



HAL
open science

PHOTOGRAPHIC PAPER TEXTURE CLASSIFICATION USING MODEL DEVIATION OF LOCAL VISUAL DESCRIPTORS

David Picard, Ngoc-Son Vu, Inbar Fijalkow

► **To cite this version:**

David Picard, Ngoc-Son Vu, Inbar Fijalkow. PHOTOGRAPHIC PAPER TEXTURE CLASSIFICATION USING MODEL DEVIATION OF LOCAL VISUAL DESCRIPTORS. IEEE Int. Conf. on Image Processing, Oct 2014, Paris, France, France. 5 p. hal-01063123

HAL Id: hal-01063123

<https://hal.science/hal-01063123v1>

Submitted on 11 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PHOTOGRAPHIC PAPER TEXTURE CLASSIFICATION USING MODEL DEVIATION OF LOCAL VISUAL DESCRIPTORS

David Picard, Ngoc-Son Vu, Inbar Fijalkow

ETIS, UMR 8051 / ENSEA, Université de Cergy-Pontoise, CNRS, F-95000, Cergy

ABSTRACT

This paper investigates the classification of photographic paper textures using visual descriptors. Such classification is called fine grain due to the very low inter-class variability. We propose a novel image representation for photographic paper texture categorization, relying on the incorporation of a powerful local descriptor into an efficient higher-order model deviation where texture is represented by computing statistics on the occurrences of specific local visual patterns. We perform an evaluation on two different challenging datasets of photographic paper textures and show such advanced methods indeed outperforms existing descriptors.

Index Terms— Image texture, Image classification, Image texture analysis

1. INTRODUCTION

Texture classification is an active research topic in the field of computer vision and image processing, playing a significant role in many applications such as medical image analysis, remote sensing, object recognition, document analysis, environment modeling, content-based image retrieval and many more. As the demand of such applications increases, texture classification has received considerable attention over the last decades and numerous novel methods have been proposed [1, 2, 3, 4, 5].

Recently, the orderless Bag-of-Words (BoW) approach [6], representing texture images statistically as histograms over a discrete texon dictionary, has proven successful in texture classification tasks. The strength of the BoW for texture classification comes from its two complementary components: the powerful *local* texture descriptors and *global* statistical histogram characterization. The former attempts to extract a set of robust and discriminative features from local patches to encode local patterns; while the latter relies on the fact that the texture image is a combination of those local patterns viewed as a random sampling process.

In this paper, we tackle the identification of photographic paper textures with papers having similar physical properties (same sheet, same package, same manufacturer, etc). In such context, we must consider features allowing to discriminate between small local differences in textures. To that end, we investigate the recent advances in BoW approaches, which we argue are able to capture such tiny local variations. More formally, we aim at incorporating a recent powerful local descriptor, Patterns of Oriented Edge Magnitudes (POEM) [7], into an efficient second order model deviation method, Vectors of Locally Aggregated Tensors (VLAT) [8], and then evaluate the gain of such approaches compared to global descriptors alone. By experiments carried out on two very challenging photographic paper texture datasets [9] with very low inter-class variability, we show significant improvements over using the descriptors alone.

The rest of the paper is organized as follows: in the next section we present the related work on texture descriptors and BoW approaches. We then detail in Section 3 the proposed combination of descriptors and signatures. In Section 4, we evaluate these approaches and analyze the results before we conclude.

