15], SURF [16], and other, while area based similarity cue were based on Normalized Cross Correlation (NCC) [18] and Mutual Information (MI) [11]. In the context of this paper, keypoint features are abundant in images (typically more than 10K features occur in 5 Mpixel images). Moreover, it is shown that feature based methods tend to perform more accurately than area based approaches.

Most MSI registration methods in the context of remote sensing assume a planar terrain [9, 13, 12, 14]. When this assumption is valid a homography suffices as the transformation model between images. Moreover, the RANSAC method [20] is utilized to filter spurious matches and robustly estimate the homography [9, 19, 21, 13, 14]. The need for filtering of spurious matches is eminent due to the variability of image content at different spectra. In this context, the work in [19] classified all spectral bands into three categories called Main Spectral Groups, where terrain objects usually pose the same stable spectral characteristics and behavior. After registering images within each group, the groups were then registered to each other. All registration tasks were carried out using SIFT features and RANSAC based registration.

The structure of the terrain surface cannot be always assumed as planar (i.e. when a hill is imaged), particularly when imaging altitude is low (i.e. as in aerial imaging [10, 21]). Accounting for non-planar terrain not only creates a need for the 3D reconstruction of the terrain but, also, casts the use of model-based robust techniques more difficult. For this reason more complicated methods of eliminating such matches have been developed [12, 14, 13].

2.2 Cultural heritage

In [22], a comprehensive review of the challenges in the registration of MSI is presented in the domain of cultural heritage for paintings and manuscripts where images from different modalities or wavelengths often contain both similar and unique information. In the domain of cultural heritage image registration has been employed in image
mosaics and the alignment of multispectral images.

Image mosaic, the process of obtaining a wider field-of-view of a scene from a sequence of partial views by stitching multiple overlapping snapshot images, has received considerable attention in this domain. The main reason is that multispectral images of high resolution can be acquired through the stitching of partial images of an artwork. Typically, in such cases, image overlap is small so that a few views suffice for the generation of the mosaic. In several works, manual or semi-automatic approaches were employed (i.e. [23]). In these works, reference points were mostly manually selected and corresponded across images to achieve registration. In [24], a cost-efficient and open source system is proposed for this purpose, where a mosaic is achieved semi-automatically using the public domain software ImageJ to correspond SIFT Features along with a RANSAC strategy for filtering the established matches.

In more relevance to this work is the registration of images that portray the same region at different spectra, in order to create a multispectral cube. Similarly to mosaic, manual or semi-automatic techniques have been employed, with the use of manually selected control points or correlation (i.e. [25, 26, 27, 28, 29]); moreover, only translation (shift) is considered in the alignment process, in these works. In the review paper [1], the need for translation and scale transformations for the alignment of multispectral images for cultural applications to form an image cube is acknowledged. Similarly in [30, 31] correlation based registration of images, implemented by the VIPS image analysis system [32] is proposed. Cross correlation has been also employed in [22], utilizing wavelet response instead of image intensities. In [33] and [34] SIFT keypoint features and cross correlation are combined. Moreover, in [35], spurious matches are reduced using the RANSAC.

A general observation in the majority of approaches in the field of cultural heritage, with the exception of [22] where a painted grid of blue lines was used, is that evaluation of the proposed algorithms, is conducted using mainly qualitative, rather than quantitative, methods. As a result, there is lack of comparative evaluation among the proposed methods for multispectral image registration.

2.3 Contributions of this work

The proposed approach is based on keypoint correspondences. Utilizing these correspondences homography transformations are estimated that enable the registration of images into a multispectral cube. It is, furthermore, shown that keypoint correspondences yield higher accuracy than correlation based methods, because they cope better with the radiometric differences between images acquired at different wavelengths. Moreover, it is shown that the robust, non-linear homography estimation approach employed provides higher accuracy from the conventional RANSAC based approaches. To quantitatively and comparatively validate the proposed method MSI datasets imaging conventional artworks, annotated with ground truth regarding the registration of images, were created. Utilizing those datasets, comparative, qualitative

\[ \text{http://imagej.net/} \]
and quantitative evaluation of the proposed method is achieved. These datasets are publicly availed as a benchmark dataset.

