Intensity histogram equalisation, a colour-to-grey conversion strategy improving photogrammetric reconstruction of urban architectural heritage

Andrea Ballabeni and Marco Gaiani

Department of Architecture, Alma Mater Studiorum, University of Bologna, Italy
Emails: marco.gaiani | andrea.ballabeni@unibo.it

This paper presents a novel technique for the conversion of the colour signal into a grey-level signal tailored to the 3D reconstruction of urban scenes. The presented approach, named IHE from Intensity Histogram Equalisation, derives from previous methods. The proposed approach takes as input a set of images showing the same urban object possibly captured under different conditions and by different cameras. Then it processes the input images by maximising the peaks of their chromatic channel distributions, in order to preserve the chromatic in-formation as much as possible. IHE has been evaluated by comparing its performance with that of other state-of-the-art algorithms in terms of 3D reconstruction. The experiments, carried out on two datasets, show that IHE generally outperforms the other approaches.

Received 08 April 2016; revised 06 June 2016; accepted 16 June 2016

Published online: 3 August 2016

Introduction

In the field of Architectural Heritage (AH), 3D model construction and visualisation, using photogrammetric techniques, is increasingly becoming a key approach, ensuring ease of use and efficient results, even for non-professionals. This fast increase of use of photogrammetric techniques is mainly due to the great advance, in the last few years, of the digital photogrammetric pipeline for 3D reconstruction leading to fully automated methodologies able to process large image datasets and delivering 3D products with a level of detail and precision, which change according to the applications [1-3].

The integration of computer vision algorithms with reliable and precise photogrammetric methods is nowadays producing (commercial and open) successful solutions (often called Structure from Motion (SfM), firstly introduced by Ullman [4]) for automated 3D reconstructions from large image datasets [5-7]. A significant progress has been recently achieved in the areas of efficient algorithms for scalable image matching [8], large-scale bundle adjustment [9], and in generating dense and well-calibrated clouds of point as an output [10], the core components of the photogrammetric pipeline. As a result, it is nowadays possible to easily reconstruct large scenes from image sequences and at low costs [11]. This fast technical advances led to the availability of many freeware, on-line, and commercial software (e.g. Autodesk ReCap [12], ARC3D [13], VisualSFM [14] and Agisoft PhotoScan [15]). This software can perform a semi-automatic 3D reconstruction starting from a collection of images, in a context in which different people may have taken these images at different times and with different cameras. Similar images and points have to be recognised and merged to produce a model, by means of matching algorithms that allow the identification of accurate correspondences. These correspondences are then used by the SfM algorithms to estimate the precise camera pose, which are finally used as an input into multi-view-stereo (MVS) methods that produce dense 3D models with an accuracy comparable to laser scanners [16].

The standard pipeline for geometry reconstruction from images involves four major algorithmic steps (Figure 1):

- Image capture & pre-processing to improve the image quality for successful photogrammetric processing. It consists mainly in colour enhancement, image de-noising, colour-to-grey conversion and image content enrichment;
- SfM infers the extrinsic camera parameters (position and orientation) and the camera calibration (focal length and radial distortion) by finding sparse but stable correspondences among images. A sparse point-based 3D representation of the subject is created as a further product of camera reconstruction;
- MVS reconstructs dense 3D geometry by finding visual correspondences in the images using the estimated camera parameters. These correspondences are triangulated yielding dense 3D information;
- Surface Reconstruction & Texturing: takes as an input a dense point cloud and produces a globally consistent textured surface mesh.

In detail, in this pipeline, the camera self-calibration and the orientation of the whole images dataset heavily depend on the number of image features that can be detected, on the compact representation of the visual appearance of the extracted region, and, in the end, on the reliability of the matched extracted region throughout the entire image collection.

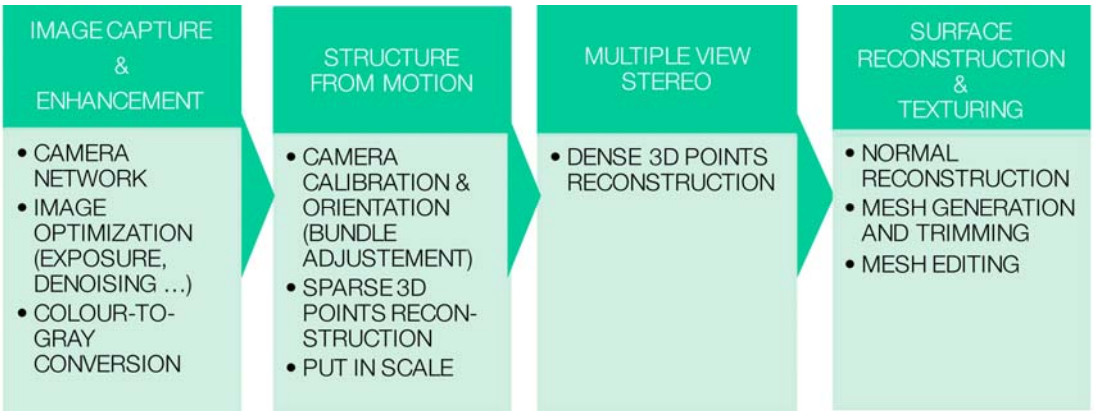
IMAGE CAPTURE & ENHANCEMENT
• CAMERA NETWORK
• IMAGE OPTIMIZATION (EXPOSURE, DENOISING ...)
• COLOUR-TO-GRAY CONVERSION
STRUCTURE FROM MOTION
• CAMERA CALIBRATION & ORIENTATION (BUNDLE ADJUSTEMENT)
• SPARSE 3D POINTS RECON-STRUCTION
• PUT IN SCALE
MULTIPLE VIEW STEREO
• DENSE 3D POINTS RECONSTRUCTION
SURFACE RECONSTRUCTION & TEXTURING
• NORMAL RECONSTRUCTION
• MESH GENERATION AND TRIMMING
• MESH EDITING

Figure 1: The standard pipeline aimed at the geometry reconstruction from images.

Feature identification and matching is at the base of many automated photogrammetric and computer vision problems and applications like 3D reconstruction, dense point cloud generation, object recognition or tracking, etc. A feature detector (or extractor) is an algorithm that takes an image as an input and delivers a set of local features (or regions) while a descriptor computes on each extracted region a specific representation of the extraction. Good image features should be independent from any geometric transformation applied to the image, robust to illumination changes and they should have a low feature dimension in order to perform a quick matching. Although these features can be edges, ridges or regions of interest, the image features used in most SfM approaches comprise scale- and affine-invariant region (or blob) detectors [17], where detection is usually performed with a Difference of Gaussian (DoG) or Histograms of Gradients (HoG) or Hessian methods. Once points and regions (invariant to a class of transformations) are detected, (invariant) descriptors are computed to characterise the feature. Currently, efficient blob description methods are merely invariant to linear changes in apparent brightness and to the geometric trans-formations of translation, rotation, and scale. Combinations of these basic geometric transformations are valid approximations only to small perspective changes and hence, large perspective changes cannot be handled. Nowadays the most popular and used operator is the Scale-invariant feature transform (SIFT) method, developed by Lowe [8]. SIFT has good stability and invariance and it detects local key points with a large amount of information using the DoG method. As reported in literature [18-21], the typical failure cases of the SIFT algorithm are changes in the illumination conditions, reflecting surfaces (e.g. cars or windows), object/scene with strong 3D aspect, highly repeated structures in the scene and very different viewing angle between the images.

Although the source of feature extraction is generally a collection of true-colour frame images, and despite numerous attempts have made to exploit the colour information, the today most common feature detectors and descriptors have been developed to work on single-band greyscale images since it - amongst other reasons - greatly reduces the computational complexity of the algorithm, when compared to the use of the standard three channels in a full colour image. Also SIFT operates this way.

Approaches based on log-RGB colour transformations, such as the one described by Maddern *et al.* [22], are not applicable in this context: these methods showed to be effective in natural scenes, where colour wavelengths are rather uniformly distributed, while our datasets show strong colour wavelengths biases if compared to natural scenes datasets: red and yellow tones are more widely represented than other colour tones.

Multi-view stereo is the general term denoting a group of techniques that use stereo correspondence as their main cue, usually applied on more than two images [16, 23-24]. The pairwise disparity estimation allows to compute image to image correspondences between adjacent rectified image pairs, and independent depth estimates for each camera viewpoint. An optimal joint estimate will be achieved by fusing all independent estimates into a common 3D model. The fusion can be performed with low computational complexity through controlled correspondence linking. The multi-view version of stereo originated as a natural improvement to the two-view case. Instead of capturing two photographs from two different viewpoints, multi-view stereo would capture more viewpoints in-between to increase robustness, e.g. to image noise or surface texture [25-26]. All the MVS algorithms assume the same input: a set of images and their corresponding camera parameters. According to Seitz *et al.* [16], stereo methods can be local or global. Local (or window-based) methods compute the disparity at a given point using the intensity values within a finite region [27]. Most of the proposed matching methods are based on similarity or photo-consistency measures; in other words, they compare pixel values between the images. These measures can be defined in either image or object space, according to the algorithms (stereo or multi-view). The most common measures (or matching costs) include squared and absolute intensity differences, normalised cross-correlation (NCC) [28], dense feature descriptors [29], census transform [30], mutual information [31], gradient-based algorithms [32] and bidirectional reflectance distribution functions (BRDFs) [33] and are applied to the intensity images.

