

# Texture based classification of images using frequency estimated pairwise MRF joint distributions on site labels from wavelet decomposed images

Markus Louw

Fred Nicolls

Department of Electrical Engineering  
University of Cape Town, South Africa

markus.louw@gmail.com

## Abstract

In this paper we demonstrate the efficacy of using joint probabilities on the values (pixel intensities/wavelet coefficients) for neighbouring sites (pixels/spatially neighbouring wavelet coefficients), to classify images based on texture. The classification capacity for this type of joint distribution, used as a feature, is tested using a first nearest neighbour (NN1) method, which counts the number of errors when comparing the labeled texture to the label calculated by assigning the texture to the class of its nearest neighbour, calculated using our method. We compare our classification results to another method based on histogram comparison. Our classification methodology is simple, general, extensible and fast to calculate.

## 1. Introduction and Literature Review

There has been a number of attempts at image texture classification in the past, many of which use Markov Random Field parameter estimation to extract texture based features. In [5], images are segmented into regions of different textures by doing a multispectral Karhunen-Louve expansion based factorization. Parameters on the image regions are estimated using a Gauss Markov Random Field (MRF) model. Segmentation is done by detecting texture boundaries, which exist when the MRF texture parameter changes too much after which segmentation is done on the image from within the MRF parameter space. In [7], image texture patterns are specified through MRF parameter estimation. The MRF parameter for textures are found through MCMC sampling. The use of these MRF parameters as texture features is explored using a nearest linear combination (NLC) classifier. In [12], a method is found for comparing 3D volumetric MRI data using texture. The data is assumed to be modeled by a Gaussian MRF. Distances between the texture classes using the Kullback Leibler divergence (KLD) are found elegantly by deriving a closed form expression for the probability of data given class labels using sparse symmetric block circulant matrices, calculated using a 3D FFT. Wavelet signal and image decomposition was introduced in [8], and was soon used for image texture classification and image retrieval. A comparison of wavelets for texture based image retrieval was done in [16], where it was found that Gabor wavelets gave slightly better classification than orthogonal and biorthogonal wavelets, at a much greater computational cost. Our work also resembles that of [4], in which the marginal distributions of wavelet coefficients are modeled with a generalized Gaussian density (GGD). Textures are then compared by measuring the KLD between Gaussian distribution parameter sets obtained by fitting Gaussian models to histogrammed wavelet subband coefficients. The spatial interactions between pixels in an image are

therefore modeled implicitly through the wavelet decomposition. More recently in [6], a texture segmentation method which combines morphological operators with wavelet decomposition is developed, with good segmentation results. The method we have developed, in contrast to the previous methods, models the probabilistic joint interactions between pixels and between wavelet coefficients explicitly (i.e. not through a parameterized MRF potential energy interaction). A joint distribution for pairwise neighbouring pixels or wavelet subband coefficients is derived, which allows us to use the many probabilistic distances that are available, for classification. We prove the classification power of these joint distributions used as features using a first-nearest-neighbour (NN1) classification scheme, where an image is assigned to the class of the image which it most closely resembles. Classification errors are counted for each data set used. We also test the joint pixel/coefficient probabilities for use as features using our worst-nearest-neighbour “WNN statistic” (described later), which is more rigorous than the NN1 statistic. Our method does not require the usual estimation of MRF parameters (which is known to be generally time consuming and inaccurate), and is therefore faster and not prone to classification errors resulting from bad MRF parameter fitting.

## 2. Data Sets

The data sets we use to compare our algorithm are of froth images taken from video feeds of flotation cells. There are three data sets (samples shown in Figs. 1, 2, 3), and in each data set the froth images have been grouped into different classes. The froths from the first data set contains classes of very different froths, where in some of the classes there is camera blurring. The second data set consists of froth classes where the froths are different, but there is no camera blurring. The third data set consists of froth surfaces which are only slightly different. We would expect any classification algorithm, including ours, to perform better on the first two data sets than on the third one.

## 3. Histogram based probabilistic inter-class distance

We can form an intensity histogram for each colour channel in the image by counting the number of pixels in the image  $\mathbf{Y}$  that have a certain intensity level  $l$ . The equation for the red channel follows:

$$H_R(\mathbf{Y}, l) = \sum_{i=1} I(\mathbf{Y}_R(i) = l), \quad (1)$$

where  $I()$  is the indicator function which returns 1 when the argument is true.

