Statistical hypothesis test for robust classification on the space of covariance matrices

Ioana Ilea, Lionel Bombrun, Christian Germain, Romulus Terebes, Monica Borda

STATISTICAL HYPOTHESIS TEST FOR ROBUST CLASSIFICATION ON THE SPACE OF COVARIANCE MATRICES

Ioana Ilea[1,2], Lionel Bombrun[1], Christian Germain[1], Romulus Terebes[2], Monica Borda[2]

[1]: Université de Bordeaux, Laboratoire IMS, Groupe Signal et Image. *{ioana.ilea, lionel.bombrun, christian.germain }@ims-bordeaux.fr*
[2]: Technical University of Cluj-Napoca. *{romulus.terebes, monica.borda }@com.utcluj.ro*

ABSTRACT

This paper introduces a new statistical hypothesis test for robust image classification. First, we introduce the proposed statistical hypothesis test based on the geodesic distance and on the fixed point estimation algorithm. Next, we analyze its properties in the case of the zero-mean multivariate Gaussian distribution by studying its asymptotic distribution under the null hypothesis H_0. Then, the performance of the proposed classifier is addressed by analyzing its noise robustness. Finally, the robust classification method is employed for the classification of simulated Polarimetric Synthetic Aperture Radar images of maritime pine forests.

Index Terms— Robust image classification, statistical hypothesis test, geodesic distance, fixed point estimate, SAR imagery.

1. INTRODUCTION

Multiscale approaches have been found to be effective for many image processing applications including filtering [1], segmentation [2] or classification [3]. In the classification context, the image processing workflow consists in two steps. First of all, the image is decomposed into a set of wavelet subbands, each of them being modeled by a probability density function with a specific parameter vector. Next, for each subband, the estimated parameter vector composes the signature of the image. Then, a similarity measure based on a probabilistic metric is computed between the signature vectors.

Simple but effective methods have been proposed to characterize wavelet detail statistics based on univariate models such as the generalized Gaussian distribution [3]. Nonetheless, they do not take into account the dependencies existing in the image. To overcome this difficulty, multivariate distributions, including elliptical models [4, 5] and copula based approaches [6, 7], have been proposed to model the spatial and spectral dependencies in the images. Once the feature vectors are computed for each texture image, a distance (or at least a divergence) is calculated in order to measure the degree of similarity between two images. A well-known choice is the Kullback-Leibler (KL) divergence [8], or its symmetric version: the Jeffreys divergence [9]. Recently, some authors have

proposed to consider the geodesic distance, which has shown superior retrieval rate compared to the KL divergence [5].

In a classification or texture retrieval experiment, a nearest neighbor classifier is generally considered. In such case, a test image is labeled to the class of the closest training image. But, nothing tells that the test image is well classified, especially for noisy datasets. An hypothesis test should be performed to regulate the false alarm rate. Inspired from previous works on the KL divergence [10] and on the family of (h, ϕ) divergences [11, 12], we introduce a new statistical hypothesis test based on the geodesic distance. The advantage of using the geodesic distance lies in its property of being a distance measure, which is symmetric and respects the triangle inequality. The main objective of the paper is the construction of a robust classification algorithm.

The robustness of an algorithm can be analyzed by taking into account the decision-making strategies but also the parameter estimation techniques. For this latter, we propose an extension of the statistical hypothesis test to robust estimators [13]. A comparison between the sample covariance matrix (SCM) and the fixed point (FP) estimates is conducted in various experiments.

The paper is structured as follows. Section 2 introduces the proposed statistical hypothesis test based on the geodesic distance. Section 3 presents the robust FP estimates for the zero-mean multivariate Gaussian distribution (MGD). The performance of the proposed statistic is analyzed in terms of efficiency and noise robustness on simulated dataset. Comparisons with the SCM estimator are also carried out. Section 4 introduces an application for the classification of maritime pine forests based on simulated Polarimetric Synthetic Aperture Radar (PolSAR) images. Conclusions and future works are finally reported in Section 5.