Summarising, we could state that almost all these reconstruction methods are conceptually de-signed to work on greyscale images meaning that, sooner or later in the processing, for a given spatial location, the algorithm will only consider a single intensity value instead of the RGB triple. In practice, this means that the algorithm can be applied onto each colour channel separately, or that a standard conversion (conventionally the computation of the luma Y' 601, luma Y' 709 or the luminance CIE Y component) provides the necessary single-band input. That is also our case where algorithms both for sparse and dense phase work using only the luminance channel.

Currently the existence of an optimal algorithm to convert a colour image into its greyscale variant is poorly documented for the photogrammetric pipeline and the conversion is commonly assumed to have little, if any, impact on the final results. However, since many methods for converting to greyscale have been employed in the different solutions for SfM and MVS, we believe it is prudent to assess whether this assumption is warranted. Related work has shown that illumination conditions and camera parameters can greatly influence the properties of several recent image detector/descriptor types [20-21]. And colour-to-grey conversion is factually a dimensionality reduction problem. This process should not be undervalued, since there are many different properties that need to be preserved. Mainly, in our case, isoluminant colour changes are usually not preserved with commonly used colour-to-grey conversions. In any case, the 3D to 1D dimension reduction leads to information loss (display has only Y-range for contrast condensing) and that the appearance of this loss is related to the method. A major cause of this loss is that display has less than {0,100} Y-range whereas a colour image has over 200 colour differences.

Furthermore, many conversion methods have been proposed in recent years; however, these methods mainly focus on perceptual accuracy in terms of the fidelity of the converted image when reproduced from colour-to-greyscale tones. These kinds of approaches are not designed to fulfil the needs of visual stereo matching and image matching algorithms, where local contrast preservation is crucial in the process of matching by local operators. This is well demonstrated by the SIFT operator, where the candidate key-points with low contrast are rejected in order to decrease the number of ambiguous points in the matching process [8]. Related work has shown that illumination conditions and camera parameters can greatly influence the properties of several recent image descriptor types [34]. This suggests greyscale algorithms that are less sensitive to illumination conditions may exhibit superior performance when illumination is variable. Moreover [35], evaluating the impact of conversion algorithms on tracking and homography calculation results based on SIFT, in the context of augmented reality applications, allowed the authors to conclude that different grayscale conversion techniques can cause significant discrepancies in the overall performance.

Secondly, an enhancement of the images forming the convergent photogrammetric network through pre-processing could be an important step for subsequent feature extraction and image matching and finally for dense stereo matching. Research has found that a filter is guaranteed to provide detail in shadowed and saturated areas simultaneously, and thus to allow a greater number of interest points to be detected [36].

In this paper we evaluated many different state-of-the-art algorithms for colour-to-grey conversion with the goal to understand what could improve the quality and the accuracy of results applying a grey scale conversion as a pre-processing step in the context of image matching and multi-view stereo matching.

In general, three approaches are used to evaluate the correctness of different colour-to-greyscale conversion algorithms:

- Perceptual evaluation, such as the one employed by Ma *et al.* [37], is best suited for greyscale printer reproduction and other human-related tasks.
- Objective Quality Assessment using objective quality model like the structural similarity (C2G-SSIM) index, which evaluates the luminance, contrast, and structure similarities between the reference colour image and the converted image. The three components are then combined depending on image type to yield an overall quality measure [38].
- Approaches tailored to measure the results of the subsequent image processing algorithms [39].

We use the third approach, by evaluating the effectiveness of different greyscale conversions with respect to the image-based reconstruction problem in three contexts: image matching, image orientation, and semi-global matching. We compare the output of a complete, robust and reliable solution (a calibrated version SIFT as detector/descriptor [8], VisualSfM as SfM solution [40], and nframes SURE [41] for dense stereo matching) for two different collections of AH images using five colour-to-grey conversion techniques.

Starting from these trials we adapted the most promising algorithm in order to create an ad-hoc algorithm able to optimise the conversion process by simultaneously evaluating the whole set of images. The developed solution, the so-called IHE (Intensity Histogram Equalisation), is a colour-to-grey conversion technique that combines the idea developed in the Multi-Image Decolorise (MID) [39] of evaluating a whole set of images instead of a single one in order to preserve tonal coherence during the matching phase, with a new specifically developed measurement criterion used to evaluate the decolourisation quality. Our frameworks specify the statistical properties of the input data with the help of a representative collection of image patches provided by the same images that were converted. Differently from MID, which is an adaptation of the algorithm described by Grundland and Dodgson [42], our conversion is a generalisation of the MATLAB rgb2gray algorithm, and simultaneously takes as input the whole set of images to be matched. IHE bears some similarity to the algorithm previously introduced by Song *et al.* [43]. However, a significant critical difference lies in the measurement criterion used to evaluate the decolourisation quality. In brief, Song *et al.* [43] employs the bilateral filtering with high computational complexity; on the contrary, IHE propose a different dominant colour hypothesis and aims to maximise the tonal range on the image set. IHE is part of an efficient image pre-processing workflow [44] aiming to increase the quality of the results in two central steps of the photogrammetric pipeline:

- The automated alignment of image datasets by:
  a. increasing the number of correct correspondences, particularly in textureless areas;
  b. tracking features along the largest number of possible images to increase the reliability of the extracted correspondences;
  c. correctly orienting the largest number of images within a certain dataset;
  d. delivering sub-pixel accuracy bundle adjustment results.
- The generation of denser and less noisy 3D point clouds.

The developed methodology for image pre-processing and enhancement consists of colour balancing, image denoising, RGB to grey conversion and image content enhancement, starting from RAW images, i.e. images as close as possible to the direct camera output.

We move then in the well-delineated context in which Stamatopoulos, Fraser and Cronk [45] demonstrated that a pre-processing approach for RAW imagery (i.e. images containing unprocessed pixels saved as it was captured by the sensor since it contains the values just after the analog-to-digital conversion, without any of the camera processing enhancements applied) can yield significant photogrammetric accuracy improvements over those obtained with JPEG. In our consideration, only the basic in-camera processing was retained: black point subtraction; bad pixel removal; dark frame, bias subtraction & flat-field correction; green channel equilibrium correction; Bayer interpolation. To avoid as much as possible uncontrolled modification of the RAW pixel values we did not allow on-camera: denoising; colour scaling; image sharpening; colour space conversion; Gamma correction; format conversion.

Three background papers related to our study are focused on colour-to-grey conversion to improve image matching using SIFT and SIFT-like operators [46-48]. Another related paper aims to develop a new method to improve stereo-matching and multi-stereo matching results [39]. Our development confirmed the main results illustrated in the above papers (e.g. efficiency of colour-to-grey methods for SIFT feature-matching) and shows the efficiency of the IHE technique to improve multi-view stereo-matching.

Colour-to-grey conversion problem and existing techniques

Over the past decades, several simple and more complex colour-to-greys algorithms have been developed to derive the best possible decolourised version of a colour input image. In essence, the decolourisation problem is one of dimension reduction since multiple input channels have to be mapped onto a single channel. In doing so, the colour contrast and image structure contained by the colour image should be preserved in the greyscale image. These conversions can be done in Colour Space (linear or nonlinear) or in Image Space converting pixels (RGB) using colours in the image and assigning different grey for different colour.

Among these conversions in Colour Space the CIE Y method is a widely used conversion that is based on the CIE 1931 XYZ colour space. It takes the XYZ representation of the image colour and uses Y as the grey value. For images having isoluminant regions, the luminance channel will fail to represent structures or features in the colour image, since the linear combination using fixed weights can produce the same result for some different groups of R, G and B values.

Image Space colour-to-grey conversions (also called functional) are image-independent local functions of every colour, e.g., for every pixel of the colour image a grey value is computed using a function whose only parameters are the values of the corresponding colour pixel.

Following Benedetti's method [39] Image Space conversions can be subdivided into three groups:

- trivial methods
- direct methods
- chrominance direct methods.

Trivial methods are the most basic and simple ones, as they do not take into account the spectral power distribution (SPD) of the colour channels. Precisely, only the mean value of the RGB channels is considered. They lose a lot of image information because for every pixel they discard two of the three colour values, or discard one value averaging the remaining ones, not taking into account any colour properties. Despite the loss of information these colour-to-greyscale conversions are commonly used for their simplicity.

Typical example of trivial method is the *RGB Channel Filter* that selects a channel among R, G or B and uses this channel as the greyscale value. The green filter gives the best results (this method is afterwards called *green2gray*), due to the sensors Bayer pattern configurations and because the green channel is typically very similar to the luminance channel, and the blue filter gives the worst results in terms of lightness resemblance. For images having many colours in the green field we have a lot of information missed.

The *Value HSV* method takes the *HSV* representation of the image and uses Value *V* as the greyscale value. This is equivalent to choosing for every pixel the maximum colour value and using it as the greyscale value. This method loses the information relative to which colour value is kept for a pixel. Another problem is that the resulting image luminance is heavily biased toward white.

Direct methods are standard methods where the conversion is a linear function of the pixel's colour values. Typically, this class of functions takes into account the spectrum of different colours. The simplest solution is obviously the *Naive Mean* that takes the mean of the colour channels or non-linear, gamma-corrected R', G', and B' components to form a greyscale image; but the true advantage of direct methods compared to the trivial ones is that, because they take information from every channel, it's possible to have different weights for different colours means. This allows taking into account factors such as the relative spectral distribution of the colour channels and the human perception. Many of the most used greyscale conversion are based on a method of this family, aiming with a simple weighted sum to represent luminance such as the CIE Y channel or the luma channels Y' 601 and Y' 709, the latter two differing by the weighting coefficients used since they are based on the Rec. 601 NTSC primaries and the Rec. 709 sRGB primaries, respectively.

