

AN AUTOMATIC METHOD TO LEARN AND TRANSFER THE PHOTOMETRIC APPEARANCE OF PARTIALLY OVERLAPPING IMAGES

Marco Zuliani

Mayachitra, Inc.
5266 Hollister Ave
93110 Santa Barbara, CA
zuliani@mayachitra.com

Luca Bertelli, B. S. Manjunath

Electrical and Computer Engineering Department
University of California, Santa Barbara
93106 Santa Barbara, CA
{lbertelli, manj}@ece.ucsb.edu

ABSTRACT

The first major contribution of this paper is a robust method to learn the photometric mapping between the overlapping portions of two registered images acquired either under different lighting conditions or different sensor modalities. Then, once such mapping is learnt, we demonstrate how it generalizes so that the photometric appearance can be transferred from one image to the other out of their overlapping area.

This task is fundamental in several different contexts, such as image colorization, seamless mosaicking or change detection. After introducing the theory and discussing the algorithms, we will present several examples that confirm the efficacy of the proposed method dealing with different types of images.

Index Terms— Radial Basis Function, Approximation, RANSAC, Photometric variation.

1. INTRODUCTION

In this paper we propose a novel, robust method to learn the function that describes the photometric mapping between two images that, *after being registered*, share a common overlapping area. The basic idea is to first use the channel n -tuples (such as the RGB values) at corresponding pixel positions to learn robustly an approximant which is valid on the overlapping portion of the images. Such mapping can then be used to modify the photometric appearance of the images outside their overlapping area.

This problem is relevant for many different applications, such as photographic manipulation (e.g. colorization of aerial images into maps), image blending and change detection. We will present examples of the previously listed applications in the experimental part of this paper. For the moment being, we will introduce some related work in the field of image registration and change detection.

In [1] Brown et al. notice that after a set of images have been aligned to the same coordinate system, it may be necessary to compensate for the gain adjustment performed au-

tomatically by the camera(s) in order to render seamless mosaics. They propose to minimize a quadratic objective function in the gain parameters obtained using the intensity channel values at corresponding pixel locations. This formulation has two main limitations: first it accounts only for a simple photometric transformation and secondly it is not robust to the presence of outliers (such as points where the intensity saturates and therefore it cannot be transformed just using a gain adjustment). This event, which can be quite common, can lead to biased estimates of the gains. Our method tackles both problems by approximating the function that describes the photometric transformation (for which we may not know the analytic expression in closed form) using a set of radial basis functions (RBF) [2]. We will demonstrate the flexibility of this approach in learning the color mappings between electro-optical vs. infrared imagery (EO vs. IR) and electro-optical vs. artificially colored imagery. Moreover, our learning scheme is robust to the presence of outliers, since the weights of the RBFs are estimated within a RANSAC framework. This allows us to discard points whose photometric variation cannot be learnt using the RBF approximant. Such situations may arise in presence of structural differences between the images, originated by landscape remodeling, by a non modeled motion of the image objects or by intensity saturations. In addition to the improved robustness, the distinction between inliers and outliers (a byproduct of the estimation carried out within the RANSAC framework) can also be used as a robust change detector [3].

2. LEARNING COLOR MAPPING VIA RADIAL BASIS FUNCTIONS

2.1. Establishing Spatial Mapping Via Registration

The technical details of the image registration system we use in our approach are beyond the scope of this paper. It is enough to mention that the image correspondences are automatically established using a (point) feature based approach. The image pair is aligned to a common reference system us-

ing a homographic mopping [4], which is general enough to model the perspective distortion of planar surfaces seen by a pin-hole camera.

