ACCELERATION OF THE EXPECTATION-MAXIMIZATION ALGORITHM FOR A TWOFOLD GAUSSIAN MIXTURE MODEL BY USING THE HISTOGRAM OF THE OBSERVATIONS INSTEAD OF THE OBSERVATIONS

Evaluation of its Accuracy by Generated Histograms

J. Bruijns

Philips Research, High Tech Campus 36, 5656 AE, Eindhoven, The Netherlands

Abstract: Volume representations of blood vessels acquired by 3D rotational angiography are very suitable for diagnosing a stenosis or an aneurysm. For optimal treatment, physicians need to know the shape of the diseased vessel parts. Binary segmentation by thresholding is the first step in our shape extraction procedure. Assuming a twofold Gaussian mixture model, the model parameters (and thus the threshold for binary segmentation) can be extracted from the observations (i.e. the gray values) by the Expectation-Maximization (EM) algorithm. Since the EM algorithm requires a number of iterations through the observations, and because of the large number of observations, the EM algorithm is very time-consuming. Therefore, we developed a method to apply the EM algorithm to the histogram of the observations, requiring a single pass through the observations and a number of iterations through the much smaller histogram. This variant gives almost the same results as the original EM algorithm, at least for our clinical volumes. We have used this variant for an evaluation of the accuracy of the EM algorithm: the maximum relative error in the mixing coefficients was less than 7%, the maximum relative error in the parameters of the two Gaussian components was less than 2.5%.

1 INTRODUCTION

Volume representations of blood vessels acquired by 3D rotational angiography after injection with a contrast agent (Kemkers et al., 1998; Moret et al., 1998) have a clear distinction in gray values between tissue and vessel voxels. Therefore, these volume representations are very suitable for diagnosing a stenosis, a local narrowing of a vessel caused for example by cholesterol (see Figure 1), or an aneurysm, a local widening of a vessel caused by a weak vessel wall (see Figure 2).

For optimal treatment of a stenosis or an aneurysm, physicians need to know the shape parameters (e.g. vessel diameter, aneurysm volume) in the neighborhood of the diseased vessel parts. The starting point for many shape extraction methods is a segmented volume. Such a segmented volume can be created by thresholding. Voxels with a gray value below the threshold are classified as tissue voxels. Voxels with a gray value above the threshold are classified as vessel voxels.

To eliminate inter- and intra-operator variations this threshold should be extracted automatically from

Figure 1: A stenosis inside the white rectangle.

the data. A commonly used starting point for automatic threshold extraction is a twofold Gaussian mixture model (Kittler and Illingworth, 1986; Gan et al., 2004a; Gan et al., 2004b; Gan et al., 2005). After the parameters of this model are extracted from the observations (i.e. the gray values) by the Expectation-Maximization algorithm (Bilmes, 1997; Dempster et al., 1977), the optimal threshold is given by the in-

Figure 2: An aneurysm inside the white rectangle.

tersection point of the two Gaussian components.

Since the Expectation-Maximization (EM) algorithm requires a number of iterations through the observations, and because of the large number of observations (e.g. 512x512 pixels or 512x512x512 voxels), the EM algorithm is very time-consuming. To accelerate the extraction of the parameters of a twofold Gaussian mixture model by the EM algorithm, we developed a method to apply the EM algorithm to the histogram of the observations, requiring a single pass through the observations and a number of iterations through the much smaller histogram. The original EM algorithm is described in Section 3.1, our histogram variant in Section 3.2. Our histogram variant gives almost the same results as the original EM algorithm, at least for our clinical volumes (reported in Section 3.3). We have used our histogram variant for an evaluation of the accuracy of the EM algorithm (given further detail in Section 4).

2 RELATED WORK

Otsu (Otsu, 1979) proposed a method for the selection of a threshold for segmentation of a gray-value image in two classes based on the between-class variance. Given a gray-level histogram the threshold is chosen so that the between-class variance is maximized.

Kittler et al. (Kittler and Illingworth, 1986) presented a method for the selection of a threshold for segmentation of a gray value image in two classes based on the assumption that the gray values of the two classes have a Gaussian distribution. Given a gray-level histogram the threshold is chosen so that the Bayesian classification rule is minimized.

Wilson et al. (Wilson and Noble, 1999) presented a method for adaptive statistical segmentation of the vessel voxels in time-of-flight MRA data.

