# Fusing Vocabularies for Texture Categorisation

Fahad Shahbaz Khan, Joost van de Weijer, and Maria Vanrell

*Computer Vision Centre/Computer Science Department Building O, Campus UAB, 08193 Bellaterra (Barcelona), Spain*
*E-mail:fahad@cvc.uab.es*

**Abstract** Color and texture have been studied extensively in the field of computer vision and image processing. This paper discusses the extraction of knowledge from visual data for texture categorisation. The objective is to use color and texture attributes to classify the different categories. Two different fusion approaches, $early fusion$ and $late fusion$ at feature level have been investigated to combine both vocabularies (color and texture) in a flexible manner in order to achieve better performance. Support Vector Machines have been used for classification task and experiments have been conducted on a large dataset of ten different texture categories. In our case late fusion performs better than early fusion. Different combinations of color and texture vocabularies are then tested for late fusion. The results show that adding color information with an appropriate texture vocabulary during late fusion increases the overall performance significantlly.

*Keywords*: Color vocabulary, Texture Vocabulary, Texture Categorisation.

## 1 Introduction and Related Work

Images play a fundamental part in our daily communication and the large amount of pictures digitally available are not manageable by humans anymore [2]. Visual categorization is a difficult task, interesting in its own right, due to large variations between images belonging to the same class. Many features such as color, texture, shape, and motion have been used to describe visual information for visual categorization [3]. This paper focuses on the difficult problem of texture categorisation.

There has been a large amount of success in using ”bag of visual words” models for object and scene classification [9], [10], [11], [12], [13], [14],[4], and [7]. The first stage in the method involves selecting keypoints or regions followed by representation of these keypoints using local descriptors. The descriptors are then vector quantized into a fixed-size codebook. Finally the image is represented by a histogram of the code-book. The image classifier receives histogram representation as an input. The local features play the same role as played by words in traditional document analysis techniques [1], as they are local and have high discriminative power. In object detection tasks, the descritized features play the role of ”visual words” predictive of certain object class [4]. A classifier is then trained to recognize the categories based on these visual words. Thus image categorization means extracting features and finding the corresponding words and applying classifiers to the histogram representation of the image. Due to its success, we will use the ”bag of visual words” approach to texture categorisation problem.

Creating a visual vocabulary is a challenging task as some classes have very distinctive color, some have very characteristic texture patterns and some are characterized by combining both features. Our work is based on building a visual vocabulary that explicitly represents the various aspects (color and texture) to distinguish one class from another. To this end a dataset has been introduced which consists of 40 images of each class. Ten classes (marble, wood, beads, brick, foliage, graffiti, lace, clouds, fruit, and water) have been taken for experiments. A separate vocabulary is developed for color and texture. The vocabularies created are then used for texture categorisation. The texture vocabulary is based on LBP (Local Binary Patterns) and SIFT (Scale-Inavariant Feature Transform) whereas for color three different options have been considered namely RGB color vocabulary, Hue vocabulary and Color Naming values based on the work of [15]. Two different fusion approaches at feature level have been investigated to combine both vocabularies (color and texture) in a flexible manner in order to achieve better performance. The first approach, called $early fusion$, involves fusing local descriptors together and creating one representation of joint texture-color vocabulary. The second approach, called $late fusion$, aims at concatenating histogram representation of both color and texture, obtained independently. Different combinations of color and texture vocabularies are tested for better performance. Machine learning methods are often used to tackle the problem of detection and classification of objects in high level categories. In our approach SVM has been used for classification since it is known to produce state-of-the-art results in high dimensional problems.

To this end, the report has been organized as follows. In section 2 Vocabularies for texture and color are discussed. Afterwards, in section 3 a combined vocabulary is proposed. Section 4 presents the experimental details like the classification algorithm, the dataset used and the classification settings. Detailed experiments are shown in section 5. Finally, we sum up the conclusions.

## 2 Vocabulary for Texture and Color

Visual features color and texture are used to characterise visual keywords. In case of late fusion, texture and color vocabularies are developed separately and then combined later on for categorisation. For early fusion, vocabularies based on color and texture features are fused together to create one representation of joint texture-color vocabulary for categorisation. In our approach LBP and SIFT are used to create a texture vocabulary. Three options are considered to create a color vocabulary namely RGB, Hue and Color Naming values. The main essence of our work lies in the combination of both vocabularies. In the next sections both texture and color vocabulries will be discussed in detail.

