

Content-Based Tile Retrieval System

Pavel Vácha and Michal Haindl*

Institute of Information Theory and Automation of the ASCR,
182 08 Prague, Czech Republic
{vacha,haindl}@utia.cas.cz

Abstract. A content-based tile retrieval system based on the underlying multispectral Markov random field representation is introduced. Single tiles are represented by our approved textural features derived from especially efficient Markovian statistics and supplemented with Local Binary Patterns (LBP) features representing occasional tile inhomogeneities. Markovian features are on top of that also invariant to illumination colour and robust to illumination direction variations, therefore an arbitrary illuminated tiles do not negatively influence the retrieval result. The presented computer-aided tile consulting system retrieves tiles from recent tile production digital catalogues, so that the retrieved tiles have as similar pattern and/or colours to a query tile as possible. The system is verified on a large commercial tile database in a psychovisual experiment.

Keywords: content based image retrieval, textural features, colour, tile classification.

1 Introduction

Ceramic tile is a decoration material, which is widely used in the construction industry. Tiled lining is relatively long-lived and labour intensive, hence a common problem to face is how to replace damaged tiles long after they are out of production. Obvious alternative to costly and laborious complete wall retiling is finding of the tile replacement from recent production which is as similar to the target tiles as possible. Tiles can differ in size, colours or patterns. We are interested in automatic retrieval of tiles as the alternative to usual slow manual browsing through digital tile catalogues and the subsequent subjective sampling. Manual browsing suffers from tiredness and lack of concentration problems, leading to errors in grading tiles. Additionally, gradual changes and changing shades due to variable light conditions are difficult to detect for humans. The presented computer-aided tile consulting system retrieves tiles from a tile digital database so that the retrieved tiles are maximally visually similar to the query tile. A user can demand either similar patterns, colours or a combination of both. Although the paper is concerned with the problem of automatic computer-aided content-based retrieval of ceramic tiles, the modification for defect detection or product quality control is straightforward.

* Corresponding author.

Textures are important clues to specify surface materials as well as design patterns. Thus their accurate descriptive representation can be beneficial for sorting and retrieval of ceramic tiles. Without textural description the recognition is limited to different modifications of colour histograms only and it produces unacceptably poor retrieval results. Image retrieval systems (e.g.[4,13]) benefit from combination of various textural and colour features. Frequent features are colour invariant SIFT [3], Local Binary Patterns (LBP) [10], Gabor features [8], etc. A tile classifier [6] uses veins, spots, and swirls resulting from the Gabor filtering to classify marble tiles. The verification is done using manual measurement from a group of human experts. The method neglects spectral information and assumes oversimplified normalized and controlled illumination in a scanner. Similar features were used for tile defect detection [9]. A promising method for object/image recognition based on textural features was recently introduced [12].

Unfortunately, the appearance of natural materials is highly illumination and view angle dependent. As a consequence, most texture based classification or segmentation applications require multiple training images [18] captured under all available illumination and viewing conditions for each material class. Such learning is obviously clumsy and very often even impossible if the required measurements are not available. Popular illumination invariant features include LBP variants [10], however, they are very noise sensitive. This vulnerability was addressed [7], but used patterns are specifically selected according to the training set. Recently proposed LBP-HF [1] additionally studies relations between rotated patterns. Finally, the MR8 texton representation [18] was extended to be colour and illumination invariant [2].

We introduce a tile retrieval system, which takes advantage of a separate representation of colours and texture. The texture is represented by efficient colour invariant features based on Markov Random Fields (MRF), which are additionally robust to illumination direction and Gaussian noise degradation [16]. The performance is evaluated in a psychovisual experiment.

The paper is organised as follows: the tile analysis algorithm is introduced in Section 2, Section 3 describes a psychovisual evaluation and discusses its results. Section 4 summarises the paper.