First, they estimate the parameters of a convex model of three probability distributions (Gaussian for tissue and uniform for blood flow) by the Expectation-Maximization (EM) algorithm. Next, the resulting distributions are used to classify the vessel voxels. This process is applied to a hierarchical subdivision so that local thresholds are used for the classification step.

Liao et al. (Liao et al., 2001) introduced a method for multilevel thresholding. They modified the method of Otsu (Otsu, 1979), so that the thresholds can be computed much faster. First, they develop a criterion for maximizing a modified between-class variance. Next, a recursive algorithm is designed based on this modified between-class variance which gives the same thresholds as the method of Otsu. Finally, a look-up table is used for efficient computation.

Frederix et al. (Frederix and pauwels, 2004; Frederix, 2005) proposed a method for estimating the Gaussian mixture model with a variable number of Gaussian probability density functions. The number of Gaussian components is increased until the distribution of the Gaussian mixture model is compatible with the empirical data distribution.

Gan et al. (Gan et al., 2004a; Gan et al., 2004b; Gan et al., 2005) introduced a method for vascular segmentation of 3D rotational angiography volumes based on MIP (Maximum Intensity Projections) images. Applying the Expectation-Maximization (EM) algorithm for a twofold Gaussian mixture model, they estimate a threshold from a MIP image because the standard deviation of the individual classes in a MIP image is smaller. This threshold is used to classify the voxels belonging to this MIP image (i.e. the voxels with the maximum gray value in the projection direction). After the resulting vessel voxels are labelled in a corresponding binary volume, their gray values in the original volume is set to zero, and the procedure is repeated until over fitting of the non-vessel class is detected. This iterative algorithm is applied in the three axis directions and the resulting segmentations are combined by minimizing an energy functional for a spatially smooth result.

3 THE EM ALGORITHM FOR A TWOFOLD GAUSSIAN MIXTURE MODEL

A twofold Gaussian mixture model is given by the following convex combination:

$$f(x) = \sum_{k=1}^{2} \pi_k G(x; \mu_k, \sigma_k) \tag{1}$$

with $0 \le \pi_k \le 1$, $\pi_1 + \pi_2 = 1$ and $G(x; \mu_k, \sigma_k)$ the Gaussian probability density function.

The mean and variance of this twofold Gaussian mixture model are also convex combinations of the means and variances of the two Gaussian components:

$$\mu = \sum_{k=1}^{2} \pi_k * \mu_k$$
$$\sigma^2 + \mu^2 = \sum_{k=1}^{2} \pi_k * (\sigma_k^2 + \mu_k^2) \tag{2}$$

3.1 The EM Algorithm using the Observations

Given a set of observations $\{x_i, i \in [1, N]\}$ the parameters of a twofold Gaussian mixture model can be estimated by the EM algorithm (Bilmes, 1997). This is an iterative algorithm consisting of an Expectation step followed by a Maximization step. In the Expectation step the membership functions $\{P_{k,i}, k \in [1, 2], i \in [1, N]\}$ are computed:

$$P_{k,i} = \frac{\pi_k G(x_i; \mu_k, \sigma_k)}{\sum_{k=1}^{2} \pi_k G(x_i; \mu_k, \sigma_k)} \tag{3}$$

From this equation the following properties of the membership functions can be derived:

$$0 \le P_{k,i} \le 1$$
$$\sum_{k=1}^{2} P_{k,i} = 1 \tag{4}$$

These membership functions are used to compute new values for the parameters in the Maximization step (similar as for a discrete probability distribution):

$$\pi_k = \frac{\sum_{i=1}^{N} P_{k,i}}{N}$$
$$\mu_k = \frac{\sum_{i=1}^{N} P_{k,i} x_i}{\sum_{i=1}^{N} P_{k,i}} = \frac{\sum_{i=1}^{N} P_{k,i} x_i}{N * \pi_k} \tag{5}$$
$$\sigma_k^2 + \mu_k^2 = \frac{\sum_{i=1}^{N} P_{k,i} x_i^2}{\sum_{i=1}^{N} P_{k,i}} = \frac{\sum_{i=1}^{N} P_{k,i} x_i^2}{N * \pi_k}$$

The new values for the parameters are used to compute new values for the membership functions in the following Expectation step and so on. The algorithm stops when the change in the membership functions is very small.

