MDL region-based image registration

Maria-João RENDAS Adrien BARRAL
Lab. I3S, CNRS - Université de Nice Sophia Antipolis
{*rendas,barral*}*@i3s.unice.fr*

Abstract

We propose a new image registration criterion that uses segmented versions of the images to exploit both the geometric and photometric characteristics of the distinct regions they contain. It is inspired on the Minimum Description Length principle, considering the alignment parameters as side information, in a coding framework. The criterion is robust with respect to unstable segmentation, a frequent problem in close observation of natural scenes. Results on real images illustrate the performance attained.

1. Introduction

Estimating the map that puts in correspondence different views of scene points obtained with distinct cameras, or with the same camera from different positions/orientations is an important goal in a wide variety of applications, ranging from medical imaging to robot navigation [9, 3], and is known as the problem of *image registration* or *image mosaicing*. The majority of the proposed techniques, see e.g. [8], rely on the identification of a set of landmarks (or features) in each image, and numerical estimation of the map on the basis of the pair-wise association of features present on both images. These methods have the advantage of being relatively fast, but breakdown when many similar features are dispersed in the observed area, as it is often the case for local views of natural scenes. Global methods, based either on minimisation of the entropy of image residuals [6] or on the maximisation of mutual information, have also been proposed, in particular in the medical imaging domain [9]. These methods have very large computational requirements, involving the estimation of a probability distribution (joint distribution, when mutual information is used) for each possible map between the two images. Both approaches are based on local comparison of pixel intensities, and are thus unable to efficiently exploit the *macroscopic* structure of the images.

In this paper we propose a new criterion for image registration that exploits the geometric and photometric characteristics of the distinct regions that compose the observed images. The criterion is inspired on the Minimum Description Length (MDL) [4], and selects the map between the two images by casting the registration problem in a sequential coding perspective. The map is chosen as the optimal value of "side-information" that enables the most efficient "predictive coding," in a sense that is made precise in Section 3. Before introducing the new criterion, we define notation and the input information used by the algorithm (in particular, the fact that the images are first segmented), in Section 2. Section 4 presents results using real images, both for a simulated camera motion, and for real images acquired by a mobile observer (video camera installed in an underwater robot) that demonstrate the validity of the criterion proposed.

2. Pre-processing

Let I_n be the image acquired at sampling time $t_n, n = 1, 2, \ldots$:

$$I_n = \{I_n(s),\ s \in \mathcal{S}\}\,, \qquad I_n(s) \in \mathcal{L}\ ,$$

where $\mathcal{S}$ denotes the image support (defined in a sensor centric coordinate system), and $\mathcal{L}$ is the *finite* set of pixel values. Typically, $\mathcal{S}$ is a rectangular discrete grid, and $\mathcal{L} = \mathcal{L}_{b\&w} \equiv \{0, ..., 2^8 - 1\}$ or $\mathcal{L} = \mathcal{L}^3_{b\&w}$.
Our image registration criterion requires prior segmentation of the acquired images, and considers alignment of a set of labeled images I^s_n:

$$I^s_n = \{I^s_n(s),\ s \in \mathcal{S}\}\,, \qquad I^s_n(s) \in \{c_1, \ldots, c_{K^{I_n}}\}\,, \qquad ,$$

where $K^{I_n} \ll |\mathcal{L}|$ is the number of distinct regions in I_n. We denote by $\{v^n_\ell\}_{\ell=1}^{K^{I_n}}$ the set of empirical distributions[1]

$$\forall s \in \mathcal{S},\ \ I^s_n(s) = c_\ell \Leftrightarrow I_n(s) \sim v^{I_n}_\ell\ \ .$$

[1]Notation $x \sim p$ indicates that variable x is drawn from probability law p.

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

We will also use the following notation:

$$\mathcal{R}_\ell^{I_n} : \{s \in \mathcal{S} \ : \ I_n^s(s) = c_\ell\} \ \ ,$$

for the support of region ℓ, and $T_\ell^{I_n} = |\mathcal{R}_\ell^{I_n}|$ for its size. Clearly, the collection $\left\{\mathcal{R}_\ell^{I_n}\right\}_{\ell=1}^{K_n}$ is a partition of the image support $\mathcal{S}$. When the image to which these regions (or distributions) refer is obvious from the context, we will often omit the upper indice: $\mathcal{R}_\ell^I \equiv \mathcal{R}_\ell$.
We do not address here the segmentation problem, and concentrate on the problem of determining image correspondence, that is formulated in the next section. The results in section 4 use a slightly modified version of the K-mean segmentation algorithm [5], but the criterion proposed in the paper is robust with respect to the choice of segmentation algorithm.