### 2.1 Texture Vocabulary

For human perception texture is an important visual category. Texture is one of the most common low level features and plays an important role for the character of region for digital images. There are many different ways of solving the problem of texture analysis. In this regard we investigate the use of LBP and SIFT for creating texture vocabulary since they are known to yield very good good performance in recent texture studies [6], [5], [4], and [7].

The LBP operator can be regarded as a unifying approach to the structural and statistical models of texture analysis. LBP detects local patterns in an image. In LBP a binary pattern is obtained for each pixel in an image by thresholding a certain neighborhood with the center pixel. A histogram is then build out of these binary patterns. The most important properties of LBP operator are its tolerance to illumination and its computational simplicity. The operator is usually applied to gray-scale images and the derivatives of intensities.

There are several variations with LBP operators such as the $rotation\ invariant$ LBP operator $\left(LBP_{P,R}^{ri}\right)$, Uniform patterns LBP $\left(LBP_{P,R}^{u2}\right)$, the $joint\ distribution$ of rotation invariant LBP operator with its local variance $\left(LBP_{P,R}^{riu2}/VAR_{P,R}\right)$. For our experiments we have used all these variations while creating a texture vocabulary.

Scale-invariant feature transform (SIFT) is used in computer vision to detect and describe local features. The SIFT algorithm was published by David Lowe in [8]. The idea behind the algorithm was to find stable image features that can be used for object recognition. A SIFT vocabulary is constructed by applying the K-means algorithm to the set of local descriptors extracted from the images. Euclidean distance is used in clustering. The vocabulary size will be later optimized on the performance score.

### 2.2 Color Vocabulary

A color vocabulary is created to represent the color aspects of an image. The measured color values vary significantly due to large amount of variations. In this work color histogram approach is used in the Hue, Saturation, Value (HSV) color space, Red, Green, Blue (RGB) color space and Color Naming values mentioned in the work of [15].

For RGB vocabulary, RGB values for each pixel in an image are clustered using k-means clustering. The number of clusters is optimized on the dataset. In case of Hue vocabulary, the Hue descriptor proposed by [16] is used to compute the Hue and Saturation at each position. The work in [16] is principled approach to extend the SIFT shape descriptor with a color descriptor. The third method used to create a color vocabulary is the color naming values based on the work of [15]. Color naming involves the assignment of linguistic color labels to image pixels. [17] proposed to learn color names form real-world images and used PLSA model for learning. The 11 colors names used are black, blue, brown, grey, green, orange, pink, purple, red, white and yellow. For clustering K-means method is used. The soft assignment of colors have been used to make the distribution more uniform which we found to improve the results.

# 3 Combined Vocabulary

After creating the color and texture vocabulary, both vocabularies are then combined in a flexible manner to achieve better performance. The discriminative power of each vocabulary varies for different classes. Some classes are distinguished by color and some by texture. The two texture vocaularies created from LBP and SIFT features are combined with the three color vocabularies namely RGB, HUE and Color Naming values. Two techniques, early fusion and late fusion have been used to combine the color and texture vocabularies. As a next step both early and late fusion have been analysed to decide which one performs better in our case.

## 3.1 Early Fusion and Late Fusion

In early fusion, the local features of color and texture are combined before quantization. The fusion involves concatenating the texture features and color features and as a result of this concatenation a joint vocabulary of both features is obtained using euclidean distance in the K-means algorithm.

In late fusion the two features color and texture are computed independently. The two features are then fused together in one representation. Here the different vocabularies are concatenated after quantization. A weight vector $\alpha$ is introduced to obtain a combined histogram $n(w|I)$ of color and texture vocabularies for an image $I$.

$$n(w|I) = \begin{bmatrix} \alpha \; n(w_{color}|I) \\ (1-\alpha) \; n(w_{texture}|I) \end{bmatrix} \quad (1)$$

where $w$ is the number of total vocabulary words, $w_{color}$ are Color words and $w_{texture}$ are texture words. The weight vector $\alpha$ is learned through cross-validation on training data. A 4 fold cross validation is used where the dataset is divided into training, validation and test set.

The use of weight vector is neccessary to allow the weighting of different feature representation in the SVM classifier. Different combinations are tried in order to achieve better performance.

# 4 Experimental Setup

The performance of combined vocabularies will be tested on the classification task using SVM. Details of the proposed procedures are outlined in this section.

