

MRF-BASED AUTOMATIC IMAGE ORDERING AND ITS APPLICATION TO MOSAICING

Ran Song* Yonghuai Liu* Yitian Zhao* Ralph R. Martin† Paul L. Rosin†

* Department of Computer Science, Aberystwyth University, UK

† School of Computer Science & Informatics, Cardiff University, UK

ABSTRACT

A fast and robust auto-sorting method for image ordering based on Markov Random Fields (MRF) is proposed. We present a specific MRF model for the ordering problem and use pairwise phase correlation for the formulation. The MRF is inferred by a modified belief propagation (BP) method. Experimental results prove that the new method can reorder a disorganised collection of images without human input, prior information or restrictions, as just the first stage of a multi-stage mosaicing process, but also provides information that can be used to guide a mosaicing process in order to reduce both local mismatch and global error accumulation.

Index Terms— MRF, image ordering, image mosaicing

1. INTRODUCTION

The requirement to construct a large, composite mosaic image from a set of smaller images arises in a large number of applications. These depend on image mosaicing techniques [1, 2] which attempt to stitch a set of overlapped images together in an optimal fashion. Sometimes, there is no guarantee either that the source image set is ordered in any predictable fashion, or that any support information is available regarding imaging positions. Typical examples of such image sets are given in Fig. 1. However, most of the existing mosaicing techniques assume that the source images are ordered and rely strongly on this assumption to estimate local and global registrations.

[3] infers the ordering by solving for specific reliability using a quality measure where a coarse-to-fine local registration is needed for all images. The computational cost is usually very high for such ‘registration-based’ ordering methods. [4] proposes an algorithm based on viewpoint tracking. But it needs to calibrate the camera. In contrast, we develop a fully automatic method relying only on the information within the captured images. [5] proposes a comprehensive multi-stage approach which uses image patches and multi-view matching to deliver estimates of camera positions and corrected image clusters. The process is slow for high resolution images as it is based on features. It also requires large overlaps between images to offer a sufficient number of corresponding points.

Thanks to the RIVIC project supported by HEFCW/WAG for funding.

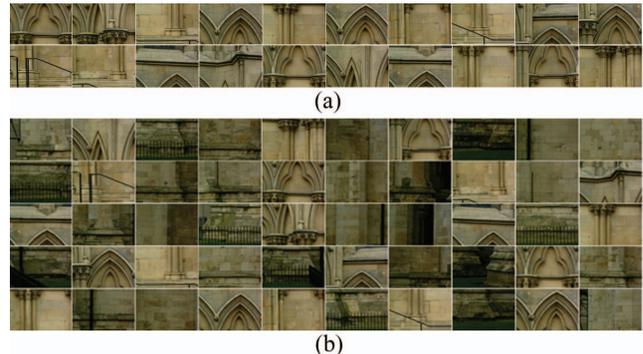


Fig. 1. (a) 20 images from the outer wall of the south transept of York Minster, UK. (b) 50 mixed images: 20 are the ones in (a) and the other 30 are from the wall of the north transept.

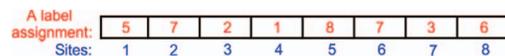


Fig. 2. The output of the labeling is an order: 5-7-2-1-8-7-3-6

We propose a fast and robust image ordering method based on MRF. The production of a full mosaic would be a later, separate process. Since the new method is independent of any other process, it has a wide range of applications.

2. MAP-MRF MODELING FOR IMAGE ORDERING

In an image ordering problem, the use of multiple camera positions means that the exact relationship between any two overlapped images within a multi-row image sequence is effectively random. There is not only one ordering which can guarantee that two successive images in the order are overlapped. The idea here is to utilize the MRF model to introduce extra constraints that lead to one optimal and unique ordering.