3 Overview of approach

The core points of this approach are illustrated in Figure 1. In the context of this work, the imaging distance is sufficiently large and, so, the imaged surface is adequately approximated by a single planar surface, both in term of structure as well as depth of light perversion to the surface.

Ideally, Figure 1(a), the scene is imaged by a stable camera and a wheel of filters whose surface is perpendicular to the camera’s optical axis. All images are acquired from the same camera pose (location and orientation), at the same focus distance and are, thus, subject to the same scaling effect. Then, images $I_i$ would be aligned and stacking them would suffice to obtain an aligned spectral cube $S$.

Typically, the sets of filters used for MSI are manufactured to exhibit the same optical path, that is the distance the light travels inside the filter is the same for all filters. Consequently the interference of the filters in the light path has the same impact for all wavelengths and switching between them does not affect the focus. Still, in practice the focus may have to be readjusted especially when switching between spectral bands, i.e. from visible to NIR. Furthermore, the focus is affected by the camera lens. The lens cannot provide perfect focus for all wavelengths and refocusing is required as the wavelength changes. The refocusing between image acquisitions, in turn, changes the camera FOV. In addition, during refocusing, the camera may be accidentally vibrated or even moved, which is common when acquisition takes place in the field rather than in a controlled laboratory environment. The effects of these two factors are modeled as a change in image scaling and a change of camera pose (location and orientation), respectively. The outcome of this acquisition is a stack of misaligned images, as shown in Figure 1(b), acquired upon image planes with relative differences in 3D pose and size among them.

In the above context, it is assumed that the registration of the coordinate frames of any two images, so that pixels with the same coordinates image the same physical point, is sufficiently modeled by a homography transformation. A homography is a perspective transformation between two images of the same planar surface in space, has 8 Degrees Of Freedom (DOF) and is represented by a $3 \times 3$ matrix $H$. A homogeneous 2D point upon the first image, let $p = [x \ y \ 1]^T$ is transformed to occur at its corresponding location $q$ on the second image as $q = H \cdot p$, where $\cdot$ denotes matrix multiplication. This transformation fully covers the effects due to changes in camera pose and image scaling, assuming that a planar surface is imaged. Establishment of point correspondences across the two images, enables the estimation of this homography transformation, Figure 1(c). Using this estimate an image can be warped to be registered to the other.

Usually, the content of two arbitrary spectral images from the spectral sequence is
Figure 1: Method overview (see Section 3). Geometry of ideal (a) and realistic (b) image acquisition. (c) Correspondence establishment among two images can provide an estimate of the homography that registers one image to the other. (d) Registration of all images in a sequence is used to align multiple spectral images.
sufficiently different for feature matching to be fruitful. In contrast, images acquired at neighboring, spectra are more similar and, thus, correspondences are searched only in consecutive images. By registering consecutive images, all images can eventually be registered to the first image (or any other of the sequence). Henceforth, the homography transformation between images $I_i$ and $I_j$ will be denoted as $H_{i,j}$. Warping the images according to the registrations, creates an aligned $S$, Figure 1(d). The accuracy by which these registration transformations are estimated, determines the accuracy by which images $I_i$ are aligned to form $S$. It is noted that that as image posture and focal length varies, the imaged physical area varies too. Therefore, albeit registered, the warped images do not have a complete overlap. As a result, physical points near the boundary of the field of view may not be imaged in all images of the sequence.

4 The proposed approach

In this section, the proposed approach is formulated, followed by a method for the evaluation of its accuracy.

4.1 Method formulation

The input to the proposed method is the spectral image sequence $I$, which is comprised of the original, misaligned, spectral images $I_i$, $i \in [1, N]$. The computational steps of the proposed method are the following.

4.1.1 Feature detection

Initially, keypoint features were detected, in all input images $I_i$. The SIFT [36] features were selected for this purpose. Keypoint feature locations in $I_i$ are noted: $q_{i,k}$, $k \in [1, \nu_i]$, where $\nu_i$ is the number of detected features in $I_i$. If available, all the 16 bits of pixel depth (see Section 5.1) are utilized in feature extraction. For this reason, the library in [37] was slightly modified to treat 16-bit images, as its original version supports only 8-bit images. In preliminary experiments, it was confirmed that using the entire dynamic range increases the number of detected features and their robustness of matching. This is attributed to the better, more characteristic, keypoint descriptions obtained from the additional information.