## 4.1 Feature Detection

Out of the five descriptors discussed so far, we used dense detection in LBP, RGB, HUE and Color Naming values. Dense detection means that the descriptor is computed for every pixel in the image. This is unattainable in case of SIFT due to its computational cost. For SIFT, we use grid detection where the extracted patches have radius 10 and are located 10 pixels apart.

## 4.2 Dataset

The approach outlined above is tested on a dataset with 10 classes (Marble, Wood, Beads, Foliage, Graffiti, Lace, Clouds, Fruit, and Water) with 40 images for each class. The images in the dataset have been collected from Google, Flickr, and Corel image collection. Figure 1 shows some of the images from the dataset. The dataset is very challenging due to wide range of textures and color in it. For example, the Foliage class in our dataset has mostly green color, however there are few images in this class that have red color in it. Similarly there is a wide range

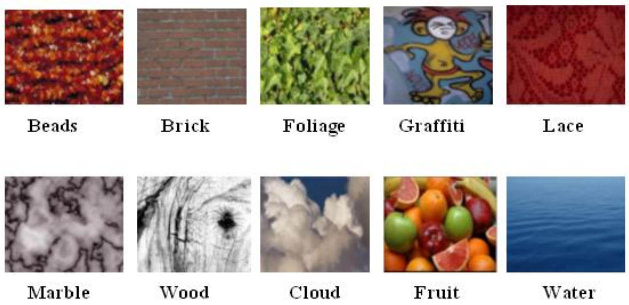

Figure 1: Typical Examples of Each Class from the dataset.

of different texture patterns and colors in Marble and Graffiti class. Furthermore, the lace class contains a lot of variations in scale.

### 4.3 Classification Settings

The dataset has been divided into train set, validation set and test set. A 4 fold cross validation has been performed and as a result the train set contains 225 images whereas validation set comprises of 75 images. The test set contains 100 images. In our experiments multiclass SVM is used for classification. To evaluate the classification performance we use the classification score. The classification score gives the percentage of correctly classified instances in the testset. To robustify the results we repeat the procedure a hundred times for each of the experiments and report the averaged classification score.

## 5 Experiments

This section explains in detail the creation of color and texture vocabularies and the proposed methodology used for combining these vocabularies. In experiment 1 early fusion will be compared with late fusion. Experiment 2 is about optmizing the individual vocabularies of texture. Experiment 3 provides an insight of three color vocabularies. Experiment 4 deals with combining both texture and color vocabularies in order to optimize the classification performance.

### 5.1 Experiment 1: Early Fusion versus Late Fusion

The first experiment is about the use of late fusion versus early fusion. In order to decide we used one texture and one color vocabulary. More precisely SIFT and RGB vocabularies have been used to compare the performance of these two methodologies. In Table 1 the results of these experiments are summerised. The results using late fusion show a better classification score as compared to early fusion. Therefore, from this point on only late fusion will be considered to combine color and textue vocabularies.

| Vocabulary | Classification Score |
|---|---|
| $SIFT$ | 60.25 |
| $RGB$ | 58.75 |
| $SIFT/RGB(EarlyFusion)$ | 62.73 |
| $SIFT/RGB(LateFusion)$ | 67.69 |

Table 1: Classification Score (percentage) using SIFT and RGB during Early Fusion.

### 5.2 Experiment 2: Texture Vocabularies

This section provides detailed results obtained using only the texture information, while ignoring color information. As a first step a texture vocabulary has been created using local binary patterns. There are different variations related to LBP such as Rotation Invariant LBP $\left(LBP^{ri}_{P,R}\right)$, Uniform LBP $\left(LBP^{u2}_{P,R}\right)$, Rotation Invariant with Uniform LBP $\left(LBP^{riu2}_{P,R}\right)$, Rotation Invariant Variance LBP $(VAR_{P,R})$ , and Joint distribution of Rotation Invariant Uniform LBP with its Variance $\left(LBP^{riu2}_{P,R}/VAR_{P,R}\right)$. As shown in Table 2, the uniform LBP gives the better performance. From the table it is also obvious that rotation invariance with uniform patterns, $LBP^{riu2}_{P,R}$, does not help to improve the over all performance. In fact it is the joint distribution of $LBP^{riu2}_{P,R}$ operator with the $VAR_{P,R}$ operator that improves the performance as compared to $VAR_{P,R}$ and $LBP^{ri}_{P,R}$ alone.