We denote $s = \{1, \dots, n\}$ representing the sites in the network and $L = \{1, \dots, n\}$ is the label set where each label corresponds to an image number. We define the label assignment $x = \{x_1, \dots, x_n\}$ to all sites as a realisation of a family of random variables defined on s . We also define an observation field $y = \{y_1, \dots, y_n\}$. Fig. 2 illustrates a possible labeling for ordering 8 images. An optimal labeling should satisfy the *maximum a posteriori probability* (MAP) criterion [6]: $x^* = \arg \max_x (P(x|y)) = \arg \max_x (p(y|x)P(x))$.

According to Hammersley-Clifford theorem [6]: the prior probability $P(x)$ satisfies a Gibbs distribution: $P(x) =$



Fig. 3. 10 input images obtained from an aerial imaging system within Wells Cathedral, UK. The intervening beam and hanging cable in I_2 and I_3 are framed with red rectangles.

$Z^{-1} \times e^{-U(x)}$, where Z is a normalising constant.

Let the likelihood densities $p(y|x)$ be expressed in the exponential form: $p(y|x) = Z_y^{-1} \times e^{-U(y|x)}$. Then the posterior probability is a Gibbs distribution: $P(x|y) = Z_U^{-1} \times e^{-U(y|x)+U(x)}$. The MAP-MRF problem is converted to a minimisation problem for the posterior energy $U(x|y)$:

$$x^* = \arg \min_x (U(x|y)) = \arg \min_x (U(y|x) + U(x)) \quad (1)$$

$U(y|x)$ is computed by the sum of the *likelihood potentials* at all sites: $U(y|x) = \sum_{i \in s} V_1(y_i|x_i)$. $U(x)$ is computed by the sum of binary *clique energies*: $U(x) = \sum_i \sum_{j \in \mathcal{N}(i)} V_2(x_i, x_j)$ where $\mathcal{N}(i)$ denotes the neighbourhood of i .

We formulate the MAP-MRF model using phase correlation. Phase correlation is not robust for mosaicing, but when only an approximate answer is required it can offer exactly what is needed for automated image ordering—rapid answers to the questions ‘Are these two images overlapping with each other?’ and ‘Is the degree of overlap large or small?’.

Given that there is a translation parameter t between image I_1 and I_2 , the value of pixel r in I_2 is: $I_2(r) = I_1(r+t) = I_1(r) * \delta(r-t)$. After a Fourier transform Ψ , the phase correlation is: $\Psi(I_1) = \Psi(I_2)e^{j2\pi fx} \Rightarrow e^{j(\varphi_1 - \varphi_2)} = e^{j2\pi fx}$. By performing an inverse Fourier transform Ψ^{-1} , we obtain

$$d(r) = \delta(r-t) = \Psi^{-1}(e^{j(\varphi_1 - \varphi_2)}) \quad (2)$$

Therefore, we can derive the translation t by looking for the peak of the impulse function d . When the two input images suffer from noise, pseudo-periodic structures, and/or any other unknown transformations, there will be multiple peaks, but there will normally be a single principal peak if the images do overlap. Fig. 3 and 4 show an example. A stained glass window was imaged from multiple positions. An intervening beam and hanging cable obscure some views. They do not appear in I_1 , but both objects appear in I_2 , while I_3 includes just the beam, but in a different position. These are sources of misinformation and could confuse feature-based approaches. However, as shown in Fig. 4, phase correlation produces a principal peak with an impressively high relative contrast. In general, the larger the principal peak, the larger the size and quality of the overlapping region.

Furthermore, many mosaicing approaches are based on local misfit measures between pairs of adjacent images and, owing to accumulated errors and unknown/incorrect corrections for lens and positional distortions, the process is highly

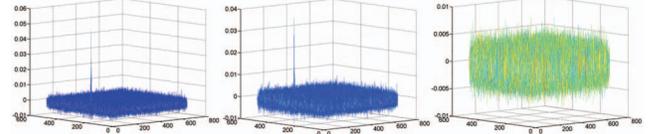


Fig. 4. L: The principal peak of I_1 and I_2 is 0.0521; M: the principal peak of I_1 and I_3 is 0.0398 (top right); R: there is no principal peak for I_1 and I_8 as the maximum is only 0.0085