4.1.2 Feature matching

In the next step, keypoint correspondences were established in consecutive images. Matching is performed as in [36], requiring a high ($> 0.8$) Nearest Neighbor Distance Ratio to establish a match. To increase the robustness of matching, feature correspondences are additionally filtered to be symmetric or, otherwise, satisfy the “left-right consistency” constraint [38]. Let $M_{i,i+1}$ the matches between consecutive images $I_i$ and $I_{i+1}$, where $i \in [1, N - 1]$. Let also $\lambda_{i,i+1}$ the number of these matches. The
result of this matching process is stored in matrix $M_{i,i+1}$. Then $M_{i,i+1}$ is a $\lambda_{i,i+1} \times 2$ matrix containing the indices of the corresponding features.

### 4.1.3 Transform model estimation

In this work, the homography transformations between any two images with established keypoint correspondences between them is performed based on the work in [39]. We choose to estimate transforms between consecutive images, as their content is typically more similar.

More specifically, each homography $H_{i,i+1}$ is estimated robustly to eliminate remaining spurious matches. Homography estimation utilizing the implementation in [39] and based on the SIFT established correspondences, is henceforth referred as Method 1. In the estimation, point coordinates are normalized to improve conditioning [40]. Then, Least Median of Squares (LMedS) linear fit is applied to reject spurious matched or, otherwise, matching outliers [41]. Finally, non-linear refinement of the linear homography estimate is performed by minimization of the transfer error. The minimization is performed using the Levenberg-Marquardt algorithm [42], as implemented in [43]. It ought to be noted that the minimized error is symmetric for the two images and, thus, the ordering of the two images is not of concern.

A conventional order of registration is after the nominal order of images in the sequence. Transformations $H_{1,i}$ are computed for $i \in [1, N]$, that map all images of the sequence to the 1st, reference image. The transformations are computed from the output of the previous step as: $H_{1,i} = H_{1,2} \cdot H_{2,3} \cdots H_{i-2,i-1} \cdot H_{i-1,i}$, where $H_{1,1}$ is the $3 \times 3$ identity matrix $I_{3x3}$.

An alternative order of image registration was also tested, motivated by the observation that images near the outer limits of the spectrum exhibit greater noise levels, thus when the first image is chosen as the reference, sequential images are registered to one of the most noisy images of the sequence and registration error will be greater. By choosing the reference image to be in the middle of the spectrum an attempt is made to reduce this effect. In addition, using one of the middle images as the reference reduces error propagation, as shorter sequences of homographies are utilized. Let image $v$ be the reference image then the transformations $H_{v,i}$ are computed as:

$$H_{v,i} = \begin{cases} H_{v,v-1} \cdot H_{v-1,v-2} \cdots H_{i+2,i+1} \cdot H_{i+1,i} & \text{if } i < v \\ H_{v,v+1} \cdot H_{v+1,v+2} \cdots H_{i-2,i-1} \cdot H_{i-1,i} & \text{if } i > v \\ I_{3x3} & \text{if } i = v, \end{cases}$$

(1)

where $H_{i+1,i} = H_{i+1,i}^{-1}$. Apart from a fully projective (8 Degrees of Freedom) homography, the method above can be used to estimate affine homographies. In these homographies, the third row is fixed to [0, 0, 1] and thus depend upon 6 Degrees of Freedom. For the framework of the utilized experimental setup we found that the estimation of an affine homography is more stable compared to estimating a fully projective homography Section 5.6.
Besides the planarity of the surface, the relatively large focal length (narrow FOV) typically utilized in multispectral apparatuses, further justifies this approximation. Based on these experiments, the proposed approach estimates affine homographies. It is noted that, in the case of affine homography estimation, there is no non-linear refinement step in the optimization. The reason is that, for affine homographies, the algebraic error is equal to the geometric one. Therefore, the optimization of the algebraic distance, also minimizes the geometric distance and no subsequent refinement is necessary.

4.1.4 Image resampling and transformation

The original images are warped using the corresponding homography transformations that were estimated in the previous step, using bilinear interpolation. That is, image $S_i$ is computed by warping image $I_i$ according to homography $H_{1,i}$. Stacking the warped images $S_i$, comprises the spectral cube $S$, which is the output of the method.