2. RELATED WORK

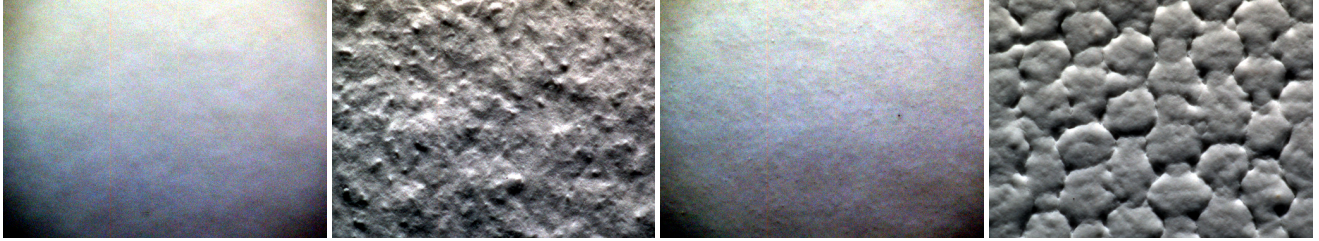
Most of earlier work on texture analysis focused on the development of filter banks and on characterizing the statistical distributions of their responses. Davis [10] exploited polarograms and generalized co-occurrence matrices to obtain rotation invariant statistical features. In [11], Duvernoy proposed Fourier based descriptors to extract texture feature on the frequency domain. In [12] Eichmann and Kasparis presented texture descriptors based on line structures extracted by the Hough transform. Although many efforts have been carried out along this direction, the supremacy of filter bank-based descriptors for texture analysis has been challenged by several authors [13, 14] who showed that it is possible to discriminate between textures using the intensities or differences of pixel within small scale neighborhoods. They demonstrated that despite the global structure of the textures, very good discrimination could be achieved by exploiting the distributions of such pixel neighborhoods. Two particularly important works along these lines are the VZ-Joint classifier [14] and the Local Binary Pattern (LBP) method [13]. The efficient texture descriptor LBP may be the preferable choice over the VZ-Joint classifier due to its simplicity.

Due to its computational efficiency and good discriminative property, the LBP descriptor [13] has gained considerable attention, and has already been used in many other applications, e.g., face recognition [15, 16]. Despite its great success, the conventional LBP operator comes with limitations, such as small spatial support regions, loss of local textural information, and sensitivities to rotation and noise. A lot of effort has been devoted to overcome these limitations, and many different LBP like descriptors have been presented such as the completed LBP model of Guo et al. [17] including both the magnitudes of local differences and the pixel intensity itself, the Dominant LBP in [18], the Local Ternary Patterns of Tan and Triggs [19], the extended LBP model of Liu et al. in [20], the circular symmetric LBP of Heikkila et al. [21], as well as the Patterns of Oriented Edge Magnitudes (POEM) features of Vu and Caplier [7].

More recently, several authors proposed a texture model based on visual dictionaries [6, 22, 14]. In these approaches, texture are considered to be sampled from a distribution of characteristic visual patterns or textons (named *visual words* in this context). This distribution is estimated by computing the histogram of occurrences of the visual words, hence the name Bag of Words. The similarity between textures is then measured with standard metric on histograms (e.g., χ^2 distance or Canberra distance). Robust and discriminative *local* texture descriptors and *global* statistical histogram characterization



(a) Images from the *bw* dataset.



(b) Images from the *inkjet* dataset.

Fig. 1: Images extracted from the texture datasets. The first row contains images from the *bw* dataset, while the second row contains images from the *inkjet* dataset. Each image is from a different category.

have supplied complementary components toward the BoW feature extraction of texture images.

In fact, the BoW approach is originally borrowed from the object recognition field [23], where local distinctive features are extracted from regions of interest, quantized in a dictionary (or *visual codebook*) of *visual words*. The histogram of occurrences of the visual words is then computed and used as a signature. Recently in object recognition, the BoW approach has been superseded by methods encoding more efficiently the statistical information in the set of local descriptors. In those approaches, the visual codebook is considered as universal model of the descriptors space, and the images are modeled as deviations from this model. Such methods include the popular Fisher Vectors [24] that achieve state of the art performances in object recognition challenges, the first order method [25] as well as the higher order deviations using a tensor framework [26].

