ROTATION-INVARIANT FEATURES BASED ON STEERABLE TRANSFORMS WITH AN APPLICATION TO DISTRIBUTED IMAGE CLASSIFICATION

Baltasar Beferull-Lozano

Swiss Federal Institute of Technology-EPFL,
School of Communications and Computer Science,
AudioVisual Communications Laboratory,
Lausanne, CH-1015, Switzerland
Email: baltasar.beferull@epfl.ch

Hua Xie, Antonio Ortega

Department of Electrical Engineering,
Signal and Image Processing Institute,
University of Southern California,
Los Angeles, CA, 90089,
Email:{huaxie,ortega}@sipi.usc.edu

ABSTRACT

In this paper, we propose a new rotation-invariant image retrieval system based on steerable pyramids and the concept of angular alignment across scales. First, we define energy-based texture features which are steerable under rotation, i.e., such that features corresponding to the rotated version of an image can be easily obtained from the features of the original (non-rotated) image. We also propose an approach to measure similarity between images that is robust to rotation; images are compared after being aligned in angle. The retrieval process is performed by means of a Decision Tree Classifier where the angular alignment is performed at each node in the tree. To demonstrate the effectiveness of our system we consider a distributed image classification system, where the feature encoder and the classifier are physically apart and thus features are compressed before being transmitted. Our results of retrieval performance versus rate show a clear gain with respect to a wavelet transform (as an example, for the same rate, the retrieval precision is increased from 40% to 65%).

1. INTRODUCTION AND MOTIVATION

Feature extraction and matching are very important components in searching multimedia databases. Texture information is useful for this purpose and several approaches have been proposed to extract texture-related features based on various linear transforms, such as the wavelet transform. Basically, most of the well-known texture feature extraction methods measure the energies of the subbands obtained from a wavelet transform as texture discriminating features. One drawback of using critically sampled transforms for this purpose is that the features are not rotation or shift invariant. In this paper we address the problem of designing efficient rotation-invariant texture features and demonstrate their use in the context of decision tree classifier. Our goal here is to enable locating similar images in the database, even if the image captured is rotated with respect to those most similar to it in the database.

Two main previous approaches have been proposed for rotation invariant texture matching. In the first one (i.e. [1]), an stochastic model (i.e. Hidden Markov model) is assumed for features derived from wavelet transforms and the training is performed

using samples with different orientations and identified as belonging to the same class. In the second one (i.e. [2]), given the outputs of a transform, some specific rotation-invariant quantities are defined. In our work, instead, we achieve rotation-invariance by using the concept of angular alignment, i.e., the features obtained from two images are aligned before being compared. the two images being compared. To achieve this, we define a set of features which are *steerable* in the sense that given the features of an image sample, it is possible to obtain the features corresponding to any rotated version of it. These features are obtained from the subbands of a steerable pyramid [3, 4]. We also propose a new similarity measurement which measures the distance between two feature vectors only after they have been aligned. This angular alignment can be performed efficiently using simple constrained steepest descent algorithms. Although several features obtained from a steerable pyramid have been proposed in previous work [5], the property of achieving angular alignment by using steerability in the feature space, has not been considered. On the other hand, approximate rotation invariance has been tried to be achieved through features based on oriented Gabor filters [6], but real rotation invariance is not achieved because Gabor filters do not provide the steerability property.

One drawback of steerable transforms in many image representation applications comes from the fact that they are oversampled and thus result in a significant storage penalty with respect to critically sampled transforms. For this reason, in this paper we choose a distributed image classification application, where features have to be compressed before transmission, to demonstrate that no rate penalty exists and features based on steerable representations outperform wavelet-based features even when operating at the same rate. In our comparisons we use three different quantization algorithms: a) simple uniform quantization, b) quantization with optimal bit allocation and c) a classified vector quantization scheme optimized from a rate-distortion-complexity point of view proposed by Xie and Ortega [7]. In order to decrease the complexity of the retrieval, a Decision Tree Structure Classifier (DTC) is often used in practical applications; images having similar features are clustered together in the nodes of the classifier. In this paper, we show how the rotation-invariance can be incorporated in the DTC based retrieval by performing angular alignment at each node in the tree and defining an appropiate distance between an image and a tree node, which ensures that a best-first-search method works correctly. Our experimental results show a substantial gain in retrieval performance versus rate of our method with respect to

a retrieval system based on a wavelet transform.

