Fitting Gaussian Copulae for Efficient Visual Codebooks Generation

Miriam Redi
EURECOM, Sophia Antipolis
2229 route des crêtes
Sophia-Antipolis
redi@eurecom.fr

Bernard Merialdo
EURECOM, Sophia Antipolis
2229 route des crêtes
Sophia-Antipolis
merialdo@eurecom.fr

Abstract

The Bag of Words model is probably one of the most effective ways to represent images based on the aggregation of locally extracted descriptors. It uses clustering techniques to build visual dictionaries that map each image into a fixed length signature. Despite its effectiveness, one major drawback of this model is the codebook informativeness and its computational complexity. In this paper we propose Copula-BoW (C-BoW), namely an efficient local feature aggregator inspired by the Copula theory. In C-BoW, we build in a quadratic time an efficient codebook for vector quantization, based on the correlation of the marginal distributions of the local features. Our experimental results prove that the C-BoW signature is much more efficient and as discriminative as traditional BoW for scene recognition and video retrieval (TRECVID [14] data). Moreover, we also show that our new model provides complementary information when combined to existing local features aggregators, substantially improving the final retrieval performance.

1 Introduction

Image signatures, namely low-dimensional image representations, are of crucial importance for the development of discriminative Content-Based Multimedia Retrieval (CBMR) systems. Among the existing image signatures, image representations based on the aggregation of locally extracted descriptors (LEDs) have proved to be very effective for CBMR. LED-based signatures can be performed using two opposite approaches: the Bag of Words models (BoW) and Marginal Alphabets Aggregator(MEDA).
The BoW model [1] is probably one of the most effective LED aggregators. First, local descriptors (such as SIFT [8]) of length k are computed around a set of image salient [4] points or a dense grid. A "visual codebook" that quantizes the k-dimensional LED space in a set of "visual words" is then computed. Each image is then mapped into a fixed length signature by collecting the occurrences of such visual words. In order to minimize the quantization error, the words of the visual dictionary need to follow the joint distribution of the LEDs: the majority of the visual words should lie in the more densely populated regions of the LEDs space, while few visual words should lie in the sparse areas. To define the more appropriate visual words, the general approach is to perform a search in the k-dimensional space: this is achieved by clustering the LEDs using traditional k-means [1], mean-shift [7], hierarchical clustering [10], or by quantizing the LEDs space using a fixed lattice [15]. Despite its high discriminative ability, one of the major drawback of BoW model is its ability to estimate the density of the LEDs in the feature space. Moreover, the operations in the k-dimensional space determine a high computational cost, that is *polynomial* with the number of dimensions of the LED and the number of words produced.

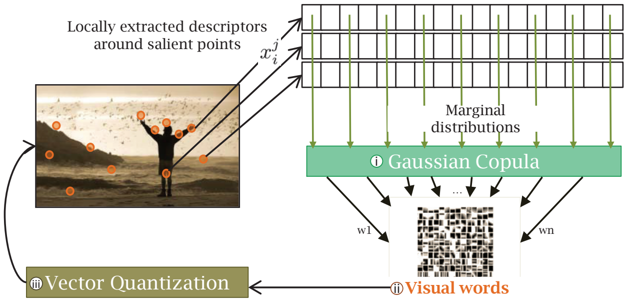


Figure 1. C-BOW: codebook of $k-$d words is estimated from the marginals of the LEDs

The MEDA model was proposed in [12] as an alternative approach to BoW for LED aggregation. In MEDA, the compact image representation is based on a visual codebook ("alphabet") of 1-dimensional "letters", generated by quantizing into n bins each dimension of the LED based on its marginal distributions. The MEDA signature is then computed by collecting in a $k \times n$ histogram the letter frequencies for each component of the LED. Since MEDA performs vector quantization in a 1-d space, such alphabet-based approach is very efficient $(O(nk))$. However, one major issue of the MEDA model is that the marginal-based quantization eliminates the relations between the LED components, resulting in a loss of precious discriminative information for image representation.

