A JOINT MODEL OF COMPLEX WAVELET COEFFICIENTS FOR TEXTURE RETRIEVAL

Roland Kwitt and Andreas Uhl

Department of Computer Sciences,
University of Salzburg, Jakob-Haringer Str. 2, A-5020 Salzburg, Austria

ABSTRACT

We present a Copula-based statistical model of complex wavelet coefficient magnitudes for color texture image retrieval. Our model is based on two-parameter Weibull distributions and a multivariate Student *t* Copula. For similarity measurement we employ a Monte-Carlo approach to approximate the Kullback-Leibler divergence between two models. The experimental retrieval results show that the incorporation of the dependency structure between subbands significantly improves retrieval accuracy compared to previous approaches.

1. MOTIVATION

Content-based image retrieval (CBIR) systems mainly consist of two building blocks: a feature extraction (FE) block and a similarity measurement (SM) block, illustrated in Fig. 1. According to [1], both blocks are strongly related to each other. In the feature extraction block we compute a set of image descriptors, which are subsequently used for similarity measurement between a given query image and a number of candidate images in a database. In this paper we present

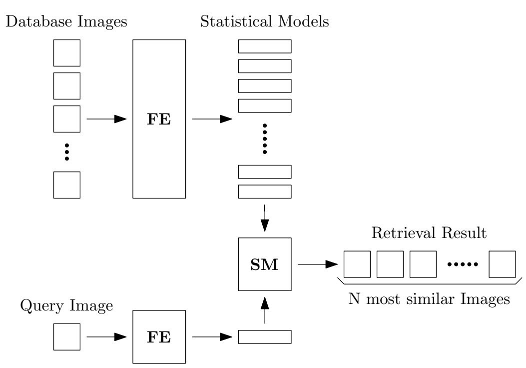

Fig. 1. Image Retrieval Framework

an extension of recent work [2] on probabilistic texture image retrieval where we have exclusively modeled the marginal distributions of complex wavelet coefficient magnitudes. Our contribution is to additionally incorporate the dependency structure across complex wavelet detail subbands of a specific decomposition level and equal subbands of different color bands. Our approach is based on multivariate modeling by means of Copulas. Specifically, we show that a Student *t* Copula with Weibull margins leads to a significant improvement in retrieval accuracy. This model is equally convenient for both grayscale and color image processing where we observe strong dependencies between equal subbands of different color bands. In the similarity measurement block we use an artificially symmetrized version of the Kullback-Leibler (KL) divergence. Due to the lack of a closed-form solution of the KL divergence between two of our statistical models, we employ a pragmatic Monte-Carlo approach. The retrieval experiments on a widely-used texture image database show superior results compared to previous works [1, 2].

The remainder of this paper is structured as follows: in Section 2 we introduce the proposed statistical model. Section 3 then provides the experimental results of our work and Section 4 recapitulates the main contributions together with an outlook on open research problems.

This research project was funded by the Austrian Science Fund (FWF) under project no. L366-N15.

2. STATISTICAL MODELING & IMAGE SIMILARITY

Statistical modeling resides in the feature extraction block of the CBIR system and consists of two steps: in the first step, we compute wavelet transform coefficients of each RGB color band. Second, we estimate the parameters of our statistical model. In the similarity measurement block of the CBIR system, the statistical models are then used to compute a measure of image similarity.

2.1. Image Representation

As in [2], we work with the Dual-Tree Complex Wavelet Transform (DT-CWT) due to the advantages over the pyramidal DWT w.r.t. image analysis. In contrast to the DWT, the DT-CWT is four-times redundant in 2-D and has six complex detail subbands per decomposition level, capturing image details oriented along $\pm 15^\circ$, $\pm 45^\circ$ and $\pm 75^\circ$. Further, the transform is approximately shift-invariant and is efficiently implemented by four parallel 2-D DWTs. The schematic frequency-tiling of the DT-CWT is shown in Fig. 2 for frequencies $w_2 \geq 0$ together with our subband enumeration scheme $S_1, \ldots, S_6$.