This paper is organized as follows: Section 2 concentrates on the feature extraction process and Section 3 describes the basic similarity measurement which is the main novelty of our work. In Section 4, we describe briefly the quantization schemes that are considered and also the DTC based retrieval system incorporating the angular alignment. Finally, Section 5 shows the experimental results of our method.

2. FEATURE EXTRACTION

Since we are interested in achieving rotation invariance, the feature extraction we consider is based on the subbands obtained from a steerable pyramid [3]. We then should choose features that are as "steerable" as possible, that is, given the features of an image oriented at an angle ϕ, it should be possible to obtain the features corresponding to the same image but oriented at an angle ϕ', by direct manipulation of the features at angle ϕ, i.e., without actually having to recalculate the features after rotating the image. In our work, we try to achieve good retrieval performance using energy-based features which are simple to manipulate.

Let $c(x_o, \phi)$ represent the value of a transform coefficient corresponding to the output of a rotated steerable filter with orientation ϕ for a certain spatial location x_o. In a steerable pyramid with J basic orientations and L levels, at each level l, given the J basic coefficients $\{c^l(x_o, \phi_1), c^l(x_o, \phi_2), \ldots, c^l(x_o, \phi_J)\}$, the transform coefficient $c^l(x_o, \phi)$ for any angle (orientation) ϕ of that same spatial location will be given by:

$$c^l(x_o, \phi) = \sum_{i=1}^{J} \alpha_i(\phi) c^l(x_o, \phi_i) \quad \forall \phi, \quad l = 1, \ldots, L \quad (1)$$

where $\{\alpha_1(\phi), \alpha_2(\phi), \ldots, \alpha_J(\phi)\}$ is the set of J steering functions which allow to perform the (exact) interpolation at any level $l = 1, \ldots, L$.

Let $E^l(\phi)$ represent the average energy of a subband oriented at an arbitrary angle ϕ in a level l, that is, $E^l(\phi)$ is given by $E^l(\phi) = \left(\frac{1}{N_l}\right) \sum_{k=1}^{N_l} (c^l(x_k, \phi))^2$, where N_l is the number of pixels of each of the subbands in level l and the subscript k goes through all the spatial locations of the subband. It is very simple to show that $E^l(\phi)$ can be calculated from the energies (sampled autocorrelations) of the basic J subbands and all the sampled cross-correlations between each pair of basic subbands:

$$E^l(\phi) = \alpha^T(\phi) C^l \alpha(\phi), \quad \alpha(\phi) = (\alpha_1(\phi) \cdots \alpha_J(\phi))^T \quad (2)$$

where C^l is the (symmetric) sampled correlation matrix with elements $C_{ij}^l = \left(\frac{1}{N_l}\right) \sum_{k=1}^{N_l} c^l(x_k, \phi_i) c^l(x_k, \phi_j) = C_{ji}^l, l = 1, \ldots, L$. Each diagonal element of C^l corresponds to $C_{ii}^l = E^l(\phi_i)$, that is, the average energy at the basic angle ϕ_i, while the off-diagonal elements correspond to sampled cross-correlations between the subbands corresponding to each pair of basic angles.

Notice that since $c(x_o, \phi + \pi) = -c(x_o, \phi)$, clearly, $E^l(\phi + \pi) = E^l(\phi)$, that is, $E^l(\phi)$ is a periodic function with period equal to π. Given a perfectly homogeneous image I with energy profile $E_I^l(\phi)$ at level l, if this image is rotated counter-clockwise by an angle θ, obtaining an image I_θ, then, we will have that $E_{I_\theta}^l(\phi) = E_I^l(\phi - \theta)$, that is, a rotation of an image corresponds to a shifted version of the energy profile.