In this paper we present Copula-BoW (C-BoW), a LED aggregator that stands in an intermediate and complementary point between the two mentioned approaches. In C-BoW, we exploit the marginal distributions (as in MEDA) to perform vector quantization in the k-dimensional space (as in BoW), taking into account the relations between LED dimensions, but without introducing exponential complexity. In our approach, the codebook is built through a fast and efficient method inspired by the Copula theory [9] and based on the correlation between the marginal distributions of the LED components. As a result, the joint distribution of the LED components, and the corresponding codebook, is modeled in a quadratic time $O(k^2)$, without involving any clustering or operation in the high dimensional space, leading to an image signature that is much more efficient and as discriminative as traditional BoW.

How is such Copula-based codebook computed? Copulae are statistical tools for linking the marginals of the variables in a random vector with their multivariate joint distribution. They first appeared in [13] for probabilistic metric spaces, and they then became popular in finance and actuarial sciences. The main advantage of Copulae is their efficiency: since they model separately marginal distributions and their dependence structure, they allow to estimate a joint probability in a quadratic time, overcoming many computational problems of traditional multivariate modeling. As a matter of fact, Copulae have been effectively used in literature for clustering ([3, 2]), and for vector quantization in image coding [5]. Surprisingly however, the use of Copula has not been explored for LED-based Image representation.

In our approach, we use a particular type of Copula, namely the Gaussian copula, to estimate the LED joint distribution and compute a codebook accordingly. The proposed approach can be summarized as follows (see Fig. 1):

(i) First, we fit a Gaussian Copula given the marginals of the LED vectors extracted from the training set. Such Copula represents the multivariate probability density function of the LED components, and it can be computed in a quadratic time $(O(k^2))$.

(ii) We now want to compute a codebook whose visual words are distributed according to the density in the LED space. In order to achieve this representation, we fill the codebook with random samples drawn from the copula-based distribution[1].

(iii) Finally the C-BoW representation is computed as a w-dimensional histogram collecting the occurrences of the copula-based words.

An important aspect of C-BoW is that, by using Gaussian Copulae, we model the joint LED distribution as a multivariate normal, assuming a Gaussian space. While the codewords in C-BoW follow a Gaussian-like distribution, the codewords in BoW reflect the real LED joint distribution. The corresponding BoW and C-BoW models represent therefore complementary sources of information regarding the distribution of the image LEDs, and their combination leads to an improvement of the retrieval performance. Moreover, both BoW and C-BoW models are in turn different from the model generated by the MEDA signatures, based on marginal distributions only. Therefore, by introducing C-BoW as a LED aggregator, we do not only provide a fast way to generate signatures based on local features, but we also add a complementary view over the LED space, that can be combined with existing approaches to increase the discriminative ability of a CBMR system.

The remainder of this paper is organized as follows: first, we show the basic notions of the Copula Theory in Section 2; we will then give a brief summarization of the BoW traditional model for image representation (Section 3), and see in Section 4 how this method can be efficiently improved through the introduction of Copulae. Finally, in Sec. 5 we test the effectiveness of C-BoW on two different datasets for indoor scene recognition [11] and video retrieval (TRECVID data [14]), and show its discriminative ability both as a stand-alone descriptor and in combination with other LED aggregators.

2 Copula theory: Estimating Multivariate Distributions from Marginals

Copulae are statistical models to couple marginal distributions to joint distributions. According to this theory [13], the joint distribution of the variables in a random vector X of length k can be decomposed into k marginal distributions and a Copula function. The Copula function is designed to define the dependencies between the marginals. Unlike traditional multivariate analysis that combines joint distribution and variable dependencies, Copulae allow to study separately the marginal distributions and their dependencies.

In order to introduce the Copula theory in a simple way, we will show the bivariate case ($k = 2$), that is easily extendable to the multivariate scenario. We define X_1 and X_2 as the two variables from the random vector X, and $u = F_1(x_1) = [P(X_1 \leq x_1)]$ and $v = F_2(x_2) = [P(X_2 \leq x_2)]$ as the their respective marginal cumulative distribution functions (CDFs), normalized over the range [0,1]. The joint distribution of X_1 and X_2 is given by $F(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2]$.