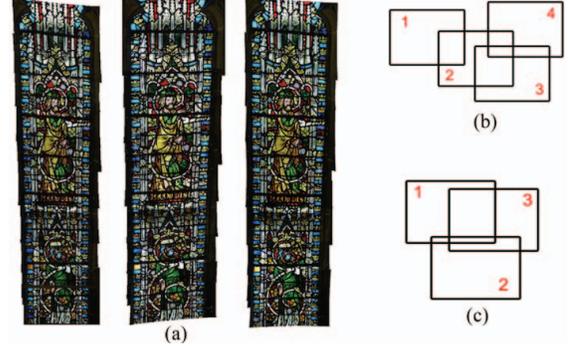


Fig. 5. (a) L: Stitching order from I_1 to I_{10} ; M: From I_{10} to I_1 ; R: First stitch from I_5 to I_1 and, independently, stitch from I_6 to I_{10} and finally stitch two partial mosaics together. (b) The illustration of *Principle 1*. (c) The illustration of *Principle 2*

nonlinear—different stitching orders lead to different mosaiced results (Fig. 5(a)). Essentially, our method is not only to look for realistic positional relationships for the collection of input images but also to define an optimal order in which to later stitch them. Thereby, we propose two principles:

Principle 1: An optimal ordering should be in line with the number of overlapping images of each image.

Principle 2: An optimal ordering should maximise the overlapping area between two successive images in the order.

In Fig. 5(b), the numbers of overlapping images of I_1 , I_2 , I_3 and I_4 are 1, 3, 2 and 2 respectively. According to *Principle 1*, the order should be 2-3-4-1 or 2-4-3-1. Fig. 5(a) also illustrates that usually a better alignment can be achieved due to the reduction of error accumulation if the alignment starts from a ‘central’ image. For a multiple image mosaicing, the more overlapping images one image has, the more ‘central’ it is. An optimal order should guarantee that these ‘central’ images can first be aligned and then the marginal images are stitched. Hence, it can be seen that this principle targets on the global optimisation of alignment. In Fig. 5(c), the stitching order should be 1-3-2 rather than 1-2-3 according to *Principle 2* as the overlapping area between I_1 and I_3 is larger than that between I_1 and I_2 . In essence, *Principle 2* tries to achieve the local optimisation of alignment. Broadly, larger overlaps should offer more information and more matching points for the estimation of positional information in image mosaicing. So, constructing the mosaic using those images with higher correlations should produce smaller local registration errors. We apply the idea of *Principle 1* to $V_1(y_i|x_i)$:

$$V_1(y_i|x_i) = \sum_{y_i \in L \setminus x_i} \min(1/P(y_i, x_i), F) \quad (3)$$

where F is a truncation constant and $L \setminus x_i$ denotes the labels in L other than x_i . $P(y_i, x_i)$ denotes the principal peak of the two images with sequence number y_i and x_i , derived from the maximum of the impulse function d in Eq. (2). For one image with a sequence number x_i , considering another image with a sequence number y_i , if $\frac{1}{P(y_i, x_i)} > F$, we think there is no overlap between them. F is set as 50 in this work. As we hope to minimise $V(y_i|x_i)$, Eq. (3) actually encourages a label assignment in compliance with **Principle 1**. On the other hand, we apply the idea of **Principle 2** to the *clique energy*:

$$V_2(x_i, x_j) = 1/P(x_i, x_j), \quad j \in \mathcal{N}(i) \quad (4)$$

By combining Eq. (3) and (4), $U(x|y)$ is computed as

$$U(x|y) = \sum_{i \in \mathcal{N}(x)} \sum_{y_i \in L \setminus x_i} \min\left(\frac{1}{P(y_i, x_i)}, F\right) + \lambda \sum_{i \in \mathcal{N}(x)} \sum_{j \in \mathcal{N}(i)} \frac{1}{P(x_i, x_j)}$$

where λ is a weighting parameter.