4.2 Accuracy evaluation

To objectively evaluate the accuracy of a registration method, ground truth correspondences between images can be used. Assuming such correspondences, registration error for image $i$, is measured as the distance of landmarks points in the transformed images to the corresponding points in the chosen reference image $v$. Landmarks are placed on the same physical points in all images; ideally, in the registered images, their locations should be coincident.

Registration accuracy is measured as follows. Let $\mu$ the number and $u_{v,k}, (k = 1, \ldots, \mu)$ the locations of landmark points in the $v^{th}$ image. Accordingly, let $u_{i,k}$, the locations of landmarks in the rest of the images, where $i$ enumerates images and $k$ landmarks. Let the locations of the transformed landmarks of image $I_i$ to the chosen reference image $u$, be: $v_{i,k} = H_{u,i} \cdot u_{i,k}$. As locations of transformed landmarks should ideally be coincident to landmarks in the $v^{th}$ image, the measured error is the mean registration error of all correspondences:

$$E = \frac{\sum_{i=1}^{N} \sum_{k=1}^{\mu} \left\| u_{v,k} - v_{i,k} \right\|}{(N - 1) \cdot \mu}$$

(2)

The error without the effect of registration, that is creating $S$ by stacking the original images $I_i$, is referred as original error and denoted as $E_0$. It is computed by substituting $v_{i,k}$ with $u_{i,k}$ in (Equation 2).

Albeit the proposed method matches images sequentially, error is measured according to the reference image. Thereby, residual errors that are propagated through registrations are accounted and the overall error is measured. In this way, the effective error for the end-user of the result is assessed.
5 Experiments

5.1 Image acquisition

IRIS II is lightweight portable system comprising of a high resolution camera, a novel filter wheel able to interchange 28 filter positions and fast electronics. The sensor, in combination with the band-pass filters, allows sensitivity to the UV region (350 nm up to 400 nm, BP10nm), to the visible (400 nm-700 nm, BP25nm) and to the Near IR region (700 nm-1200 nm, BP50nm). For the needs of this experiment an Electrophysics 25mm f/1.3 lens was used for the acquisition. The dynamic range of the camera is set to be 8 bits for each data point. With this system 28 quasi-monochromatic high resolution images (5 MPixel), are collected, one for each chosen transmission band. The final 3D spectral cube can have maximum dimensions of $2560 \times 2048 \times 28$ data points. The system is portable so that it can be transferred to the objects location, for measurement and analysis.

For the acquisition of the spectral cube the camera is placed against the object under investigation. The illumination sources are placed at approximately 45 degrees relatively to the object surface axis. For each transmission band a standard acquisition process is followed. The focus is properly adjusted. A white highly reflecting target (Spectalon, $R = 99\% @ 250 - 2500$ nm) is placed in front of the object and the camera-lens are adjusted in order to reach the maximum possible mean intensity value. A white image is captured. Next, keeping all settings unchanged, the white target is removed and an image (spectral image) of the object is captured. Finally, still keeping all settings unchanged, light entrance is blocked in front of the camera and a black image is captured. This black image corresponds to the electronic noise of the camera.

Data acquired from the IRIS are initially analyzed by viewing the spectral images in a series starting from the lower, and moving to the higher wavelengths. The technique enables the stratigraphic analysis of the surface layers, since light increases its penetration with wavelength. Results like the existence of a varnish layer, the pigment layers, the in-depth location of the signature can be extracted. Furthermore, under-drawings and over-paintings can be discovered when studying the near infrared images. Furthermore the acquired spectral cube can be normalized by originally subtracting the noise from the spectral and the white images, separately for each transmission band and following by taking the division of the resulting images. The resulting quotient for each pixel is a value between 0 and 1 and is correlated to the reflectance of the corresponding point/area of the object. In order to re-transform these values in image values the quotient is multiplied by a constant. The value of the constant depends on the final image dynamic range. The whole process is performed automatically via custom made software programmed in Labview (version 2013). The precision of the extracted reflectance (quotient) is 5 decimal digits. Although the dynamic range of the originally acquired images was 8-bit the quotient is converted to 16-bit in order to maintain the quotients precision that is the new dynamic range of the image. Next, image registration is applied and the spectral cube is extracted. Finally the spectral
cubic is analyzed by means of imaging spectrometry, where each image point and/or image area can be expressed as a reflectance spectrum.