Following Sklar's theorem [13], a Copula C is a unique

[1] the resulting k-d samples are indeed be more concentrated in the densely populated areas of the feature space, and less numerous in its sparse area

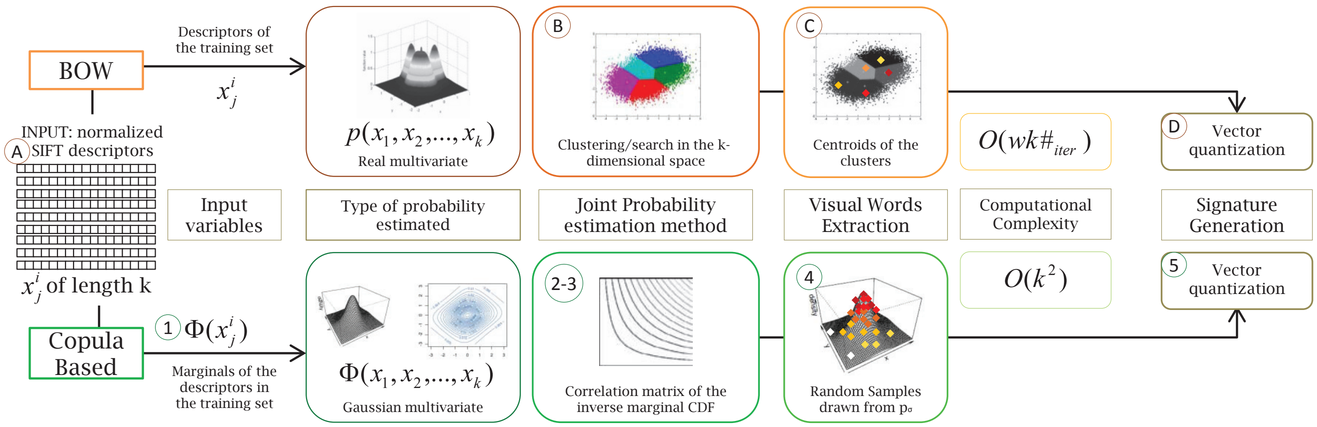

BOW
Descriptors of the training set
x_j^i
INPUT: normalized SIFT descriptors
A
x_j^i of length k
Copula Based
1
$\Phi(x_j^i)$
Marginals of the descriptors in the training set
Input variables
$p(x_1, x_2, ..., x_k)$
Real multivariate
Type of probability estimated
$\Phi(x_1, x_2, ..., x_k)$
Gaussian multivariate
B
Clustering/search in the k-dimensional space
Joint Probability estimation method
2-3
Correlation matrix of the inverse marginal CDF
C
Centroids of the clusters
Visual Words Extraction
4
Random Samples drawn from pₒ
$O(wk\#_{iter})$
Computational Complexity
$O(k^2)$
D
Vector quantization
Signature Generation
5
Vector quantization

Figure 2. Differences between Bag of words and C-BoW: while the former needs polynomial time $O(nw\#_{iter})$ for codebook construction, our method uses Copulae, with quadratic complexity, to model the LED space using a multivariate Gaussian distribution.

mapping[2] that assigns the values of the joint distribution function of two variables given each ordered pair of values of their marginal distributions, namely

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) = C(u, v). \qquad (1)$$

Therefore, if we know the mapping C, the joint distribution $F(x_1, x_2)$ can be inferred from the marginal CDFs u and v. One important type of Copulae is the Gaussian Copula, very popular for finance-related, civil engineering and medical computations. A Gaussian Copula C_Σ is defined for the two-dimensional random vector X if:
(I) The marginals of the variables in X, namely u and v follow a Gaussian distribution.
(II) The joint distribution of the variables in X, namely C_Σ, is a multivariate Gaussian.
In the formulation of the Gaussian Copula, u and v are assumed to be continuous functions, therefore Eq. (1) can be expressed as:

$$F(x_1, x_2) = F(F_1^{-1}(u), F_2^{-1}(v)) = C(u, v). \qquad (2)$$

The resulting bivariate distribution is defined as:

$$C_\Sigma(u, v) = \Phi_\Sigma(\Phi^{-1}(u), \Phi^{-1}(v)), \qquad (3)$$

being Φ_Σ the bivariate standard with covariance Σ and mean zero, and Φ^{-1} the inverse standard univariate normal.The Gaussian Copula is therefore defined as:

$$C_\Sigma = \frac{1}{\sqrt{det(\Sigma)}} exp\left(-\frac{1}{2} \cdot \begin{pmatrix} \Phi^{-1}(u) \\ \Phi^{-1}(v) \end{pmatrix}^T \Sigma^{-1} \mathbf{I} \begin{pmatrix} \Phi^{-1}(u) \\ \Phi^{-1}(v) \end{pmatrix} \right) \qquad (4)$$

3 The Bag of Words Model

In this Section we will outline the processing step involved in one of the most popular models for local visual descriptor aggregation, the Bag of Visual Words model (see Fig. 2 for a visual explanation).

(A) For an image I, first, m salient points are detected in the image. The surrounding region of each point is then described by a k-dimensional normalized vector x such as SIFT [8], obtaining the vectors in the set $x^i = (x_1^i, \ldots, x_k^i), i = 1, \ldots, m$.

(B) In order to map each image into a fixed length-signature, vector quantization needs to be performed in the k-dimensional space defined by the LED. The general approach is to compute a shared visual codebook to support the generation of a compact image representation by clustering [10, 1] the LEDs of the training set into w groups. As said, since information loss needs to be minimized during this mapping, the codebook needs to properly reflect the joint distribution of the LED components $p(x) = p(x_1, x_2, \ldots, x_k)$.

(C) The w centroids of the defined clusters are then taken as visual codewords.

(D) Finally, the image is represented by a w-dimensional histogram collecting the number of image LEDs that can

[2] In order to be defined as a two-dimensional Copula, C needs to fulfill the following requirements (see [9]):

- It is defined over the interval $[0, 1]$
- $\forall t \in [0, 1]$, then $C(t, 0) = C(0, t) = 0$ and $C(t, 1) = C(1, t) = 1$
- $\forall u_1, u_2, v_1, v_2 \in [0, 1]$, with $u_1 \leq u_2$ and $v_1 \leq v_2$, $C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0$

be approximated by each visual word.

In general, at the end of this process, the BoW signature is then used as input for traditional classifiers such as Support Vector Machines, that builds models able to predict the presence of a concept for a given visual input.

4 Using Copulae to Create Visual Words: C-BoW

As mentioned, one of the major drawbacks of the BoW model is its associated complexity, due to the clustering in the high dimensional space that is essential to approximate the joint distribution of the LEDs. On the other hand, the Gaussian Copula theory gives us a quadratic solution to estimate the distribution of a random vector based on the shape of its marginals. In this Section we show how to use Gaussian Copulae to construct visual codebooks, and build the final C-BoW signature. First, we fit a Gaussian Copula with the LEDs of the training set; we then draw random samples from the obtained Copula: these will become the visual words for the C-BoW.

In order to estimate the LED joint distribution using a Gaussian Copula C_Σ, similar to Eq. 3 we should fulfill the requirements (I) and (II) detailed in Sec. 4, namely (I) the SIFT components must follow a Gaussian distribution, and (II) the SIFT vector's joint distribution should be Gaussian. . Regarding Requirement (II), we do not know the real $p(x)$ of the SIFT components. Therefore, in this paper we take the simplistic assumption that x follows a multivariate normal distribution $\Phi(x) = \Phi(x_1, x_2, \ldots, x_k)$. Since in practice $p(x) \neq \Phi(x)$, the model generated under such assumption looks at the feature space under a different, complementary point of view compared to traditional BoW models.