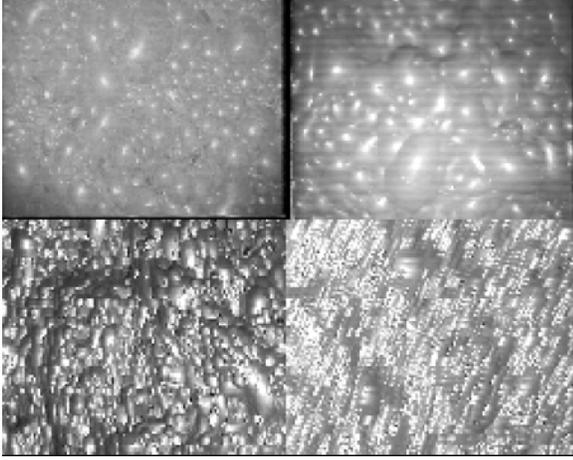


Figure 1: Four froth surface images from the first set of froth classes. Each image belongs to a different class.

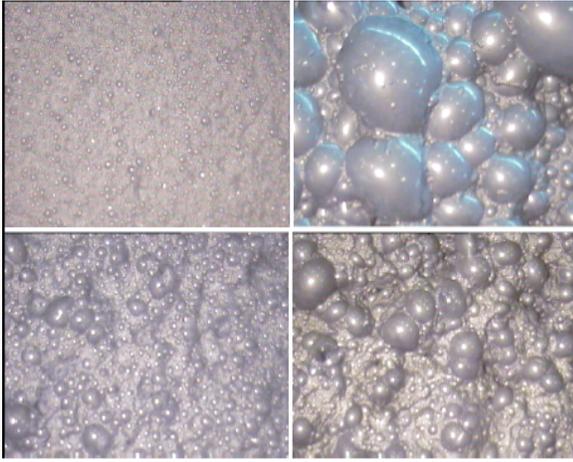


Figure 2: Four froth surface images from the second set of froth classes. Each image belongs to a different class.

#### 4. MRF model for pixels on a lattice

If  $\mathbf{Y}$  is the image data,  $\mathbf{X}$  is the set of random variables on the image lattice  $S$ , and  $x_i$  is the label on a particular site/node, and if the Markov neighbourhood of site  $i$  is  $\mathcal{N}(i)$ , then the usual MRF model for pixel labels on a lattice is expressed simply as

$$p(x_i|\mathbf{X}, \mathbf{Y}) = p(x_i|x_{s \in \mathcal{N}(i)}, \mathbf{Y}_i) \quad (2)$$

i.e. the probability of a site  $i$  having label  $x_i$  is dependent only on the labels of its Markov neighbours ( $\mathcal{N}(i)$ ) and its image pixel. Since we are not using Gibbs random field style potential energies, we do not need to specify energy terms of any kind. We do however assume that the same joint distribution between any neighbouring pair of pixels/sites is the same over the whole image, i.e.

$$p(x_i, x_j) = p(x_j, x_i) \forall i \in S, j \in \mathcal{N}(i). \quad (3)$$

The equation above implicitly assumes an isotropic pdf, i.e. the joint probability of neighbouring vertical pixels is the same as

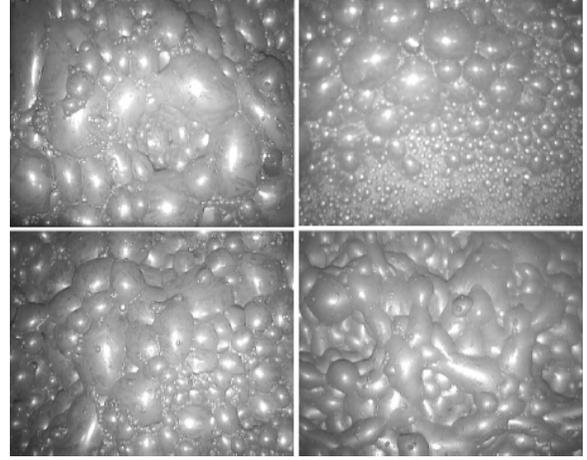


Figure 3: Four froth surface images from the third set of froth classes. Each image belongs to a different class.

that of neighbouring horizontal pixels. This allows us to estimate the joint probability through a histogramming method. In many MRF optimization problems ([13],[14]), it is assumed that the joint distribution given two random variables on a lattice applies to all pairs of random variables on the lattice (and any pair of pixel labels in the image data is a sample from this joint distribution). These methods all specify the MRF on a lattice via parametric energy functions. However, in [10] it is noted that such a joint density may be approximated explicitly using a frequentist approach (i.e. an occurrence counting methodology). We have adapted this approach via a histogramming step.