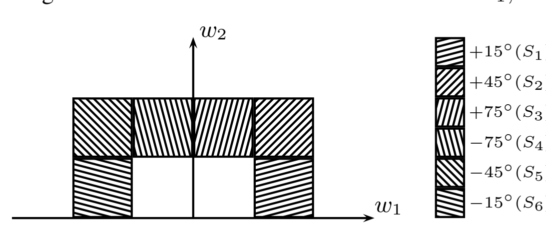

Fig. 2. Schematic frequency tiling of the DT-CWT

In previous works [1, 2], the assumption of subband independency has led to both a simple statistical model and a computationally attractive way to measure image similarity. Although we can argue that the independency assumption is valid for the DWT subbands, it is certainly violated in case of the overcomplete DT-CWT. In case of RGB color images where each color band is decomposed separately, assuming independent detail subbands is even more questionable for both the DWT and DT-CWT.

To quantify the degree of dependency we employ a graphical tool known as *Chi-plot* [3], since it is often hard for the human eye to judge a random scatter plot. Figure 3 shows Chi-plots for a selection of subband combinations from two example images (see Section 3) on decomposition level two together with the classic Pearson correlation coefficient r, Spearman's ρ and Kendall's τ. The subscript indices $c \in \{R, G, B\}$, i and j in the subband notation $S_{c,ij}$ denote the j-th subband (see Figure 2) of color channel c at decomposition level i. In case of independence we expect the points to lie in the central (shaded) region of the plot. The control limits (the width of the shaded region) are set to enclose $\pm 1.78/\sqrt{n}$ (n denotes the number of transform coefficients) which is a common choice in literature (see [4]). As we can see, there are strong deviations from the shaded region especially for subbands of opposite angle and equal subbands of different color bands. This behavior can be observed for all images of our database and confirms our concerns about the validity of independency assumption. To construct the dataset for statistical model estimation, we transform each RGB color band by the DT-CWT and arrange the absolute values of the mid-frequency subband coefficients (i.e. level two) in a $n \times p$ matrix $\mathbf{X}$. Since we have six subbands per scale and three color bands we obtain $p = 18$.

2.2. Modeling the Marginal Distributions

In [2] we have shown that the marginal distributions of complex wavelet coefficient magnitudes of DT-CWT transformed texture images can be fairly well modeled by two-parameter Weibull distributions. We recapitulate that the probability density function (PDF) of a Weibull distribution with shape parameter α and scale parameter β is given by

$$\mathfrak{p}(x; \mathbf{w}) = \frac{\alpha}{\beta} \left(\frac{x}{\beta} \right)^{\alpha - 1} \exp \left\{ - \left(\frac{x}{\beta} \right)^{\alpha} \right\} \tag{1}$$

with $x \geq 0$ and $\mathbf{w} = (\alpha, \beta)$. The parameters can be estimated either by using the method of moments (MM) or by using Maximum-Likelihood (ML) estimation. Unfortunately, both the MM and ML approach require numerical root finding algorithms. Another approach is to exploit the fact that the logarithm of a Weibull-distributed random variable follows an Extreme Value distribution of Type I (Gumbel) [5]. In that case the ML approach still requires numerical root finding, however there exist closed-form solutions for the MM estimates. Since the MM estimates can be used as starting values in a Newton-Raphson iteration for example we favor this approach due to less computational effort. We note that the parameter estimates $\hat{\alpha}, \hat{\beta}$ are calculated based on the column data in $\mathbf{X}$ (i.e. for each subband).

2.3. Copula Modeling

In multivariate modeling, Copula theory is a very important tool, especially when the dependency structure of the underlying data is of particular interest. A Copula is a multivariate distribution with uniform marginals [6]. Given a p-dimensional vector $U = (u_1, \ldots, u_p)$ on the unit cube $[0, 1]^p$, the Copula function C is defined as

$$C(u_1, \ldots, u_p) = \mathbb{P}(U_1 \leq u_1, \ldots, U_p \leq u_p). \tag{2}$$