The most popular of direct methods is the MATLAB *rgb2gray* that converts from RGB to greyscale, using the NTSC CCIR 601 luma weights, with the formula.

$$Y = 0.2989R + 0.5870G + 0.1140B \tag{1}$$

Other weights can be used, according to the users' applications and software (e.g., Adobe Photoshop uses these specific weights for the channels R, G, and B: 0.4, 0.4, 0.2). RGB channel filters are not affected at all by gamma compression problems, since they do not manipulate colour values but only choose one of them; this is one of reason of its attractiveness also in our field.

Chrominance direct methods are based on more advanced colour spaces and are able to mitigate the problem related to isoluminant colours. These conversions are still local functions of the image pixels, but they assign different greyscale values to 'isoluminant' colours. To achieve this result, the luminance information is slightly altered using the chrominance information. In order to increase or decrease the 'correct' luminance to differentiate isoluminant colours, these methods exploit a result from studies on human colour perception as the Helmholtz-Kohlrausch (H-K) effect [49]. The H-K effect states that the perceived lightness of a stimulus changes according to a function of the chroma. This phenomenon is predicted by a chromatic lightness term that corrects the luminance based on the colour's chromatic component and on starting colour space.

Chrominance direct methods can be performed either locally or globally.

Local chrominance direct methods [50] make pixels in the colour image not processed in the same way and usually rely on the local chrominance edges for enhancement. Local mapping methods generate output greyscale pixel values, which are dependent on the local distribution of colour pixel values. For example, Smith *et al.* [51] employ a local sharpening step after obtaining the greyscale image by global mapping: an adaptively weighted multiscale unsharp masking enhances chrominance edges.

In the field of automatic photogrammetry, the use of local methods presents many problems: local changes and contradictions could appear, and they have high computational costs. Mainly these techniques might distort the appearance of constant colour regions (e.g. the same colour may output a different grey value) and, using colour contrast map to enhance grey image, may produce haloing artefacts, as discussed by Kim *et al.* [52]. These artefacts constitute a major problem in our case because

SIFT blob features will be differently altered in various images presenting different point of view and exposure, preventing the correct matching.

Global methods attempt to produce one mapping function for the whole image. In the global mapping method, an input colour pixel value, independent of its location, is mapped to an output greyscale pixel value. In this way, the method grants the same luminance for the same RGB triplets and high-speed conversion. Mostly, colour order (i.e. a scale organisation of the colour perception) is strictly satisfied, also it might be ambiguous for the human perception. For example, Grundland and Dodgson [42] proposed a fast linear mapping algorithm that adds a fixed amount of chrominance to the lightness, where the original lightness and colour order can be better preserved by restraining the added chrominance. Benedetti *et al.* [39] achieved the best results by global techniques, but our studies show that this approach fails for RGB triplets having the same chrominance but different luminance (e.g. in the case of RGB signals taken with different exposure times). In general, global techniques typically involve complex optimisation steps, which make the resulting decolourisation methods orders of magnitude slower than trivial and direct methods.

Finally, some notations about gamma correction that is a nonlinear operation used to control signal amplitude. In digital image processing, this operation is used to control brightness and reduce gradient variation in video or still image cameras. The main problem is that we have not information about the image's gamma. Moreover, many applications/algorithms ignore this issue. With regard to the trivial methods, Value HSV and RGB channel filters are not affected at all by the gamma, since they do not manipulate colour values but only choose one of them. Also direct methods techniques are robust from this point of view, although applying these conversions to gamma pre-compensated values is not theoretically correct. It is difficult to predict the impact of this issue for chrominance techniques, although from practical experience, Benedetti *et al.* [39] states that techniques basically weighting colour values with the spatial driven perturbations that enhance them, seem to be the most robust. We underline that are both very sensible to this issue, since they use saturation and the proportions between the image chromaticities to choose the mapping of a colour hue to increases or decreases in the basic lightness. If the values are not linear, these ratios change significantly and the resulting mapping is very different.

Colour-to-grey conversion methods compared with the *IHE*

We investigated different colour-to-grey methods to explore a wide range of approaches.

A preliminary evaluation of the aforementioned methods (Figure 2), was done applying a pixel-by-pixel Lp difference method with an offset of 127 levels of brightness to better identify visually the differences, to different conversion of a set of five images representing most common issues. This technique is the most appropriate to evaluate a method's efficiency for machine readable process. The simple image subtraction can rapidly provide visual results rather than using CIELAB ΔE*ab or other perceptually-based image comparison methods.

Concerning Image Space conversions, as trivial method we chose the *green2gray*, where the green channel is extracted from a RGB image and used to create the final greyscale image.

As direct method we evaluated the MATLAB *rgb2gray* a technique based on the above mentioned weighted sum of the three separate channels because of its relationship with human vision and its popularity in the Computer Vision community.

Among direct methods we tested also Adobe Photoshop conversion using predefined settings. Adobe Photoshop devised also custom non-linear projections, but these require users to set image-dependent parameters by trial and error [53]. For this reason, were left out of our tests.

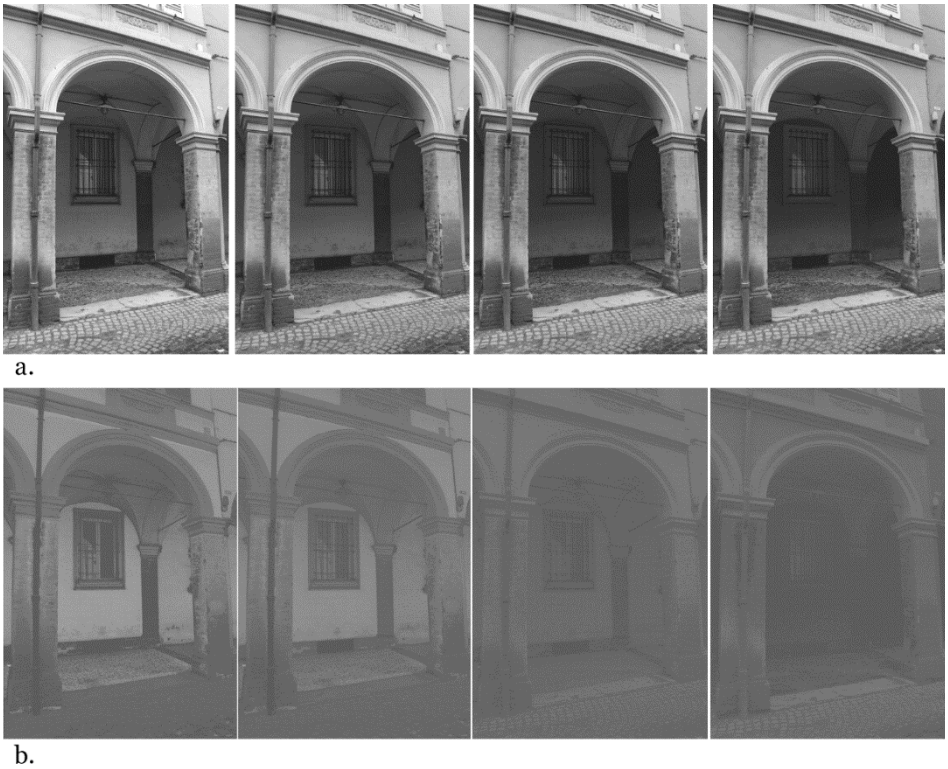

Figure 2: (a) from left to right: grey level conversions of a colour image obtained with Adobe PhotoShop, Decolourise, Realtime, MATLAB rgb2gray, IHE; (b) image differences between the results obtained by using Adobe Photoshop and the other techniques listed above.

Among Chrominance direct methods, we discarded all methods tested accurately without success [39] and also the methods discarded by this study with significant motivations. Precisely Gooch *et al.* [54] was not implemented because Ĉadík [37] demonstrates that, although its gradient-preserving nature could improve features discriminability, in practice it does not improve the quality of the results because of its inherent problems with the input parameter selection and its inconsistent spatial locality. We implemented the local methods [51] technique but we abandoned them soon due to the problems in MVS software where its adaptively-weighted multi-scale unsharp mask generates large problems (no models produced in the ground dataset): it's well known that the unsharp masking filter enhances the fine details of the image and colours are mapped inconsistently among different parts of the images depending on the surrounding neighbourhoods. While the algorithm can use spatial information to determine the mapping, the same colour should be mapped to the same greyscale value for every pixel in the image. Finally, we implemented Grundland and Dodgson's method [42] and two techniques of the same authors: *Contrast Preserving Decolorisation* [55] and *Real-time Contrast Preserving Decolorisation* [56]. Even if they are both based on the same theoretical framework and algorithm, they differ only for a simplified version of the algorithm. Preliminary tests with both the methods prove the efficiency of the second one. A detailed description of the chrominance direct methods considered here follows.

