

Image Matching and Retrieval by Repetitive Patterns

Petr Doubek, Jiri Matas, Michal Perdoch and Ondrej Chum
 Center for Machine Perception, Department of Cybernetics
 Czech Technical University
 Prague, Czech Republic
 {doub,matas,perdom1,chum}@cmp.felk.cvut.cz

Abstract—Detection of repetitive patterns in images has been studied for a long time in computer vision. This paper discusses a method for representing a lattice or line pattern by shift-invariant descriptor of the repeating element. The descriptor overcomes shift ambiguity and can be matched between different views. The pattern matching is then demonstrated in retrieval experiment, where different images of the same buildings are retrieved solely by repetitive patterns.

Keywords-repetitive patterns, image retrieval

I. INTRODUCTION

Man-made environments contain many repeating elements, e.g. windows on a facade, tiles on the floor or bars of a railing. These repetitive patterns are distinctive for humans. However, they pose a problem even for state-of-the-art image matching and retrieval algorithms, because the repeating elements are treated independently and since they are individually indistinguishable they increase the number of tentative correspondences and possible mismatches, see the top row in Fig. 1. Our goal is to detect repetitive patterns and match the entire pattern and thus turn a problem – ambiguity of individual elements that are difficult to match – into a strength, i.e. the distinctiveness of the whole pattern.

Different classes of repetitive patterns can be encountered in images: repetition of the basic building block – tile – on a 2D lattice, repetition along 1D line, scattered tiles. In this paper, we consider tiles repeating on a regular 2D lattice with possible perspective distortion or a regular repetition along a line. Bottom row of Fig. 1 shows examples of such repetitive patterns.

In one of the first papers on the subject, Leung and Malik [1] grow the pattern from local seed window by SSD registration into a possibly deformed 2D lattice. Schaffalitzky and Zisserman [2] use a very similar approach, investigating deeper the geometric transformations that generate the pattern – they defined perspectively distorted line repetition as conjugate translation and lattice repetition as conjugate grid.

Tuytelaars et al. [3] took a global approach by clustering repeating elements using a cascaded Hough transform. They focus on detecting symmetries, repetitive patterns play only a minor role.



Figure 1. Motivation and examples

A computational model for periodic pattern was proposed by Liu et al. [4] using the theory of crystallographic groups. Detection is performed on frontoparallel images of textures, patterns are classified based on their geometric structure.

Park et al. [5], [6] present impressive results on deformed lattice discovery focusing on detecting a complete pattern. The evaluation metric is the percentage of tiles detected in a pattern. For matching and retrieval, detection of the entire pattern is of minor importance and the percentage of detected tiles is not our objective.

To our knowledge, only Schindler et al. [7] attempted matching or retrieval by repetitive patterns. However, the matching is not inter-image, but against a manually prepared groundtruth database of facades. The database is small, containing only nine patterns.

In contrast to the previous work, we focus not on the detection itself, but on retrieving images of the same object by detected repetitive patterns. There is an inherent shift ambiguity in the repetitive pattern detection which we address by proposing a shift-invariant descriptor of the pattern.

II. REPETITIVE PATTERN DETECTION

The detection of lattice and line repetitive patterns is described in report [8]. Output of any of the cited repetitive pattern detection methods could be used if the implementation meets the following requirements: it can detect perspectively distorted lattice and/or line patterns; multiple patterns per image are handled; it returns representative frontoparallel

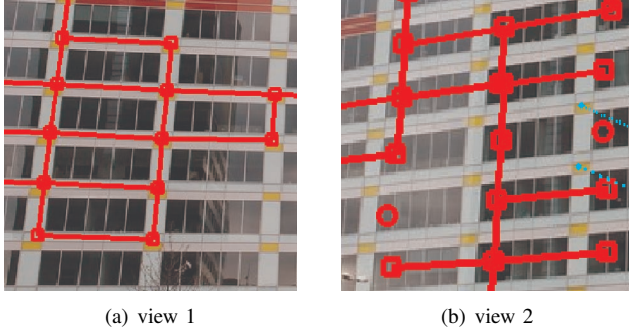


Figure 2. Tile shift-ambiguity

image of repeating element – *tile* – for each pattern; and the returned patterns are sorted by strength.