2.2. Radial Basis Function Networks

Supervised learning, also known as non-parametric regression, is the problem of estimating a function f , given only a training set of N input-output mappings ($\mathbf{x}_i \mapsto y_i$, where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$ for $i = 1, \dots, N$). The only assumption made about the function f is that is likely to be smooth. Mathematically this turns out to minimize a functional $H(f)$ composed of two parts: a data fitting term $E(f)$ and a regularizing term $\Phi(f)$:

$$H(f) = E(f) + \lambda\Phi(f) \quad (1)$$

The data fitting term can be expressed as:

$$E(f) = \frac{1}{2} \sum_{i=1}^N (f(\mathbf{x}_i) - y_i)^2 \quad (2)$$

The general solution for minimizing (1) can be derived using calculus of variations as:

$$f(\mathbf{x}) = \sum_{i=1}^N w_i h(\mathbf{x}, \mathbf{x}_i) + \sum_{j=1}^q d_j \psi_j(\mathbf{x}) \quad (3)$$

where h is the Green's function for the differential operator ϕ , ψ_j with $j = 1, \dots, q$ are the basis functions of the null space of ϕ and w_i is a set of weights. If we require rotation invariance for the function ϕ (i.e. $\phi(f(\mathbf{x})) = \phi(f(R\mathbf{x}))$, with R being a rotation matrix), we obtain that the Green's functions $h(\mathbf{x}, \mathbf{x}_i)$ are *Radial Basis Functions* (RBFs). This means that they depend only upon the radial distance from the centroid, i.e. $h(\mathbf{x}, \mathbf{x}_i) = h(\|\mathbf{x} - \mathbf{x}_i\|)$. The choice of the regularization function yields different types of radial basis function (cubic splines, thin plate splines, gaussian kernels). In this paper we restrict our attention to the particular type of regularization which gives Gaussian RBFs, whose null space is empty (see [2], for more details). Hence the second term in the right hand side of eq. (5) vanishes. In this case, solving for the weights w_i yields:

$$\mathbf{w} = (H + \lambda I)^{-1} \mathbf{y} \quad (4)$$

where \mathbf{w} and \mathbf{y} are column vectors containing respectively w_i and y_i , H is a $N \times N$ matrix such that $h_{ij} = h(\|\mathbf{x}_i - \mathbf{x}_j\|)$ and I is the $N \times N$ identity matrix. In most practical applications N is very large, making the approach in (4) impractical. Starting with H we can select $M < N$ basis vectors (columns) and form the $N \times M$ matrix \hat{H} . This is equivalent to using only M RBFs, centered in a subset of cardinality M of the data points. Note that now the centroids of the RBFs are not required anymore to belong to the data points, even though this is a common choice. Assuming for now the regularization

coefficient $\lambda = 0$, the problem of estimating the weights becomes a Linear Least Squares Estimation problem, in which we want to find \mathbf{w} such that the squared magnitude of the error $\mathbf{e} = \mathbf{y} - \hat{H}\mathbf{w}$ is minimized (i.e. $\mathbf{e}^T \mathbf{e}$ is minimized). The LLSE optimal solution is given by the orthogonal projection:

$$\mathbf{w} = \hat{H}^\dagger \mathbf{y} \quad (5)$$

where \hat{H}^\dagger is the pseudo-inverse of \hat{H} : $\hat{H}^\dagger = (\hat{H}^T \hat{H})^{-1} \hat{H}^T$. Regularization can be obtained by adding $\lambda \hat{I}$ to \hat{H} , where $\hat{I} = [I \ 0]^T$ (I is the $M \times M$ identity matrix and 0 is a $M \times (N - M)$ zero matrix).

2.3. Robust Estimation of the RBF weights

The least square estimator (5) that yields the RBF weights can be biased by the presence of outliers. In this specific context we regard as outliers the pairs of n -tuples for corresponding pixels (in the overlapping portion of the images) for which the RBFs cannot learn a satisfactory photometric mapping. To overcome this problem we propose to carry out the weight estimation within the RANSAC framework [5] (more specifically we use the MSAC variation suggested in [6]). Note that the size of the *minimal sample set* (MSS) is equal to the number of the RBFs used to approximate the photometric mapping. In fact, since each n -tuple produces an equation and since there are M RBFs, the cardinality of the minimal sample set is M . At each iteration, RANSAC selects a MSS which is used to instantiate the model parameters. Then, among the remaining points, those who fit the model within a given noise threshold T_η will be detected and the quality of the solution is evaluated in terms of the goodness of the fit. The algorithm will iterate enough times to reduce the probability of *never* sampling a minimal sample set composed *only* by inliers. This probability is less than ε as long as the the number of iterations is:

$$\hat{T}_{iter} \geq \left\lceil \frac{\log \varepsilon}{\log(1 - q)} \right\rceil$$

where:

$$q = \frac{\binom{N_I}{k}}{\binom{N}{k}} = \frac{N_I!(N - M)!}{N!(N_I - M)!} = \prod_{i=0}^{M-1} \frac{N_I - i}{N - i} \approx \left(\frac{N_I}{N} \right)^M$$

and N_I denotes the number of inliers (also estimated online as the largest consensus set found so far). This number can rapidly grow very large. To mitigate this problem we sub-sample the n -tuples in the overlapping portion of the images (i.e. we reduce N). The error threshold, a parameter quite critical in RANSAC, is set accordingly to the dynamic range of the intensity values of the images: in all our experiments it was set to $T_\eta = 13$.

3. EXPERIMENTAL EVALUATION

In this section we demonstrate the effectiveness of the proposed approach in learning the photometric mapping between

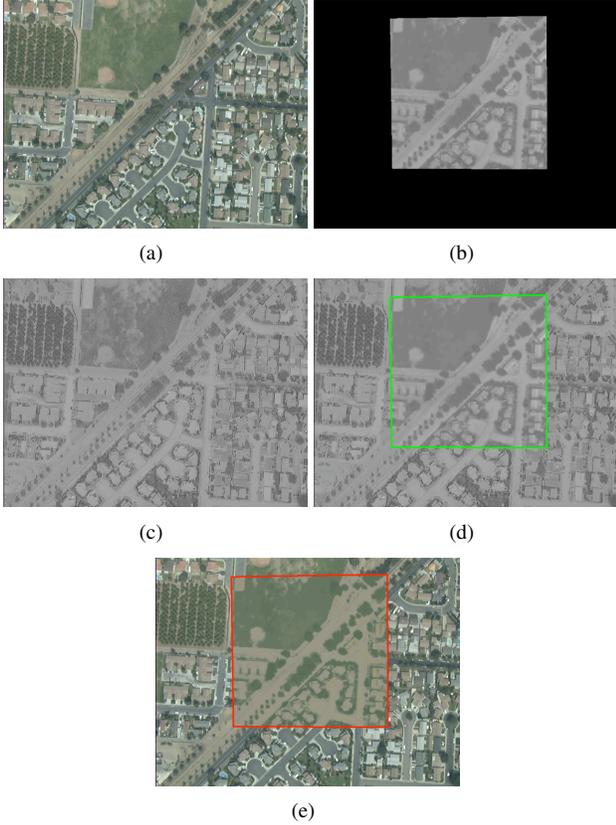


Fig. 1. Aerial view acquired under different sensor modalities. (a) Electro-Optical (EO) image. (b) Infra-Red (IR) image. (c) Photometric mapping learnt from EO to IR. (d) Comparison of the original IR (overlapping area) and the learnt IR (non-overlapping area) to show the consistency in the mapping. (e) Photometric mapping learnt from IR to EO.

the overlapping portions of two registered images (often acquired either under different lighting conditions or different sensor modalities) and its capability of colorizing areas that do not share any overlap. In all examples the registration is obtained in a fully automatic manner (except for the case in Fig. 3, where the satellite image and the map were already registered). The number of RBFs is chosen between 10 and 50 and the λ parameter (coefficient of the regularizing term) is kept *constant* for all the experiments ($\lambda = 0.001$). The position of the centroids is chosen randomly selecting a subset of the data points. We also would like to mention that we tested our framework with other criteria for selecting the centroids, such as k -means clustering on the data points or uniform sampling of the feature space. A preprocessing step, in which the two images are smoothed using a Gaussian kernel, is used in order to mitigate the effect of the registration errors which could give raise to incompatible n -tuple pairs.