In [7], Sklar showed that given a p-dimensional distribution function F with margin distribution functions $F_1, \ldots, F_p$, there exists a p-dimensional Copula C such that

$$F(x_1, \ldots, x_p) = C(F_1(x_1), \ldots, F_p(x_p)), \tag{3}$$

exploiting the fact that every random variable can be transformed to a uniform random variable by its probability integral transform. In our work, we use the Student *t* Copula, which is a member of the class of elliptical Copulas, to capture the dependency structure between the transform coefficients of different subbands. We have evaluated other Copulas such as the Multivariate Normal or Archimedian Copulas w.r.t. both Akaike's and Bayes' Information Criterion (AIC/BIC) (see [8]). Throughout all of our experiments the Student *t* Copula showed the lowest AIC and BIC for our image data. The Student *t* Copula is specified by the multivariate Student *t* distribution. Given that $\mathbf{\Sigma}$ denotes a symmetric positive definite matrix with $\mathrm{diag}(\mathbf{\Sigma}) = (1, \ldots, 1)$, then the Multivariate Student *t* Copula is defined as

$$C(u_1, \ldots, u_p; \mathbf{\Sigma}, \nu) = T_{\mathbf{\Sigma},\nu}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_p)) \tag{4}$$

where $T_{\mathbf{\Sigma},\nu}$ denotes the standard multivariate Student *t* distribution with ν degrees of freedom and $t_\nu^{-1}(\cdot)$ denotes the inverse Student *t* distribution function. The corresponding PDF is given by

$$\mathfrak{c}(u_1, \ldots, u_p; \mathbf{\Sigma}, \nu) = |\mathbf{\Sigma}|^{-1/2} \frac{\Gamma\left(\frac{\nu+p}{2}\right) \left[\Gamma\left(\frac{\nu}{2}\right)\right]^p}{\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^p \Gamma\left(\frac{\nu}{2}\right)} \frac{\left(1 + \frac{1}{\nu} \xi^T \mathbf{\Sigma}^{-1} \xi\right)^{-\frac{\nu+p}{2}}}{\prod_{i=1}^p \left(1 + \frac{\xi_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}} \tag{5}$$

with $\xi_i = t^{-1}(u_i)$ and $\xi = (\xi_1, \ldots, \xi_p)$. Without specifying any specific copula or margin distribution, the joint PDF $\mathfrak{f}$ of the whole Copula model can be written as

$$\mathfrak{f}(x_1, \ldots, x_p) = \mathfrak{c}(F_1(x_1), \ldots, F_p(x_p)) \cdot \prod_{i=1}^p \mathfrak{f}_i(x_i). \tag{6}$$

Given a Student *t* Copula and Weibull margins, the log-likelihood $l(\theta)$ of observing $\{(x_{i1}, \ldots, x_{ip}) : i = 1, \ldots, n\}$ follows as

$$l(\theta) = \sum_{i=1}^n \log\left(\mathfrak{c}(F_1(x_{i1}; \mathbf{w}_1), \ldots, F_p(x_{ip}; \mathbf{w}_p); \mathbf{\Sigma}, \nu\right) + \sum_{i=1}^n \sum_{j=1}^p \log \mathfrak{p}_j(x_{ij}; \mathbf{w}_j) \tag{7}$$

with $\theta = (\theta_C, \mathbf{w}_1, \ldots, \mathbf{w}_p)$, $\theta_C = (\mathbf{\Sigma}, \nu)$ and $F_i(\cdot, \mathbf{w}_i)$ denoting the cumulative distribution functions of the Weibull distribution. Regarding parameter estimation issues, direct maximization of the log-likelihood equation (7) to estimate the model parameters is a cumbersome procedure. For that reason, we use a two-stage estimation procedure which is known as the *Inference from Margins* (IFM) [9] method. The IFM method works as follows: first, we estimate the shape and scale parameters of the Weibull marginal distributions using ML estimation as explained in Section 2.2. Second, using the estimated Weibull parameters $\hat{\mathbf{w}}_i, 1 \leq i \leq p$ we maximize

$$\hat{\theta}_C = \arg\max_{\mathbf{\Sigma},\nu} \sum_{i=1}^n \log\left(\mathfrak{c}(F_1(x_{i1}; \hat{\mathbf{w}}_1), \ldots, F_p(x_{ip}; \hat{\mathbf{w}}_p); \mathbf{\Sigma}, \nu\right). \tag{8}$$

Hence we obtain a joint model for the absolute values of the complex wavelet coefficient magnitudes on a specific decomposition level for all three color bands.