Multiplying the formulae for μ_k and σ_k^2 in Equation 5 with π_k, summing over the components and using Equation 4 gives the following properties similar to those of Equation 2:

$$\sum_{k=1}^{2} \pi_k * \mu_k = \frac{\sum_{i=1}^{N} x_i}{N} = \bar{x}$$
$$\sum_{k=1}^{2} \pi_k * (\sigma_k^2 + \mu_k^2) = \frac{\sum_{i=1}^{N} x_i^2}{N} = s^2 + \bar{x}^2 \tag{6}$$

with $\bar{x}$ the arithmetic mean and s the standard deviation computed from the observations.

The EM algorithm is started by using the following somewhat arbitrary but for our clinical volumes effective initial values for the parameters:

$$\pi_k = 1/2$$
$$\mu_1 = \bar{x} - \alpha * s$$
$$\mu_2 = \bar{x} + \alpha * s \tag{7}$$
$$\sigma_k = s\sqrt{1 - \alpha^2}$$

This equation gives real values for the parameters, fulfilling the conditions of Equation 6, as long as $0 < \alpha < 1$. Which values for α have been tested and which one is selected for the final experiments, is given further detail in Section 4.1.

3.2 The EM Algorithm using the Histogram

In order to reduce computational time, we propose to use the histogram of the observations. If we have a histogram $\{h_l, l \in [1, M]\}$ instead of a set of observations $\{x_i, i \in [1, N]\}$, we can generate a set of observations from this histogram. For each bin we have h_l observations with the observation value l (the bin number). To avoid troubles caused by possible zero parts at the beginning and/or at the end of the histogram, the first bin B and the last bin E with a nonzero value are first computed:

$$(h_B > 0) \wedge ((h_l = 0)\forall(l \in [1, B-1]))$$
$$(h_E > 0) \wedge ((h_l = 0)\forall(l \in [E+1, M])) \tag{8}$$

In the Expectation step the membership functions $\{P_{k,l}, k \in [1, 2], l \in [B, E]\}$ are computed as follows:

$$P_{k,l} = \frac{\pi_k G(l; \mu_k, \sigma_k)}{\sum_{k=1}^{2} \pi_k G(l; \mu_k, \sigma_k)} \tag{9}$$

From this equation the following properties of the membership functions can be derived (similar as in Equation 4):

$$0 \leq P_{k,l} \leq 1$$
$$\sum_{k=1}^{2} P_{k,l} = 1 \qquad (10)$$

These membership functions are used to compute new values for the parameters in the Maximization step:

$$\pi_k = \frac{\sum_{l=B}^{E} h_l P_{k,l}}{\sum_{l=B}^{E} h_l} = \frac{\sum_{l=B}^{E} h_l P_{k,l}}{N}$$

$$\mu_k = \frac{\sum_{l=B}^{E} h_l P_{k,l} l}{\sum_{l=B}^{E} h_l P_{k,l}} = \frac{\sum_{l=B}^{E} h_l P_{k,l} l}{N * \pi_k} \qquad (11)$$

$$\sigma_k^2 + \mu_k^2 = \frac{\sum_{l=B}^{E} h_l P_{k,l} l^2}{\sum_{l=B}^{E} h_l P_{k,l}} = \frac{\sum_{l=B}^{E} h_l P_{k,l} l^2}{N * \pi_k}$$

Multiplying the formulae for μ_k and σ_k^2 in Equation 11 with π_k, summing over the components and using Equation 10 gives the following properties similar to those of Equation 6:

$$\sum_{k=1}^{2} \pi_k * \mu_k = \frac{\sum_{l=B}^{E} h_l l}{N}$$

$$\sum_{k=1}^{2} \pi_k * (\sigma_k^2 + \mu_k^2) = \frac{\sum_{l=B}^{E} h_l l^2}{N} \qquad (12)$$

The EM algorithm is started by using the same initial values for the parameters as used in the previous section (see Equation 7) but now with $\bar{x}$ and s estimated from the histogram instead of from the observations.

3.3 Observations versus Histogram Results

To test the EM algorithm based on histograms, we have computed the parameters of a twofold Gaussian mixture model from a set of observations, i.e. the 16 bits gray values of a MIP (Maximum Intensity Projection) image and from the corresponding gray value histogram (256 bins between the minimum and maximum gray value of the MIP image). We have used 84 clinical volume datasets (58 of them with a resolution of 128x128x128 voxels, 25 of them with a resolution of 256x256x256 voxels and 1 with a resolution of 512x512x512 voxels), acquired with the 3D Integris system (Philips-Medical-Systems-Nederland, 2001). The voxel size varies between 0.2 and 1.2 millimeter.