2. HYPOTHESIS TEST FOR ROBUST CLASSIFICATION

2.1. Context

Let $X_1 = (\mathbf{x}_1^1, \dots, \mathbf{x}_m^1)$ and $X_2 = (\mathbf{x}_1^2, \dots, \mathbf{x}_n^2)$ be two sets of m and n independent and identically distributed random variables (vectors) $\mathbf{x}$ according to the parametric models

$p(\mathbf{x}|\theta_1)$ and $p(\mathbf{x}|\theta_2)$. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be the maximum likelihood (ML) estimates computed on these sets. In a classification problem, the aim is to determine if X_1 and X_2 are issued from the same parametric model. Let consider the following hypothesis test

$$\begin{cases} H_0 : \theta_1 = \theta_2, \\ H_1 : \theta_1 \neq \theta_2. \end{cases} \quad (1)$$

Under the regularity conditions discussed in [11], it has been proved in [10, 11, 12] that under the null hypothesis H_0 and for sample sizes $n, m \longrightarrow \infty$, the test statistic S_{GD} follows a chi-square distribution:

$$S(\widehat{\theta}_1, \widehat{\theta}_2) = \frac{2mnv}{m+n} D(\widehat{\theta}_1, \widehat{\theta}_2) \xrightarrow[n,m \to \infty]{} \chi_M^2, \quad (2)$$

where the degree of freedom M is equal to the dimension of the parameter space. v is a constant depending on the considered similarity measure D ($v = 1$ for the KL divergence). This hypothesis test has been first introduced in [10] for D being the KL divergence and further generalized in [11] for the class of (h, ϕ) divergences. In the paper, we extend this definition to the Rao geodesic distance, which is the shortest path in the parametric manifold. Indeed, under the null hypothesis H_0, distributions are lying infinitesimally close on the probabilistic manifold and in such case the KL divergence equals half of the squared geodesic distance (GD). Hence, when $\theta_1 = \theta_2$, the test statistic

$$S_{GD}(\widehat{\theta}_1, \widehat{\theta}_2) = \frac{mn}{m+n} GD^2(\widehat{\theta}_1, \widehat{\theta}_2) \quad (3)$$

is asymptotically chi-square distributed with M degrees of freedom for sufficiently large values of m and n. Note that under H_0, the distribution of the statistic S_{GD} is independent of θ_1 and θ_2.

2.2. Application to zero-mean multivariate Gaussian distributions

In the following, we suppose that X_1 and X_2 are two zero-mean multivariate Gaussian distributions (MGDs), having the parameter vectors represented by the covariance matrices $\mathbf{M}_1$ and $\mathbf{M}_2$. In this case, the Rao geodesic distance is computed according to:

$$GD(\hat{\mathbf{M}}_1 || \hat{\mathbf{M}}_2) = \left[\frac{1}{2} \sum_i (\ln \lambda_i)^2 \right]^{\frac{1}{2}}, \quad (4)$$

where $\hat{\mathbf{M}}_1$ and $\hat{\mathbf{M}}_2$ are the SCM estimators of the covariance matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ and λ_i, $i = 1 \ldots d$ are the eigenvalues of $\hat{\mathbf{M}}_2^{-1} \hat{\mathbf{M}}_1$. It yields that the null hypothesis $\mathbf{M}_1 = \mathbf{M}_2$ can be rejected at a level α if $Pr(\chi_M^2 > S_{GD}(\hat{\mathbf{M}}_1, \hat{\mathbf{M}}_2)) \leq \alpha$. Here, the degree of freedom M is equal to $d(d+1)/2$ where d is the dimension of the covariance matrix.
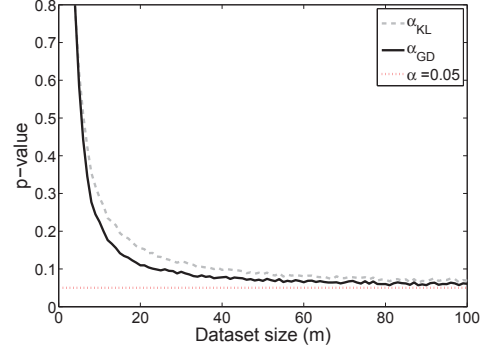
Fig. 1. Convergence of the estimated p-value as a function of the dataset size.