Based on this we choose the correlation matrices $\{C^l\}_{l=1}^{L}$ as the energy-based texture features in our system. Notice that since each matrix C^l is symmetric, the total number of features will be $J(J+1)L/2$. Therefore, the interdependencies between different orientations in terms of cross-correlations are necessary in order to characterize the energy profile of an arbitrary rotation of a given image. We do not consider the use of the energy of the low-pass residual subband as a feature in our proposed system. Obviously, as the number J of basic orientations increases, the resolution in angle (angular bandwidth of basic filters) increases and the energy profile $E^l(\phi)$ will be therefore more accurate, but on the other hand, the number of raw features may become substantially larger than in the case of a wavelet-based texture representation.

3. SIMILARITY MEASUREMENT

In the similarity measurement, we are interested in making use of the steerability property present in the features in order to identify equivalent features, where equivalency will correspond to having different rotated versions of a unique image. The next proposition shows that the sampled correlation matrix C_I^l for an image at a given level l and the sampled correlation matrix $C_{I_\theta}^l$ for the same image but rotated counter-clockwise by an angle θ, are related in a simple way.

Proposition 1 *Given a steerable representation with J basic angles, the correlation matrices $C_{I_\theta}^l$ and C_I^l, both evaluated with respect to the same set of basic angles $\{\phi_1, \ldots, \phi_J\}$, are related as follows:*

$$C_{I_\theta}^l = R(\theta) C_I^l R^T(\theta),$$

$$R(\theta) = \begin{pmatrix} \alpha_1(\phi_1 - \theta) & \alpha_2(\phi_1 - \theta) & \cdots & \alpha_J(\phi_1 - \theta) \\ \alpha_1(\phi_2 - \theta) & \alpha_2(\phi_2 - \theta) & \cdots & \alpha_J(\phi_2 - \theta) \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1(\phi_J - \theta) & \alpha_2(\phi_J - \theta) & \cdots & \alpha_J(\phi_J - \theta) \end{pmatrix}$$
$$(3)$$

In the particular case where the J basic angles are taken to be equiespaced, then $R(\theta)$ becomes an orthogonal matrix for any θ, and therefore, $C_{I_\theta}^l$ and C_I^l become orthogonally equivalent.

Proof: The proof is given in [8].

This property holds for every level independently. However, notice that when a image is rotated, all the decomposition levels will be equally rotated. This means that given an image I and a rotated version I_θ of it, the Frobenius norms:

$$\|C_I^l - R(-\theta) C_{I_\theta}^l R^T(-\theta)\|_F, \quad l = 1, \ldots, L$$

(same rotation angle for all the levels), will tend to be small.

Taking all this into account, the similarity measurement $D(I_1, I_2)$ between 2 different images I_1 and I_2 that we propose is the following:

$$D(I_1, I_2) = \text{Min}_\theta \left(\sum_{l=1}^{L} \|C_{I_1}^l - R(-\theta) C_{I_2}^l R^T(-\theta)\|_F \right)$$
$$(4)$$

Clearly, those levels containing more energy will influence more in the minimization of (4) and those levels with small energy will have little influence in it.
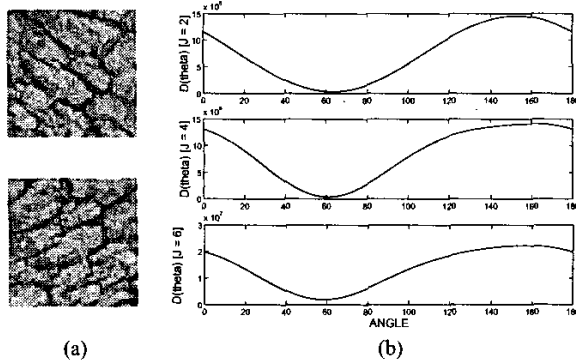
(a)   (b)

Fig. 1. (a) "bark" physically rotated at 60 and 120 degrees; (b) $D(\theta)$ for $J = 2, 4, 6$. Notice how in all three cases, the minimum is achieved for $\theta = 60$ degrees, which is the exact relative angle between the two texture image samples.