The first two examples show the learning of the mapping between Electro-Optical (EO) and Infra-Red (IR). Fig. 1(c) and Fig. 2(c) show the mapping from EO to IR, learnt based on the overlapping portions of the two views. As shown

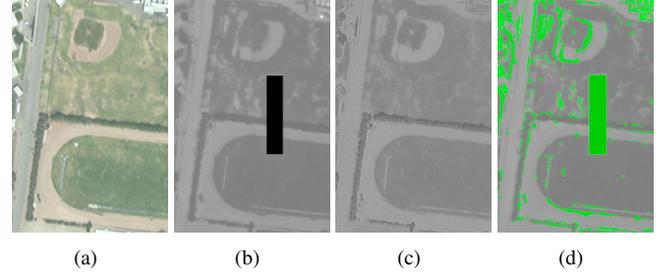


Fig. 2. Another aerial view acquired under different sensor modalities. (a) Electro-Optical (EO) image. (b) Infra-Red (IR) image with synthetically generated occlusion. (c) Photometric mapping learnt from EO to IR. (d) Outliers (in green) detected by RANSAC during the robust estimation of the RBF weights.

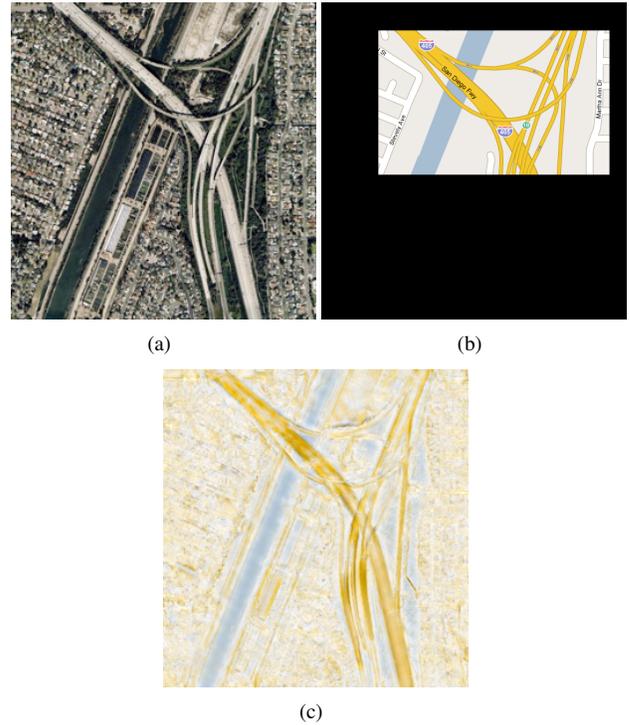


Fig. 3. (a) Satellite image. (b) Map. (c) Photometric mapping learnt from Satellite to Map (learnt from the overlapping region and then extended to non-overlapping area).

in Fig. 1(d), where we compare the original IR image (on the overlapping area) and the learnt IR image (on the non-overlapping area) the coloring is consistent. Fig. 1(e) demonstrates that it is possible to learn the inverse mapping (IR to EO) as well. In addition, Fig. 2 shows that estimating the RBF weights using RANSAC, the algorithm becomes extremely robust to the presence of outliers (i.e. black patch in Fig. 2(b)). This can be used as a robust change detector when the images to be compared are acquired under different sensor modalities or lighting conditions.

In the extremely challenging example of Fig 3, maps are learnt from satellite images. In order to make more discriminant the description of the points in the first view (i.e. the

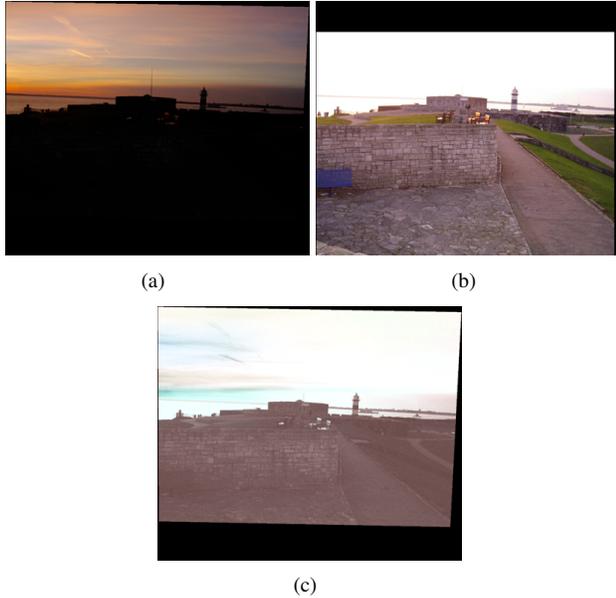


Fig. 4. Two images of the same scene taken under different lighting conditions. (a) Sunset scene. (b) Day light scene. (c) Photometric mapping learnt from Sunset to Day light.