Some simulation results are displayed to evaluate the potential of the proposed statistical hypothesis test on a simulated dataset. The sets X_1 and X_2 are generated as m and n independent and identically distributed random vectors distributed according to a zero-mean MGD, having the covariance matrix of the form $\mathbf{M} = (M_{i,j})$ with

$$M_{i,j} = \rho^{|i-j|}. \quad (5)$$

For each set X_1 and X_2, the covariance matrix is estimated according to the maximum likelihood principle by using the SCM estimates. The significance level α is set to 0.05 and 10^4 Monte Carlo iterations are considered. Fig. 1 draws the evolution of the estimated p-value as a function of the dataset size ($m = n$ in this experiment) for the SCM estimate with $\rho = 0.5$. The solid line corresponds to the geodesic distance while the dashed line corresponds to the Kullback-Leibler divergence. As expected, the estimated p-value converges to the significance level α for sufficiently large m. In addition, it can be observed that the convergence is faster for the geodesic distance than for the Kullback-Leibler divergence. In the following, only the geodesic distance will be considered.

In the next section, we propose an extension of this hypothesis test to robust covariance matrices estimators.

3. EXTENSION TO ROBUST ESTIMATORS

3.1. Robust M-estimators

The M-estimators, a family of estimators, have been introduced in the context of robust theory to tackle the presence of outliers in the dataset or errors in the model. For zero-mean observations, the M-estimator of the covariance matrix is defined as the solution of [13, 14]:

$$\hat{\mathbf{M}} = \frac{1}{N} \sum_{i=1}^{N} u(\mathbf{x}_i^T \hat{\mathbf{M}}^{-1} \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T. \quad (6)$$

where $u(\cdot)$ is a positive-valued function which gives a weight to each observation $\mathbf{x}_i$ in the computation of the covariance

matrix. Obviously the weight function $u(\cdot)$ should decrease to zero to ensure that outliers have a smaller contribution to the covariance matrix estimate than other observations. Note that the SCM estimator gives an uniform weight (*i.e.* $u(\cdot) = 1$) to each observations, hence confirming its non robustness.

In literature, this family of M-estimators has been extensively studied and found to be a generalization of covariance matrix ML estimates for the family of elliptical distributions. Many authors have focused to the particular case where $u(t) = 1/t$. In such case, the M-estimator of covariance reduces to the so-called fixed point (FP) estimator given by [15, 16]:

$$\hat{\mathbf{M}}_{it+1} = \frac{1}{N} \sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^T}{\mathbf{x}_i^T \hat{\mathbf{M}}_{it}^{-1} \mathbf{x}_i}, \qquad (7)$$

with it being the iteration. Practically, this recursive algorithm can be initialized with the identity matrix and converges in about 10 iterations [16, 17].

The FP estimator has a unique solution $\hat{\mathbf{M}}$ up to a scale factor. Indeed, for any positive scalar $c \neq 0$, if $\hat{\mathbf{M}}$ is a solution of (7), $c\hat{\mathbf{M}}$ is also a solution of (7). In the following, the covariance matrix is normalized such that $tr(\hat{\mathbf{M}}) = d$. This FP estimator can be interpreted as the ML estimate of the normalized covariance matrix for a Gaussian scale mixture model[1] where the multipliers τ_i are assumed to be unknown deterministic parameters [15].

In [16], Pascal *et al.* have analyzed the properties of the FP estimator. In particular, the FP estimate is the unique solution of (7) and it is an unbiased and consistent estimate. Moreover, the FP estimate has a Wishart distribution behavior with $(d/(d+1))N$ degrees of freedom. This latter is used in the hypothesis given in (3) to adapt the degrees of freedom M of the statistic's asymptotic distribution.