3. Registration of segmented images

We present in this section the main result of the paper, which is a novel criterion for *global* registration of *segmented* images. Let us first state in general terms the image registration (or alignment) problem. Given images I and M, whose field of view is partially overlapping, we want to estimate the map H such that $I(s)$ and $M(H(s))$ represent the same point:

$$I(s) \leftrightarrow M\left(H(s)\right) \ \ . \tag{1}$$

The exact expression of the operator H depends on the camera model and environment geometry. For pin-hole cameras and flat scenes, H is an homography matrix [2].

MDL is a principled approach to model selection problems that, in intuitive terms - for a recent and formal presentation of MDL see [4, 7] - tells that the best model for the data is the one that "allows its shortest coding." Inspired by MDL, our criterion $\mathcal{C}(H)$ for finding the map H is also motivated by a coding/communication framework, in which the images are sequentially sent to a distant receiver over a noiseless channel. It measures the increase in coding efficiency when map H is sent as side information to the receiver, allowing it to use knowledge contained in the previous images to reconstruct the incoming image, see Fig. 1.
I is the image that is being coded, and M the image that is already known by the receiver when it receives the coded version of I. Our registration criterion associates to each map H the difference of the length of the messages of the two coding systems sketched in Fig. 1:

$$\mathcal{G}'(H) = \mathcal{C}(I) - \mathcal{C}(I|M,H) - \mathcal{C}(H) \ \ , \tag{2}$$

In this expression: *(i)* $\mathcal{C}(I)$ is the number of bits required to code image I alone (see Section 3.1); *(ii)*

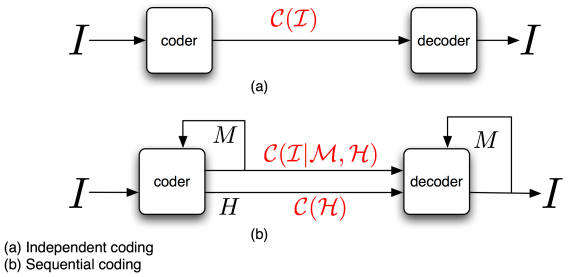

Figure 1. Coding paradigm: (a) Isolated Coding; (b) Sequential coding.

$\mathcal{C}(I|M,H)$ is the coding length when M is known to the receiver and H is sent as side information (Section 3.2); *(iii)* $\mathcal{C}(H)$ is the number of bits required to code the alignment information H.

If we consider a constant length code for H, $\mathcal{C}(H) = C$, we can drop it in (2), and estimate H as

$$\hat{H} = \arg\max_H \mathcal{G}(H), \ \ \mathcal{G}(H) \equiv \mathcal{C}(I) - \mathcal{C}(I|M,H) \ \ . \tag{3}$$

3.1. Independent coding of each image, $\mathcal{C}(I)$

We can use the models $\{v_\ell^I\}_{\ell=1}^{K_I}$ to built an *optimal code* for image I. Neglecting rounding errors, this optimal code will code image intensity $p \in \mathcal{L}$ in region ℓ with a number of bits[2] given by

$$L(p|v_\ell^n) = -\log v_\ell^I(p) \ \ . \tag{4}$$

(See [1] for all issues related to coding.) It is easy to show that the number of bits needed to code region $\mathcal{R}_\ell^I$, $L(\mathcal{R}_\ell^I)$, is given by:

$$L(\mathcal{R}_\ell^I) = \sum_{s \in \mathcal{R}_\ell^I} L(p|v_\ell) = T_\ell^I \mathbf{H}(v_\ell^I) \ \ ,$$

where $\mathbf{H}(v)$ is the Shanon entropy of distribution v:

$$\mathbf{H}(v) = -\sum_{p \in \mathcal{L}} v(p) \log(v(p)) \ \ .$$

Use of this optimal code, adapted to the image I, requires that the receiver be informed of the distributions $\{v_\ell^I\}_{\ell=1}^{K_I}$ as well as the image contours $\{\delta\mathcal{R}_\ell\}_{\ell=1}^{K^I}$. Denoting by $L(v_\ell^I)$ the cost of coding v_ℓ^I and by $L(\delta\mathcal{R}^I)$ the cost of coding the image contours, the total number of bits required to code I is

$$\mathcal{C}(I) = \sum_{\ell=1}^{K_I} L(v_\ell^I) + L(\delta\mathcal{R}^I) + \sum_{\ell=1}^{K_I} T_\ell^I \mathbf{H}(v_\ell^I) \ \ . \tag{5}$$

[2] All logarithms are base 2.