C-BoW defines Copula-based words as follows (see Fig. 2 for a visual explanation):

(1) First, given the set of t training LEDs $x^i = (x^i_1, \ldots, x^i_k), i = 1, \ldots, t$, we compute the corresponding CDF $u_1 = \Phi(x_1), u_2 = \Phi(x_2), \ldots u_k = \Phi(x_k)$, normalized in the interval $[0, 1]$.

(2) According Eq. (3), we then have to compute the inverse of the gaussian CDF of the resulting vectors, namely

$$\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_k) \tag{5}$$

(3) Extending Equations 3 and 4 for $k >> 2$, $\Phi_\Sigma(x)$ is the normal multivariate distribution with covariance equal to the correlation matrix Σ and mean zero, where Σ can be computed as the correlation matrix between the vectors in (5):

$$\Sigma(a,b) = \frac{cov(\Phi^{-1}(u_a), \Phi^{-1}(u_b))}{\sigma(\Phi^{-1}(u_a))\sigma(\Phi^{-1}(u_b))} \tag{6}$$

$cov(x, y)$ corresponds to the covariance between variables x and y, and $\sigma(x)$ is the standard deviation of variable x.

(4) Given the correlation matrix, we draw a set of w random samples from the multivariate distribution with correlation Σ and mean zero[3]. After recovering the original scale of the data, we obtain our visual dictionary $D = w_1, w_2, \ldots, w_w$.

(5) Finally, we use the codebook D for vector quantization as traditional BoW models: each image is represented through a histogram that collects the visual word occurrences.

5 Experimental Validation

In this Section, we evaluate the proposed theory with a set of experimental results, testing the accuracy and effectiveness of C-BoW against traditional BoW and MEDA models. We first propose a set of experiments that compare the three models on a scene recognition task. We then test the effectiveness of C-BoW (as a stand-alone descriptor and combined with BoW and MEDA) in a content-based video retrieval system on the TRECVID [14] dataset.

Our Results show that by adding this efficient and fast signature to the pool of existing LED aggregators we can substantially improve the performances of the categorization (+ 25%) and retrieval (+ 50 %) systems built in our experiments.

5.1 Scene Recognition Task

For this task a labeled image dataset is provided, and a classifier needs to categorize a set of test images with their corresponding class, based on the image signatures used as input.

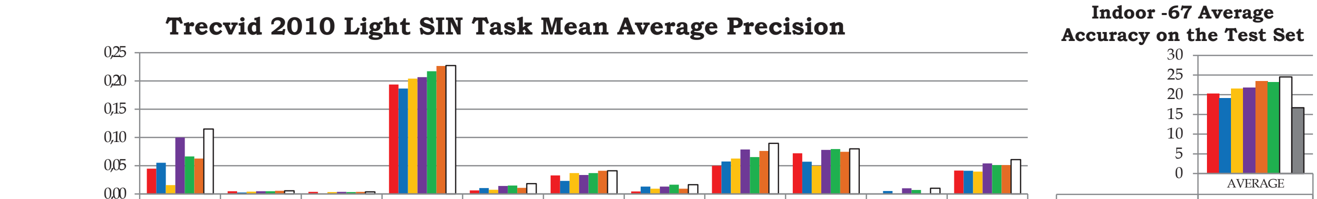
Experimental Setup

For the Scene Recognition Task we chose a challenging database, namely the Indoor-67 database. This database was first introduced in [11] for indoor scene recognition, and it is composed of around 15000 images categorized in 67 classes. In order to compare our approach with the existing ones, we compute BoW, C-BoW and MEDA, that are all based on locally extracted descriptors.

Therefore, we first extract LEDs from each image in the dataset, using PCA-SIFT [6] ($k = 36$). We then use the following experimental setup:

• For BoW, we use k-means clustering over the LEDs in the

[3] In practice, those examples have to be rescaled in order to recover the original range of data. We obtain the values compatible with the original LEDs by first computing the Gaussian CDF over each randomly drawn sample, and then taking the inverse of the normalized CDF computed in step 1