Notice that when I_1 and I_2 are two rotated versions of the same image, the angle θ^* for which the minimum is achieved in (4) should be close to the relative angle between I_1 and I_2, that is, the angle one needs to rotate (clock-wise) I_1 in order to get I_2. Thus, one way to see the goodness of our similarity measurement (4) is to check whether the estimated angle θ^* is actually close to the real relative angle between 2 physically rotated versions of the same image. Moreover, it might also be useful in some practical applications to find out approximately this relative angle. Fig. 1 illustrates this by showing the function $D(\theta) = \sum_{l=1}^{L} \|C_{I_1}^l - R(-\theta)C_{I_2}^l R^T(-\theta)\|_F$, for the case where I_1 and I_2 are rotated versions of "bark" texture from the Brodatz set [11].

As explained in Section 4, this angular alignment has to be performed many times in the retrieval process and thus it is important to devise fast algorithms to find the minimizing angle θ^* in (4). In [8], it is shown that for $J = 2$, θ^* can be found analytically, and for $J > 2$, it is possible to design low complexity constrained steepest descent algorithms. This is because it can be proved that the number of stationary points of the function being minimized in (4) is upper bounded and at the same time, the angular distance between any two contiguous stationary points is lower bounded [8] making it simple to search for these points in a few non-overlapping angular intervals.

4. QUANTIZATION AND RETRIEVAL PROCESS

In this work, we have tested our proposed scheme when the feature vectors are quantized using a set of scalar quantizers with three different quantization algorithms: (i) Simple uniform quantization (same stepsize); (ii) Non-uniform quantization with optimal bit allocation in a rate-distortion sense [9]; (iii) Classified quantization optimized in a rate-distortion-complexity sense, proposed by Xie and Ortega [7]. The block diagram of this system is illustrated in Fig. 2. First, a complete K-means binary DTC tree T is designed and then this tree is pruned in order to get an optimal subtree $S^* \preceq T$ which serves as a pre-classifier and whose leaves correspond to different classes. Each class is associated with a different encoder containing a set of stepsizes selected from a previously predefined set. The optimal subtree S^* and encoders are
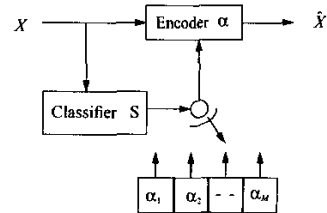

Fig. 2. Block diagram of a classified vector quantization system. Separate encoders $\{\alpha_i\}$ are designed for the classes $i = 1, \cdots, M$. The input feature vector X is first classified and then encoded with an encoder specifically designed for the corresponding class.

optimized using the G-BFOS algorithm [10] so that the MSE is minimized under both rate and complexity constraints. For more details, see [7].

The retrieval process is always performed with the DTC T using the best-first-search and branch-and-bound[1]. Let $\underline{D}(Q, t)$ denote the distance of the query Q with node t in the tree. In order to ensure that this search algorithm finds the correct closest matches, we need to define a distance $\underline{D}(Q, t)$ satisfying the property that $\underline{D}(Q, t)$ is a lower bound of the distances of Q to all the images in node t, and we need to take into account the angular alignment process. Notice that:

$$\underline{D}(Q, I) = \min_{\theta} d(Q_\theta, I) \tag{5}$$
$$\geq \min_{\theta} d(Q_\theta, I_c) - d(I_c, I) \text{ (triang. inequality)}$$
$$\geq \min_{\theta} d(Q_\theta, I_c) - R(t) \text{ (upper bound)}$$