3. ENERGY MINIMISATION USING BP

There are several methods minimising the posterior energy. [7] shows that Graph-Cuts (GC) [8, 9] and BP [10, 11] are both efficient and powerful. GC is regarded fast and has some desirable theoretical guarantee on the optimality of the solution it can find. However, since our clique energy term is neither ‘metric’ nor ‘semimetric’ in terms of the definition in [8], GC is not suitable here. BP is thus a natural choice for us.

The BP works by iteratively passing messages across sites in a MRF network, briefed as follows:

- (1) For all neighbour site pairs $(i, j) \in \mathcal{N}$, initialising message m_{ij}^0 to zero, where m_{ij}^t is a vector of dimension given by the number of possible labels and denotes the message that site i sends to a neighbouring site j at iteration t .
- (2) For $t = 1, 2, \dots, T$, updating the messages as

$$m_{ij}^t(x_j) = \min_{x_i \in L \setminus x_j} \left(\sum_{y_i \in L \setminus x_i} \min\left(\frac{1}{P(y_i, x_i)}, F\right) + \frac{\lambda}{P(x_i, x_j)} + \sum_{h \in \mathcal{N}(i) \setminus j} m_{hi}^{t-1}(x_i) \right)$$

- (3) After T iterations, computing a belief vector for each site,

$$b_j(x_j) = \sum_{y_j \in L \setminus x_j} \min\left(\frac{1}{P(y_j, x_j)}, F\right) + \sum_{i \in \mathcal{N}(j)} m_{ij}^T(x_j)$$

From site $j=1$ to $j=n$, one by one, determining labels at each site:

$$x_j^* = \arg \min_{x_j \in L \setminus \{1, 2, \dots, j-1\}} (b_j(x_j)) \quad (5)$$

This strategy minimises the error accumulation early on in the mosaicing process and reduces the accumulative effect of errors in the final mosaic. This labeling scheme is applicable for the image sequences like Fig. 1(a) which contain only one complete image sequence. For the disorganised image sequences like Fig. 1(b), we set a threshold M for recognizing inappropriate input images or images from different sequences when determining x_j^* . Given that site j is labeled with x_j by Eq. (5), if $\sum_{k=1}^{j-1} P(x_k, x_j) < M$, the system

Table 1. Pairwise phase-correlation peaks of the 20 images. For example, in the top left cell, the value 0.0451 corresponds to the phase-correlation peak of I_1 and I_2 and the bottom right cell is the peak value 0.0117 produced by I_{19} and I_{20}

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	0.0451	0.0097	0.0648	0.0988	0.012	0.009	0.0088	0.0097	0.0121	0.0093	0.0097	0.0099	0.0106	0.0526	0.0346	0.0086	0.0149	0.0128	0.0698	
2		0.0097	0.0139	0.0109	0.0094	0.0093	0.0091	0.0094	0.0104	0.0107	0.0092	0.0094	0.0099	0.0142	0.0794	0.0099	0.0098	0.0143	0.1176	
3			0.0093	0.0183	0.0095	0.0091	0.0091	0.0094	0.0098	0.0092	0.0093	0.0099	0.0099	0.0099	0.0099	0.0099	0.0099	0.0099	0.0099	0.0094
4				0.0095	0.027	0.0094	0.0087	0.0131	0.015	0.0091	0.0094	0.0094	0.0155	0.0243	0.0376	0.0153	0.0097	0.0108	0.0099	
5					0.0093	0.011	0.0143	0.0101	0.0092	0.0096	0.0371	0.009	0.0101	0.0794	0.0097	0.0094	0.0264	0.0112	0.0191	
6						0.0086	0.0104	0.0208	0.0254	0.0106	0.0097	0.0165	0.0133	0.0575	0.0194	0.0829	0.0098	0.0213	0.0121	
7							0.0194	0.009	0.009	0.0463	0.0099	0.0102	0.009	0.0255	0.0091	0.0087	0.035	0.0606	0.011	
8								0.0109	0.0088	0.0323	0.0119	0.0088	0.0089	0.0097	0.0092	0.0097	0.0355	0.0107	0.0087	
9									0.0099	0.0096	0.0092	0.0132	0.0119	0.0098	0.0137	0.0213	0.0098	0.0097	0.0088	
10										0.0094	0.0094	0.0129	0.0114	0.0152	0.0157	0.0201	0.0101	0.0434	0.0097	
11											0.0101	0.0099	0.0084	0.0092	0.0094	0.0104	0.0098	0.0093	0.0102	
12												0.0109	0.0089	0.0097	0.0091	0.0088	0.0098	0.0096	0.009	
13													0.0131	0.0094	0.0158	0.0172	0.0097	0.0102	0.0095	
14														0.0106	0.0547	0.0108	0.0094	0.0114	0.0097	
15															0.0107	0.0091	0.0772	0.0519	0.0104	
16																0.0124	0.0093	0.0099	0.0097	
17																	0.0091	0.0091	0.0092	
18																		0.0653	0.0111	
19																			0.0117	