Decolorise

The technique, described by Grundland and Dodgson [42], performs a global greyscale conversion by expressing greyscale as a continuous, image-dependent, piecewise linear mapping of the primary RGB colours and their saturation. Authors designed a linear combination method, which aims at maximally preserving the original colour contrast by using a strict order constraint for colour mapping based on the human vision system. The algorithm, called *Decolorise*, works in the YPQ colour opponent space and aim to contrast enhance. The Grundland *Decolorise* algorithm has five steps:

1. The colour image is converted into a colour opponent colour space;
2. The colour differences are measured using a Gaussian sampling;
3. The chrominance projection axis is found by predominant component analysis;
4. The luminance and chrominance information are merged;
5. The dynamic range is adjusted using the saturation information to balance between the original range and the desired amount of enhancement.

The process is controlled by three parameters: the degree of image enhancement, the typical size of relevant image features in pixels, and the proportion of image pixels assumed to be outliers. The luminance channel Y is obtained with the NTSC CCIR 601 luma weights. A main step concerns the use of predominant component analysis. Unlike principal component analysis, which optimises the variability of observations, predominant component analysis optimises the differences among observations. The predominant chromatic axis aims to capture, with a single chromatic coordinate, the colour contrast information that is lost in the luminance channel.

Decolorise is very sensitive to the issue of gamma compression with some risks of decrease of the quality of the results mainly in light areas or dark areas where many features will be lost because the saturation balancing interacts incorrectly with the outlier detection.

Real-time contrast preserving decolorisation

Starting from the observation that the human visual system does not univocally perceive chrominance and lightness, while their relationship to the adjacent context play a crucial role, and that the order of different colours also cannot be uniquely defined by people, Lu, Xu and Jia [55] relaxes the colour order constraint and presents a new method seeking better preservation of colour contrast and significant enhancement of visual distinctiveness for edges. For colour pairs without a clear order in brightness, authors propose a bimodal distribution, i.e. a mixture of two Gaussians, to automatically find suitable orders with respect to the visual context in optimisation. This strategy enables automatically finding suitable grey scales and preserves significant colour change. Practically authors, use a global mapping scheme where all colour pixels in the input are converted to greyscale using the same mapping function (a finite multivariate polynomial function), to automatically find suitable brightness orders with respect to the visual context. Therefore, two pixels with the same colour will have the same grey scale. Mathematically, they define the polynomial space of each input RGB vector $c = (r, g, b)$ with its degree n as:

$$\prod_n = span\, \{r^{d_1} g^{d_2} b^{d_3} \colon d_i = 0, 1, 2, \dots, d_1 + d_2 + d_3 \leq n\} \qquad (2)$$

where Π_n is a polynomial space spanned by a family of monomials. The mapping function is thus expressed as:

$$f(r, g, b; \omega) = \sum_i \omega_i\, m_i \qquad (3)$$

where m_i is the ith monomial basis of Π_n. The mapping function is uniquely determined by weights $\{\omega_i\}$. Empirically, we use degree n=2, which means that the total number of $\{\omega\}$ is 9 and the mapping function is a linear combination of elements in $\{r, g, b, rg, rb, gb, r^2, g^2, b^2\}$. Authors note that the polynomial form is actually a generalisation of common linear and nonlinear colour-to-grey mapping functions.

The technique is today implemented in OpenCV 3.0 [57]. In order to achieve real-time performance, authors further devise a discrete searching optimisation which takes advantage of a linear parametric greyscale model as well as a sampling based *P*-shrinking process [56]. Specifically, they approximate their previous optimisation-based method and achieve real-time performance by confining the polynomial colour model into a constrained, discrete linear colour model. To further speedup the decolourisation process, they down sample the high-resolution input to a 64 × 64 small scale. This is valid due to the inherent colour redundancy of natural images. Experiments show this approximated solution can achieve real-time performance for high-resolution images, without obvious quality degradation. Unsatisfactory results could be produced in special cases as demonstrated e.g. also by Song *et al.* [43]. However, these fails are outside of our case. Moreover, the new algorithm is not based on the human perception but on a simple numerical order in brightness, enforcing the appropriateness to our field of application. Main drawback is that in two different images same colour could be converted in different brightness values.

The new *IHE* technique

Based on the results achieved with the aforementioned methods, a new decolourisation technique, named Intensity Histogram Equalisation (*IHE*), was developed. The aim of *IHE* is to preserve the consistency between different images considering the following requirements:

- *Feature Discriminability*: the method should preserve the image features discriminability to be matched as much as possible;
- *Chrominance Awareness*: the method should distinguish among isoluminant colours;
- *Global Mapping*: while the algorithm can use spatial information to determine the mapping, the same colour should be mapped to the same greyscale value for every pixel in the image;
- *Colour Consistency*: besides Global Mapping, the same colour should also be mapped to the same greyscale value in every image of the set to be matched;
- *Greyscale Preservation*: if a pixel in the colour image is already achromatic it should maintain the same grey level in the greyscale image;
- *Unsupervised algorithm*: it should not need user tuning to work properly, in particular for large datasets.

Basically four considerations are at the heart of our development:

- combining the channels of a multi-band image with the help of a pixel-wise weighted sum usually results in weights that are given by some standard values or chosen heuristically. This does not take into account neither the statistical nature of the image source nor the intended further processing of the scalar image;
- an image does not often contain the full tonal range, but a subset of tones relatively small, and in the original version and the converted version;
- almost all the previous methods pay more attention to maximise local difference for decolourisation but ignore global colour distribution, and quantitative evaluation used in these methods only consider neighbouring pixel/region pairs;

- the image matching and multi-stereo existing colour-to-grey conversion algorithms work considering the images individually. this way, it is possible that the same colour is converted into different grey values, in different images.

The new method takes the idea of the *Multi-Image Decolourise* method of evaluation of the whole set of images in order to match them simultaneously [39]; but it uses a different algorithm for the colour-to-grey conversion. While *Multi-Image Decolourise* is an adaptation of Grundland and Dodgson's method [42], our conversion is a generalisation of the MATLAB *rgb2gray* algorithm. MATLAB algorithm takes each channel value and multiply it for a constant value. Instead of using fixed channels multipliers, *IHE* method adapts them accordingly to the image dataset, so that the *IHE* colour-to-grey equation can be generalised as:

$$IHE = \omega r\mathrm{R} + \omega g\mathrm{G} + \omega b\mathrm{B} \qquad (4)$$

where R, G, B are the red, green and blue colour components and *ωr, ωg* and *ωb* are the channel multipliers, whose value varies in each dataset.

IHE method consists of three steps, in which:

- it merges the images of the whole dataset;
- it adapts channel multiplier values in order to maximise the tonal representation;
- it converts each single RGB image accordingly to the found multipliers.

In particular, the whole set of images is first bound together placing the images side by side as in a *patchwork*. Although the procedure could be performed on each single image, in this context this first step is necessary because intensities values have to be preserved across multiple images. Similarly to [43], *IHE* employs the standard colour-to-grey conversion model with 66 sets of weights (*ωr, ωg, ωb*). However, a significant critical difference lies in the measurement criterion used to evaluate the decolourisation quality. In brief, Song *et al.* [43] employs the bilateral filtering with high computational complexity; on the contrary, *IHE* is characterised by a simple and straightforward algorithm. *IHE* has no claim to realistically convert images from colour-to-greyscale. It rather aims to preserve as much as possible the amount of information conveyed in the rgb-to-grey conversion process.

IHE has its foundation in the statistics of extreme-value distributions of the considered images and presents a more flexible strategy, adapting dynamically channel weights depending on specific input images, in order to find the most appropriate weights for a given colour image and aiming to preserve as much as possible the amount of conveyed information. In *IHE,* luminance is computed by weighting in different ways the red, green and blue components, until the distribution of the converted image is close as much as possible to a uniform distribution. In particular, the algorithm compute 5051 permutations of RGB weighting triplets whose sum is 1 ([0, 0, 1], [0, .01, 0.99], … [0.99, 0.01,0], [1, 0, 0]). Each triplet is used as weight in the *IHE* equation so that, for each cycle:

$$\begin{aligned}
&IHE1 = 0\mathrm{R} + 0\mathrm{G} + 1\mathrm{B}\\
&IHE2 = 0\mathrm{R} + 0.01\mathrm{G} + 0.99\mathrm{B}\\
&\dots\\
&IHE5050 = 0.99\mathrm{R} + 0.01\mathrm{G} + 0.01\mathrm{B}\\
&IHE5051 = 1\mathrm{R} + 0\mathrm{G} + 0\mathrm{B}
\end{aligned}$$

For each cycle, an intensity histogram is calculated and the intensity distribution is evaluated by calculating the fitting goodness of the distribution with respect to a rectangular one.

To calculate the best rectangular fitting, we assumed a 0 slope regression line. The general equation of the regression line is:

$$\beta = \bar{y} - m\bar{x} \tag{5}$$

where β is equivalent to the average of the histogram points.

After calculating the average, the minimum error within all the calculated combinations of channel mixing is sought. The error is calculated as least squares error:

$$S = \sum_{i=1}^{n}(y_i - \beta)^2 \tag{6}$$

where y_i are the actual points, while β is the best linear fitting of the histogram. Then, *IHE* chooses the mixing that maximises the number of tones obtained in the converted image. Finally, similarly to Song *et al.* [43], *IHE* uses a measurement criterion to evaluate the decolourisation quality, i.e. the newly defined dominant colour hypothesis.