satellite image), we added to the color information a description based on texture (Gabor feature descriptors are computed using 6 orientation and the smallest 2 scales, in order to limit edge delocalization). Therefore the learnt mapping goes from a feature space with 15 components (3 colors plus 12 texture features) to the standard 3 component color space of the map.

In Fig. 4 we apply the proposed framework to learn the mapping between two different lighting condition (sunset to day light). Similarly in Fig. 5, we demonstrate that it possible to learn the mapping between two different DOQQ images. Mapping from the first modality to the second one is shown in Fig 5(c), while the reverse mapping is shown in Fig. 5(d). The colorization remains consistent despite the structural changes in the DOQQs due to landscape remodeling.

4. CONCLUSIONS

In this paper we presented an approach to learn the photometric mapping between partially overlapping images and to extend such mapping out of the overlapping area. The function that realizes the mapping is learnt using a set of radial basis functions whose weights are robustly estimated within a RANSAC framework. The efficacy of our approach has been shown on several image pairs acquired under different sensor modalities and lighting conditions. We have also experimentally proved that the discrimination between inliers and outliers can serve as a robust change detector.

Future work will explore the possibility of introducing a spatial component in the learning process so that the same color can be mapped differently according to the spatial location. We also plan to explore further the possibility developing a change detector framework based on the learnt photo-

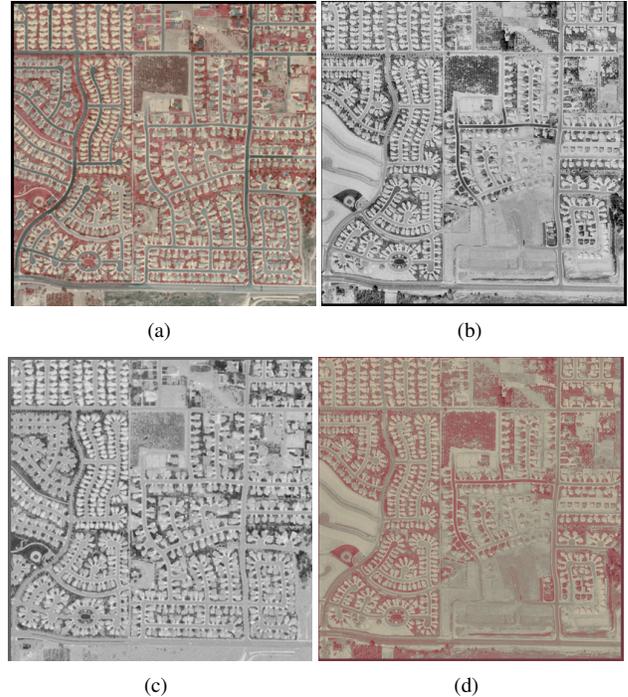


Fig. 5. (a,b) Two different DOQQ images of the same aerial scene. (c) Photometric mapping learnt from DOQQ1 to DOQQ2. (d) Learnt mapping from DOQQ2 to DOQQ1.

metric mapping between two images.

5. REFERENCES

- [1] M. Brown and D. Lowe, “Automatic panoramic image stitching using invariant features,” *International Journal of Computer Vision*, 2006, Accepted for publication.
- [2] T. Poggio and F. Girosi, “Networks for approximation and learning,” *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1481–1497, 1990.
- [3] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, “Image change detection algorithms: a systematic survey,” *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, March 2005.
- [4] Barbara Zitová and Jan Flusser, “Image registration methods: a survey,” *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, October 2003.
- [5] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, pp. 381–395, 1981.
- [6] P.H.S. Torr and A. Zisserman, “MLESAC: A new robust estimator with application to estimating image geometry,” *Journal of Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.