III. SHIFT-INVARIANT TILE REPRESENTATION

For matching and retrieval, we have to be able to compare mean tiles M between patterns. If the two compared patterns are images of the same real-world pattern, their mean tiles should be similar, except for possible 90,180 or 270 degrees rotation, different scale and translation. The rotation ambiguity arises due to different ordering or sign of the two lattice vectors. The translation ambiguity can be caused by different choice of tile centers. A tile can contain multiple different interest regions and any of them can be picked as a seed of the pattern. In different images of the same pattern, the choice of the same interest region is not guaranteed, see Fig. 2.

Fourier transform magnitude descriptor. In the frequency domain, image shift affects only the phase component. Therefore, for the shift and scale invariant representation we can use a normalized vector \mathbf{d} of the first 10×10 real coefficients of discrete 2D Fourier transform of the grayscale tile $F = \mathcal{F}(M)$, omitting the first coefficient (zero frequency).

Fourier transform with zero phase of the first harmonic. In an idea similar to [9], we adjusted the Fourier transform phase coefficients $\phi[F]$ to reach zero for the phase of the first harmonic coefficients in both horizontal and vertical direction ($\phi[F(1,0)]$ and $\phi[F(0,1)]$)

$$F^* = \mathcal{F}(M(m,n))e^{-j\phi[F(0,1)]l}e^{-j\phi[F(1,0)]k} \quad (1)$$

The “zero-phased” tiles $M^* = \mathcal{F}^{-1}(F^*)$ corresponding to the zero-phased descriptor \mathbf{d} are shown in Fig. 3 – notice their similarity.

The comparison of the zero-phased Fourier transform descriptor with the simpler Fourier transform magnitude descriptor ended with a tie as each prevailed on one dataset, see Fig. 4 for details. We believe that in some cases the Fourier transform magnitude descriptor might prove useful because of its higher robustness, despite its lower discriminative power. We also tested the description by SIFT descriptor of the zero-phase tile M^* , however it showed the

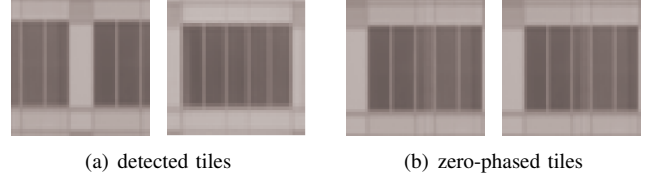


Figure 3. Shift-invariant mean tile. The same pattern can result in shifted lattice detections in different views – see Fig. 2. The mean tiles are therefore shifted (a). Our zero-phased mean tile representation is shift-invariant (b).

least successful, probably due to an increased sensitivity to inaccuracies in the tile size. As a baseline method, pixel intensities of 10×10 tile were tested – this descriptor is not shift-invariant and as expected, it has the lowest detection rate on both datasets.