where I_c is the centroid in node t, $R(t)$ the radius of node t given by $R(t) = \max_{I \in t} d(I_c, I)$ and $d(Q_\theta, I)$ is given by the expression inside the parenthesis in (4). Thus, defining $\underline{D}(Q, t)$ as $\underline{D}(Q, t) = \min_\theta d(Q_\theta, I_c) - R(t)$, then, it is guaranteed that the best-first-search method will find the correct closest match. Thus, we see that a crucial differential point in our work is that in the retrieval process using the DTC, at each node of the tree, alignments between the query (quantized) feature vector and each of the two representing vectors (corresponding to the two branches) have to be performed using (4). After these two alignments, two distance measurements are performed and a branch is chosen.

5. EXPERIMENTAL RESULTS

We have evaluated the performance of our proposed method applied to the Brodatz texture images [11] and have also compared with a standard wavelet transform. The features we have considered in the wavelet case are also correlation matrices obtained from the corresponding four wavelet subbands. We use 2 collections of texture samples of size 128 × 128. The first collection, which forms the *non-rotated* image database, is obtained by partitioning each of the 13 Brodatz (512×512) non-rotated texture images [11] into 16 non-overlapping texture subimages of size 128 × 128 with a total of 208 texture samples. This set is used in training of the DTC for retrieval. The second collection, which forms the *rotated*

---

[1]This algorithm has a complexity of $O(\log M)$ (M is the number of feature vectors in the database) as compared to $O(M)$ in a linear search.

set, is obtained by partitioning (for each of the 13 texture classes) 4 large texture images oriented at 30, 60, 90 and 120 degrees also into non-overlapping subimages of size 128 × 128 and taking the 4 central subimages. In this way, in the second database, there are also 16 textures for each class and therefore, also the same total number of 208 textures. A query texture sample is taken from the *rotated* set and the feature vector is extracted and quantized using the three quantization schemes described in Section 4. We assume that each quantized component of the feature vector is independently entropy coded. The $M = 16$ closest textures from the *non-rotated* set are obtained and the average retrieval precision over all the rotated texture samples is measured.

Without compressing the features, the average retrieval performance for the steerable case is of 67.03% and 66.55% for $J = 2$ and $J = 4$ respectively, while in the wavelet case, the performance is of 41.85%, an improvement of about 25% is obtained. Fig. 3 shows the retrieval performance of compressed steerable feature vectors for $J = 2$ and $J = 4$. We can clearly see that the classified quantizer achieves the best performance among the three quantization schemes. By using the classified quantizer with expected tree length $l = 2$ (complexity constraint), the retrieval performance degrades very gracefully. Even with the bit rate reduced to around 1 bit/element, we can still achieve about the same precision as using uncompressed feature. Fig. 4 shows the comparison of the retrieval performance with compressed features between steerable transform with $J = 4$ and a standard wavelet transform. The reason why we compare these two cases is that the dimension of the feature vector under both cases is the same ($N = 48$), so that they will result in a comparable bit rate. Again, we see that steerable achieves much better retrieval precision than wavelet over all bit rates. With respect to the retrieval complexity reduction by employing a DTC instead of a linear search, we have computed the number of distance computations that have to be performed to find the $M = 16$ closest matches. Instead of 208 distance computations as in the case of linear search, the DTC requires on average 121.97 for $J = 2$ and 39.82 for $J = 4$.

6. REFERENCES

[1] J. L. Chen and A. Kundu. Rotational and gray-scale transform invariant texture identification using wavelet decomposition and hidden Markov model. *IEEE Trans. on Patt. Recog. and Mach. Intell.*, vol. 16, pp. 208-214, 1994.

[2] W. R. Wu and S. C. Wei. Rotation and gray-scale transform-invariant texture classification using spiral resampling. *IEEE Trans. on Image Proces.*, vol. 5, pp. 1423-1434, 1996.