| LBP operator | P,R | Bins | Classification Score |
|---|---|---|---|
| $LBP^{ri}_{P,R}$ | $8,1$ | 36 | 54.16 |
| $LBP^{u2}_{P,R}$ | $16,2$ | 243 | 62.30 |
| $LBP^{riu2}_{P,R}$ | $8,1$ | 10 | 47.27 |
| $LBP^{riu2}_{P,R}/VAR_{P,R}$ | $16,2/8,1$ | 328 | 58.10 |
| $LBP^{ri}_{P,R}+LBP^{u2}_{P,R}$ | $8,1+16,2$ | 279 | 64.27 |

Table 2: Classification Score (percentage) using LBP.

Moreover another texture vocabulary has been created based on SIFT features. SIFT descriptor describes local features detected in the images as a 128 dimension vector. A visual vocabulary is then learned from these descriptors using the k-means algorithm. The results have also been calculated using angle invariant SIFT descriptor but there has been no gain in performance. Table 3 shows the classification scores based on SIFT with different vocabulary size. The optimal performance is achieved when the size of the vocabulary is 600. After this, the performance again drops as shown in the case of vocabulary with the size of 800.

| Vocabulary Size | Classification Score |
|---|---|
| 200 | 55.30 |
| 400 | 58.17 |
| 600 | 60.25 |
| 800 | 59.13 |

Table 3: Classification Score (percentage) using SIFT.

### 5.3 Experiment 3: Color Vocabularies

The different options considered for developing a color vocabulary are RGB, Hue, and Color Naming values. Table 4 shows the classification scores based on these 3 color vocabularies. The table shows that the classification score obtained using RGB vocabulary is better than the other two vocabularies. But there is no significant differences in the classification score of all three vocabularies.

| Vocabulary | Vocabulary Size | Classification Score |
|---|---|---|
| $RGB$ | 50 | 58.75 |
| $HUE$ | 50 | 58.14 |
| $CN$ | 11 | 56.18 |

Table 4: Classification Score (percentage) using Color Vocabularies.

### 5.4 Experiment 4: Combined Texture and Color Vocabularies

The texture and color vocabularies are now combined into one vocabulary by concatenating the histogram representation (of color and texture), obtained independently from each local feature. In the first step the texture vocabulary based on SIFT is combined with 3 color vocabularies. Table 5 shows the classification results obtained by combining SIFT with 3 color vocabularies. Table 6 shows the classification results obtained by combining LBP with 3 color vocabularies. Table 7 shows the classification results obtained by combining SIFT and LBP together with 3 color vocabularies. The final class-wise score of best vocabulary of texture alone, best vocabulary of color alone and best combined vocabulary (LBP, SIFT and color) is shown in Figure 2.

| Vocabulary | Voc Size | Classification Score |
|---|---|---|
| $SIFTandRGB$ | 650 | 71.45 |
| $SIFTandHUE$ | 650 | 70.34 |
| $SIFTandCN$ | 611 | 71.13 |

Table 5: Classification Score (percentage) by combining SIFT with Color Vocabularies.

| Vocabulary | Voc Size | Classification Score |
|---|---|---|
| $LBPandRGB$ | 110 | 67.86 |
| $LBPandHUE$ | 110 | 66.54 |
| $LBPandCN$ | 71 | 65.88 |

Table 6: Classification Score (percentage) by combining LBP with Color Vocabularies.

| Vocabulary | Voc Size | Classification Score |
|---|---|---|
| $SIFTandLBPandRGB$ | 710 | 76.45 |
| $SIFTandLBPandHUE$ | 710 | 74.51 |
| $SIFTandLBPandCN$ | 671 | 70.39 |

Table 7: Classification Score (percentage) by combining LBP and SIFT together with Color Vocabularies.

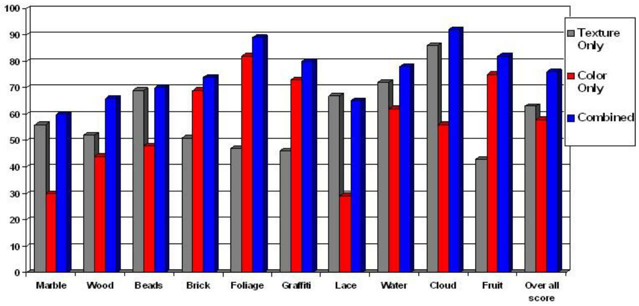

Figure 2: Classification score of each class using the best Texture, Color and Combined vocabularies. Note that the performance of combined vocabulary is significantlly better than the performance of individual texture and color vocabularies.