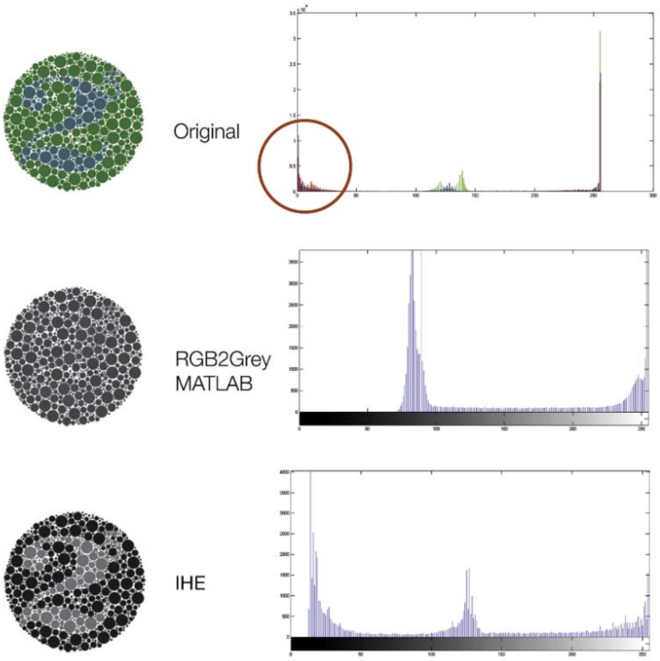

Figure 3: Greyscale conversion of an image characterised by isoluminant colours. Image and its histogram (top to bottom): original; converted using the MATLAB rgb2gray; converted using the IHE for a single image.

We try to clarify *IHE* behaviour with the example in Figure 3. The conversion of this image using trivial or direct methods leads to an image similar to the one in the centre of the same figure, where blue and green are isoluminant and are converted using the same shade of grey. The resulting conversion totally lost the meaningful original information. Representing the distribution of t.

One of the original image and comparing it with the distribution of the converted image, we would notice that one of the three peaks of the distribution of the original image is not present in the converted version.

Main disadvantage of IHE is the high computational pre-processing cost that we alleviate using sampled copies of our dataset at 25%. As said this is valid due to the inherent colour redundancy of natural images. In this way the pre-processing time is just some minutes, in any case a little time compared to those of the entire pipeline.

Figures 4 and 5 report examples of *IHE* results with respect to MATLAB *rgb2gray* method. The main disadvantage of the developed method is the high computational pre-processing time due to the sampled patches on each image of the dataset.

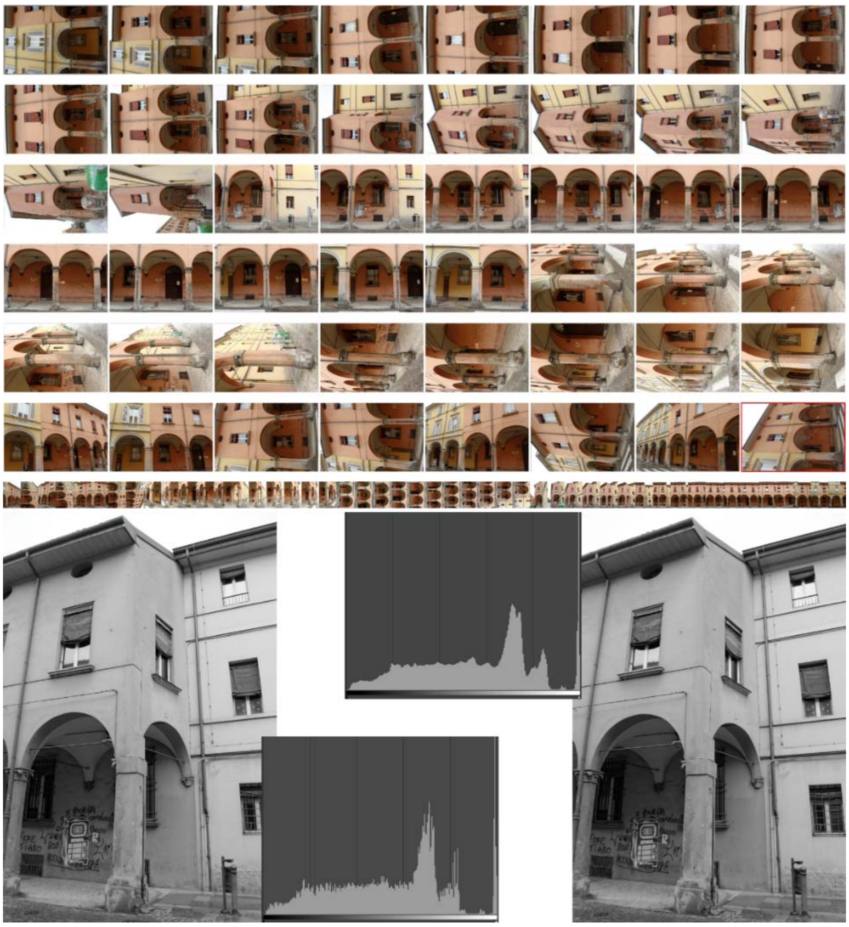

Figure 4: IHE technique processing and results: top original dataset; middle: the mosaic of all the images; bottom left: image converted using MATLAB rgb2gray; bottom right: image converted via IHE.

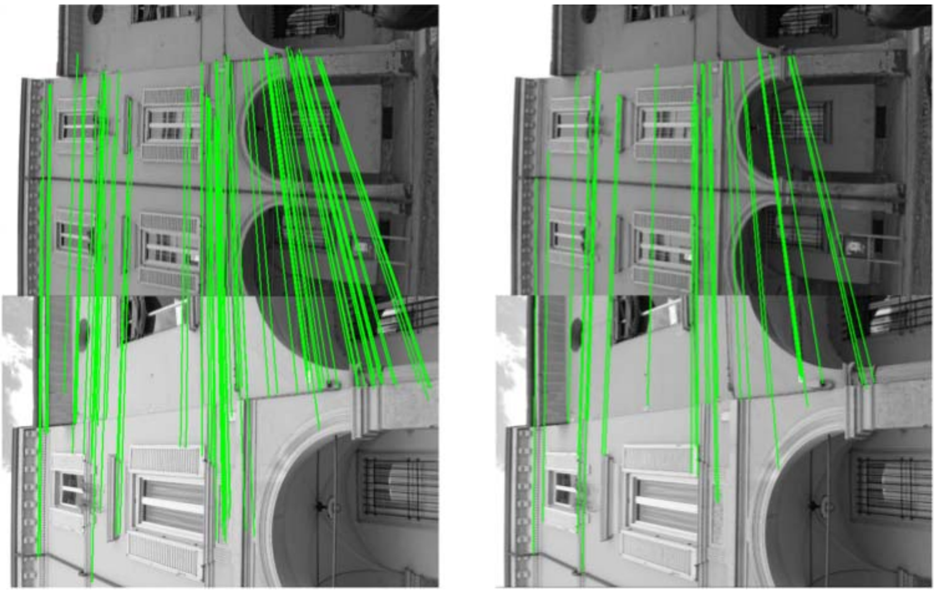

Figure 5: Two images converted into grey tones with IHE approach (on the left) and MATLAB rgb2gray (on the right). Matches appear to be quantitatively larger and qualitatively more effective, as they are evenly distributed across the whole image.

Experimental setup and evaluation results

To fit our needs, we used, conversely, an approach that is tailored to measure the results of the subsequent image processing algorithms, by evaluating the effectiveness of different greyscale conversions with respect to the image-based reconstruction problem in three contexts: images matching, feature tracking and camera poses, and semi-global matching.

As detector/descriptor we used a calibrated version of VLFeat implementation [58-59] of the very popular SIFT [8]. SIFT derives a large set of compact descriptors starting from a multi-scale representation of the image (i.e. a stack of images with increasing blur simulating the family of all possible zooms). In this multi-scale framework, the Gaussian kernel acts as an approximation of the optical blur introduced by a camera. The detection and location of keypoints is done by extracting the 3D extrema with a DoG operator. Since there are many phenomena that can lead to the detection of unstable keypoints, SIFT incorporates a cascade of tests to discard the less reliable points. Only those that are precisely located and sufficiently contrasted are retained. Calibration of parameters that control the detection of points is fundamental for the efficiency of the detection mechanism.

As SfM solution we used the Changchang Wu VisualSfM [37], a GUI application for 3D reconstruction using SfM. The reconstruction system integrates: SIFT on GPU (SiftGPU), Multicore Bundle Adjustment, and Incremental SfM. This package can reconstruct large scenes from multi-views and users can set parameters and receive feedback (as graphs and reports) at each stage; the package exploiting multicore parallelism has the following functionalities: a) feature detection, b) feature matching, c) sparse 3D reconstruction, d) dense 3D reconstruction, e) coordinate transformation, f) mesh generation, and g) texture mapping. We exploited mainly the bundle adjustment phase [9]. For dense stereo matching we used nframes SURE [41] (photogrammetric SUrface REconstruction from imagery) a MVS technique [60-61] where a reference image is matched to a set of adjacent images using a semi-global matching (SGM) type of stereo algorithm [62]. For each pair, a disparity map is computed; then all disparity maps sharing the same reference view are merged into a unique final point cloud capitalising on the redundancy across the stereopairs. Within a premodule, a network analysis and selection of suitable image pairs for the reconstruction process is performed. Epipolar images are then generated and a time- and memory-efficient SGM algorithm is applied to produce depth maps. All these maps are then converted into 3D coordinates using a fusion method based on geometric constraints that both help in reducing the number of outliers and increase precision (Figure 6 and 7).