	Airplane	Boat_ship	Bus	Cityscape	Classroom	Demonstration	Hand	Nighttime	Singing	Telephones	MAP
🟥 BOW	0,044812	0,0046951	0,00342944	0,19360842	0,00612153	0,03278292	0,00443164	0,04974289	0,07204775	0,00041005	0,041208174
🟦 C-BOW	0,05513722	0,00260917	0,00030123	0,18663297	0,01030324	0,02319784	0,01298522	0,05740641	0,05712605	0,00525574	0,041095509
🟨 MEDA	0,01548938	0,00405293	0,00324665	0,20423441	0,00719979	0,03676535	0,00918763	0,06243025	0,04957104	0,00072935	0,039290678
🟪 C-BOW + BOW	0,099995	0,004695	0,003429	0,206594	0,014013	0,03332	0,012859	0,078555	0,077924	0,009891	0,0541275
🟩 C-BOW +MEDA	0,066369	0,004737	0,003247	0,21732	0,014726	0,036909	0,016191	0,065101	0,07941	0,006904	0,0510914
🟧 MEDA+BOW	0,06245	0,005454	0,00344	0,226512	0,010783	0,041124	0,009188	0,07605	0,07459	0,00073	0,0510321
□ MEDA+BOW+C-BOW	0,114728	0,005608	0,003441	0,227289	0,01835	0,041124	0,016293	0,089471	0,079622	0,009891	0,0605817

	AVERAGE
🟥 BOW 360	20,2309197
🟦 C-BOWULA 350	19,08301364
🟨 MEDA	21,4532
🟪 BOW+ C-BOW	21,81607576
🟧 MEDA+BOW	23,44665606
🟩 MEDA+C-BOW	23,22963636
□ MEDA+ C-BOW +BOW	24,53042424
■ MARGINAL	16,72058182

Figure 3. Results for Trecvid SIN Task and scene categorization on the Indoor-67 database

training set. We compute a dictionary of 360 visual words.
● For MEDA, we use the percentile approach proposed in [12] and we compute an alphabet of 10 bins per dimensions, resulting in a 360-dimensional signature.
● For C-BoW, we estimate the Copula with the LEDs of the training set. We then randomly draw 360 examples, namely the words of the codebook, from the obtained CDF.
In order to verify that the Copula and the correlation between the LED coefficients are in fact contributing to the discriminative power of our signature, for sake of completeness we also compare the C-BoW signature with a marginal-based randomly generated signature (*marginal* in Fig. 3. This marginal-based signature is generated by fitting one univariate gaussian to model the marginal of each dimension of the LED. The resulting codebook is built by randomly drawing the value of each component from the corresponding marginal. By doing so, we use the same idea as C-BoW, but we eliminate the correlation between the LEDs: the comparison of the performances of this signature and BoW performances gives us the idea of the importance of modeling the entire joint distribution using Copulae. The resulting signatures are then used as input for a multi-class SVM classifier with a chi-square kernel of degree 2. A one vs all SVM is built to separate each class from the others, test images are classified, and finally performances are evaluated with multiclass prediction accuracy. Signature contributions are then combined using posterior linear fusion.
Results
Results on this task show that our Copula-based solution actually reaches comparable performances compared to k-means based BoW, and that actually our k-dimensional probability estimation brings substantial improvements compared to the marginal-based fitting (*marginal* vs *C-BoW*). Moreover, as shown in Table 1, C-BoW is much more efficient than BoW (about 800 time faster), due to the quadratic complexity achieved by our formulation.
Moreover, we can see how C-BoW brings complementary information compared to the other existing LED aggregators. By just combining MEDA with C-BoW (namely, the two methods that do not need any search in the k-d space), we achieve an improvement of about 15% over the BoW-only based classification. When we combine the contributions of the three models, the performances on the test set improve by more than 20%.

