$$\hat{C}_{N+1} = \frac{N}{N+1}\hat{C}_N + \frac{N}{(N+1)^2}(x_{N+1} - \hat{\mu}_N)(x_{N+1} - \hat{\mu}_N)^T$$

[0058] By using the above expressions for the estimates of the mean and covariance and the matrix inversion lemma, the equation D^2 for an N-sample reference set and an (N+1)th test point becomes:

$$D^2 = \delta_N + 1^T \hat{C}_N + 1^{-1} \delta_N + 1$$

[0059] where

$$\delta_{N+1} = \frac{N}{N+1}(x_{N+1} - \hat{\mu}_N)$$

and

$$\hat{C}_{N+1}^{-1} = \frac{N+1}{N}\hat{C}_N^{-1} - \frac{C_N^{-1}\delta_{N+1}\delta_{N+1}^T\hat{C}_N^{-1}}{N + \frac{N}{N+1}\delta_{N+1}^T\hat{C}_N^{-1}\delta_{N+1}}$$

Hence,

$$D^2 = \frac{N+1}{N}D_{N+1,N}^2\left\{1 - \frac{D_{N+1,N}^2}{D_{N+1,N}^2 + N + 1}\right\}$$

[0060] where $D^2_{N+1,N}$ is $\delta_N+1^T\hat{C}_N^{-1}\delta_N+1$. Hence, a new point x_N+1 can be tested against an estimated and assumed normal distribution for a common estimated mean and covariance. It should be noted that employing the log-likelihood ratio for the one-class hypothesis test, the test statistic can be derived directly.

[0061] In practice, when the parametric D^2 test is used, N reference notes are scanned and segmented. Each of these segments is classified by determining a value for D^2 for every one of the reference notes and using this to set a reference threshold value for each segment. This threshold value is used to determine whether corresponding segments of subsequent test notes are valid.

[0062] The above analysis is based on the assumption that the feature vectors are distributed as multivariate Gaussian. Often this does not hold in practice, although may be an appropriate pragmatic choice in many applications. However, this assumption can be relaxed and arbitrary densities can be considered. The density under a mixture model has the following standard form:

$$p(x) = \sum_{j=1}^{M} p(x \mid j)P(j)$$

[0063] where P(j), j=1, . . . , M are the mixing parameters. These are chosen to satisfy the constraints

$$\sum_{j=1}^{M} P(j) = 1$$

[0064] and $P(j)\geqq 0$. The component density functions $p(x|j)$ are normalized so that $\int p(x|j)dx=1$. As a specific example, Gaussian mixture models are used, so that:

$$p(x \mid j) = \frac{1}{(2\pi)^{p/2}|\sum j|^{1/2}}\exp\left\{-\frac{1}{2}(x - \mu_j)^T \sum_j^{-1} (x - \mu_j)\right\}$$

[0065] The required parameters can be estimated using the Expectation Maximization (EM) algorithm. A technique for doing this is described in "Neural Networks for Pattern Recognition" by Bishop, Oxford University Press, New York (1995). This density can then be employed in computing the log-likelihood ratio. Unlike the case for the multivariate Gaussian distribution there is no analytic distribution for the test statistic λ under the null-hypothesis. To obtain the otherwise non-analytic null distribution under the mixture of Gaussian density, bootstrap methods can be employed. By doing this, the various critical values of λ_{crit} can be established from the empirical distribution obtained. It can be shown that in the limit as N tends to infinity, the likelihood ratio can be estimated by the following:

$$\lambda = \frac{\sup_{\theta\in\Theta} L_0(\theta)}{\sup_{\theta\in\Theta} L_1(\theta)} \to p(x_{N+1}; \hat{\theta}_N)$$

[0066] where $p(x_{N+}1;\hat{\theta}_N)$ denotes the probability density of x_{N+1} under the model estimated by the original N samples. After generating B bootstrap samples from the reference data set and using each of these to estimate the parameters of the mixture distribution $\hat{\theta}^i{}_N$, B bootstrap replicates of the test statistic $\lambda^i{}_{crit}$ (i=1, . . . B) can be obtained by randomly selecting an (N+1)th sample and computing $p(x_N+1;\hat{\theta}^i{}_N)\approx\lambda^i{}_{crit}$. By ordering $\lambda^i{}_{crit}$ in ascending order, the critical value α can be defined to reject the null-hypothesis at the desired threshold or significance level if λ is less than or equal to λ_α, where λ_α is the jth smallest value of $\lambda^i{}_{crit}$ and $\alpha=j/(B+1)$. By scanning N different notes; segmenting the images and calculating λ for each segment to determine a reference threshold, it is possible to validate subsequent notes by determining whether one or more of the test segments is within the tolerance level of what is acceptable.

[0067] Any of the above classifiers could be used to validate banknotes or other image rich documents. In any case, when the required one-class classifier is selected, it is trained on each segment or sub-region of the note, so that boundary conditions for each segment can be defined, that is limits that define what is regarded as being an acceptable or unacceptable variation from the reference segment. The classifier for the ith sub-region is denoted as D_i, where i 1,