3.2. Classification efficiency

Here, the proposed hypothesis test based on the statistics S_{GD} is used for implementing a two-classes classification algorithm for simulated data. The experiment consists in defining two independent and identically distributed random set of vectors X_1 and X_2 of size m_1 and m_2 distributed according to the MGD and having the covariance matrices $\mathbf{M}_1$ and $\mathbf{M}_2$. A third dataset X_t of size m_t and covariance matrix $\mathbf{M}_t$ has been defined in the same manner. The implemented algorithm has to classify X_t in one of the two available groups, by choosing the one with the most similar covariance matrix. Here, we consider that X_t should be of class 2, by generating it using the same parameters as for X_2. Under these assumptions, the hypothesis test consists in verifying if the

---

[1]Recall that a Gaussian scale mixture model admits the stochastic representation $\mathbf{x} = \sqrt{\tau}\mathbf{z}$ where τ is a scalar random variable called multiplier ($\tau \in \mathbb{R}^+$) and $\mathbf{z}$ a Gaussian random vector with zero-mean and covariance matrix $\mathbf{M}$.
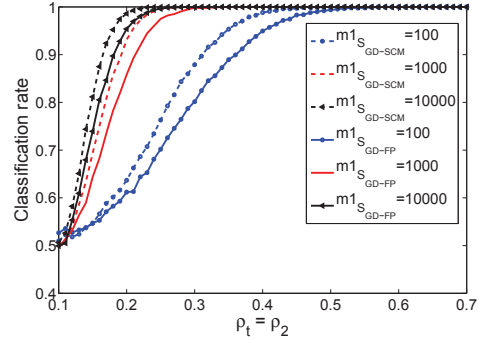
Fig. 2. Correct classification rate of X_t in class 2 by using the S_{GD} with the SCM and FP estimates, if $m_1 = 100, 1000, 10000$, and $m_2 = m_t = 1000$.
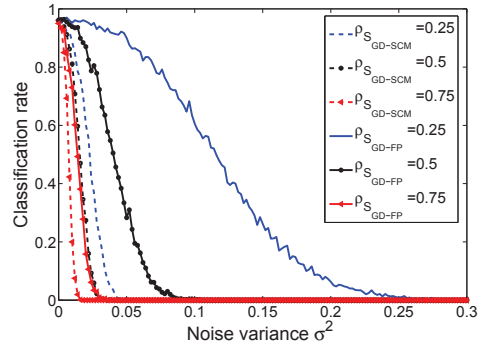
Fig. 3. Evolution of the performances of S_{GD} as a function of the noise variance σ^2, by using the SCM and FP estimate for various covariance matrices: $\rho = 0.25, 0.5, 0.75$ and $d = 3$.

distribution X_t has the same parameter vector as X_2, or in other words, if X_t is of class 2.

Fig. 2 draws the influence of the estimation algorithm and the influence of datasets' size on the classification performance. The simulations are carried out for a 3-dimensional dataset ($d = 3$) with $m_1 = 100, 1000$ and 10000 and $m_2 = m_t = 1000$. Several values for the covariance matrix $\mathbf{M}$ are tested (ρ_2 and ρ_t ranging from 0.1 to 0.7, while ρ_1 is fixed to 0.1). Each time $\mathbf{M}$ is estimated by the SCM (dashed lines) and FP (solid lines) algorithms. 10^4 Monte Carlo iterations are performed to obtain average performances.

As observed, the best performances are obtained for the SCM estimate compared to the FP one, illustrating the efficiency of this former. This observation is natural since this experiment has been carried out in a purely Gaussian context. The next subsection will analyze the robustness by considering some noisy data.