3. PROPOSED METHOD

In this paper, we propose to combine the powerful local descriptor, Patterns of Oriented Edge Magnitudes (POEM) [7], with an efficient second order model deviation method, Vectors of Locally Aggregated Tensors (VLAT) [8] for texture classification. For the sake of completeness, this section details these two components of the proposed texture classification pipeline.

3.1. POEM feature extraction

The key idea of the POEM algorithm is to characterize object appearance by the relations between the distributions of local gradients of neighboring patches. The POEM feature extraction process consists of the three following steps:

(s1) Gradient computation and orientation quantization: let $\theta(p)$ and $m(p)$ denote the gradient orientation and magnitude at pixel p . The gradient orientation at each pixel is evenly discretized over $0-\pi$ (*unsigned*) or $0-2\pi$ (*signed*). A pixel feature is represented as a d -dimensional vector with *only one non-null element*

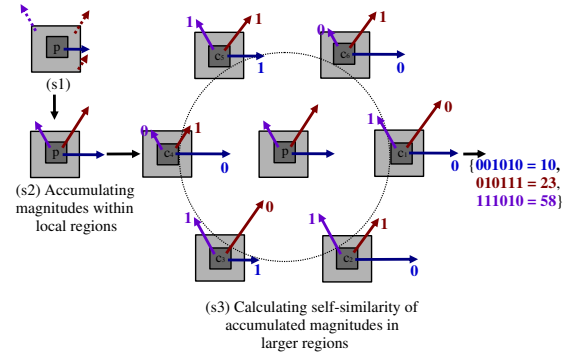


Fig. 3: POEM feature extraction for one pixel p . In (s3), the algorithm considers the length of vectors.

$[\hat{m}_1(p), \dots, \hat{m}_d(p)]$ where d is the number of discretized orientations, the i -th element of this vector takes the original magnitude ($\hat{m}_i(p) = m(p)$) if the discretized orientation of the current pixel belongs to this bin, and $\hat{m}_i(p) = 0$ otherwise (see Fig. 3 where only one non-null vector emits from pixel p).

(s2) Magnitude accumulation: to incorporate information from neighboring pixels (Fig. 3), a local histogram of orientations over all pixels within a local image patch (also called “cell”) is computed. As a result, a pixel conveys now richer information, i.e., the distribution of edge direction of a local image structure. At pixel p , the feature vector is $[\hat{m}_1(p), \dots, \hat{m}_d(p)]$ where $\tilde{m}_i(p) = \sum_{p_j \in C} \hat{m}_i(p_j)$ and C refers to cell centered on the considered pixel p .

(s3) Self-similarity calculation: the accumulated magnitudes are encoded using the self-similarity LBP-based operator within a more extended image patch (called “block”). The original LBP operator labels the pixels of an image by thresholding the neighborhood surrounding the pixel with the intensity value of central pixel, and considering the sequence of 8 result bits as a number. The similar procedure is applied on the accumulated magnitudes and across dif-