The maximum relative differences between the parameters of the twofold Gaussian mixture models,

computed using the observations, and those, computed using the gray value histograms, are given in Table 1.

Table 1: Maximum relative differences between the parameters of the twofold Gaussian mixture models, computed using the observations, and those, computed using the histograms. k is the number of the Gaussian component.

k	π	μ	σ
1	0.1684%	1.0099%	0.8629%
2	0.3093%	0.5420%	0.1797%

These results show that even with a histogram of 256 bins (i.e. 16 bits gray values are reduced to 8 bits gray values), the parameters of the twofold Gaussian mixture models, derived from the 8 bits histogram, have almost the same value as the parameters of the twofold Gaussian mixture models, derived from the original 16 bit gray values of the MIP images.

Since we have used the gray values of the MIP images (a 2D dataset with $\Theta(n^2)$ observations) instead of the gray values of the original volumes (a 3D dataset with $\Theta(n^3)$ observations), we can only give a lower bound for the saving in computing time. The EM algorithm based on histograms (including the time to generate the histogram) was at least 1.8 times faster, on average 3.9 times faster and at most 8.2 times faster.

4 EVALUATION OF THE ACCURACY OF THE EM ALGORITHM

4.1 Experiments

The small relative differences (see Table 1) do not guarantee that the parameters of the twofold Gaussian mixture models, computed by the EM algorithm, are accurate. Our histogram variant facilitates the evaluation of the accuracy of the EM algorithm. Therefore, we have generated histograms (256 bins) for a series of twofold Gaussian mixture models with the following parameters:

$$\pi_1 = \{0.50, 0.52, ..., 0.68, 0.70\}$$
$$\pi_2 = 1 - \pi_1$$
$$\mu_1 = 0.30$$
$$\mu_2 = \{0.40, 0.42, ..., 0.58, 0.60\}$$
$$\sigma_1 = 0.05$$
$$\sigma_2 = \{0.050, 0.055, ..., 0.095, 0.100\}$$

We have combined all possible parameters, giving a total of 1331 Gaussian mixture models. The chosen parameters for the Gaussian probability functions (i.e. μ and σ) are multiplied with 256 before the generation of the histogram so that the domain of these parameters is the same as the bin numbers. The generated histograms are comparable with gray value histograms derived from a MIP image or a gray value volume.

Two examples are shown in Fig. 3 and Fig. 4. In the first example the Gaussian mixture model results in a single wide cluster. In the second example the Gaussian mixture model results in two overlapping clusters. Note that in this last case the intersection of the two Gaussian probability density functions does not coincide with the position of the minimum between the two clusters.
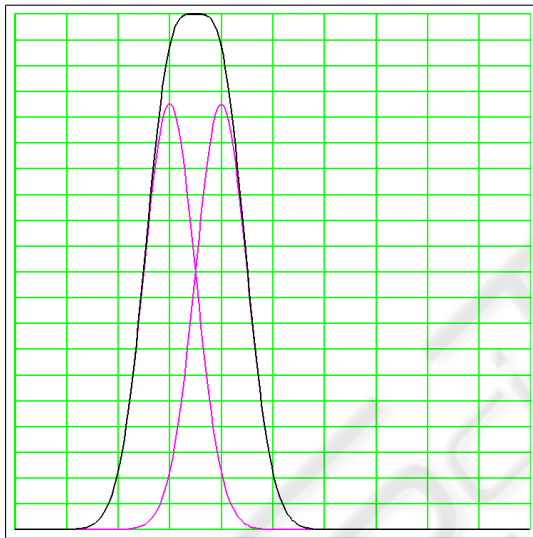
Figure 3: Twofold Gaussian mixture model resulting in a single wide cluster. The black curve is the generated histogram. The two magenta curves are the Gaussian density functions.