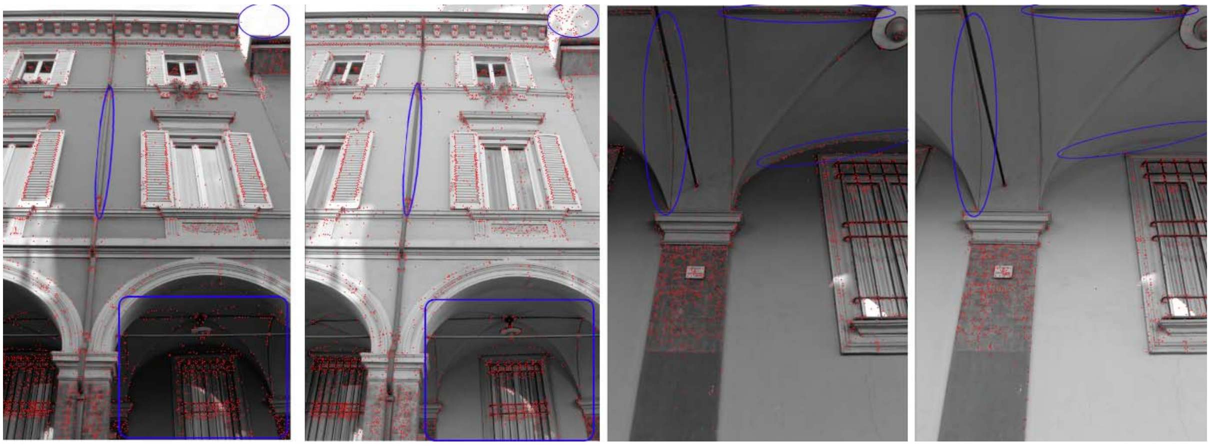

Figure 6: Points are more uniformly distributed across multiple images using IHE approach (left) column than using MATLAB rgb2gray conversion (right). Important feature points are detected only using IHE.

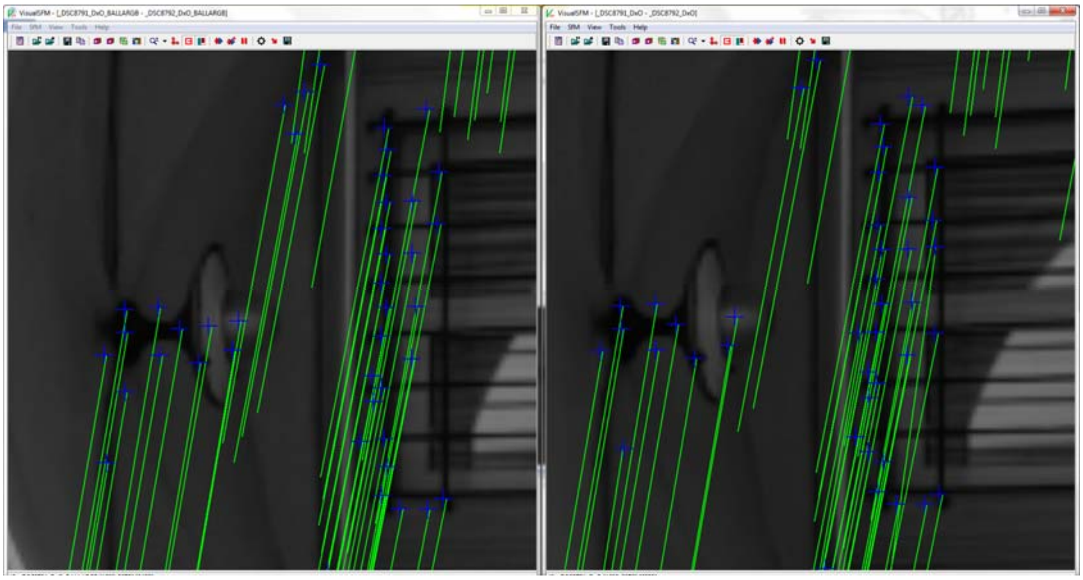

Figure 7: The number of correct SIFT matches is significantly larger (left) and more robust in different lightning conditions using IHE approach than using MATLAB rgb2gray conversion (right).

Our experiments are conducted with relatively few instances, since classifier performance is much more sensitive to the quality of the descriptors in this setting. This has been obtained through previous calibration works of descriptors and SFM system, which allowed extremely limited fluctuations in results. Following outcomes were analysed:

- *pairwise matching efficiency*: using a set of images (Figure 8) featuring illumination differences, textureless surfaces, possible loss of information in the colour-to-grey conversion and elements with strong 3D features, we tested pairwise matching efficiency of the operators with respect to three camera movements: (i) parallel with limited baseline (00-01); (ii) rotation of 90° (00-03); (iii) tilt of more than 30° (01-02). The number of correct inlier matches is normalised with all putative correspondences (Table 1).

$$efficiency = \frac{\#\, inliers}{\#\, putative\ corrispondances}$$

- *number of oriented cameras*: this parameter allows to understand the completeness of the final reconstruction. A greater number of oriented cameras corresponds to a more complete final reconstruction;
- *root mean square error of the Bundle Adjustment*: it expresses the re-projection error of all computed 3D points;
- *visibility of 3D points in more than 3 images*: this parameter allows to understand the reliability of the reconstruction. The greater the number of cameras from which it is observed a point, the greater is its accuracy;
- number of points in the dense reconstruction using a unique camera orientation for all the datasets: it is a quantitative parameter which allows to understand the final point cloud density.

Obviously, the first metric that comes to mind when varying the input of an SfM chain is the total amount of Interest Points that are extracted in a typical image for the considered subject. This number however, does express neither the strength and reliability of those features nor the camera's exterior and interior orientation upon which they are based.

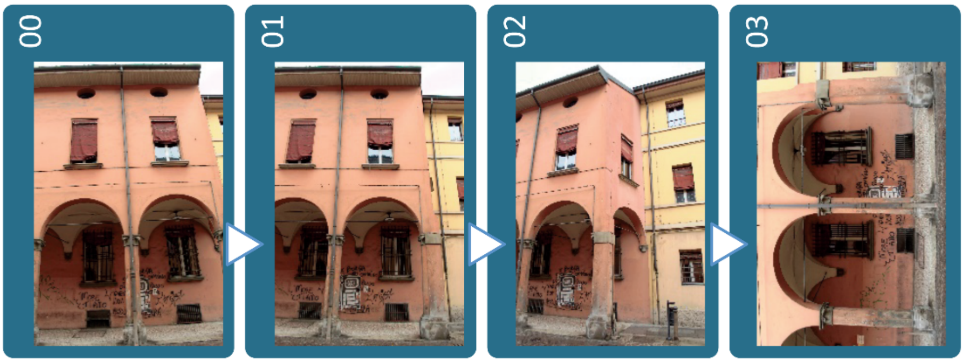

Figure 8: Images used to test pairwise matching efficiency.

	PARALLEL	ROTATE 90°	TILT 45°
	00 - 01	00 - 03	01 - 02
green2grey	0.992	0.786	0.640
rgb2gray	0.980	0.690	0.329
Adobe Ps	0.992	0.821	0.630
Decolorise	0.992	**0.863**	0.626
Realtime	0.992	0.827	0.618
IHE	**0.993**	0.825	**0.676**

Table 1: Efficiency of each operator.

In order to have a common evaluation procedure, once the feature points are extracted and described with the aforementioned algorithm implementations, the descriptors matching procedure, the outlier detection phase and the final bundle adjustment are run inside the same software environment. In particular, correct image correspondences were obtained following [63-64] and then a Random sample consensus (RANSAC) was used to eliminate possible mismatches [65].

The various colour-to-grey techniques were evaluated on two image networks featuring different imaging configurations, textureless areas and repeated pattern/features. The two datasets represent an urban test framework and show a typical historical urban scenario. These data allow to verify the efficiency of different techniques in different situations (scale variation, camera rotation, affine transformations, etc.). The datasets contain convergent images, some orthogonal camera rolls and a variety of situations emblematic of failure cases, i.e., 3D scenes (non-coplanar) with homogeneous regions, distinctive edge boundaries (e.g., buildings, windows, doors, cornices, arcades), repeated patterns (recurrent architectural elements, bricks, etc.), textureless surfaces and illumination changes.

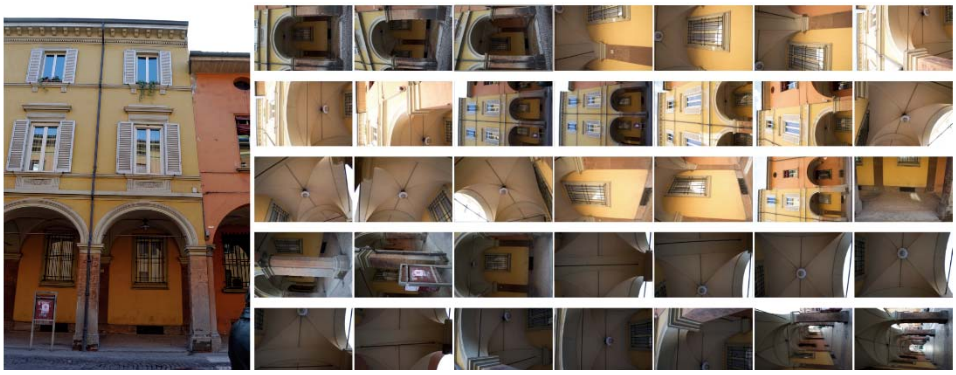

Figure 9: The porticoes dataset.