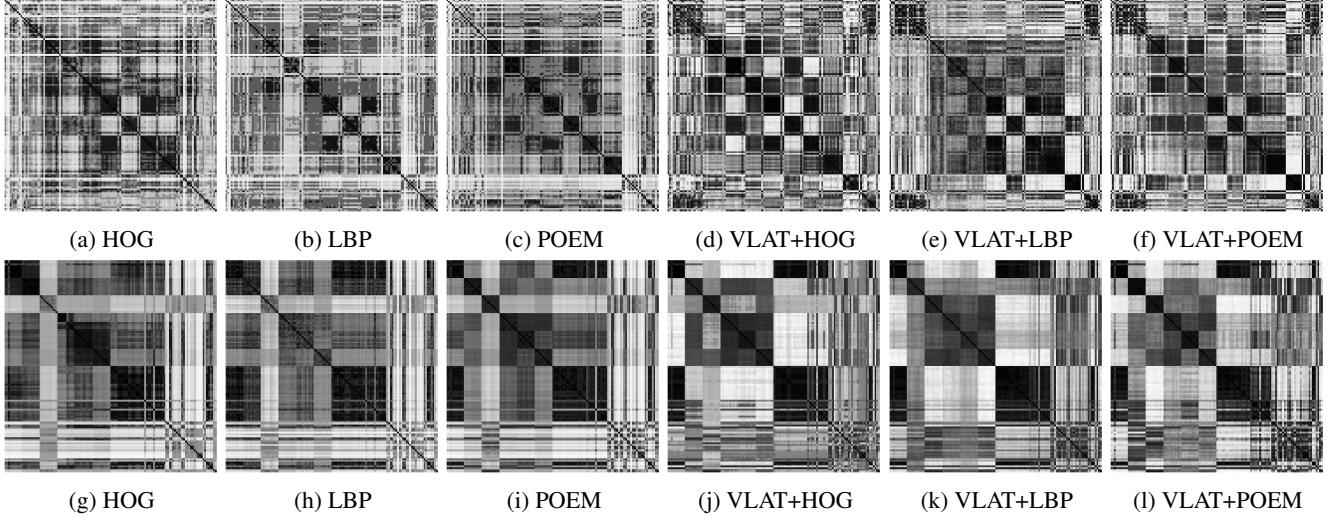


Fig. 2: Distance matrix for the different descriptors and signatures on both datasets: first row for the *bw* dataset, second row for the *inkjet* dataset. Dark colors indicate small distances, while lighter colors indicate greater distances.

ferent directions to build the POEM features. Formally, at pixel p , for the discretized direction i , the POEM feature is calculated as:

$$POEM_{L,w,n}^i(p) = \sum_{j=1}^n f(\tilde{m}_i(p) - \tilde{m}_i(c_j))2^j \quad (1)$$

where c_j are surrounding pixels; L, w refer to the size of blocks and cells, respectively; n is number of pixels surrounding the considered pixel p ; and f is defined as: $f(x) = 1$ if $x \geq \tau$, $f(x) = 0$ otherwise where the value τ is slightly larger than zero to provide some stability in uniform regions. Then, the POEM feature set at that pixel p is the concatenation of these d unidirectional POEM values: $POEM_{L,w,n}(p) = [POEM^1, \dots, POEM^d]$.

3.2. VLAT Signature

The POEM feature extraction is performed on a dense grid basis, using several scale of regions (usually in the order of 5% to 15% of the whole image). The descriptors \mathbf{d}_{r_i} extracted from an image i are stored in an unordered set $\mathcal{A}_i = \{\mathbf{d}_{r_i}\}$.

Using a large set of descriptors $\mathcal{A} = \{\mathbf{d}_{r_i}\}$ sampled from the entire texture collection, we compute a clustering of the descriptors space using the k-means algorithm. For each cluster k , we compute the first and second order moments using the partition \mathcal{C}_k of \mathcal{A} corresponding to the descriptors belonging to cluster k :

$$\mu_k = \frac{1}{|\mathcal{C}_k|} \sum_{\mathbf{d}_{r_i} \in \mathcal{C}_k} \mathbf{d}_{r_i} \quad (2)$$

$$\Sigma_k = \frac{1}{|\mathcal{C}_k|} \sum_{\mathbf{d}_{r_i} \in \mathcal{C}_k} (\mathbf{d}_{r_i} - \mu_k)(\mathbf{d}_{r_i} - \mu_k)^\top \quad (3)$$

μ_k is thus the mean of all descriptors belonging to cluster k and Σ_k is its covariance matrix. The set of (μ_k, Σ_k) for all k is a model of the whole descriptors space. The set of $\{\mu_k\}$ alone is akin to the classical visual words dictionary of the BoW approach.