As already told in Section 3.1, using Equation 7 results in initial values for the parameters which fulfill the conditions of Equation 6, provided that $0 < \alpha < 1$. We have tested the following values for α:

$$\alpha = \{0.1, 0.2, ..., 0.9, 0.99, 0.999, 0.9999\}$$

In all cases, the mixing coefficient π_2 gave the greatest relative error. The relative difference between the maximum relative errors was less than 0.02%. So, all tested values of α gave the same results for the final parameters. The only difference was the maximum number of iterations. Higher values of α gave a lower number of iterations, except $\alpha = 0.9999$. To avoid possible future problems with values for α very
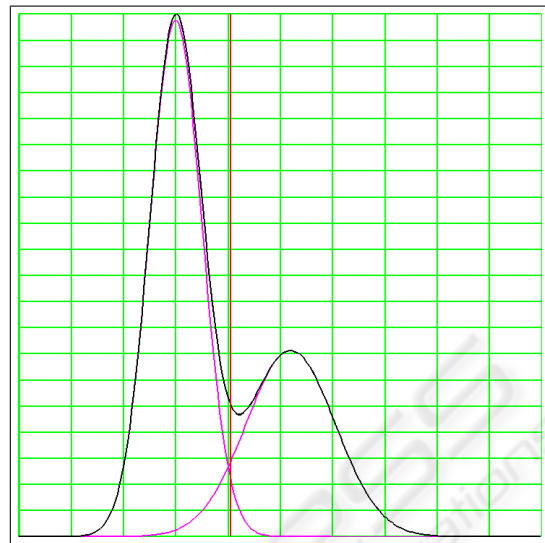

Figure 4: Twofold Gaussian mixture model resulting in a double cluster.

close to 1, we have used $\alpha = 0.9$ for all experiments (including for the comparison between observations and histograms described in Section 3.3).

4.2 Results

The extracted versus the chosen parameters π_1, π_2, μ_2 and σ_2 of the twofold Gaussian mixture models are shown in Fig. 11 upto and including Fig. 15. The straight diagonal line from left under to upper right gives the ideal relation (i.e. the extracted parameter is exactly equal to the chosen parameter). As is clear from these pictures, the EM algorithm results in a small deviation of the least-square line through the extracted parameters π_1 and π_2 compared to the ideal relation line.

Table 2: The maxima of the relative deviations between the chosen and extracted parameters of the twofold Gaussian mixture models. Initial values derived from histograms using Equation 7. k is the number of the Gaussian component.

k	π	μ	σ
1	3.3716%	0.3594%	0.8881%
2	6.5987%	1.1691%	2.2362%

The maxima of the relative deviations between the chosen and extracted parameters of the twofold Gaussian mixture models are given in Table 2. In this case, the initial values for the iterative EM algorithm are derived from the histograms using Equation 7. These maxima reveal that the EM algorithm for a twofold Gaussian mixture model results in small relative errors for all parameters.

To verify our EM algorithm we have also used the chosen parameters as initial values. The maxima of the relative deviations between the chosen and extracted parameters of the twofold Gaussian mixture models are given in Table 3. So, if the chosen parameters of the twofold Gaussian mixture models are used as initial values, the EM algorithm result in very small relative errors.

Table 3: The maxima of the relative deviations between the chosen and extracted parameters of the twofold Gaussian mixture models. Chosen parameters used as initial values. k is the number of the Gaussian component.

k	π	μ	σ
1	0.0015%	0.0000%	0.0000%
2	0.0020%	0.0031%	0.0249%

The number of iterations for the experiments with the initial values derived from the histograms, varied between 4 and 910 with a mean value of 77.3. As can be expected, the number of iterations for the experiments with the initial values equal to the chosen parameters was much lower, namely always 2.

4.3 Discussion

As is clear from Fig. 5 upto and including Fig. 10, the relative errors in the extracted parameters of the twofold Gaussian mixture models are related to the differences between the chosen mean coefficients μ_1 and μ_2. So, as can be expected, the EM algorithm may give better results if the mean coefficients μ_1 and μ_2 of the two Gaussian components differ more.

But for our gray value segmentation algorithms the only important parameter is the segmentation threshold (i.e. the intersection point of the two Gaussian components). The extracted versus the "chosen" threshold is shown in Fig. 13. The maxima of the absolute and relative deviations are 0.3922 (also a bin number between 1 and 256) respectively 1.1111%. So, the deviations in the threshold are also small.

To facilitate value judgment of the EM threshold deviations, we extracted the Kittler threshold (Kittler and Illingworth, 1986) from the generated histograms. The Kittler versus the "chosen" threshold is shown in Fig. 16. The maxima of the absolute and relative deviations are 2.7451 (also a bin number between 1 and 256) respectively 7.3684%. Comparing Fig. 16 with Fig. 13, and comparing the Kittler deviations with the EM deviations, reveals that the EM threshold is more accurate than the Kittler threshold.