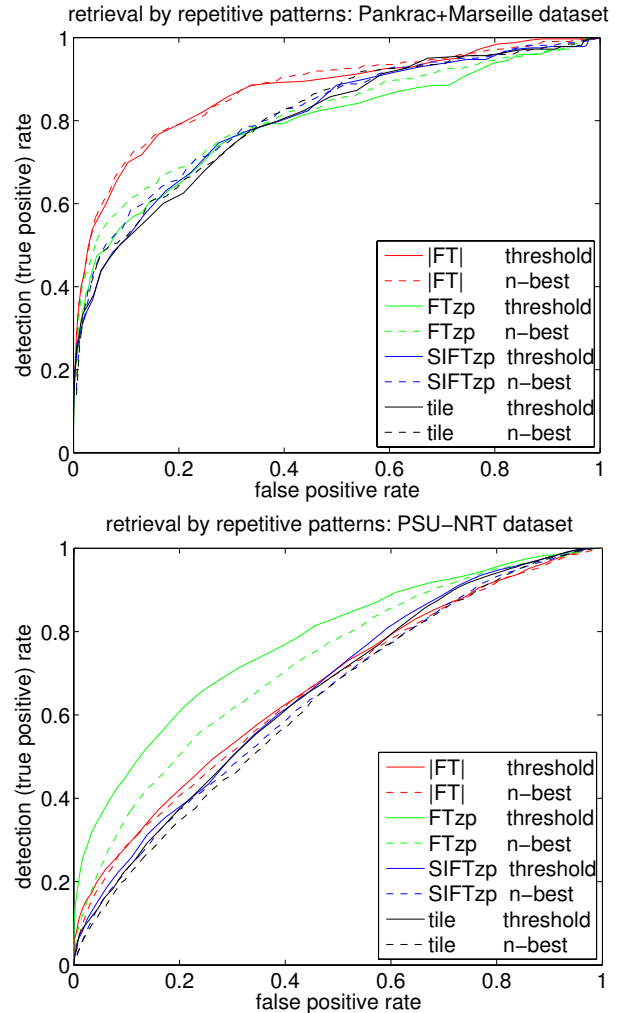


Figure 4. Comparison of descriptors: |FT| labels Fourier transform magnitude descriptor, FTzp is Fourier transform descriptor with zero first harmonic, SIFTzp is SIFT descriptor of the “zero-phased” tile, and “tile” descriptor is a simple vector of intensities of 10×10 tile.

Peaks in color RGB histogram. The above mentioned descriptors take into account only image intensity. To take

advantage of the color information, we also calculate and store two largest peaks in the color RGB histogram of the mean tile. However, it should be noted that for our main test objects, facades, color is only a minor cue.

IV. IMAGE RETRIEVAL BY REPETITIVE PATTERNS

We formulated the task of image retrieval as follows: for a given image query I_q (and set of its repetitive patterns \mathcal{C}_q), retrieve from the dataset \mathcal{I} set of images \mathcal{R}_q containing objects most similar to some object in the query image. Figure 5 shows an example of successful retrieval: the pattern correspondence which had the greatest influence on the match between the query and the retrieved image is marked by red, green and cyan for the first, second and third retrieved image respectively.

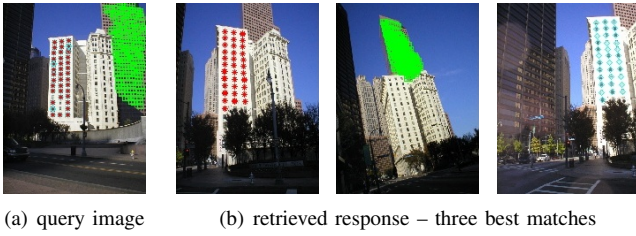


Figure 5. Retrieval by image query

The list of repetitive patterns \mathcal{C}_q of the query image is compared to repetitive patterns \mathcal{C}_i , where $i \in \{1 \dots |\mathcal{I}|\}$, of each image in the dataset and match score $S_{q,i}$ is calculated, see Sec. IV-A. The retrieved images are chosen as the ones with the highest match score.

We are aware that this naive each-to-each matching with complexity of a single query linear in the dataset size is acceptable only for a limited size of the dataset.

A. Match Score of an Image Pair

The match score $S_{i,j}$ of an image pair (I_i, I_j) is calculated solely from their repetitive patterns $(\mathcal{C}_i, \mathcal{C}_j)$, the images themselves are not used at this stage. For each pattern $C_k^i \in \mathcal{C}_i$, we find the best matching pattern $C_l^j \in \mathcal{C}_j$ with match score s_k , see Sec. IV-B for calculation of the match score between two patterns. Figure 6 shows an example of matching sets of patterns from two images.

The matching score s_k between patterns C_k^i and C_l^j is adjusted by ranking of the patterns inside their respective lists $\mathcal{C}_i, \mathcal{C}_j$

$$s'_k = \left(1 - \frac{1}{2} \frac{k+l}{|\mathcal{C}_i| + |\mathcal{C}_j|}\right) s_k . \quad (2)$$

This adjustment reflects the intention that large and strong patterns should have greater influence on the match score between images than small local patterns. The ranking inside lists can be used as a measure of the pattern strength because the pattern list in each image is required to be sorted as described in Sec. II.

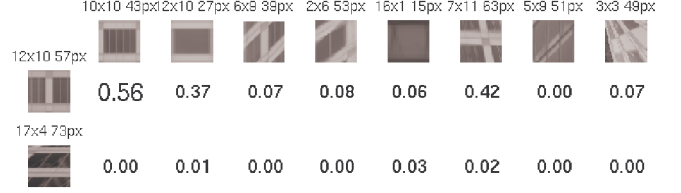


Figure 6. Matching sets of patterns of two images. First column shows mean tiles of patterns of the first image, first row contains mean tiles from the second image. We show tiles before the zero-phase normalization. Numbers in the table are match scores between each pattern pair. From each row, the maximum is selected, i.e. $s_1 = 0.56$ and $s_2 = 0.03$.