2 Tile Analysis

The tile image analysis is separated into two independent parts: colour analysis and texture analysis. Advantage of this separation is ability to search for tiles with similar colours, texture, or both — according to user preference. Colours are represented by histograms, which discard any spatial relations. On the other hand, the texture analysis is based on spatial relation modelling by means of MRF type of model, which is followed by computation of colour invariants. Colour invariants are employed instead of texture analysis of grey-scale images, because colour invariants are able to distinguish among structures with same luminance.

The texture representation with MRF colour invariants was chosen, because this representation is invariant to changes of illumination colour and brightness

[15], robust to variation of illumination direction [16] and combinations of previous conditions [17]. Moreover the MRF colour invariants are robust to degradation with an additive Gaussian noise [15] and they outperformed alternative textural features such as Gabor features or LBP in texture recognition experiments [15,16,17], especially, with variations of illumination conditions. Such illumination variations are inevitable, unless all images are acquired in a strictly controlled environment.

2.1 Colour Histograms

Colour information is represented by means of cumulative histograms [14], which we compute for each spectral plane separately. The cumulative histogram is defined as the distribution function of the image histogram, the i -th bin H_i is computed as

$$H_i = \sum_{\ell \leq i} h_\ell , \quad (1)$$

where h_ℓ is the ℓ -th ordinary histogram bin. The distance between two cumulative histograms is computed in L_1 metric.

2.2 CAR Textural Features

The texture analysis is based on the underlying MRF type of representation, we use efficient Causal Autoregressive Random (CAR) model. The model parameters are estimated and subsequently transformed into colour invariants, which characterize the texture.

Let us assume that multispectral texture image is composed of C spectral planes (usually $C = 3$). $Y_r = [Y_{r,1}, \dots, Y_{r,C}]^T$ is the multispectral pixel at location r , which is a multiindex $r = [r_1, r_2]$ composed of r_1 row and r_2 column index, respectively. The spectral planes are mutually decorrelated by the Karhunen-Loeve transformation (Principal Component Analysis) and subsequently modelled using a set of C 2-dimensional CAR models.

The CAR representation assumes that the multispectral texture pixel Y_r can be modelled as linear combination of its neighbours:

$$Y_r = \gamma Z_r + \epsilon_r , \quad Z_r = [Y_{r-s}^T : \forall s \in I_r]^T \quad (2)$$

where Z_r is the $C\eta \times 1$ data vector with multiindices r, s, t , $\gamma = [A_1, \dots, A_\eta]$ is the $C \times C\eta$ unknown parameter matrix with square submatrices A_s . In our case, C 2D CAR models are stacked into the model equation (2) and the parameter matrices A_s are therefore diagonal. Some selected contextual causal or unilateral neighbour index shift set is denoted I_r and $\eta = \text{cardinality}(I_r)$. The white noise vector ϵ_r has normal density with zero mean and unknown diagonal covariance matrix, same for each pixel.

The texture is analysed in a chosen direction, where multiindex t changes according to the movement on the image lattice. Given the known history of CAR

process $Y^{(t-1)} = \{Y_{t-1}, Y_{t-2}, \dots, Y_1, Z_t, Z_{t-1}, \dots, Z_1\}$ the parameter estimation $\hat{\gamma}$ can be accomplished using fast and numerically robust statistics [5]:

$$\begin{aligned} \hat{\gamma}_{t-1}^T &= V_{zz^{(t-1)}}^{-1} V_{zy^{(t-1)}} , \\ V_{t-1} &= \left(\begin{array}{cc} \sum_{u=1}^{t-1} Y_u Y_u^T & \sum_{u=1}^{t-1} Y_u Z_u^T \\ \sum_{u=1}^{t-1} Z_u Y_u^T & \sum_{u=1}^{t-1} Z_u Z_u^T \end{array} \right) + V_0 = \left(\begin{array}{cc} V_{yy^{(t-1)}} & V_{zy^{(t-1)}}^T \\ V_{zy^{(t-1)}} & V_{zz^{(t-1)}} \end{array} \right) , \\ \lambda_{t-1} &= V_{yy^{(t-1)}} - V_{zy^{(t-1)}}^T V_{zz^{(t-1)}}^{-1} V_{zy^{(t-1)}} , \end{aligned}$$

where the positive definite matrix V_0 represents prior knowledge.