5.2 Video Retrieval Task

In order to test the effectiveness of our feature for video retrieval, we select the Light Semantic Indexing Task (SIN) of TRECVID 2010 edition. For this evaluation campaign, a database of videos and a list of semantic concepts is provided. The participating groups are required to build a content-based retrieval system based on visual features, that, for a given concept, retrieves a list of shots ranked according to their relevance.
Experimental setup
The TRECVID 2010 development database is composed of 3200 videos, that we split in two halves, one for training and one for test. The 2010 Light SIN task involves the modeling of 10 semantic concepts. We test the effectiveness of the three models (BoW, MEDA and C-BoW) for a complete retrieval system over this database, using traditional SIFT descriptors extracted from the central keyframe of each shot and then aggregated with:
● BoW, generating with k-means clustering a codebook of 500 visual words.
● MEDA, using uniform quantization with a number of bins adapted for each concept.
● C-BoW, randomly drawing 500 examples from the fitted Copula.
A concept-specific SVM is learnt for each concept of the task, using chi-square kernel of degree 2. The classified examples of the test set are then ranked according to their concept score. To show the contribution of the fused descriptors, posterior linear combination is used. Finally, the ranked results are evaluated using mean average precision.
Results
In Figure 3 we show the performances of the C-BoW, BoW

	BoW	Copula BoW
Trecvid 2010	2 days	≅950s
Indoor 67	62563s	81 ± 5 s

Table 1. Computational times for codebook generation in Bow and C-Bow

and MEDA for the TRECVID SIN Task. C-BoW outperforms the MEDA model and achieves, with much less computation, the same performances as the BoW model. For some concepts for which the vocabulary of BoW was less discriminative, e.g. Telephones, C-BoW achieves good performances due to the better estimation of the joint LED distribution. This improved informativeness of the codewords is of particular importance when we combine the descriptors together. When we fuse the contributions from C-BoW and BoW the mean average precision on the test set increases by almost 35% over the BoW-only retrieval. As said, this is due to the different assumptions on the joint distribution of the LEDs. Same complementarity, due to the marginal vs joint distribution estimation, can be seen when combining C-BoW and MEDA (same behavior can be noticed of course with BoW and MEDA combined). Therefore, when we consider the combined contribution of the three models together, each of them brings a different point of view regarding the distribution of the LEDs and the resulting mean average precision increases of about 50% compared to BoW only.

6 Conclusions

In this paper we presented C-BoW, an efficient, Copula-based method for local image descriptors aggregation. We showed that our model, can build in a quadratic time a discriminative codebook for LED quantization, based on the correlation between the marginal distributions of the components of the local image descriptors. Our experimental results showed that the resulting C-BoW signature is as discriminative as BoW when used for both image categorization and video retrieval. Moreover, we verify that, by adding this efficient and fast signature to the pool of existing LED aggregators, we can significant improve the final retrieval (+50%) and categorization performances (+20%) when we combine the different approaches together.

References

[1] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, page 22. Citeseer, 2004.

[2] E. Cuvelier and M. Noirhomme-Fraiture. Clayton copula and mixture decomposition. *ASMDA 2005*, pages 699–708, 2005.

[3] F. Di Lascio and S. Giannerini. A new copula–based clustering algorithm.

[4] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Computer Vision and Pattern Recognition, 2003 IEEE Computer Society Conference on*.

[5] X. Guo, L. Wang, J. Zeng, and X. Zhang. Vq codebook design algorithm based on copula estimation of distribution algorithm. In *RVSP 2011*, pages 178–181. IEEE, 2011.

[6] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. 2004.

[7] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *International Journal of Computer Vision*, 77(1):259–289, 2008.

[8] D. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[9] R. Nelsen. *An introduction to copulas*. Springer Verlag, 2006.

[10] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2161–2168. Ieee, 2006.

[11] A. Quattoni and A. Torralba. Recognizing indoor scenes. In *Computer Vision and Pattern Recognition, IEEE Conference on*, 2009.

[12] M. Redi and B. Merialdo. Marginal-based visual alphabets for local image descriptors aggregation. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1429–1432. ACM, 2011.

[13] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8(1):11, 1959.

[14] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In *MIR '06*, New York, NY, USA, 2006. ACM Press.

[15] T. Tuytelaars and C. Schmid. Vector quantizing feature space with a regular lattice. In *ICCV*, pages 1–8, Rio de Janeiro, Brazil, 2007. IEEE Computer Society.