5.2 Datasets

5.2.1 Mock-up painting

To approximate a typical artwork and comprise the needs of the registration research, initially a mock-up painting was created consisting of pencil drawing and various layers of paint and varnishes that produce different images for the different wavelengths acquired, as it can be seen in Figure 2. Initially, a wooden panel was coated with a white tempera preparation layer. A sketch of a woman’s head, in cubistic technique, was then drawn with pencil and over it, a draft drawing of a landscape with a pair of houses was sketched. The paint layer that followed after the second drawing (the landscape), was created using a variety of Giotto tempera paints (red, yellow, blue, white, green, black and brown) fully covering the pencil drawings. To finish the paint, two different varnish layers were then applied, a dammar varnish on the left side of the painting and a mastic varnish on the right side of the painting, leaving a vertical band in the middle of the painting with no vanish as a color depended reference area.

5.2.2 Byzantine icon

A Byzantine icon from a private collection has been also used for the purposes of our work. This icon, painted in egg tempera with gold leaf, on a wooden panel surfaced, depicting the “Saint John the Baptist”. The half frontal figure, is set against a gold background and wears a green mantle and blue fore tunic. The subject wears a pair of wings, holds a tied scroll in his left hand and raises his right in blessing. The halo is defined by two concentric circles.

5.3 Dataset and ground truth annotation

To quantitatively assess the registration’s accuracy, multispectral image sequences annotated with ground truth correspondences were required. As such datasets were not publicly available, they were created using the sample artworks of Section 5.2 and provided as a pertinent benchmark. Introduced markers were used for the μ landmarks establishing ground truth correspondences across all images of the sequence, as visual features in artwork (e.g. paintings), are typically wavelength dependent, and thus, manual annotation of ground truth could be subjective.

In particular, μ = 16 small markers were printed using a Lexmark E250dn laser printer, with regular Q-Connect black toner upon white A4 laser printer paper using the design template of Figure 5(left). Each target was a, square, small, black and white 2 × 2 checkerboard, Figure 5(right), which was made sure that was clearly distinguishable in each wavelength. The targets were placed in a 4×4 grid arrangement covering almost entirely the field of view of the sample artworks as shown in Figure 4.
Figure 2: Spectral image registration experiment for the landscape with houses described in Section 5.2.1. Shown are 9 out of the 23 images of the sequence, in particular those at 425, 525, 750, 800, 850, 900, 1000, 1100, 1200 nm. Wavelength increases top to bottom and left to right.
Figure 3: Spectral image registration experiment for the Byzantine icon described in Section 5.2.2. Shown are 9 out of the 23 images of the sequence, in particular those at 360, 380, 425, 500, 525, 750, 900, 1000, 1100 nm. Wavelength increases top to bottom and left to right.
MSI registration for cultural heritage

Figure 4: Sample images from the benchmark dataset for the evaluation of the proposed algorithm, corresponding to the 400, 500, 750 and 1000 nm wavelengths (top to bottom and left to right), using sample artwork of Section 5.2.1. Upon the painting, μ = 16 markers are arranged in a 4 × 4 grid, facilitating annotation with ground truth correspondences.

Marker size was decided to 25 mm² so that the markers were imaged by ≈ 40 × 40 pixels in the final images.

Using the sample artworks augmented with the targets, datasets were acquired using the apparatus of Section 5.1 consisting of 23 spectral images, instead of the 28 that the IRIS II system can acquire. The reason is that due to deterioration 5 filters have been excluded from the apparatus and, hence, the datasets are comprised of 23 images each. In effect, this makes the matching process more difficult as some of the image pairs exhibit now greater spectral and, thus, visual differences. The wavelengths that the 23 images were acquired are the following: 360 nm, 370 nm, 380 nm, 400 nm, 425 nm, 475 nm, 500 nm, 525 nm, 550 nm, 575 nm, 600 nm, 625 nm, 650 nm, 700 nm, 750 nm, 800 nm, 850 nm, 900 nm, 950 nm, 1000 nm, 1050 nm, 1100 nm and 1150 nm, covering a field of view that corresponds to an area of 190 mm by 142 mm upon the surface of the artworks. The resulting datasets for sample artworks of Section 5.2.1 and Section 5.2.2 will be thereafter referred to as Dataset 1 and Dataset 2 respectively.