A linear combination of the highest three pattern matches $s'_1, s'_2, s'_3 \in \{s'_k\}$ form the match score of an image pair (I_i, I_j) : $S_{i,j} = 0.65s'_1 + 0.3s'_2 + 0.05s'_3$, where the coefficients were obtained as weights of linear classifier on feature vector (s'_1, s'_2, s'_3) learned by SVM.

B. Match Score of two Repetitive Patterns

The match score $s_{k,l} = s_{TSC}$ of two repetitive patterns C_k^i and C_l^j is a product of two measurements: similarity $s_T \in \langle 0, 1 \rangle$ of grayscale mean tiles and similarity $s_C \in \langle 0, 1 \rangle$ of color histograms.

Tile similarity s_T . The similarity between normalized vector descriptors of grayscale tiles is calculated as $s_T = 1 - |\mathbf{d}_k^i - \mathbf{d}_l^j|$.

Color similarity s_C . The color of both tiles is represented by two peaks in RGB histogram $\mathbf{p}_k^i = (\mathbf{p}_{k,1}^i, \mathbf{p}_{k,2}^i)$ and \mathbf{p}_l^j , where $\mathbf{p} = (r, g, b, n) \in \langle 0, 1 \rangle^4$, n being the relative count of pixels in the peak bin. The similarity formula has to take into account the possibility that the peaks are switched

$$s_C = 1 - \frac{\min(|\mathbf{p}_k^i - \mathbf{p}_l^j|, |\mathbf{p}_k^i - (\mathbf{p}_{l,2}^j, \mathbf{p}_{l,1}^j)|)}{\max(|\mathbf{p}_k^i|, |\mathbf{p}_l^j|)} . \quad (3)$$

Lattice size similarity s_L . We also tested a third factor, lattice size similarity. Each lattice has its width and height expressed in a number of tiles as $\mathbf{l} = (w, h)$. The similarity is computed as

$$s_L = 1 - \frac{1}{2} \left(\frac{|w_k - w_l|}{\max(w_k, w_l)} + \frac{|h_k - h_l|}{\max(h_k, h_l)} \right) . \quad (4)$$

We neglect the possibility that the width and height are switched, because in the prevailing type of images in our test data – building facades – the orientation remains almost always constant (sky up).

However, the lattice size similarity showed as misleading in experiments due to image crop or occlusion by other buildings the lattice size often differs significantly.

V. EXPERIMENT

The method was tested on two dataset: 1) our dataset Pankrac+Marseilles¹ with 106 images of app. 30 buildings

¹<http://cmp.felk.cvut.cz/data/repetitive>

and 2) the building subset of near-regular texture dataset PSU-NRT [6], containing 117 images with over 20 buildings.

For each image $I_i \in \mathcal{I}$, the groundtruth is labeled as a set of images $\mathcal{G}_i \subseteq \mathcal{I}$ that contains some object from I , i.e. \mathcal{G}_i is the groundtruth response to the query by image I_i . The response \mathcal{R}_i is either set of n images with the highest matching score $S_{i,j}$ or thresholded set $\mathcal{R}_i = \{I_j : S_{i,j} \geq \theta\}$, where θ is the threshold on the image match score. Figure 7 shows example responses \mathcal{R}_i with three best matching images. The trade-off between detection rate and false positive rate can be adjusted by n or by θ , see Fig. 4.