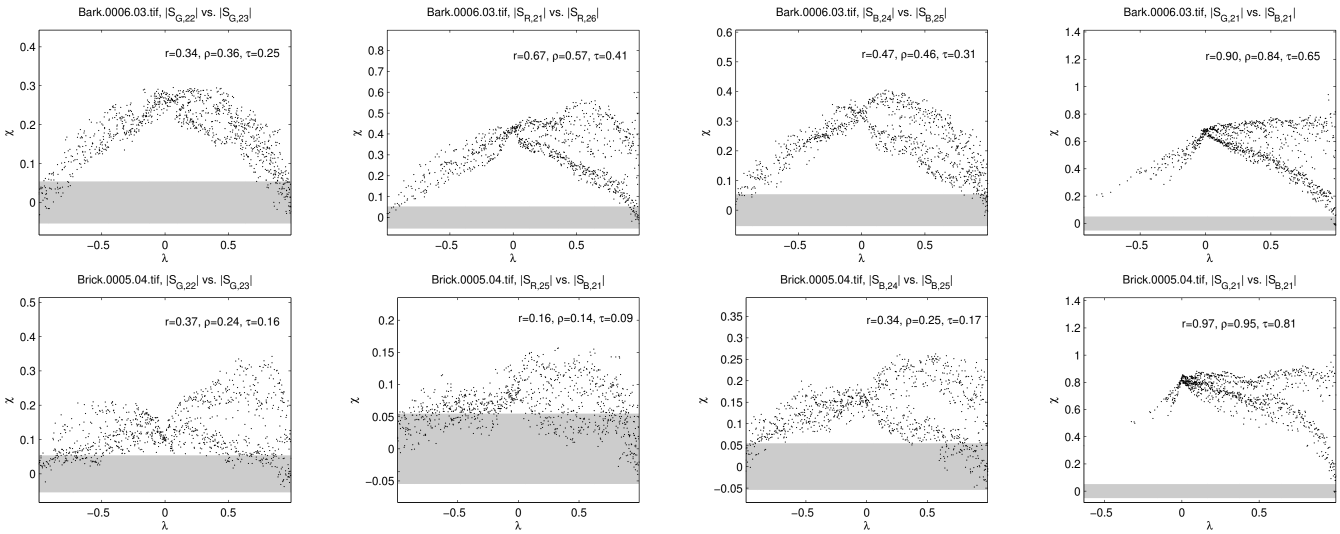

Fig. 3. Chi-plots for different combinations of subband coefficients

2.4. Similarity Measurement

As we have mentioned in Section 1, similarity measurement is the second integral part of a CBIR system. In the context of statistical image retrieval, Do and Vetterli [1] introduced the KL divergence as a reasonable similarity measure with theoretical foundation in decision theory. However, in contrast to [1, 2], the KL divergence between two Copula models has no closed-form solution. Hence, we choose a pragmatic Monte-Carlo (MC) approach and exploit the fact that the KL divergence between two PDFs $\mathfrak{f}$ and $\tilde{\mathfrak{f}}$ can be written as $D(\mathfrak{f}||\tilde{\mathfrak{f}}) = \mathbb{E}_{\mathfrak{f}}[\log \mathfrak{f}(x) - \log \tilde{\mathfrak{f}}(x)]$. We can approximate $D(\mathfrak{f}||\tilde{\mathfrak{f}})$ using Monte-Carlo simulation by drawing a random sample $x_1, \ldots, x_n$ from the model density $\mathfrak{f}(x)$ and then calculate

$$D_{MC}(f||\tilde{\mathfrak{f}}) \approx \frac{1}{n} \sum_{i=1}^{n} \log \mathfrak{f}(x_i) - \log \tilde{\mathfrak{f}}(x_i) \qquad (9)$$