#### 5. Joint probabilities of neighbouring pixels

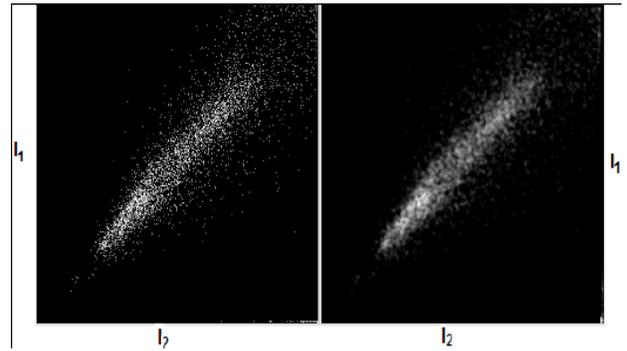


Figure 4: Joint probability distribution of discretized site values done on a wavelet decomposed images (left). The distribution is blurred with a Gaussian kernel, and renormalized (right).

First we form a histogram of the neighbouring pixel intensities. Since the intensity range is from 0..255, the histogram has dimensions  $256 \times 256$ . The histogram for the red channel

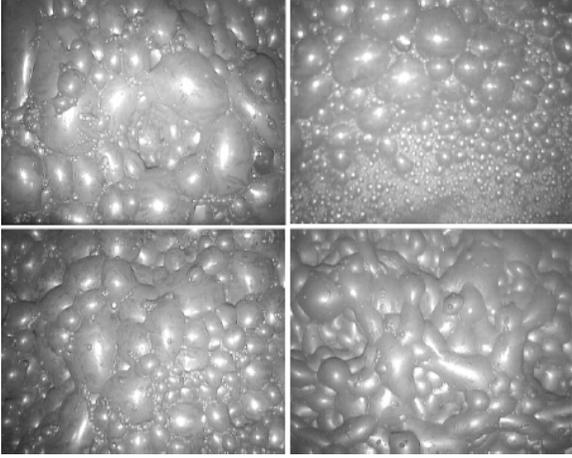


Figure 5: A froth surface image.

is formed as follows:

$$H_R(\mathbf{Y}_R, l_1, l_2) = \sum_{i=1}^N \sum_{j \in \mathcal{N}(i)} I(\mathbf{Y}_R(i) = l_1, \mathbf{Y}_R(j) = l_2) \quad \forall l_1, l_2 \in (0..255), \quad (4)$$

where in the above algorithm,  $I(\cdot)$  is a logical AND function which returns 1 when both its arguments are true, and returns 0 otherwise, and  $N$  is the number of sites (pixels) in the image. Algorithmically this is done in the following way:

1. Initialize all  $H_R(\mathbf{Y}_R, l_1, l_2) = 0$
2. For  $l_1, l_2 = 0, \dots, 255$
3. For  $i = 1..N$
4. For  $j \in \mathcal{N}(i)$
5. if  $j > i$  (ordering prevents double counting)
6.  $l_1 = \mathbf{Y}_R(i)$
7.  $l_2 = \mathbf{Y}_R(j)$
8.  $H_R(\mathbf{Y}_R, l_1, l_2) = H_R(\mathbf{Y}_R, l_1, l_2) + 1$
9. end(5)
10. end(4)
11. end(3)
12. end(2).

After this we can construct a joint distribution by normalizing the histogram:

$$p(l_1, l_2) = k \cdot H(l_1, l_2) \quad (5)$$

A visualization of such a probability distribution is shown in Fig. 4, left. It is useful to smooth the probability distribution before using the distance measures in Table 1, to compensate for measurement noise in the image capturing process, and to compensate for having too few data points to populate the histogram/joint probability distribution. Any blurring kernel can be used, provided the distribution is normalized after the operation.