3.3. Noise robustness

The performances of the SCM and FP estimates are now compared in terms of noise robustness for the statistic S_{GD}. Here, two datasets X_1 and X_2 are generated as independent and identically distributed random vectors distributed according to a zero-mean MGD of covariance matrix $\mathbf{M}$. The set X_2 is next corrupted by an independent additive white Gaussian noise of covariance matrix $\sigma^2 \mathbf{I}_d$, σ^2 being the noise variance and $\mathbf{I}_d$ is the identity matrix.

The significance level α is set to 0.05 and several values for the covariance matrix $\mathbf{M}$ are tested ($\rho = 0.25$, 0.5, and 0.75). 10^3 Monte Carlo iterations are considered to estimate the classification rate. Results are displayed in Fig. 3. The dashed and solid lines corresponds respectively to the SCM and FP estimates. Clearly, the FP estimator is much more robust than the SCM, especially for smaller values of ρ. For example, at a recognition rate of 60%, the FP estimate provides a gain of 9.4 dB for $\rho = 0.5$.

4. APPLICATION ON SAR IMAGERY

Covariance matrix estimators are used for a wide variety of applications in signal and image processing, including direction of arrival and number of signals estimation for multi-antenna receiver algorithms [18], segmentation of medical images [19], ... In this section, we present an application of the proposed statistic to the classification of maritime pine forests SAR images according to the stand age. More experiments can be found in [20].

For this experiment, the database contains 350 complex L-band simulated polarimetric SAR images of pine tree forests grouped in four classes according to the stand age (less than 10 years, between 10 and 20 years, between 20 and 30 years, over 30 years). The dataset is obtained by using the Pol-SARproSim software [21] for a platform situated at 3580 meters and having an incidence angle of $45°$ with several resolutions: 0.5, 1, 2, 3 and 5 meters. For each image, three polarizations (HH, HV and VV) are considered and merged to form a 3-dimensional array. This array is then filtered using the stationary wavelet transform with the Daubechies' db4 wavelet. Let N_{sc} and N_{or} be respectively the number of scales and orientations of the wavelet decomposition. Since the subbands of the wavelet decomposition are assumed independent, the geodesic distance between two images I_1 and I_2 can be expressed as a function of the geodesic distance computed on each subbands as:

$$GD^2(I_1 \| I_2) = \sum_{s=1}^{N_{sc}} \sum_{o=1}^{N_{or}} GD^2(\hat{\theta}_{1,s,o} \| \hat{\theta}_{2,s,o}) \qquad (8)$$

where $\hat{\theta}_{1,s,o}$ corresponds to the maximum likelihood estimate of θ_1 for the subband at scale s and orientation o.

To model the wavelet subband, various stochastic models have been proposed in the literature. Among them, the
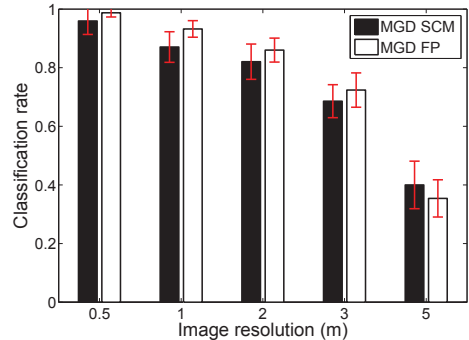
Fig. 4. Classification results by using the S_{GD} along with the SCM and FP estimates for simulated SAR data with wavelet coefficients modeled by MGDs.

univariate generalized Gaussian distribution (GGD) has been found to accurately model the empirical histogram of wavelet subband [3]. More recently, many authors have proposed to model the spatial dependency of wavelet coefficients. To this aim, the wavelet coefficients located around the spatial neighborhood 3×3 of the current spatial position are clustered in a random vector $\mathbf{x}$. The realizations of those vectors $\mathbf{x}$ can further be modeled by elliptical distributions [4, 5], copula based models [6, 7], etc. In this paper, we focus on the MGD model which admits a close-form expression for the geodesic distance recalled in (4) and the SCM and the FP estimates are analyzed for the estimation of the covariance matrices.