which converges to $D(f||\tilde{\mathfrak{f}})$ as $n \rightarrow \infty$. Unfortunately, the KL divergence is not a true metric since it is neither symmetric nor does it satisfy the triangle inequality. The lack of symmetry however can be eliminated by using the artificially symmetrized version $D^s_{MC}(\mathfrak{f}||\tilde{\mathfrak{f}}) := 0.5 \cdot \left(D_{MC}(\mathfrak{f}||\tilde{\mathfrak{f}}) + D_{MC}(\tilde{\mathfrak{f}}||\mathfrak{f}) \right)$ instead. We further note that due to the Monte-Carlo approach, the KL divergence will differ to a certain extend (depending on n) in each experimental run. To assess the impact of n on D^s_{MC}, we use a reference approximation with $n = 10^6$ samples to check the deviation for $n \in \{10^2, 10^3, 10^4, 10^5\}$. As a matter of fact, there is a trade-off between precision and computation time here, though we accept a certain degree of deviation in favor of lower computational cost. Our experiments showed that for $n \geq 10^3$ the variations in the KL divergence due to randomness did not lead to different retrieval results. Therefor, we use $n = 10^3$ in the following section.

3. RETRIEVAL EXPERIMENTS

Our experimental results are based on 40 texture images of the MIT Vision Texture (VisTex) database [10] which is commonly used for the evaluation of texture retrieval systems. All images are 512×512 pixel RGB images. To quantify the quality of the retrieval results we use the common setup of splitting each of the 40 texture images into $B = 16$ non-overlapping subimages of size 128×128 pixel. Each subimage is then used as a query once. Before wavelet decomposition we normalize the color bands of each subimage by subtracting its mean and dividing by its standard deviation. Regarding the choice of DT-CWT filters we use Kingsbury's Q-Shift (14,14)-tap filters for decomposition levels ≥ 2 in combination with (13,19)-tap near-orthogonal filters for decomposition level one.

To conduct a comparative study, we select the approaches of [1] and [2] which are designed to work with grayscale images. As a straightforward extension to color images, we separately transform each color band and sum up the all the KL divergences (i.e. assuming subband independency). Further, we compare against a recent retrieval approach [11] where the authors employ a multivariate Power Exponential Distribution as the statistical model for DWT detail subbands and measure image similarity by using the geodesic distance.[1] We remind that in order to obtain comparable results we model only the DT-CWT/DWT detail subbands on decomposition level two.

In the first part of the experiments we assess the goodness-of-fit of the Copula model. As it is noted in [4], a reasonable way of visually judging the goodness-of-fit is by drawing a random sample from the underlying model and plotting the sample against the true observations. Since this test is only feasible for bivariate data, we selected pairs of subband combinations to get an impression of the model suitability. Figure 4 shows a selection of different subband combinations for an example (grayscale) texture image, visualizing different types of dependency. As we can see, the random sample (gray dots) of size $5 \cdot 10^3$ is actually a good approximation to the original data (black crosses). Similar results can be observed for all database images.

As an evaluation criteria, we first count the number of correctly retrieved images among the top K matches. By *correct* we understand the correct texture class (i.e. the parent of the subimage). Given that set $Q := \{r_1, \ldots, r_B\}$ denotes the correct set of membership indices and $\{q_1, \ldots, q_K\}$ is the set of the top K retrieved

[1] Thanks to G. Verdoolaege for providing the MATLAB code. Our MATLAB code will be available under `http://www.wavelab.at`.

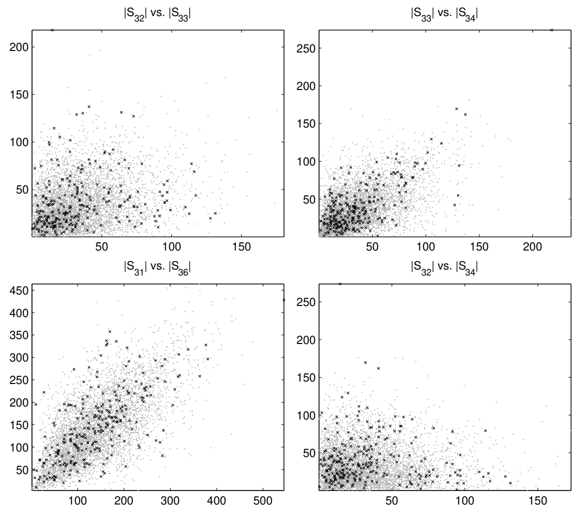

Fig. 4. Random sample of size $5 \cdot 10^3$ of fitted models (gray points) together with original transform coefficients (black crosses)