## 6. Image wavelet decomposition and wavelet coefficient discretization

A wavelet decomposition [8] generates approximation and detail coefficients: these are continuous valued and may be negative. To use our histogramming approach for estimating joint

probabilities between approximation and detail coefficients between neighbouring sites (the sites occur in subbands), it is necessary to discretize (bucket) the coefficient values. Any partitioning of the space in which the coefficients exist is acceptable. We divided the space into 256 evenly spaced partitions over the range between the minimum and maximum coefficient values over all the images in each data set (this allows us to use the same algorithm on 256 level grayscale images, without applying any wavelet decomposition). An example of the wavelet decomposition of the image shown in Fig. 5 using the biorthogonal wavelet is shown in Fig. 6. Any set of wavelet kernels will suffice.

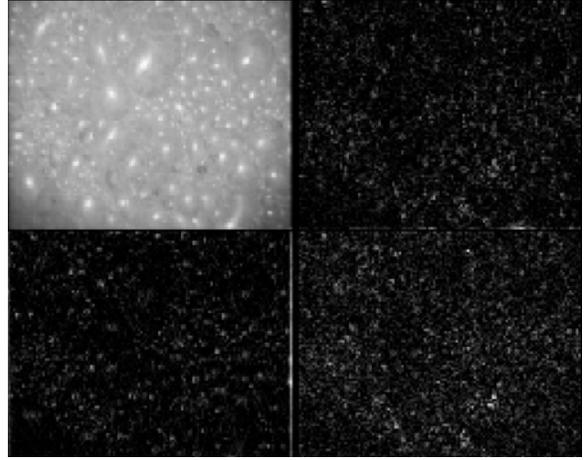


Figure 6: Wavelet decomposition of froth surface image.

Bhattacharyya:	$J_B(p_1, p_2) = -\log \left\{ \int_{\mathcal{X}} [p_1(X)p_2(X)]^{1/2} dX \right\}$
Matusita:	$J^t(p_1, p_2) = \left\{ \int_{\mathcal{X}} [\sqrt{p_1(X)} - \sqrt{p_2(X)}]^2 dX \right\}^{1/2}$
Patrick-Fisher:	$J_P(p_1, p_2) = \left\{ \int_{\mathcal{X}} [p_1(X)\pi_1 - p_2(X)\pi_2]^2 dX \right\}^{1/2}$
Kolmogorov:	$J_K(p_1, p_2) = \int_{\mathcal{X}}  p_1(X)\pi_1 - p_2(X)\pi_2  dX$
Inner product:	$J_c(p_1, p_2) = \int_{\mathcal{X}}  p_1(X)\pi_2(X)  dX$

Table 1: A list of distances between probability density functions, taken from [17], where  $0 \leq \alpha_1, \alpha_2 \leq 1$  and  $\alpha_1 + \alpha_2 = 1$ .  $\pi_1$  and  $\pi_2$  are the prior probabilities on the distributions.

	DS	$J_B^h$	$J_{PF}^h$	$J_M^h$	$J_K^h$	inner prod
WNN	1	4	4	4	4	4
NN1	1	1	1	1	1	1
WNN	2	18	18	18	18	18
NN1	2	7	10	7	9	10
WNN	3	12	16	13	15	17
NN1	3	2	6	2	6	6

Table 2: Classification results using pdfs derived by normalizing colour pixel intensity histograms (no joint distributions, no wavelet decomposition). DS indicates the data set used.

	DS	$J_B^h$	$J_{PF}^h$	$J_M^h$	$J_K^h$	inner prod
WNN	1	0	0	0	0	0
NN1	1	0	0	0	0	0
WNN	2	17	19	17	17	18
NN1	2	7	8	7	7	9
WNN	3	11	12	11	10	10
NN1	3	2	1	2	2	1

Table 3: Classification results using joint pdfs on neighbouring pixel intensities (no wavelet decomposition). DS indicates the data set used.

	DS	$J_B^h$	$J_{PF}^h$	$J_M^h$	$J_K^h$	inner prod
WNN	1	0	0	0	0	0
NN1	1	0	0	0	0	0
WNN	2	17	18	17	17	18
NN1	2	7	8	7	7	7
WNN	3	11	11	11	11	10
NN1	3	2	1	2	1	1

Table 4: Classification results using joint pdfs of discretized neighbouring site values, using approximation coefficients only (1st order wavelet decomposition). DS indicates the data set used.