To detect the landmark points accurately, an automatic corner detection algorithm
was employed, based on [44], that would accurately find the location of the checkerboard inner corner, given an initial coarse estimate. Manual annotation was greatly reduced, using this automatic refinement method. As misregistration is usually relatively small across consecutive images, the refinement result for an image can be used as the coarse estimate for the succeeding image in the sequence. In this way, only the initial estimate was provided manually for $I_1$. The resultant centers of the checkerboards were visually inspected, to ensure the accuracy of the corners location. These centers were set as the landmarks $u_{i,k}$, see (Equation 2), comprising the ground truth annotation of the dataset. The order of the centers is consistent in the annotation and avails the landmark correspondence information.


### 5.4 Computational cost

The main software modules implementing the proposed approach are the following.

- **$M_1$, Feature detection**: a modified version of the library in [37] to support feature detection in 16 bit images.

- **$M_2$, Feature matching**: is computationally complex because it extends the implementation in [37] so that only symmetric matches between the two images are utilized.

- **$M_3$, Homography estimation and warping**: receives matches from $M_2$ and provides the final result.
During the processing of the 23, 2560 × 1920 images in Dataset 1, a total of 229017 features were detected and a total of 77812 matches were established, finally estimating 22 homographies registering consecutive images of the sequence. Execution of a serial implementation of the above pipeline, for Dataset 1, lasts 1168 sec (19.48 min), on a single core of a conventional PC with an Intel CPU at 2.27 GHz with 6 GB of RAM. During this execution computational effort was shared among the 3 modules as follows: M1 11.58%, M2 87.99%, and M3 0.43%.

Modules M1 and M2 are amenable to parallel implementation while M3 is not, because it employs an iterative algorithm. Using the OpenMP API [45], a parallel implementation of the above pipeline, for M1 and M2, utilizes the multiple cores available in a CPU. For the same data, this implementation executed in 447 sec (7.45 min) on a PC with an Intel CPU at 2.27 GHz, with 4 cores, and with 6 GB of RAM. The same run on another PC with Intel CPU at 3.07 GHz, with 8 cores, and with 6 GB of RAM, lasted ≈ 37% less, that is 281 sec (4.69 min).

5.5 Qualitative evaluation

For the qualitative evaluation, the images of Dataset 1 were registered. Visual inspection indicated a greatly accurate registration and iterating the images as a video does not reveal any apparent motion. In Figure 2, 9 out of 23 registered images are shown. As observed in the figures and as discussed in Section 3, images \( I_i \) do not image exactly the same area. This is the reason that registered images are incomplete (have no value for some pixels); at the bottom rows of the figures in the example.

A more insightful assessment of the accuracy of the result is provided by tracing the results along the image sequence by slicing the multispectral cube \( S \) on a plane defined by choosing one of the two dimension of the images and the wavelength dimension as shown in Figure 6(top). Shown on the left is a sample image of Dataset 1, with an image column marked in red, while in the center and right 2D slices of \( S \), along the marked column. In the middle, shown is the slice corresponding to the original stack of images before registration; the misalignment of spectral images is evident. On the right, shown is the slice corresponding to the registered stack of images (\( S \)). A similar visualization is provided in Figure 6(bottom), where image alignment is demonstrated showing the same image region for the original and registered images. In particular, a square 301 × 301 pixel image segment is marked on the sample image of the sequence (top, left). The contents of the same segment in all images \( I_i \) and \( S_i \) is shown in the middle and bottom row, respectively. The registration improvement induced by the proposed method is clearly observed.