will conclude that this image has no overlap with any of the previous images in the sequence and will output the current order as one outcome and perform our image ordering scheme again for the rest of the images, repeating the process iteratively as long as there are still some images that have not been labeled in the input sequence.

4. EXPERIMENTAL RESULTS

In practice, because phase correlation can be realized by fast Fourier transform (FFT), this scheme performs much faster than those mentioned in Section 1. We first load the 20 disordered images shown in Fig. 1(a). The images contain repeated deceptive objects—without a larger context, some are so similar that it is difficult to differentiate and identify them by eye. Numbering them in terms of their loading order, the corresponding phase-correlation peaks are shown in Table 1. We use it as an index to obtain the principal peak P for any image pair. The strongest correlation is found between I_2 and I_{20} and the consequent stitching order produced by our ordering method is: 2–20–1–5–15–16–4–18–19–7–6–13–17–10–9–14–8–3–12–11. Then we can produce a single composite mosaic using a well-known multi-stage mosaicing method, outlined as follows: (1) Feature point extraction and matching using SIFT algorithm [12]; (2) Robust feature point matching using RANSAC [1]; (3) Transform estimation using Levenberg-Marquardt [1] algorithm and Bundle Adjustment [2]; (4) Image fusion using multi-band blending [2]. Fig. 6 illustrates this process at various stages of construction.

We compare the new method with a feature-based method [5]. In [5], the number of feature matches found between each pair of input images (the two-view matches) is calculated by feature extraction and matching where the feature detector and descriptor are not state-of-the-art. Then a spanning tree is constructed and the images are reordered according to the number of the two-view matches, subject to the constraint that the ordering will not create a cycle in the graph composed of the edges of the tree. To make our comparison more convincing, we replace the feature extraction and matching method employed in [5] with SIFT [12]. We still use the images in Fig. 1(a) for the test. The result of the comparison is shown in Table 2 quantitatively. It can be seen that the feature-based ordering is much more time-consuming than the new method (25 times more in computational time) although the mosaicing it guides is slightly better in terms of the mean registration error (MRE)). For qualitative comparison with Fig. 6, Fig. 7 illustrates the mosaicing process guided by the feature-based ordering at various stages of construction.

In Fig. 1(b), some of the 50 mixed input images have quite sim-

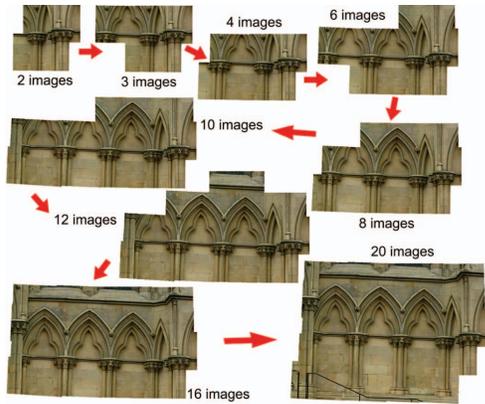


Fig. 6. Selected stages in process of constructing final mosaic

Table 2. Comparison between our method and the one in [5]