3.2. Sequential coding with side information, $\mathcal{C}(I|M,H)$

Coding penalty. If an *iid*[3] sequence $x_1^n, x_i \sim v$ is coded with the code that is optimal for a distribution $\nu \neq v$, average codelength will increase. The Kullback-Leibler divergence between the two distributions

$$D(v||\nu) = \sum_{p\in\mathcal{L}} v(p)\log\left(\frac{v(p)}{\nu(p)}\right) \ ,$$

tells us exactly how much this penalty is (see [1]). Under a misfit code $\tilde{v}_\ell \neq v_\ell$, the number of bits required to code each region $\mathcal{R}_\ell$ is larger than the Shanon entropy:

$$L'_{\tilde{v}_\ell}(\mathcal{R}_\ell) = T_\ell\left[\mathbf{H}(v_\ell) + D(v_\ell|\tilde{v}_\ell)\right] \ . \tag{6}$$

Inter-frame model extrapolation. Consider now the situation where the images are sequentially sent, as shown in the bottom of Fig. 1. Regions in the new image I and in the previous one (M) will in general not be the same: image segmentation, in special for natural scenes, is known to be prone to strong frame-to-frame variability, due to changes in illumination or to the appearance of new objects in the field of view. For this reason, we cannot simply associate distributions in the regions of I to those of the regions of M, and a more flexible manner of using the information contained in previous images is required.

To each possible H we associate new probability models $\{\tilde{\nu}_\ell^H\}$ for the regions in I that overlap M as *mixtures* of those associated to the regions of M, in the manner we describe next. Consider the finer partition of $\mathcal{S}$ determined by each map H

$$\{R_{\ell,m}^{M\downarrow I}(H)\}_{m=1,\ell=1}^{K_M,K_I} \ ,$$

where $R_{\ell,m}^{M\downarrow I}(H)$ is the support of the intersection of $\mathcal{R}_\ell^I$ with the map by H of $\mathcal{R}_m^M$:

$$R_{\ell,m}^{M\downarrow I}(H) = \mathcal{R}_\ell^I \cap H[\mathcal{R}_m^M] \subset \mathcal{S} \ . \tag{7}$$

Let $N_{\ell,m}(H) = \left|R_{\ell,m}^{M\downarrow I}(H)\right|$ (it can be zero), and denote by $\mathcal{IN} \subset \{1,\ldots,K_I\}$ the set of regions $\mathcal{R}_\ell$ of I for which not all $N_{\ell,m} = 0$ (i.e., that have a non empty overlap with $H[\mathcal{S}]$), and $\overline{\mathcal{IN}}$ its complement. These sets obviously depend on H.

We code regions $\mathcal{R}_\ell \in \mathcal{IN}$ using the best estimate of their distribution, given H and the models $\{\nu_m^M\}_{m=1}^{K_M}$ of the regions of M. They are mixtures of the empirical distributions of M

$$\tilde{\nu}_\ell^H = \sum_{m=1}^{K_M} \alpha_{\ell,m}(H)\nu_m^M, \qquad \ell \in \mathcal{IN} \ , \tag{8}$$

with coefficients $\alpha_{\ell,m}(H)$ given by

$$\alpha_{\ell,m}(H) = \frac{N_{\ell,m}}{\sum_n N_{\ell,n}}, \qquad \ell \in \mathcal{IN}; m = 1,\ldots,K_M \ .$$

Note that once the map H and the contours $\{\delta R_\ell^I\}$ are known in the receiver, the "predicted models" $\tilde{\nu}_\ell^H$ (8) can be determined. For regions $\ell \in \overline{\mathcal{IN}}$ the new image models are used, with optimal codelengths as described in the previous section, and sent to the receiver. For regions $\ell \in \mathcal{IN}$ there is no need of transmitting their distributions, but the penalized codelength (6) must be considered. The total codelength is thus

$$\begin{aligned}\mathcal{C}(I|M,H) \ = \ & L(\delta\mathcal{R}^I) + \\ & \sum_{\ell\in\overline{\mathcal{IN}}}\left[L(v_\ell^I) + T_\ell^I\mathbf{H}(v_\ell^I)\right] + \\ & \sum_{\ell\in\mathcal{IN}} T_\ell^I\left[\mathbf{H}\left(v_\ell^I\right) + D\left(v_\ell^I||\tilde{\nu}_\ell^H\right)\right] \ .\end{aligned} \tag{9}$$

3.3. The criterion $\mathcal{C}(H)$

Using (5) and (9) in (3), we obtain finally

$$\mathcal{G}(H) = \sum_{\ell\in\mathcal{IN}}\left[L(v_\ell^I) - T_\ell^I D\left(v_\ell^I\|\tilde{v}_\ell^H(\lambda)\right)\right] \ . \tag{10}$$

This final expression has an intuitive interpretation: it compares, for the map H considered, the overhead incurred by the need to specify the "new" optimal codes v_ℓ^I with the redundancy due to use of the codes (8). For H close to the true map, $\tilde{v}_\ell^H \simeq v_\ell^I$, and the penalty is negligeable compared to the cost of specifying the new distributions. Note that the *size* of the region put into correspondence does not appear explicitly in $\mathcal{G}(H)$, being indirectly reflected through the set $\mathcal{IN}$: the larger is the overlap implied by H, the larger will be this set.