## 7. Probabilistic distances

There are many probabilistic distances between probability functions which may be used to compare their similarity. These include the Chernoff [3], Bhattacharyya [2], Matusita [9], Patrick-Fisher [11], Kolmogorov [1] distances, and the Kullback-Leibler and symmetric Kullback-Leibler divergences [15]. Some of the analytical expressions are given in Table 1. Given these expressions for probabilistic distance between two pdfs, one can measure the distances between any two (sets of) joint probability distributions. If froth surface #1 has a pairwise pixel joint pdf of  $p_1(l_1, l_2)$ , and if froth surface #2 has a pairwise pixel joint pdf of  $p_2(l_1, l_2)$ , then the probability distances in Table 1 can be calculated by discrete summation. For example the Bhattacharyya distance has the summation

$$J_B(p_1, p_2) = -\log\left(\sum_{l_1, l_2} [p_1(l_1, l_2)p_2(l_1, l_2)]^{1/2}\right). \quad (6)$$

The classification results for orders of wavelet decompositions from 1 to 3 are done, and we report on the results from two different distance measures for each order of decomposition. The first,  $J^1$ , uses only the distances calculated on the discretized approximation coefficients. The distances  $J^2$  use the distance calculated using only the discretized coefficients from the 2nd subband;  $J^3$  and  $J^4$  are the distances calculated using only 3rd and 4th subbands respectively. We then define the total probabilistic distance between two wavelet decomposed images, using all 4 subbands (1 approximation and 3 detail), as

$$J^T = J^1 + J^2 + J^3 + J^4. \quad (7)$$

## 8. Classification Results

We tested the algorithm on three sets of real data: the first comprises 28 froth images taken from 7 classes, where some of the images are blurry due to bad camera focussing or where there

	DS	$J_B^h$	$J_{PF}^h$	$J_M^h$	$J_K^h$	inner prod
WNN	1	4	9	2	0	4
NN1	1	1	1	0	0	1
WNN	2	20	20	20	19	20
NN1	2	12	15	12	13	12
WNN	3	15	19	14	16	19
NN1	3	4	10	4	7	5

Table 5: Classification results using joint pdfs of discretized neighbouring site values, using all subband coefficients (1st order wavelet decomposition). DS indicates the data set used.

	DS	$J_B^h$	$J_{PF}^h$	$J_M^h$	$J_K^h$	inner prod
WNN	1	0	0	0	0	0
NN1	1	0	0	0	0	0
WNN	2	21	21	21	20	21
NN1	2	7	6	7	9	8
WNN	3	15	14	15	11	12
NN1	3	2	1	2	3	2

Table 6: Classification results using joint pdfs of discretized neighbouring site values, using all subband coefficients ( $J^T$ ) (2nd order wavelet decomposition). DS indicates the data set used.

are horizontal striping artifacts. The second data set consists of 24 images of very different froth types, divided into 6 classes. The third data set consists of 11 classes with three images per class of the same froth type, where the froth is in different states (each state corresponds to a class). We expect the third data set to have the worst classification results since the froth has similar appearance across operating states. Nearest neighbour classification results when using the probability distances using only the probability distributions derived from histogramming are shown in Table 2. Classification results using each of the probability distances operating on the joint distributions of Eqn. 5 are shown in Tables 2 to 7. To show the relative effect of conducting a wavelet decomposition, we include results of the probability distances calculated using the coefficients on the wavelet decompositions. In Tables 2 to 7, the NN1 statistic counts the number of times, for each image in each class, that the probabilistic distance to an extra-class image is less than *all* intra class distances (i.e. an erroneous classification has occurred). The WNN statistic counts the number of times for each image in each class, where the probabilistic distance to an extra class image is less than the *furthest* intra class distance. The WNN statistic is therefore the more rigorous of the two classifiability indicators (NN1 and WNN), and has correspondingly higher numbers for erroneous classification in the tables than the NN1 statistic. A score of zero in any of the tables is good since it represents the notion of zero classification errors.