REFERENCES

Bilmes, J. (1997). A gentle tutorial on the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical Report ICSI-TR-97-021, University of Berkeley, Berkeley, CA, USA.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. Series B (Methodological)*, 39(1):1–38.

Frederix, G. (2005). *Beyond Gaussian Mixture Models: Unsupervised Learning with applications to Image Analysis*. PhD thesis, Katholieke Universiteit of Leuven, Belgium.

Frederix, G. and pauwels, E. (2004). A statistically principled approach to histogram segmentation. Technical Report Report PNA-E0401, CWI, Amsterdam, The Netherlands.

Gan, R., Chung, A., Wong, W., and Yu, S. (2004a). Vascular segmentation in three-dimensional rotational angiography based on maximum intensity projections. In *Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 133–136, Arlington, VA, USA.

Gan, R., Wong, W., and Chung, A. (2005). Statistical cerebrovascular segmentation in three-dimensional rotational angiography based on maximum intensity projections. *Med. Phys.*, 32(9):3017–3028.

Gan, R., Wong, W., Chung, A., and Yu, S. (2004b). Statistical cerebrovascular segmentation in three-dimensional rotational angiography based on maximum intensity projections. In *Proc. CAR*, pages 195–200, Chicago, USA.

Kemkers, R., de Beek, J. O., Aerts, H., Koppe, R., Klotz, E., Grasse, M., and Moret, J. (1998). 3D-rotational angiography: First clinical application with use of a standard philips C-Arm system. In *Proc. CAR*, pages 182–187, Tokyo, Japan.

Kittler, J. and Illingworth, J. (1986). Minimum error thresholding. *Pattern Recognition*, 19(1):41–47.

Liao, P., Chen, T., and Chung, P. (2001). A fast algorithm for multilevel thresholding. *Journal of Information Science and Engineering*, 17(5):713–727.

Moret, J., Kemkers, R., de Beek, J. O., Koppe, R., Klotz, E., and Grass, M. (1998). 3D rotational angiography: Clinical value in endovascular treatment. *Medicamundi*, 42(3):8–14.

Otsu, N. (1979). A threshold selection method from gray level histograms. *IEEE Trans. Syst., Man, Cybern.*, 9(1):62–66.

Philips-Medical-Systems-Nederland (2001). INTEGRIS 3D-RA. instructions for use. release 2.2. Technical Report 9896 001 32943, Philips Medical Systems Nederland, Best, The Netherlands.

Wilson, D. and Noble, J. (1999). An adaptive segmentation algorithm for time-of-flight MRA data. *IEEE Trans. Med. Imag.*, 18(10):938–945.
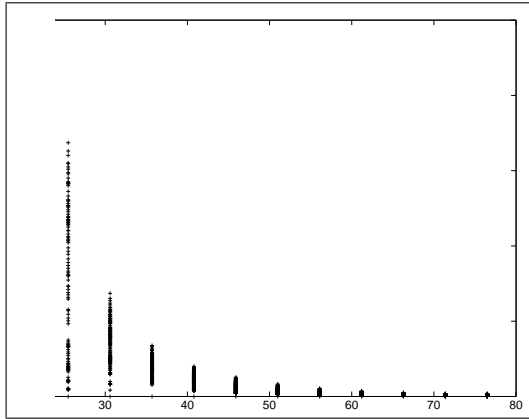
APPENDIX

Pictures

Figure 5: Relative error in % of the extracted π_1 versus the chosen $\mu_2 - \mu_1$.
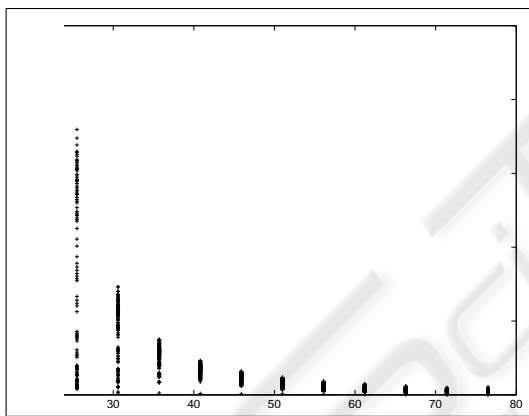