To completely specify $\mathcal{G}(H)$, we still have to indicate an expression for $L(v_\ell^I)$: the number of bits required to code the empirical distributions v_ℓ^I (of sequences of length T_ℓ^I over a finite set of size $|\mathcal{L}|$). Giving no preference to any particular distribution, we consider an uniform code with length $L(v_\ell^I) = \log(|\Xi_\ell|)$, where Ξ_ℓ is the set of possible empirical distributions of sequences of length T_ℓ^I, which has size

$$|\Xi_\ell| = \left(\begin{array}{c} |\mathcal{L}|-1 \\ T_\ell^I + |\mathcal{L}| - 1\end{array}\right) \ .$$

This completes the definition of $\mathcal{G}(H)$ in (10).

[3] *iid*≡ independent and indetically distributed. Notation x_i^j denotes the set $\{x_i, x_{i+1}, \ldots, x_{j-1}, x_j\}$.

Figure 2. Urban aerial images.

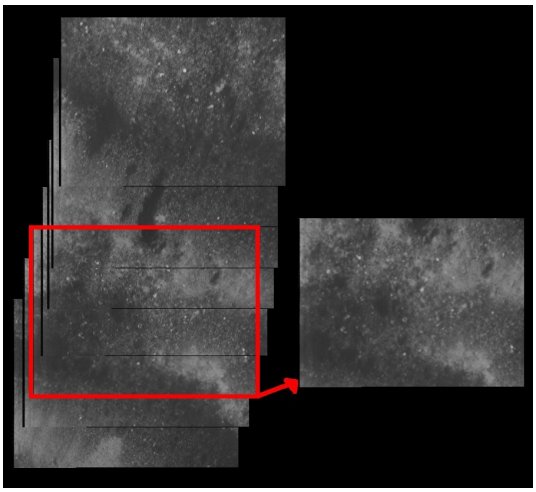

Figure 3. Underwater images.

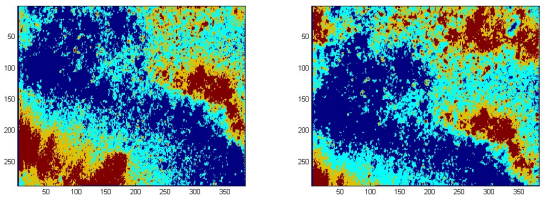

Figure 4. Segmented images (figure 3).

4. Results

We present in this section mosaics obtained superposing images using estimates of inter-image maps H (modeled as simple translations in the image plane) that maximise the coding gain (10). Fig. 2 shows the superposition of 3 aerial images of an urban area obtained from Google Earth, which have a clear geometric and structured content. The inset shows a detail of the right bottom corner, where the images are "glued" together, demonstrating that the correct maps have been identified. The second example, Figure 3, uses real images of the ocean floor taken with an underwater robot, and are typical of natural scenes, presenting gradual transitions between regions, and no clear geometric structure. The red rectangle indicates the recomposed region corresponding to one of the images, that is isolated in the right side of the figure to enable evaluation of the quality of the estimated maps. In Fig. 4 we show the segmented version of two images in the sequence, where instabilities in the exact region boundaries are noticeable, as it is often the case. Despite this variation, the criterion proposed identified the correct image correspondances. Lack of space prevents discussion of segmentation issues here, and a thourough presentation is defered to a forthcoming publication. In this example, data departs from two underlying assumptions of the estimaiton criterion: *(i)* the automatic control of the underwater camera - as well as variation of natural light - has induced intensity changes from image to image, and *(ii)* the maps H are modelled as simple translations, when in fact the robot has undergone small oscillations in orientation and altitude. This example shows that our criterion is not over-sensitive with respect to these unmodeled artifacts.

References

[1] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, 2006.

[2] O. Faugeras and Q.-T. Luong. *The Geometry of Multiple Images*. MIT Press, 2004.

[3] N. R. Gracias and J. Santos-Victor. Underwater video mosaics as visual navigation maps. *Computer Vision and Image Understanding*, 79(1):66–91, 2000.

[4] P. Grünwald. *The Minimum Description Length Principle*. MIT Press, 2007.

[5] S. P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 2:129–136, 1982.

[6] M.-J. Rendas. Construction of video mosaics using the minimum description length. In *Proc. IEEE OCEANS'06, Boston, USA*, 2006.

[7] J. Rissanen. *Information and Complexity in Statistical Modeling*. Springer, Information Science & Statistics, 2007.

[8] H. Zhang and S. Negahdaripour. Improved temporal correspondences in stereo-vision by ransac. In *ICPR (4)*, 2004.

[9] Zöllei, Lilla. *A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images*. PhD thesis, MIT, 2006.