Colour invariants are computed from the CAR parameter estimates to make them independent on colours. The following colour invariants were derived [15]:

1. trace: $\text{tr } A_s, \forall s \in I_r$,
2. diagonal: $\nu_s = \text{diag}(A_s), \forall s \in I_r$,
3. $\alpha_1: 1 + Z_r^T V_{zz}^{-1} Z_r$,
4. $\alpha_2: \sqrt{\sum_r (Y_r - \hat{\gamma} Z_r)^T \lambda^{-1} (Y_r - \hat{\gamma} Z_r)}$,
5. $\alpha_3: \sqrt{\sum_r (Y_r - \mu)^T \lambda^{-1} (Y_r - \mu)}$, μ is the mean value of vector Y_r ,

Feature vectors are formed from these illumination invariants, which are easily evaluated during the CAR parameters estimation process. The invariants $\alpha_1 - \alpha_3$ are computed for each spectral plane separately.

2.3 CAR-Based Tile Analysis

At the beginning, a tile image is factorised into K levels of the Gaussian-downsampled pyramid and subsequently each pyramid level is modelled by the previously described CAR model. The pyramid is used, because it enables models to captures larger spatial relations. Moreover, the CAR models analyse a texture in some fixed movement direction, therefore additional directions are employed to capture supplementary texture properties. More precisely, we used $K = 4$ levels of Gaussian-downsampled pyramid and the CAR models with the 6-th order hierarchical neighbourhood (cardinality $\eta = 14$). The texture was analysed in three orthogonal directions: row-wise, column-wise top-down and column-wise bottom-up. Finally, the estimated parameters for all pyramid levels and directions are transformed into colour invariants and concatenated into a common feature vector.

The dissimilarity between two feature vectors of two tiles T, S is computed using fuzzy contrast [11] in its symmetrical form FC_3 :

$$FC_p(T, S) = M - \left\{ \sum_{i=1}^M \min \left\{ \tau(f_i^{(T)}), \tau(f_i^{(S)}) \right\} - p \sum_{i=1}^M \left| \tau(f_i^{(T)}) - \tau(f_i^{(S)}) \right| \right\} ,$$

$$\tau(f_i) = \left(1 + \exp \left(- \frac{f_i - \mu(f_i)}{\sigma(f_i)} \right) \right)^{-1} ,$$



Fig. 1. Partition of tile image into five regions. The texture is analysed in the whole image and separately in these regions.

where M is the feature vector size and $\mu(f_i)$ and $\sigma(f_i)$ are average and standard deviation of the feature f_i computed over all database, respectively. The sigmoid function τ models the truth value of fuzzy predicate.

The textural representation is based on the homogeneity assumption, which is an inherent property of all textures. Unfortunately, some tiles contain insets or other violations of the homogeneity assumption. Therefore the CAR models are additionally estimated on each of five tile regions depicted in Fig. 1. The dissimilarities of corresponding image regions and whole images are combined to finally produce the dissimilarity of tiles $D(T, S)$:

$$D(T, S) = \text{Norm} \left(\sum_{\ell=1}^5 FC_3(T_\ell, S_\ell) \right) + \text{Norm}(FC_3(T, S)) \quad , \quad (3)$$

$$\text{Norm}(FC_3(T, S)) = \frac{FC_3(T, S) - \mu(FC_3)}{\sigma(FC_3)} \quad , \quad (4)$$

where T_ℓ, S_ℓ are the ℓ -th regions of images T, S , respectively. *Norm* is dissimilarity normalisation, where $\mu(FC_3)$ and $\sigma(FC_3)$ are mean and standard deviation of distances of all images. In practice, $\mu(FC_3)$ and $\sigma(FC_3)$ could be estimated on a subset of dataset, since the precise estimation is not necessary. This textural tile representation is denoted as “2D CAR 3x” in the results.