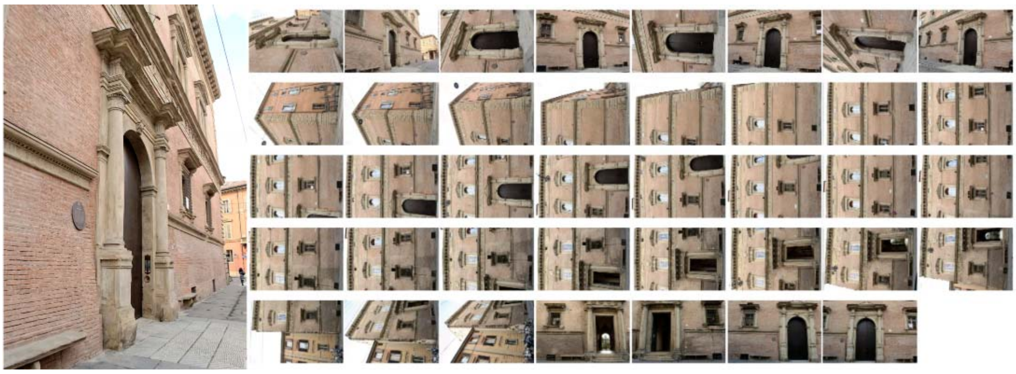

Figure 10: The Palazzo Albergati dataset.

The first dataset (35 images) pertains two spans of a three-story building (6 × 11 m) characterised by arches, pillars/columns, cross vault and plastered wall. Camera was moving along the porticoes, with some closer shots of the columns (Figure 9).

The second dataset (39 images) depict a three-story historical palace (54 × 19 m) characterised by repeated brick walls, stone cornices and a flat façade. The camera was moving along the façade of the building, with some closer shots of the entrances (Figure 10).

	Adobe PS	**Realtime**	**Decolorise**	**MATLAB rgb2grey**	**IHE**	**green2grey**
Number of oriented cameras	33	30	16	33	**35**	**35**
Bundle Adjustment RMS	0.424	0.366	0.379	0.353	0.581	**0.548**
Points from more than 3 cameras	3763	3439	524	3874	**4872**	4759
Points in dense reconstruction	1444269	1522044	1375971	1184432	**1964397**	1703607

Table 2: Dataset Portico 35 - results.

	Adobe PS	**Realtime**	**Decolorise**	**MATLAB rgb2grey**	**IHE**	**green2grey**
Number of oriented cameras	**39**	**39**	**39**	**39**	**39**	38
Bundle Adjustment RMS	0.474	**0.057**	0.048	0.137	**0.057**	0.052
Points from more than 3 cameras	28885	**32704**	32158	27418	29545	37069
Points in dense reconstruction	**27965855**	27553306	27710452	27915526	27944181	27821933

Table 3: Dataset Albergati - results.

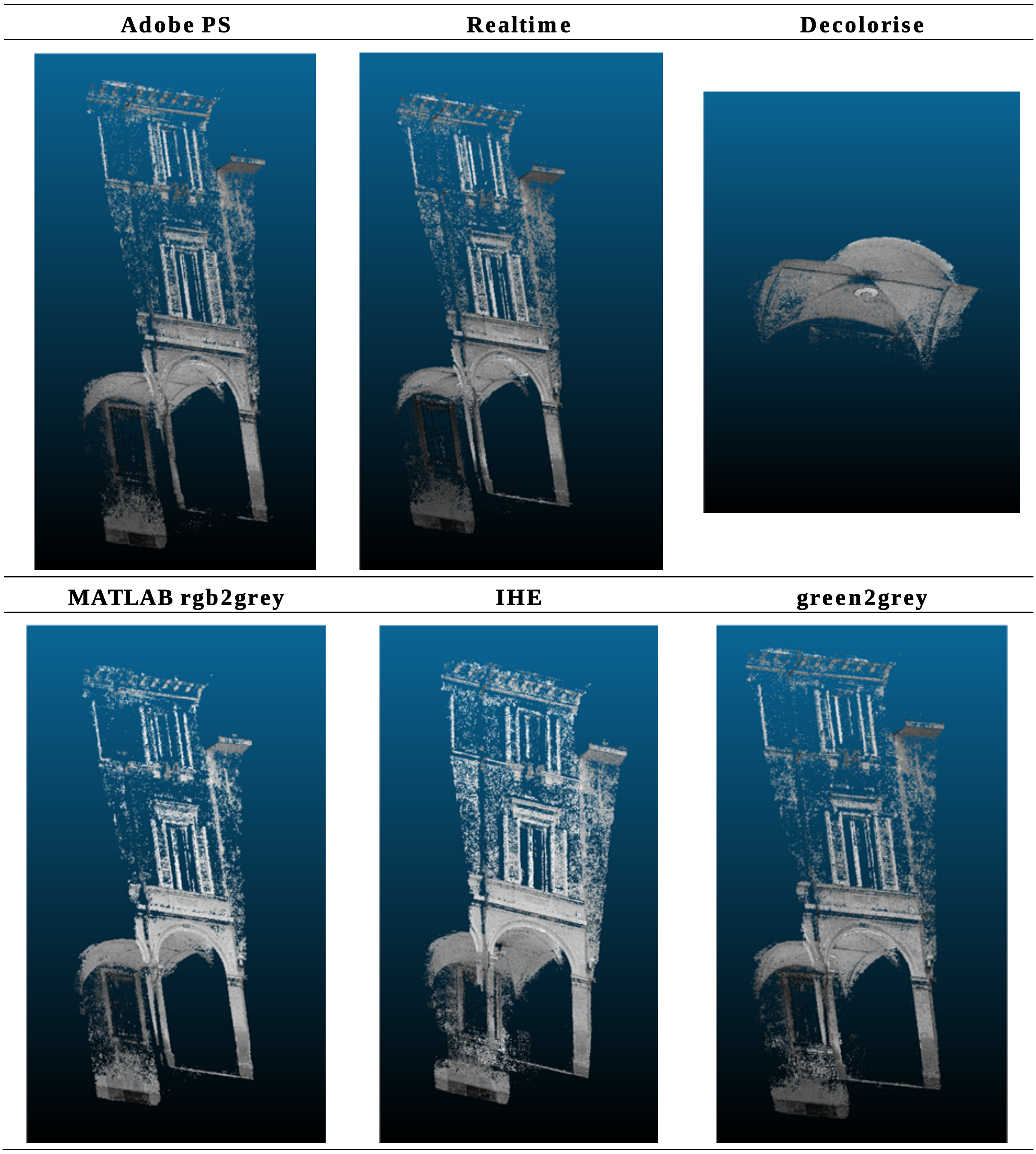

Figure 11: The porticoes dataset: dense stereo matching results, visual comparison.

Images were acquired - in both datasets - using a Nikon D3100 with an 18 mm nominal focal length. Different image scales, number of images, camera network, object texture and size, characterise them.

For every conversion method, and for the *IHE as well*, the datasets are relatively oriented using VisualSFM. Then, a dense point cloud is extracted with a unique tool (SURE) by using (fixing) the same camera parameters for all methods.

Results show how algorithms are differently affecting the Bundle Adjustment procedure as well as the dense matching results (Table 2, Table 3 and Figure 11). It can be generally noticed for the *IHE* a larger number of oriented images, better reprojection errors and denser point clouds, proving its efficiency a pre-processing technique for automatic photogrammetry.

Conclusions

In this paper we investigate if, in the automatic photogrammetric pipeline applied to AH case, the method used to convert from colour-to-greyscale matters, and we can definitively state that it does influence performance. For all datasets a positive observation is that the standard method used performs well, but there is a significant gap among the top performing and worst performing methods.

Furthermore, we observed that neither of the current solutions implemented in the automatic photogrammetry software, nor those potentially existing, are designed for a workflow based on automatic machine reading.

We presented then the IHE, a new decolourisation approach specifically designed for the automatic photogrammetric pipeline. We demonstrated that our solution outperforms existing methods for two representative datasets of AH and in each pipeline phase: namely image matching, camera orientation, dense stereo matching. We achieved this results developing a transformation from colour-to-greyscale that is along the whole dataset in use robust to changes in brightness, though by allowing the gamma value to vary per colour channel.

Also if other image pre-processing techniques of captured images could give great benefits to increase the quality of the photogrammetric pipeline (i.e. contrast enhancement) as proven e.g. in [36], we demonstrated that the advantages of an appropriate colour-to-grey conversion are not trivial, and at limited computational cost and time consumption compared to the typical of image enhancement operations.