For each texture i , we use the same clusters to compute the partitions \mathcal{C}_{k_i} of its set of descriptors \mathcal{A}_i . For each cluster, the VLAT

aggregation corresponds to the deviation between the texture second order moments centered on μ_k and the model covariance matrix Σ_k :

$$\tau_{k_i} = \sum_{\mathbf{d}_{k_{r_i}} \in \mathcal{C}_{k_i}} [(\mathbf{d}_{k_{r_i}} - \mu_k)(\mathbf{d}_{k_{r_i}} - \mu_k)^\top - \Sigma_k] \quad (4)$$

The VLAT signature is the flattening and concatenation of matrices τ_{k_i} for all k . We argue this second order deviation is well suited for fine grain identification, since not only the pattern prototypes are identified as it would in standard BoW approaches, but also the slight variations that might differentiate very closely related textures (*e.g.*, paper from the same manufacturer, but not from the same sheet).

4. EXPERIMENTS

We evaluate our pipeline on two different photographic paper texture datasets (*bw* and *inkjet*) [9] containing each 120 images of photographic paper sheets. Each dataset contains 9 categories consisting of ten images with similar but varying physical properties (same sheet, same package, same manufacturer, etc), and 30 random images not related to the other categories. Images from both the *bw* and *inkjet* datasets are presented in Figure 1. As we can see, it is very challenging for non-experts to distinguish the categories.

In our evaluation, we compared three different descriptors, namely HOG, LBP and POEM, as well as their aggregation using the VLAT signature model. For each texture, we compute the euclidean distance between its feature and the ones of every other texture. In Figure 2, we show the different distance matrices obtained for all the 6 possible combinations. The ideal distance matrix is block diagonal for the first 9 blocks of 10 rows/columns (each block corresponding to a category), and random for the last 30 rows/columns (corresponding to the distractors). As we can see, the *bw* set is far more difficult than the *inkjet*, although it is not obvious simply by looking to the images. Using the VLAT aggregation, we are able to capture more of the block structure expected in the ideal distance matrix.

To obtain more quantitative results, we computed the precision/recall curves associated with these distance matrices. Each

	HOG	LBP	POEM	VLAT+HOG	VLAT+LBP	VLAT+POEM
<i>inkjet</i>	63.8	45.7	66.3	69.7	67.8	73.6
<i>bw</i>	26.9	24.1	26.3	31.0	26.4	31.3

Table 1: mAP for all tested combinations. The first row corresponds to the *inkjet* dataset, while the second row is for the *bw* dataset

image acts as a query, and the precision/recall is measured by the rank of the other 9 images of the same category within the corresponding sorted row of the distance matrix.

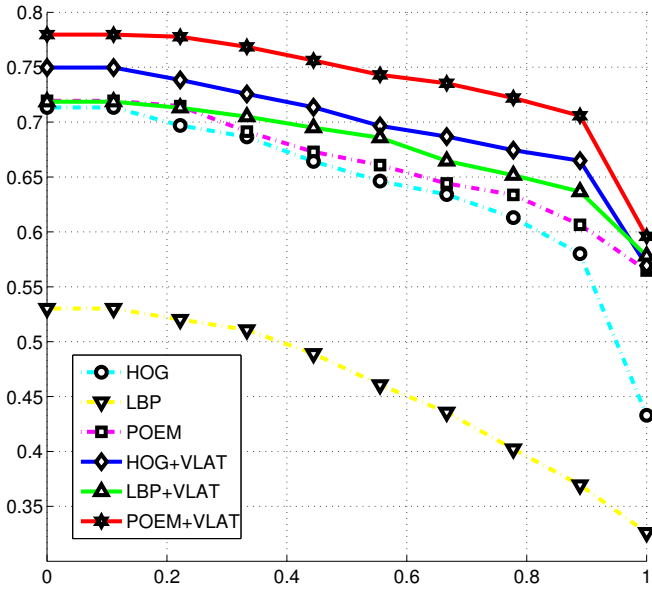


Fig. 4: Precision/Recall curves on the *inkjet* set.