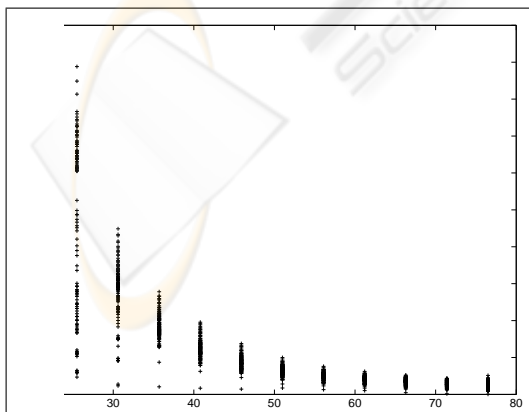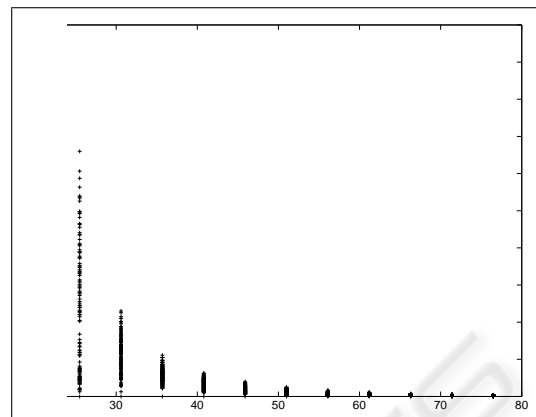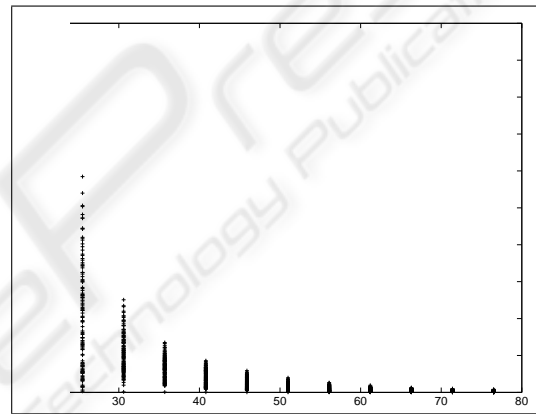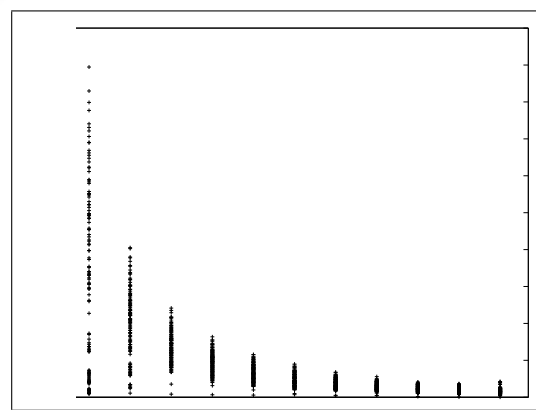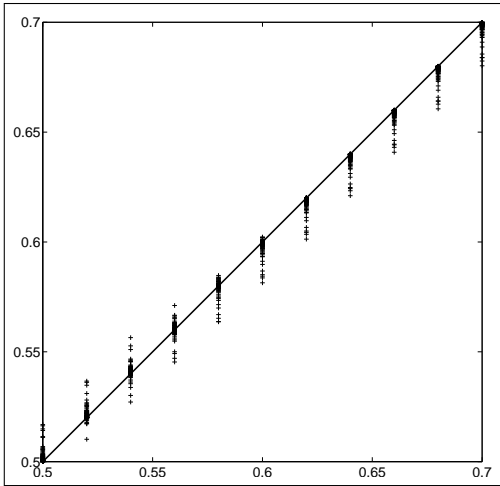Figure 6: Relative error in % of the extracted μ_1 versus the chosen $\mu_2 - \mu_1$.

Figure 7: Relative error in % of the extracted σ_1 versus the chosen $\mu_2 - \mu_1$.

Figure 8: Relative error in % of the extracted π_2 versus the chosen $\mu_2 - \mu_1$.

Figure 9: Relative error in % of the extracted μ_2 versus the chosen $\mu_2 - \mu_1$.