[3] E.P. Simoncelli, W. Freeman, E. Adelson, and D. Heeger. Shiftable multiscale transforms. *IEEE Trans. on Information Theory*, vol. 2, no. 38, pp. 587-607, 1992.

[4] B. Beferull-Lozano and A. Ortega. Coding techniques for oversampled steerable transforms. In *Int. Asilomar Conf. on Signals, Systems and Computers*, 199

[5] E. P. Simoncelli and J. Portilla. Texture characterization via joint statistics of wavelet coefficient magnitudes. In *ICIP*, 1998.

[6] Y. M. Ro, M. Kim, H. K. Kang, B. S. Manjunath and J. Kim MPEG-7 Homogeneous Texture Descriptor. In *ETRI Journal*, vol. 23, no. 2, 2001.

[7] H. Xie and A. Ortega. Entropy and Complexity constrained Classified Quantizer Design for Distributed Image Classification. *Proc. of Multimedia Signal Processing (MMSP)*, 2002.

[8] B. Beferull-Lozano. Quantization design for structured overcomplete expansions. *Ph. D. thesis, University of Southern California*, 2002.

[9] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. on Acoust. Speech and Signal Proces.*, vol. 36, pp. 1445-1453, 1988.

[10] P. A. Chou, T. Lookabough and R. M. Gray. Optimal Pruning with Applications to Tree-Structured Source Coding and Modeling. *IEEE Trans. on Information Theory*, vol. 35, pp. 299-315, 1989.

[11] Brodatz Rotated textures. http://sipi.usc.edu/services/database/Database.html.
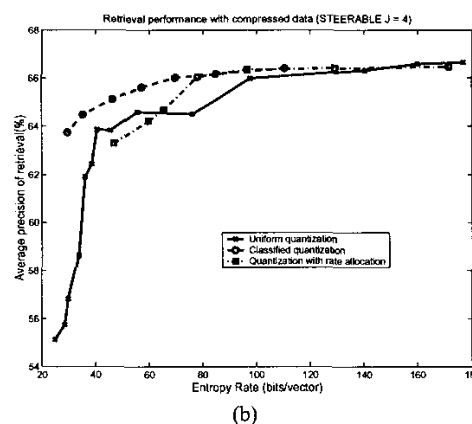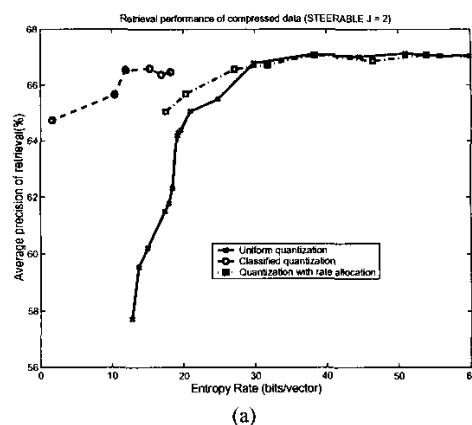
(a)

(b)

Fig. 3. Average Retrieval Performance using a 3 level steerable pyramid for the three different quantization algorithms: (a) $J = 2$ and (b) $J = 4$.
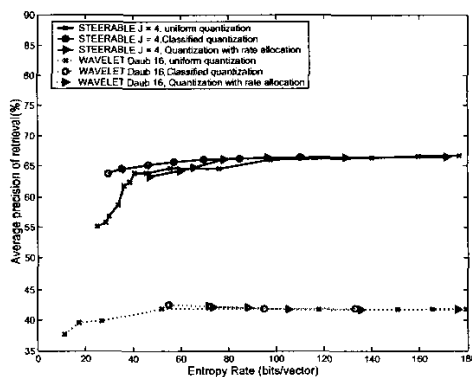

Fig. 4. Comparison between a standard 3 level wavelet pyramid ('daub16' filter bank) and a 3 level steerable pyramid with $J = 4$ for the three different quantization algorithms.