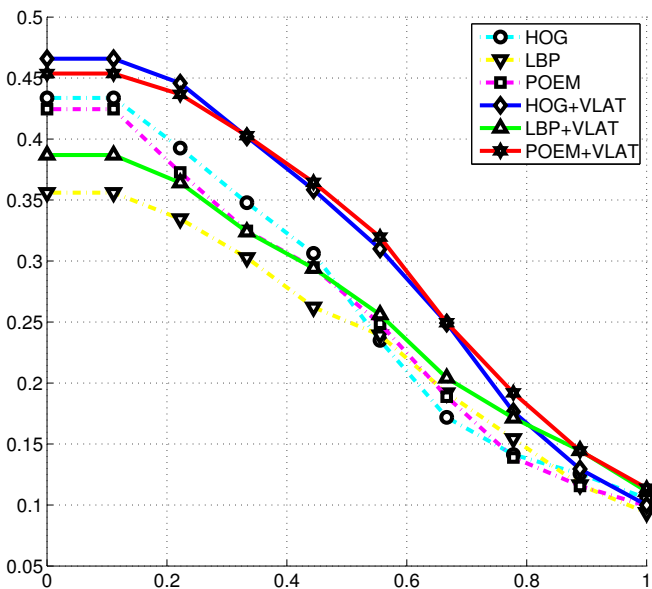


Fig. 5: Precision/Recall curves on the *BW* set.

Figure 4 shows these precision/recall curves for all possible combinations of descriptors on the *inkjet* dataset. As we can see,

single descriptors are systematically outperformed by their aggregated counterpart. The precision curve has a very slow decrease, which means that most categories are almost perfectly sorted. Figure 5 shows the precision/recall curves for the *bw* dataset. Contrarily to the *inkjet* dataset, the results are much lower, which is consistent with the observed distance matrices. The aggregation also outperforms related single descriptors in all tested combinations. It would be worth noting that on these photographic paper databases, the texture LBP descriptor surprisingly performs worst than HOG and POEM. This can be explained by the fact that these images are of very high resolution whereas LBP has too small spatial support regions. In other words, LBP cannot capture global enough information of those images.

To sum up the retrieval performances of each combination, Table 1 shows the mean Average Precision (mAp - area under the precision/recall curve). As we can see, the aggregation allows to boost the performances of the descriptors by a fair 3% to 6% of mAp, which is considerable. The VLAT+POEM combination always obtains the best performance on both datasets, which means that the model by a distribution of patterns of varying intensity is the best among those tested.

5. CONCLUSION

In this paper, we evaluated several descriptors and aggregation model in the case of fine grain texture classification. More specifically, we tackle the identification of photographic paper textures, where paper with the same physical properties have to be retrieved among a collection of many paper textures. Unlike common global approaches, we propose to model the texture as a distribution of specific visual pattern sampled from a dictionary. In that sense, we proposed to adapt efficient local descriptors with a recent model deviation based approach from the object classification community to our texture identification problem. We performed the comparison of these descriptors with their aggregated counterparts on two photographic paper texture datasets, and show that the aggregation method performs best.

6. REFERENCES

- [1] Oana G. Cula and Kristin J. Dana, "Compact representation of bidirectional texture functions," in *CVPR (1)*, 2001, pp. 1041–1047.
- [2] Haim Permuter, Joseph Francos, and Ian Jermyn, "A study of gaussian mixture models of color and texture features for image classification and segmentation," *Pattern Recognition*, vol. 39, no. 4, pp. 695–706, 2006.
- [3] Zhenhua Guo, Lei Zhang, and David Zhang, "A completed modeling of local binary pattern operator for texture classification," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1657–1663, 2010.
- [4] Jie Chen, Shiguang Shan, Chu He, Guoying Zhao, Matti Pietikainen, Xilin Chen, and Wen Gao, "Wld: A robust local