2.4 Local Binary Patterns

Local Binary Patterns (LBP) [10] are histograms of texture micro patterns. For each pixel, a circular neighbourhood around the pixel is sampled, P is the number of samples and R is the radius of circle. The sampled points values are thresholded by the central pixel value and the pattern number is formed:

$$LBP_{P,R} = \sum_{s=0}^{P-1} \text{sgn}(Y_s - Y_c) 2^s \quad , \quad (5)$$

where sgn is the sign function, Y_s is the grey value of the sampled pixel, and Y_c is the grey value of the central pixel. Subsequently, the histogram of

patterns is computed. Because of thresholding, the features are invariant to any monotonic grey-scale change. The multiresolution analysis is done by growing of the circular neighbourhood size. All LBP histograms were normalised to have unit L_1 norm. The similarity between LBP feature vectors is measured by means of Kullback-Leibler divergence as the authors suggested. We have tested features $LBP_{8,1+8,3}$, which are combination of features with radii 1 and 3 and which were computed on grey-scale images.

3 Experiments

Performance of two alternative textural retrieval methods (CAR, LBP) was evaluated in a psychovisual experiment, where the quality of retrieved images was evaluated by volunteers. The experiment was conducted on the dataset of 3301 tile images downloaded from an internet tile shop.¹ All images were resampled to the common size 300×400 pixels, the aspect ratio of rectangular images were maintained and the bigger side was resized to match the size. Thirty-four volunteers (26 males, 8 females) participated in our test. Age of participants ranged from nineteen to sixty, but majority was below forty. About one half of participants were specialist in the field of image processing. The test was administered over the Internet using a web application, so that each participant used its own computer in their environment. This setup is plausible, because we focused on significant, first glance differences, which are unlikely to be influenced by test conditions.

The test was composed of subsequent steps, where each step consisted of a query image and four test images. These four test images composed of two images retrieved by CAR method and two retrieved by LBP as the most similar to the query image, they were presented in a random order. Participants were instructed to evaluate quality of the retrieved images according to structural/textural similarity with the query image, regardless of colours. There were four ranks available: similar = 3, quite similar = 2, little similar = 1, dissimilar = 0. Subjects were also instructed that they should spend no more than one or two seconds per one test image. Because our system is intended to be a real-life application, we did not provide any examples of similar or dissimilar images, but we let people to judge the similarity in their own subjective opinion. The query images were once randomly selected and remained same for all participant in one run. They were presented in a fixed order so that the results were not influenced by different knowledge of previous images. Moreover, the first three query images were selected manually and were not counted in the results. The reason was to allow subjects to adjust and stabilise their evaluation scale.

The test was performed in two runs, where a single run consisted of the the same query and test images evaluated with different subjects. The first run consisted of 66 valid steps evaluated with 23 subjects, while the second one contained 67 valid steps ranked by 11 subjects. The evaluation of one subject was removed due to significant inconsistency with the others (correlation coefficient

¹ <http://sanita.cz>

Table 1. Subject evaluated quality of texture retrieval methods. The table contains average ranks (0 = dissimilar – 3 = similar) and corresponding standard deviations.