For the classification purpose, the database is divided into a training and testing set by a cross validation procedure. The retrieve performances are computed by using a k Nearest Neighbor classifier, with $k = 5$ and the results are shown in Fig. 4. As observed, for very high resolution images, the performance are the highest. Note also that, for high resolution images, the FP estimate improves the classification results over the SCM estimate.

5. CONCLUSION

In this paper, a statistical hypothesis test based on the geodesic distance and on the FP estimate has been introduced for image classification. Various experiments have been proposed to analyze this statistical test in the case of zero-mean MGDs on both simulated dataset and synthetic SAR images. A comparison between the SCM and the FP estimates has been performed. First, simulation results have shown that the dataset should contain at least 50 observations to ensure that the statistic follows a chi-squared distribution under the null hypothesis H_0, when using the SCM estimate. Second, the noise robustness has been analyzed as a function of the noise variance σ^2. Finally, an image processing application for the classification of SAR images has illustrated the potential of the proposed classifier. Further works will include the extension of the statistical hypothesis test to other multivariate models.

6. REFERENCES

[1] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[2] J.-F. Aujol, G. Aubert, and L. Blanc-Feraud, "Wavelet-based level set evolution for classification of textured images," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1634–1641, 2003.

[3] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, vol. 11, pp. 146–158, 2002.

[4] L. Bombrun, Y. Berthoumieu, N.-E. Lasmar, and G. Verdoolaege, "Mutlivariate texture retrieval using the geodesic distance between elliptically distributed random variables," in *IEEE International Conference on Image Processing*, 2011, pp. 3637–3640.

[5] G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized Gaussian models," *Journal of Mathematical Imaging and Vision*, vol. 43, no. 3, pp. 180–193, 2012.

[6] R. Kwitt and A. Uhl, "Lightweight probabilistic texture retrieval," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 241–253, 2010.

[7] Y. Stitou, N.-E. Lasmar, and Y. Berthoumieu, "Copulas based multivariate gamma modeling for texture classification," in *IEEE International Conference on Acoustic Speech and Signal Processing*, 2009, pp. 1045–1048.

[8] S. Kullback and R.A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[9] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 186, no. 1007, pp. 453–461, 1946.

[10] M. Kupperman, *Further applications of information theory to multivariate analysis and statistical inference*, Ph.D. thesis, George Washington University, 1957.

[11] M. Salicru, D. Morales, M.L. Menendez, and L. Pardo, "On the applications of divergence type measures in testing statistical hypotheses," *Journal of Multivariate Analysis*, vol. 51, no. 2, pp. 372 – 391, 1994.

[12] A. D. C. Nascimento, R. J. Cintra, and A. C. Frery, "Hypothesis testing in speckled data with stochastic distances," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, pp. 373–385, 2010.

[13] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[14] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 03 1987.

[15] F. Gini and M. V. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.

[16] F. Pascal, P. Forster, J. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2206–2217, June 2008.

[17] G. Vasile, J. Ovarlez, F. Pascal, and C. Tison, "Coherency matrix estimation of heterogeneous clutter in high-resolution polarimetric sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 1809–1826, April 2010.

[18] E. Ollila and V. Koivunen, "Robust antenna array processing using M-estimators of pseudo-covariance," in *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, Sept 2003, vol. 3, pp. 2659–2663.

[19] R. de Luis-García, C-F. Westin, and C. Alberola-López, "Gaussian mixtures on tensor fields for segmentation: Applications to medical imaging," *Computerized Medical Imaging and Graphics*, vol. 35, no. 1, pp. 16–30, 01 2011.

[20] I. Ilea, L. Bombrun, G. Germain, I. Champion, R. Terebes, and M. Borda, "Statistical hypothesis test for maritime pine forest SAR images classification based on the geodesic distance," in *IEEE International Geoscience and Remote Sensing Symposium*, 2015.

[21] M. L. Williams, "PolSARproSim: A Coherent, Polarimetric SAR Simulation of Forests for PolSARPro. Design Document ans Algorithm Specification (u1.0) ," 2006.