In preliminary experiments, we confirmed that utilizing all 16 available bits in the original images, is preferable from downgrading the images to the conventional 8 bit format. The additional information provides more features and more robust matches. We found this particularly important in blurry or poorly focused images, where matches are considerable fewer. Overall, the effect of the proposed approach is evident and empirically assessed as accurate enough for creating useful multispectral
Figure 6: Qualitative evaluation of multispectral registration (see Section 5.5). Top: inspections of slices of the unregistered (middle) and registered (right) spectral cube $S$, along the image column marked on the sample image (left). Middle, bottom: In the sample image above, marked is a red region. The middle and bottom rows, shown are the contents of this region for $I_i$ and $S_i$, respectively for the wavelengths of 425nm, 475nm, 525nm, 575nm, 625nm, 675nm, 750nm, 850nm, 950nm, 1050nm, 1150nm. A crosshair is overlaid upon each segment to assist comparative inspection.
5.6 Quantitative evaluation

The accuracy of the variants of the proposed approach was studied in detail and comparatively to competing methods, using the approach in Section 4.2. For each image of the input sequence SIFT feature detection yielded several thousands of features, in the range of $10^K - 20^K$, according to the image content of each spectral image.

Once the feature detection process was complete, the features in the area corresponding to the inserted landmarks were marked and specifically excluded from the registration process. This was performed so that the insertion of markers does not provide additional information to the registration process and a realistic assessment of accuracy is obtained. The images were not altered by this process, since this exclusion took place after the detection of features. In fact, this exclusion avails less information to the matching process as markers covered some image area where other keypoints could have been detected.

Since, a decision was made to use the fit of an Affine Homography for the proposed Method 1, as explained in Section 4.1.3, the verification of that decision was the first experiment performed. Using either Dataset 1 or Dataset 2, registration was performed by both a (6 DOF) Affine Homography or a full (8 DOF) Homography and the comparative results were plotted in Figure 7. For both datasets the (6 DOF) Affine Homography produced more stable and reliable results. This verifies the decision to constrain the DOF of the optimization, motivated by the a-priori knowledge of the physical experimental setup, so that a more well posed optimization, that avoids possible local minima, is possible.

In the next experiment, the proposed method was compared to another correspondence-based method for the estimation of the homography, that is found...
Figure 8: Registration errors \( (E) \) for the evaluated Method 1 and Method 2, described in Section 5.6, plot as a function of wavelength for Dataset 1 (left) and Dataset 2 (right).

To be the state-of-the-art for cultural applications in the relevant literature. Using the same input correspondences, a robust estimation of the homography using RANSAC \cite{20} was obtained, followed by robust least-squares fit of the homography to the inliers of the consensus. This was achieved by application of the Levenberg-Marquardt algorithm \cite{42} and, henceforth, this method is referred as Method 2. The results are shown in Figure 8 for Dataset 1 on the left and Dataset 2 on the right. It can be noted that error for the first image of the sequence is zero as, in this case, we chose the first image to be the reference for the sequence and that reference image is therefore registered with itself. The mean errors (and std) for the two methods for Dataset 1, in pixels, were 0.690 (0.094) for Method 1 and 1.429 (0.49) for Method 2 while for Dataset 2 were 0.883 (0.143) for Method 1 and 1.929 (0.242) for Method 2. For the particular configuration (FOV and imaging distance) their conversion in \textit{mm} is 0.078 (0.011) and 0.162 (0.055) for Dataset 1, and 0.10 (0.016) and 0.218 (0.027) for Dataset 2, respectively. Those results are also presented in the first two lines of Tables 1 and 2. The increased accuracy of Method 1 over Method 2 is attributed to, reported, better performance of the adopted homography estimation approach.

The proposed, correspondence-based, approach was also tested against indicative global matching methods utilized, based on pixel photometric differences. In the experiment, homographies were also estimated using the Sum of Absolute Differences (SAD), Mutual Information (MI), and Normalized Cross Correlation (NCC), as the image difference objective functions. Several methods for unconstrained multivariable optimization were tested for the optimization of these objective functions. In particular the: simplex search method \cite{46}, Steepest Gradient Descent \cite{47} and Limited-memory BFGS \cite{48}. Finally, BFGS Quasi-Newton method \cite{49} was chosen as the one that performed the best in the registration tests. In the experiments below, the BFGS Quasi-Newton method was utilized for the optimization of the SAD, MI, and NCC.
objective functions.