	2D CAR 3x	LBP _{8,1+8,3}
run 1	2.21 ± 0.64	2.22 ± 0.65
run 2	2.23 ± 0.62	2.21 ± 0.57

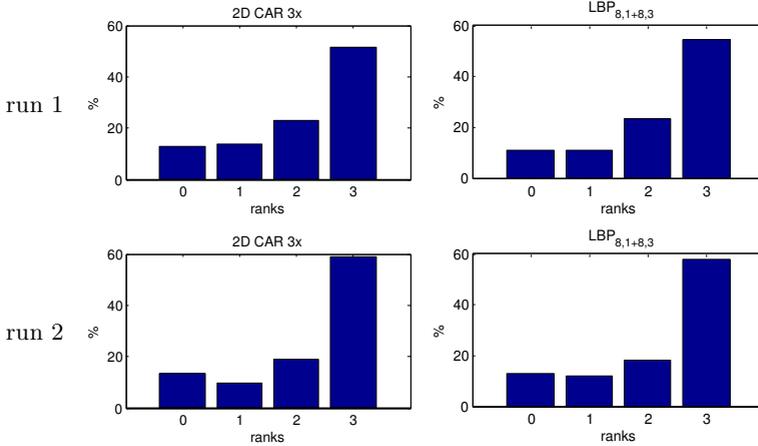


Fig. 2. Histogram of ranks (0 = dissimilar – 3 = similar) given by subjects. The first row shows histograms for the first test run, while the second row for the second run.

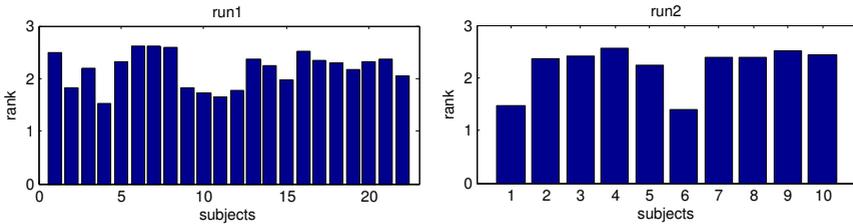


Fig. 3. Distribution of average ranks given by participants in the first and the second test run

= 0.4). Average correlation coefficients of subject evaluation were 0.64 and 0.73 for the first and the second run, respectively, which implies certain consistency in subject similarity judgements.

3.1 Discussion

The experimental results are presented in Tab. 1, which shows average ranks and standard deviations of retrieved images for CAR and LBP methods. The distribution of given ranks is displayed in Fig. 2. It can be seen that the performance



Fig. 4. Examples of similar tile retrieved by our system, which is available online at <http://cbr.utia.cas.cz/tiles/>. Query image, on the left, is followed by two images with similar colours and texture (CAR features). Images are from the internet tile shop <http://sanita.cz>

of both methods is comparable and successful. About 76% of retrieved images were considered to be similar or quite similar and only 12% were marked as dissimilar. More than two thirds of the participants ranked the retrieved tiles as quite similar or better in average, as can be seen in Fig. 3, which shows average ranks of participants. Different subject means in Fig. 3 show that the level of perceived similarity is subjective and a personal adaptation would be beneficial. Unfortunately, such an adaptation is not always possible since it requires user feedback.

As expected, the experiment revealed that LBP and CAR methods prefer different aspects of structural similarity. The LBP method is better with regular images that contain several distinct orientations of edges, while the CAR model excels in modelling of stochastic patterns. Moreover, LBP describes any texture irregularities in contrast to CAR model, which enforces homogeneity and small irregularities are ignored as errors or noise. Both approaches are plausible and it depends on a subjective view, which approach should be preferred. Moreover, according to previous experiments, the CAR features are more robust to changes of illumination direction [16] and noise degradation [15].

Based on these experiments, we decided to benefit from both these textural representations and include them into our retrieval system. The final retrieval result is consequently composed of images with colour similarity, texture similarity according to CAR, and texture according to LBP.

4 Conclusions

We designed and implemented a tile retrieval system based on two orthogonal components of visual similarity: colours and texture. The performance of the textural component was successfully evaluated in a psychovisual experiment. Example results from our interactive demonstration are shown in Fig. 4.

Our retrieval system is not limited to tile images, it can be used with other kinds of images, where the structure is important property, e.g. textiles/cloths and wallpapers.