## 9. Discussion

From the results in the previous section it is clear that this is a good classification method for froth surfaces, and outperforms the NN1 and NNE classification methods operating on pixel intensity probabilities only (i.e. when no joint pdfs are considered). The joint probability formulation implicitly takes into account high or low frequency pixel intensity changes across images, which will allow surfaces with many tiny bubbles to

	DS	$J_B^h$	$J_{PF}^h$	$J_M^h$	$J_K^h$	inner prod
WNN	1	1	7	1	3	0
NNI	1	0	2	0	0	0
WNN	2	17	19	17	18	17
NNI	2	7	10	8	8	9
WNN	3	16	20	16	17	16
NNI	3	3	12	4	6	4

Table 7: Classification results using joint pdfs of discretized neighbouring site values, using approximation coefficients only (2nd order wavelet decomposition). DS indicates the data set used.

have different appearances to surfaces with few large bubbles to be easily separated using this method. Interestingly, the incorporation of wavelet decomposition coefficients appearing in the detail subbands does not appear to improve the classification results. We believe that wavelet decomposition should nevertheless be incorporated into this methodology, since with very high resolution images, the joint distribution across pixels would be prone to super-fine scale froth surface anomalies. In addition the approximation coefficients (when used without the detail coefficients), improve the classification results, in our algorithm.

## 10. Conclusion

Using probabilistic distances such as the Bhattacharyya distance to find distances between images based on the joint probabilities between neighbouring pixels appears to be a good method for classification. Decomposing the images first using wavelets leads to better classification performance, although no classification power is added by incorporating probabilistic distance between joint distributions on the detail subband coefficients: only the approximation coefficients have benefitted the classification, in our setting.

## 11. Acknowledgements

The authors are grateful for the financial support given by the National Research Foundation of South Africa, and Anglo Platinum Ltd and Rio Tinto via the Centre for Minerals Research at the University of Cape Town.

## 12. References

- [1] B. Adhikara and D. Joshi. Distance discrimination et resume exhaustif. *Publs. Inst. Statis.*, 5:49–55, 1936.
- [2] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.
- [3] H. Chernoff. A measure of asymptotic efficiency of tests for a hypothesis based on a sum of observations. *Annals of Mathematical Statistics*, 23:493–507, 1952.
- [4] Minh N. Do and Martin Vetterli. Texture similarity measurement using Kullback-Leibler distance on wavelet subbands. *ICIP*, 2000.
- [5] Michal Haindl. Texture segmentation using recursive Markov Random Field parameter estimation. *Proceedings of the 11th Scandanavian Conference on Image Analysis*, pages 771–776, 1999.
- [6] V. Vijaya Kumar, U. S. N. Raju, M. Radhika Mani, and A.L.Narasimha Rao. Wavelet based Texture Segmentation methods based on Combinatorial of Morphological and Statistical Operators. *Int. J. of Computer Science and Network Security*, 8(8):176–181, 2008.
- [7] Jun Liu, Lei Wang, and Stan Z. Li. MRMRF Texture Classification and MCMC Parameter Estimation. *Vision Interface, Canada*, 1999.
- [8] S. Mallat. A theory for multiresolution signal decomposition. *IEEE. Transactions on Pattern Recognition and Machine Intelligence*, 11(7):674–693, 1989.
- [9] K. Matusita. Decision rules based on the distance for problems of fit, two samples and estimation. *Ann. Math. Stat.*, 26:631–640, 1955.
- [10] Kevin Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, 2002.
- [11] E. Patrick and F. Fisher. Nonparametric feature selection. *IEEE Trans. Information Theory*, 15:577–584, 1969.
- [12] E. Rangelova and A. Quinn. Difference Field Estimation for Enhanced 3-D Texture Segmentation. *British Machine Vision Conference*, 2002.
- [13] Jian Sun, Heung-Yeung Shum, and Nan-Ning Zheng. Stereo Matching Using Belief Propagation. *European Conference on Computer Vision*, 2002.
- [14] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A Comparative Study of Energy Minimization Methods for Markov Random Fields. *European Conference on Computer Vision*, 2006.
- [15] T.M.Cover and J.A.Thomas. *Elements of Information Theory*. Wiley, 1991.
- [16] W.Y.Ma and B. S. Manjunath. A comparison of wavelet transform features for texture image annotation. *ICIP*, 1995.
- [17] Shaohua Kevin Zhou and Rama Chellappa. From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel Hilbert space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006.