Here it is clear in Figure 2 that the combined vocabulary provides the best classification performance. There is only one class, lace, which shows a slight decrease in performance in case of combined vocabulary. The class performs best with texture alone. Thus it seems that the introduction of color for this class confuses the classifier and hence the classification score decreases. Except this class all other classes enjoy the addition of color and the over all performance of all nine classes increases in case of combined vocabulary.

### 5.5 Conclusions

The results shows that late fusion is better than early fusion. This could be due to the fact that in most of the categories only one of the cues, either color or texture, is constant. For example, foliage class is difficult to classify based on texture but relatively stable with respect to color appearance. In this case it is hard to find visual words based on early fusion that are consistent. As a next step late fusion has been analysed deeply by trying different combinations of both color and texture vocabularies.

Both cues, color and texture are found to be crucial to obtain a good overall classification score. Texture alone obtained $64.27\%$, color alone $58.75\%$, and the combination got $71.45\%$.When combining the vocabularies, color improves texture classification performance but best performance is achieved when texture has more influence than color. The use of cross-validation to optimize parameter settings is crucial to obtain good final performance scores. The paper contributes to texture evaluation by proposing a new dataset. The dataset contains ten classes with a wide variety in texture and color.

To our surprise SIFT and LBP together with color provided an improved performance. Adding the two texture cues gave an classification score improvement of around $5\%$. Thus this shows that the combined vocabulary of color and texture outperforms the performance of individual vocabularies. The evidence presented in this study also suggests that using a combination of color and texture attributes is powerful way of utilizing their complementary information.

Furthermore, it would be interesting to fuse the color and texture representations on the classifier level. In that case a separate classifier could be used for both features.

## References

[1] R. Baeza-Yates, B. Ribeiro-Neto, *Modern Information Retrieval*, ACM Press, 1999.

[2] Y. Chen, J. Z. Wang, "Region-based fuzzy feature matching approach to content-based image retrieval", *IEEE Trans. on Pattern Analysis and Machine Intelligence* 24(9):1252-1267, 2002.

[3] V. N. Gudivada, V. V. Raghavan, "Content based image retrieval systems", *IEEE Computer* 28(9):18-22, 1995.

[4] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object catergories: An in-depth study. A Comprehensive Study", *International Journal of Computer Vision*, 73(2): 213-238, 2007.

[5] Topi Maenpaa, Matti Pietikainen, "Classification with color and texture: jointly or separately?", *Pattern Recognition*, 37(8): 1629-1640, 2004.

[6] T. Ojala, M. Pietikinen, and T. Menp, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:971-987, 2002.

[7] S. Lazebnik, C. Schmid, and J. Ponce, "A Sparse Texture Representation Using Local Affine Regions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

[8] D. G. Lowe, "Distinctive image features from scale-invariant points", *International Journal of Computer Vision*, 60(2):91-110, 2004.

[9] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints", *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1-22, Prague, Czech Republic, 2004.

[10] A. Bosch, A. Zisserman, and J.Munoz, "Scene classification via plsa", *In Proc. ECCV*, 2006.

[11] G. Dorko, C. Schmid, "Selection of scale-invariant parts for object class recognition", *In Proc ICCV*, 2003.

[12] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories", *In Proc. CVPR*, 2005.

[13] P. Quelhas, F. Monay, J. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. Van Gool, "Modelling scenes with local descriptors and latent aspects", *In Proc. ICCV*, Beijing, China, 2005.

[14] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W.Freeman, "Discovering object categories in image collections", *In Proc. ICCV, Beijing*, China, 2005.

[15] J. van de Weijer, C. Schmid, and J.J. Verbeek, "Learning color names from real-world images", *In Proc. CVPR*, Minneapolis, Minnesota, USA, 2007.

[16] J. van de Weijer, C. Schmid, "Coloring local feature extraction", *In Proc. of the European Conference on Computer Vision*, volume 2, pages 334-348, Graz, Austria, 2006.

[17] J. van de Weijer, C. Schmid, "Applying color names to image description", *In International Conference on Image Processing*, 2007.