In Figure 9, the registration error \( E \) is plotted as a function of wavelengths for the datasets collected for this experiment. In the same plots the original error \( E_0 \) of the unregistered images was also included for reference; in some cases, as it can be seen in the plots, global methods deteriorate the original image alignment. Also, in the same figure, it is noticeable that in the case of the Dataset 2, the Mutual Information method fails when trying to register the first images of the dataset. This error is probably due to the lack of painted features in the images emitting in the lower bands of the spectrum (UV) in this sample, as it can be seen in Figure 3. In these wavelengths, registration mainly relies on keypoint features due to the surface of the artwork, such as cracks in the paint and scratches, rather than the actual painting. The results indicate the increased suitability of the proposed method for the registration of multispectral images. The inferior performance of global photometric approaches to the proposed method, is attributed to the significant photometric variation between wavelengths on the majority of the image pixels. In contrast, keypoints are based on image gradient, which is less sensitive to such changes.

In addition, the impact of changing the reference wavelength to some at the center of the spectrum, as denoted in Section 4.1 (Equation 1), was tested as follows. The experiments were repeated placing the \( v = 14^{th} \) spectral image (700 nm) as the reference. In most cases, including the proposed method, a small decrease of error \( E \) was observed.

The mean average errors (and std) results of the mentioned experiments are reported in the Tables 1 and 2. The results show a subpixel error, for the proposed method. For reference, the initial alignment error \( E_0 \) was 12.25 (5.61) pixels or 1.38 (0.63) mm.
Table 1: Mean original and registration error (and std), in pixels and \( mm \), for the compared methods of Section 5.6 for Dataset 1, described in Section 5.2.1.

<table>
<thead>
<tr>
<th>Error</th>
<th>pixels</th>
<th>( mm )</th>
<th>pixels (( v = 14 ))</th>
<th>( mm ) (( v = 14 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>0.690 (0.094)</td>
<td>0.078 (0.011)</td>
<td>0.679 (0.113)</td>
<td>0.077 (0.013)</td>
</tr>
<tr>
<td>Method 2</td>
<td>1.429 (0.490)</td>
<td>0.162 (0.055)</td>
<td>1.298 (0.552)</td>
<td>0.147 (0.062)</td>
</tr>
<tr>
<td>SAD</td>
<td>6.693 (3.275)</td>
<td>0.756 (0.370)</td>
<td>4.738 (3.192)</td>
<td>0.535 (0.361)</td>
</tr>
<tr>
<td>MI</td>
<td>8.007 (3.852)</td>
<td>0.905 (0.435)</td>
<td>6.800 (4.712)</td>
<td>0.768 (0.533)</td>
</tr>
<tr>
<td>NCC</td>
<td>9.156 (3.493)</td>
<td>1.035 (0.395)</td>
<td>5.603 (3.460)</td>
<td>0.633 (0.391)</td>
</tr>
</tbody>
</table>

Table 2: Mean original and registration error (and std), in pixels and \( mm \), for the compared methods of Section 5.6 for Dataset 2, described in Section 5.2.2.

We should note that Mutual Information fails.

6 Conclusions

A robust approach for the problem of image registration for MSI in cultural heritage was described. The proposed method is based on matching keypoints in pairs of consecutive images from the multispectral sequence to estimate a homography. Through the registration between all pairs of consecutive images, images of the sequence are registered to a chosen reference image from the sequence. Moreover, a technical procedure to quantitatively evaluate the accuracy of a registration method was utilized. In this way, a comparative evaluation of registration methods is availed that can assist in the selection of the best performing method for a specific application or visual content.

Samples, containing realistic challenges for the problem of painting MSI, were employed and experimental datasets were created to comparatively evaluate the accuracy of the method versus the state-of-the-art. It was found that approaches based on keypoint correspondences were more accurate than global matching (i.e. based on correlation of image intensities or mutual information of images) methods in the context of planar MSI datasets. In addition, it was found that the proposed approach showed subpixel accuracy and was even more accurate than the widely-employed, keypoint-based image registration approach, using RANSAC filtering.
Acknowledgments

This work was supported by the Foundation for Research and Technology—Hellas, Institute of Computer Science (FORTH—ICS) internal RTD Programme “Ambient Intelligence and Smart Environments.” Authors thank Manolis I. A. Lourakis for fruitful discussions and help with application of his library for non-linear homography estimation [39].

References