	Ordering result	Computational time(s)	MRE (pixels)
Our method	2-20-1-5-15-16-4-18-19-7-6-13-17-10-9-14-8-3-12-11	21.849535	0.2620
Feature-base method improved from [17]	2-20-1-5-15-18-7-19-10-9-17-6-4-13-14-16-3-12-11-8	549.558813	0.2449

ilar content, such as the railings and the mortar lines. The railings are an example of deceptive objects. Sometimes registration will fail if we just use a feature-based method to judge whether the two images with railings are overlapped, because the system could produce many pseudo matches located on the railings that will mislead the system. Bricks and stonework are other typical sorts of objects arising in images of architectural and heritage structures. For a disordered image set involving bricks it would be difficult to recognize and reorder them manually, because they look so similar to each other and lack distinctive features. However, our new method still correctly produce an optimal stitching order and two complete mosaics are produced. Fig. 8 shows a second outcome mosaic of 30 images produced together with the 20-image mosaic of Fig. 6.

5. CONCLUSIONS

We develop a fast method for automatic sequencing of a disordered set of images. We take an advantage of the MRF modeling and formulate it in accordance with two ordering principles. A modified is then employed to find the optimal labeling. Once each site in the MRF has been labeled, the order is found. In summary, the MRF-based method can produce an approximate ordering of images, as just the first stage of a multi-stage mosaicing process, but also provides information that can be used to guide a mosaicing process in order to reduce both pairwise registration errors and global error accumulation. Practical tests indicate that the scheme is robust and can find a good stitching order for multiple images with varying degrees of overlap, multiple deceptive features and occluding objects.

6. REFERENCES

[1] R. Szeliski, “Image alignment and stitching: A tutorial,” *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006.

[2] M. Brown and D.G. Lowe, “Automatic panoramic image stitching using invariant features,” *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.

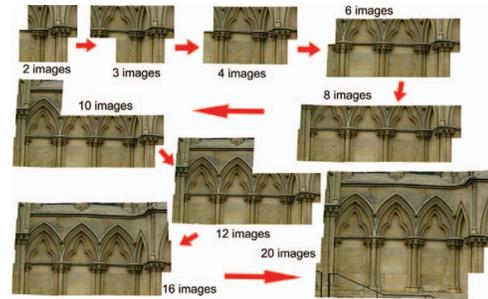


Fig. 7. Selected stages in process of constructing final mosaic using the feature-based ordering method improved from [5]



Fig. 8. As well as the 20-image mosaic, the second outcome is the 30-image mosaic of the wall of the north transept. Top left: The mosaic of the first 10 images. Top right: The mosaic with 20 images. Bottom: The final mosaic of 30 images.

[3] H.S. Sawhney, S. Hsu, and R. Kumar, “Robust video mosaicing through topology inference and local to global alignment,” in *Proc. ECCV*, 1998.

[4] R. Koch, M. Pollefeys, B. Heigl, L. Van Gool, and H. Niemann, “Calibration of hand-held camera sequences for plenoptic modeling,” in *Proc. ICCV*, 1999.

[5] F. Schaffalitzky and A. Zisserman, “Multi-view matching for unordered image sets, or how do i organize my holiday snaps?,” in *Proc. ECCV*, 2002.

[6] S.Z. Li, *Markov random field modelling in computer vision*. 2nd ed, Springer-Verlag, 1995.

[7] R. Szeliski et al, “A comparative study of energy minimization methods for markov random fields with smoothness-based priors. pattern analysis and machine intelligence,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 30, no. 6, pp. 1068–1080, 2008.

[8] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 23, no. 11, pp. 1222–1239, 2001.

[9] Vladimir Kolmogorov and Ramin Zabih, “What energy functions can be minimized via graph cuts?,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 26, no. 2, pp. 147–159, 2004.

[10] J.S. Yedidia, W.T. Freeman, and Y. Weiss, “Understanding belief propagation and its generalizations,” Technical Report TR-2001-22, MERL, 2002.

[11] P. Felzenszwalb and D. Huttenlocher, “Efficient belief propagation for early vision,” *International Journal of Computer Vision*, vol. 70, no. 1, pp. 41–54, 2006.

[12] D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.