Acknowledgements. This research was supported by grants GAČR 102/08/0593 and partially by the MŠMT grants 1M0572 DAR, 2C06019.

References

1. Ahonen, T., Matas, J., He, C., Pietikäinen, M.: Rotation invariant image description with local binary pattern histogram fourier features. In: Salberg, A.B., Hardeberg, J.Y., Jenssen, R. (eds.) SCIA 2009. LNCS, vol. 5575, pp. 61–70. Springer, Heidelberg (2009)
2. Burghouts, G.J., Geusebroek, J.M.: Material-specific adaptation of color invariant features. *Pattern Recognition Letters* 30, 306–313 (2009)
3. Burghouts, G.J., Geusebroek, J.M.: Performance evaluation of local colour invariants. *Comput. Vision and Image Understanding* 113(1), 48–62 (2009)

4. Chen, Y., Wang, J.Z., Krovetz, R.: Clue: Cluster-based retrieval of images by unsupervised learning. *IEEE Trans. Image Process.* 14(8), 1187–1201 (2005)
5. Haindl, M., Šimberová, S.: A Multispectral Image Line Reconstruction Method. In: *Theory & Applications of Image Analysis*, pp. 306–315. World Scientific Publishing Co., Singapore (1992)
6. Li, W., Wang, C., Wang, Q., Chen, G.: A generic system for the classification of marble tiles using gabor filters. In: *ISCIS 2008*, pp. 1–6 (2008)
7. Liao, S., Law, M.W.K., Chung, A.C.S.: Dominant local binary patterns for texture classification. *IEEE Trans. Image Process.* 18(5), 1107–1118 (2009)
8. Ma, W.Y., Manjunath, B.S.: Texture features and learning similarity, pp. 425–430. *IEEE*, Los Alamitos (1996)
9. Monadjemi, A.: Towards efficient texture classification and abnormality detection. Ph.D. thesis, University of Bristol (2004)
10. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(7), 971–987 (2002)
11. Santini, S., Jain, R.: Similarity measures. *IEEE Trans. Pattern Anal. Mach. Intell.* 21(9), 871–883 (1999)
12. Shotton, J.D.J., Winn, J., Rother, C., Criminisi, A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. Comput. Vision* 81(1), 2–23 (2009)
13. Snoek, C.G.M., van de Sande, K.E.A., de Rooij, O., Huurnink, B., van Gemert, J., Uijlings, J.R.R., He, J., Li, X., Everts, I., Nedovic, V., van Liempt, M., van Balen, R., de Rijke, M., Geusebroek, J.M., Gevers, T., Worring, M., Smeulders, A.W.M., Koelma, D., Yan, F., Tahir, M.A., Mikolajczyk, K., Kittler, J.: The mediamill TRECVID 2008 semantic video search engine. In: Over, P., Awad, G., Rose, R.T., Fiscus, J.G., Kraaij, W., Smeaton, A.F. (eds.) *TRECVID*. National Institute of Standards and Technology, NIST (2008)
14. Stricker, M., Orengo, M.: Similarity of color images. In: *Storage and retrieval for Image and Video Databases III*, February 1995. *SPIE Proceeding Series*, vol. 2420, pp. 381–392. *SPIE*, Bellingham (1995)
15. Vacha, P., Haindl, M.: Image retrieval measures based on illumination invariant textural MRF features. In: Sebe, N., Worring, M. (eds.) *CIVR*, pp. 448–454. *ACM*, New York (2007)
16. Vacha, P., Haindl, M.: Illumination invariants based on markov random fields. In: *Proc. of the 19th International Conference on Pattern Recognition* (2008)
17. Vacha, P., Haindl, M.: Natural material recognition with illumination invariant textural features. In: *Proc. of the 20th International Conference on Pattern Recognition* (2010) (accepted)
18. Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. *Int. J. Comput. Vision* 62(1-2), 61–81 (2005)