- image descriptor,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1705–1720, 2010.
- [5] Jaime Melendez, Domenec Puig, and Miguel Angel Garcia, “Multi-level pixel-based texture classification through efficient prototype selection via normalized cut,” *Pattern Recognition*, vol. 43, no. 12, pp. 4113–4123, 2010.
- [6] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce, “A sparse texture representation using local affine regions,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1265–1278, 2005.
- [7] Ngoc-Son Vu and Alice Caplier, “Enhanced patterns of oriented edge magnitudes for face recognition and image matching,” *Image Processing, IEEE Transactions on*, vol. 21, no. 3, pp. 1352–1365, 2012.
- [8] D. Picard and P.H. Gosselin, “Efficient image signatures and similarities using tensor products of local descriptors,” *Computer Vision and Image Understanding*, vol. 117, pp. 680–687, 2013.
- [9] P. Messier, R. Johnson, H. Wilhelm, W.A. Sethares, A.G. Klein, P. Abry, S. Jaffard, H. Wendt, S. Roux, N. Pustelnik, N. van Noord, L. van der Maaten, and E. Postma, “Automated surface texture classification of inkjet and photographic media,” in *IS&T International Conference on Digital Printing Technologies*, Seattle, USA, Sept. 2013.
- [10] Larry S. Davis, “Polarograms - a new tool for image texture analysis,” *Pattern Recognition*, , no. 3, pp. 219–223.
- [11] J. Duvernoy, “Optical digital processing of directional terrain textures invariant under translation, rotation, and change of scale,” *Appl. Opt.* 23, vol. 6, pp. 828—837, 1984.
- [12] G. Eichmann and T. Kasparis, “Topologically invariant texture descriptors,” *Comput. Vision Graphics Image Processing* 41, vol. 3, pp. 267—281, 1988.
- [13] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.
- [14] Manik Varma and Andrew Zisserman, “A statistical approach to material classification using image patch exemplars,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032–2047, 2009.
- [15] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *IEEE Trans. PAMI*, vol. 28, no. 12, pp. 2037–2041, dec. 2006.
- [16] Ngoc-Son Vu, “Exploring patterns of gradient orientations and magnitudes for face recognition,” *IEEE Trans. Information Forensics and Security*, vol. 8, no. 2, pp. 295–304, 2013.
- [17] Zhenhua Guo, Lei Zhang, and D. Zhang, “A completed modeling of local binary pattern operator for texture classification,” *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1657–1663, June 2010.
- [18] Shu Liao, Max WK Law, and Albert CS Chung, “Dominant local binary patterns for texture classification,” *Image Processing, IEEE Transactions on*, vol. 18, no. 5, pp. 1107–1118, 2009.
- [19] Xiaoyang Tan and Bill Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.
- [20] Li Liu, Lingjun Zhao, Yunli Long, Gangyao Kuang, and Paul Fieguth, “Extended local binary patterns for texture classification,” *Image Vision Comput.*, vol. 30, no. 2, pp. 86–99, Feb. 2012.
- [21] M. Heikkila, M. Pietikainen, and C. Schmid, “Description of interest regions with local binary patterns,” *Pattern Recognition*, vol. 42, pp. 452–436, 2009.
- [22] M. Varma and A. Zisserman, “A statistical approach to texture classification from single images,” *International Journal of Computer Vision*, vol. 62, no. 1–2, pp. 61–81, 2005.
- [23] J. Sivic and A. Zisserman, “Video google: A text retrieval approach to object matching in videos,” 2003, vol. 2, pp. 1470–1477.
- [24] Florent Perronnin, Jorge Sánchez, and Thomas Mensink, “Improving the fisher kernel for large-scale image classification,” in *ECCV*, 2010, pp. 143–156.
- [25] H. Jégou, M. Douze, C. Schmid, and P. Pérez, “Aggregating local descriptors into a compact image representation,” in *CVPR*, June 2010, pp. 3304–3311.
- [26] David Picard and Philippe-Henri Gosselin, “Improving image similarity with vectors of locally aggregated tensors,” in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 665–669.