Approaches	**Retrieval Rate [%]**
Do & Vetterli [1]	73.6
Kwitt & Uhl [2]	76.2
Verdoolaege et al. [11]	89.1
Proposed Copula Model	**91.6**

Table 1. Retrieval rates for $K = 16$

image indices, we compute the percentage of correctly retrieved images as

$$s_K = 100 \cdot B^{-1} \sum_{i=1}^{K} \mathbf{1}_Q(q_i) \tag{10}$$

where $\mathbf{1}_Q$ denotes the indicator function of the set Q. Table 1 lists the retrieval rates for all four approaches. As we can see, the straightforward color extension assuming subband independency does not lead to a significant increase in retrieval rate compared to the grayscale case (see original work [1, 2]). The Copula model achieves the highest overall rate followed by the multivariate model of [11] which shows quite competitive performance. As a second evaluation criterion we evaluate the retrieval performance s_K w.r.t. varying values of K. This gives us some kind of receiver operating characteristic curve (ROC) for each approach. Figure 5 shows the corresponding ROC curves for $K \geq 16$. The ROC curve comparison generally provides the same insight as the numbers in Table 1. We observe the behavior that as K increases the slope of the curves become more shallow, implying that the performance gap between the approaches increases. Regarding the behavior of [11] we notice that although the retrieval rate is lower for small K (i.e. $K \leq 40$), the ROC curve crosses the Copula curve for larger values of K. Yet, since the most interesting part for the user is for small values of K, this effect is negligible.

4. CONCLUSION

We presented a novel statistical model of complex wavelet coefficient magnitudes for color texture retrieval. By incorporating the dependency structure between DT-CWT detail subbands we achieved a significant increase in retrieval accuracy compared to straightforward color extensions of previous works. Future work includes research w.r.t. the similarity measurement block which might pose problems in terms of computation time for large image databases.

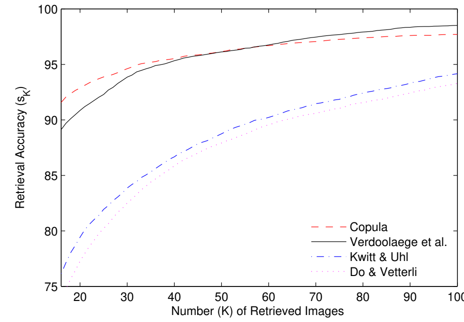

Fig. 5. ROC curve comparison

References

[1] M. Do and M. Vetterli, "Wavelet-based texture retrieval using Generalized Gaussian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[2] R. Kwitt and A. Uhl, "Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '08)*, San Diego, California, United States, Oct. 2008, pp. 933–936.

[3] P. Switzer N. Fisher, "Chi–plots for assessing dependence," *Biometrika*, vol. 72, pp. 253–265, Aug. 1985.

[4] C. Genest and A. C. Favre, "Everything you always wanted to know about Copula modeling and were afraid to ask," *Journal of Hydrological Engineering*, vol. 12, no. 4, pp. 347–368, July 2007.

[5] A. C. Cohen and B. J. Whitten, *Parameter estimation in reliability and life space models*, Marcel–Dekker, 1988.

[6] R. B. Nelsen, *An Introduction to Copulas*, Springer Series in Statistics. Springer, second edition, 2006.

[7] M. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institute de Statistique de l'Université de Paris*, vol. 8, pp. 229–231, 1959.

[8] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[9] H. Joe, *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability. Chapman & Hall, 1997.

[10] MIT Vistion Texture Database, "MIT vision and modeling group," [Online], Available from: `http://vismod.media.mit.edu/vismod/`.

[11] G. Verdoolaege and P. Scheunders S. De Backer, "Multiscale colour texture retrieval using the geodesic distance between multivariate Generalized Gaussian models," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '08)*, San Diego, California, USA, Oct. 2008, pp. 169 – 172.