References

1. Snavely N, Seitz SM and Szeliski R (2008), Modeling the world from internet photo collections, *International Journal of Computer Vision*, **80** (2), 189-210.
2. Barazzetti L, Scaioni M and Remondino F (2010), Orientation and 3D modelling from markerless terrestrial images: combining accuracy with automation, *The Photogrammetric Record*, **25** (132), 356-381.
3. Heinly J, Schonberger JL, Dunn E and Frahm J (2015), Reconstructing the world* in six days* (as captured by the Yahoo 100 million image dataset), *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3287-3295.
4. Ullman S (1979), The interpretation of structure from motion, *Proceedings of the Royal Society of London B: Biological Sciences*, **203** (1153), 405-426.
5. Pierrot-Deseilligny M, De Luca L and Remondino F (2011), Automated image-based procedures for accurate artifacts 3D modeling and orthoimage generation, *Geoinformatics FCE CTU*, **6**, 291-299.
6. Remondino F, El-Hakim S, Gruen A and Zhang L (2008), Development and performance analysis of image matching for detailed surface reconstruction of heritage objects, *IEEE Signal Processing Magazine*, **25** (4), 55-65.
7. Crandall DJ, Owens A, Snavely N and Huttenlocher DP (2013), SfM with MRFs: Discrete-continuous optimization for large-scale structure from motion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35** (12), 2841-2853.
8. Lowe DG (2004), Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, **60** (2), 91-110.
9. Wu C, Agarwal S, Curless B and Seitz SM (2011), Multicore bundle adjustment, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3057-3064.
10. Furukawa Y and Ponce J (2010), Accurate, dense, and robust multiview stereopsis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32** (8), 1362-1376.
11. Remondino F, Del Pizzo S, Kersten TP and Troisi S (2012), Low-cost and open-source solutions for automated image orientation – A critical overview, *Progress in cultural heritage preservation*, 40-54.
12. http://www.autodesk.com/products/recap-360/overview [last accessed 8 July 2016]
13. http://www.arc3d.be [last accessed 8 July 2016]
14. http://ccwu.me/vsfm [last accessed 8 July 2016]
15. http://www.agisoft.com [last accessed 8 July 2016]

16. Seitz SM, Curless B, Diebel J, Scharstein D and Szeliski R (2006), A comparison and evaluation of multi-view stereo reconstruction algorithms, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 519-528.
17. Fraundorfer F and Bischof H (2004), Evaluation of local detectors on non-planar scenes, *AAPR Proceedings*, 125-132.
18. Remondino F (2006), Detectors and descriptors for photogrammetric applications. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **36** (3), 49-54.
19. Zhao X, Zhou Z and Wu W (2012), Radiance-based color calibration for image-based modeling with multiple cameras, *Science China Information Sciences*, **55** (7), 1509-1519.
20. Morel J and Yu G (2009), ASIFT: A new framework for fully affine invariant image comparison, *SIAM Journal on Imaging Sciences*, **2** (2), 438-469.
21. Apollonio F, Ballabeni A, Gaiani M and Remondino F (2014), Evaluation of feature-based methods for automated network orientation, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **40** (5), 47-54.
22. Maddern W, Stewart A, McManus C, Upcroft B, Churchill W and Newman P (2014), Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles, *Proceedings of the Visual Place Recognition in Changing Environments Workshop, IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China.
23. Strecha C, von Hansen W, Gool LV, Fua P and Thoennessen U (2008), On benchmarking camera calibration and multi-view stereo for high resolution imagery, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.
24. Eisemann M, Frahm J, Remion Y and Ismael M (2015), Reconstruction of Dense Correspondences, in *Digital Representations of the Real World: How to Capture, Model, and Render Visual Reality*, Magnor MA, Grau O, Sorkine-Hornung O and Theobalt C (eds.), 113-133, CRC Press.
25. Tsai RY (1983), Multiframe image point matching and 3-d surface reconstruction, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **5** (2), 159-174.
26. Okutomi M and Kanade T (1993), A multiple-baseline stereo, *IEEE Transactions on Pattern Analysis and Machine* Intelligence, **15** (4), 353-363.
27. Remondino F, Spera MG, Nocerino E, Menna F and Nex F (2014), State of the art in high density image matching, *The Photogrammetric Record*, **29** (146), 144-166.
28. Goesele M, Curless B and Seitz SM (2006), Multi-view stereo revisited, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2402-2409.
29. Tola E, Lepetit V and Fua P (2010), Daisy: an efficient dense descriptor applied to wide-baseline stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32** (5), 815-830.
30. Zabih R and Woodfill J (1994), Non-parametric local transforms for computing visual correspondence, *Computer Vision - ECCV'94*, 151-158.
31. Viola P and Wells III WM (1997), Alignment by maximization of mutual information, *International Journal of Computer Vision*, **24** (2), 137-154.
32. Esteban CH and Schmitt F (2004), Silhouette and stereo fusion for 3D object modelling, *Computer Vision and Image Understanding*, **96** (3), 367-392.
33. Davis JE, Yang R and Wang L (2005), BRDF invariant stereo using light transport constancy, *Tenth IEEE International Conference on Computer Vision*, 436-443.
34. Andreopoulos A and Tsotsos JK (2012), On sensor bias in experimental methods for comparing interest-point, saliency, and recognition algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34** (1), 110-126.
35. Macêdo S, Melo G and Kelner J (2015), A comparative study of grayscale conversion techniques applied to SIFT descriptors, *SBC*, **6** (2), 31.
36. Jazayeri I and Fraser CS (2010), Interest operators for feature-based matching in close range photogrammetry, *The Photogrammetric Record*, **25** (129), 24-41.
37. Ĉadík M (2008), Perceptual evaluation of color-to-grayscale image conversions, *Computer Graphics Forum*, **27** (7), 1745-1754.
38. Ma K, Zhao T, Zeng K and Wang Z (2015), Objective quality assessment for color-to-gray image conversion, *IEEE Transactions on Image Processing*, **24** (12), 4673-4685.
39. Benedetti L, Corsini M, Cignoni P, Callieri M and Scopigno R (2012), Color to gray conversions in the context of stereo matching algorithms, *Machine Vision and Applications*, **23** (2), 327-348.
40. Wu C (2013), Towards linear-time incremental structure from motion, *IEEE 2013 International Conference on 3D Vision-3DV 2013*, 127-134.
41. Wenzel K, Rothermel M, Haala N and Fritsch D (2013), SURE – the ifp software for dense image matching, *Photogrammetric Week*, 59-70.
42. Grundland M and Dodgson NA (2007), Decolorise: Fast, contrast enhancing, color to grayscale conversion, *Pattern Recognition*, **40** (11), 2891-2896.

43. Song Y, Bao L, Xu X and Yang Q (2013), Decolorization: Is rgb2gray () out?, *SIGGRAPH Asia 2013 Technical Briefs*. ACM.

44. Ballabeni A, Apollonio F, Gaiani M and Remondino, F (2015), Advances in image pre-processing to improve automated 3D reconstruction, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **40** (5), 315.

45. Stamatopoulos C, Fraser CS and Cronk S (2012), Accuracy aspects of utilizing raw imagery in photogrammetric measurement, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B5, 387-392.

46. Ancuti CO, Ancuti C and Bekaert P (2010), Decolorizing images for robust matching, *17th IEEE International Conference on Image Processing*, 149-153.

47. Kanan C and Cottrell GW (2012), Color-to-grayscale: does the method matter in image recognition?, *PLOS ONE*, **7** (1), e29740.

48. Verhoeven G, Karel W, Stuhec S, Doneus M, Trinks I and Pfeifer N (2015), Mind your grey tones-examining the influence of decolourization methods on interest point extraction and matching for architectural image-based modelling. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **40** (5), 307-314.

49. Fairchild MD (2013). *Color Appearance Models*. John Wiley & Sons.

50. Bala R and Eschbach R (2004), Spatial color-to-grayscale transform preserving chrominance edge information. *Proceedings of the Twelfth IS&T/SID's Color Imaging Conference*, 82-86.

51. Smith K, Landes P, Thollot J and Myszkowski, K (2008), Apparent greyscale: A simple and fast conversion to perceptually accurate images and video, *Computer Graphics Forum*, **27** (2), 193-200.

52. Kim Y, Jang C, Demouth J and Lee S (2009), Robust color-to-gray via nonlinear global mapping, *ACM Transactions on Graphics (TOG)*, Art. 161.

53. Brown R (2004), *The Photoshop show starring Russell Brown*, Peachpit Press.

54. Gooch AA, Olsen SC, Tumblin J and Gooch B (2005), Color2gray: salience-preserving color removal. *ACM Transactions on Graphics*, 634-639.

55. Lu C, Xu L and Jia J (2012), Contrast preserving decolorization, *IEEE International Conference on Computational Photography*, 1-7.

56. Lu C, Xu L and Jia J (2012). Real-time contrast preserving decolorization, *SIGGRAPH Asia 2012 Technical Briefs*.

57. http://opencv.org/opencv-3-0.html [last accessed 8 July 2016]

58. http://www.vlfeat.org [last accessed 8 July 2016]

59. Vedaldi A and Fulkerson B (2010), VLFeat: An open and portable library of computer vision algorithms, *Proceedings of the 18th ACM international conference on Multimedia*, 1469-1472.

60. Haala N and Rothermel M (2012), Dense multi-stereo matching for high quality digital elevation models, *Photogrammetrie-Fernerkundung-Geoinformation*, **4**, 331-343.

61. Rothermel M, Wenzel K, Fritsch D and Haala N (2012), SURE – photogrammetric surface reconstruction from imagery. *Proceedings LC3D Workshop*, Berlin.

62. Hirschmüller H (2008), Stereo processing by semiglobal matching and mutual information, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **30** (2), 328-341.

63. Agarwal S, Snavely N, Simon I, Seitz SM and Szeliski R (2009), Building Rome in a day. *12th IEEE International Conference on Computer Vision*, 72-29.

64. Frahm J, Fite-Georgel P, Gallup D, Johnson T, Raguram R, Wu C, Jen Y-H, Dunn E, Clipp B, Lazebnik S and Pollefeys M (2010), Building Rome on a cloudless day, *Computer Vision–ECCV 2010*, 368-381.

65. Fischler MA and Bolles RC (1981), Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, **24** (6), 381-395.