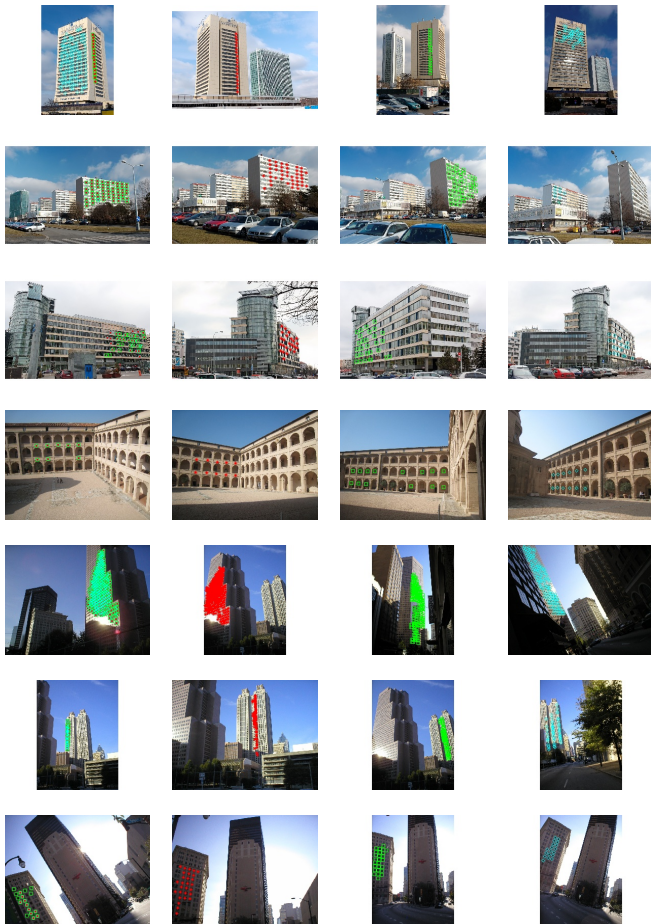


Figure 7. Retrieval experiment – queries on the left followed by 3 best matches

As expected, the retrieval by repetitive patterns performed better on the Pankrac+Marseilles dataset, where the average size of the tile image is larger and more details are observable. It corresponds to the purpose of these two datasets, authors of the PSU-NRT dataset used it solely for detection of repetitive patterns in single images, whereas our Pankrac+Marseilles dataset was created to test retrieval.

The average detection time by our Matlab implementation on 1000×700 image is 25 seconds. The time to run a single

query on 106 images dataset is 1 second, increasing linearly with the size of the dataset.

VI. CONCLUSIONS

We presented a method for image retrieval using repetitive patterns as the only feature. The contribution of the paper lies in 1) representing the pattern by a shift-invariant tile that can be matched to tiles of the same pattern detected in different views and 2) demonstrating that this repetitive pattern representation can be used to retrieve images from a dataset. Our dataset used for testing is publicly available together with the groundtruth.

Although the retrieval results of our method alone would not be sufficient especially on larger datasets, the repetitive pattern matching can be used to boost performance of standard matching methods based on single image features.

ACKNOWLEDGMENT

The authors were supported by Czech Science Foundation Project 102/07/1317 and by EC project FP7-ICT-247022 MASH.

REFERENCES

- [1] T. K. Leung and J. Malik, “Detecting, localizing and grouping repeated scene elements from an image,” in *ECCV*, 1996, pp. 546–555.
- [2] F. Schaffalitzky and A. Zisserman, “Geometric grouping of repeated elements within images,” in *BMVC*, 1998.
- [3] T. Tuytelaars, A. Turina, and L. Van Gool, “Noncombinatorial detection of regular repetitions under perspective skew,” *PAMI*, vol. 25, no. 4, pp. 418–432, April 2003.
- [4] Y. Liu, R. Collins, and Y. Tsin, “A computational model for periodic pattern perception based on frieze and wallpaper groups,” *PAMI*, vol. 26, no. 3, pp. 354–371, March 2004.
- [5] M. P. Park, R. T. Collins, and Y. L. Liu, “Deformed lattice discovery via efficient mean-shift belief propagation,” in *ECCV*, 2008, pp. 474–485.
- [6] M. Park, K. Brocklehurst, R. Collins, and Y. Liu, “Deformed lattice detection in real-world images using mean-shift belief propagation,” *PAMI*, vol. 31, no. 10, pp. 1804–1816, October 2009.
- [7] G. Schindler, P. Krishnamurthy, R. Lublinerman, Y. Liu, and F. Dellaert, “Detecting and matching repeated patterns for automatic geo-tagging in urban environments,” in *CVPR*, 2008.
- [8] P. Doubek, J. Matas, M. Perdóch, and O. Chum, “Detection of 2D lattice patterns of repetitive elements and their use for image retrieval,” FEE CTU, Prague, Research Report CTU–CMP–2009–16, 2009.
- [9] T. Pajdla and V. Hlaváč, “Zero phase representation of panoramic images for image based localization,” in *CAIP*, 1999, pp. 550–557.