Figure 10: Relative error in % of the extracted σ_2 versus the chosen $\mu_2 - \mu_1$.

235

Figure 11: The extracted versus the chosen π_1.
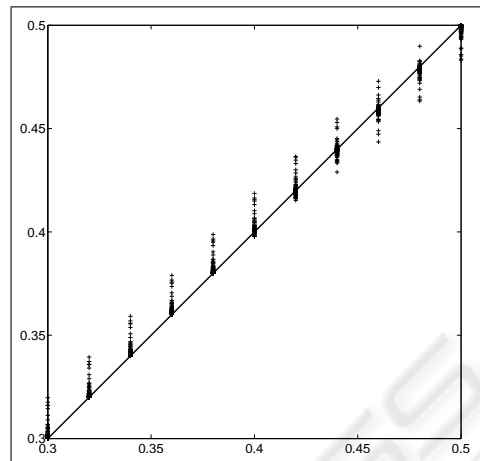
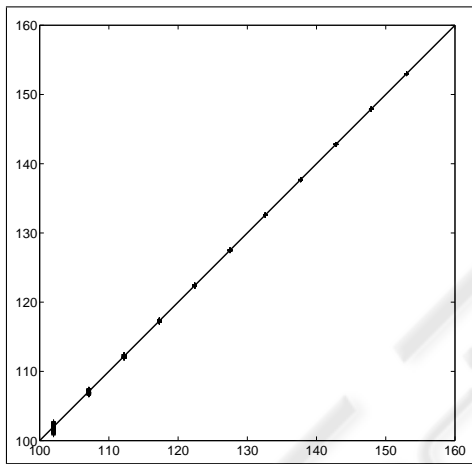Figure 14: The extracted versus the chosen π_2.

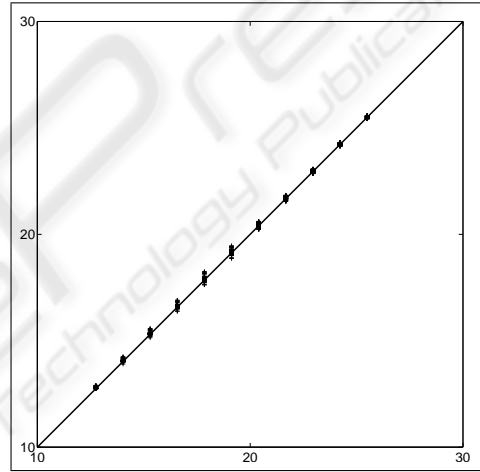Figure 12: The extracted versus the chosen μ_2 as bin number between 1 and 256.

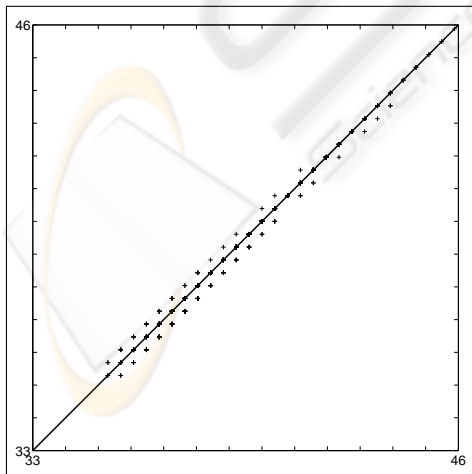Figure 15: The extracted versus the chosen σ_2 as bin number between 1 and 256.

Figure 13: Extracted versus "chosen" threshold (bin number between 1 and 256) by EM algorithm.
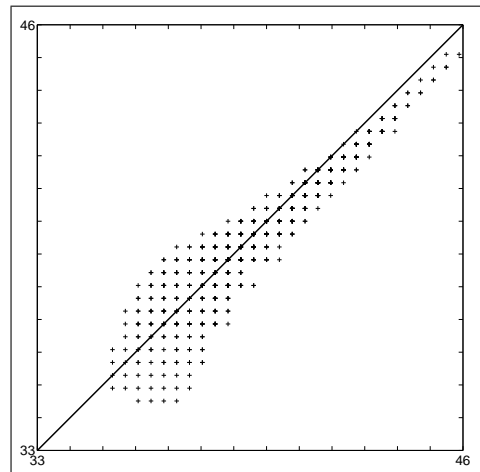

Figure 16: Extracted versus "chosen" threshold (bin number between 1 and 256) by Kittler algorithm.