

Robust scale estimate for the generalized Gaussian Probability Density Function

Rozenn Dahyot & Simon Wilson

Department of Statistics Trinity College Dublin Ireland

dahyot@mee.tcd.ie, swilson@tcd.ie

September 2005

Abstract

This article proposes a robust way to estimate the scale parameter of a generalised centered Gaussian mixture. The principle relies on the association of samples of this mixture to generate samples of a new variable that shows relevant distribution properties to estimate the unknown parameter. In fact, the distribution of this new variable shows a maximum that is linked to this scale parameter. Using non-parametric modelling of the distribution and the MeanShift procedure, the relevant peak is identified and an estimate is computed. The whole procedure is fully automatic and does not require any prior settings. It is applied to regression problems, and digital data processing.

1 Introduction

Many problems in computer vision involve the separation of a set of data into two classes, one of interest in the context of the application and the remaining one. For instance, edge detection in images requires the thresholding of the gradient magnitude to discard noisy flat areas from the edges. The challenge is then to automatically select the appropriate threshold [21].

Regression problems also involve the simultaneous estimation of the variance or standard deviation of the residuals/errors. The presence of a large number of outliers makes difficult the estimation of the parameters of interest. Performance of robust estimators are highly dependent on the setting of a threshold or scale parameter, to separate the good data (inliers) that fit the model, from the gross errors (outliers) [4]. The scale parameter, needed in M-estimation and linked to the scale parameter of the inliers residuals, is often set a priori or estimated by the Median Absolute Deviation. Applications of robust regression [4] are, for instance, line fitting [25], or camera motion estimation [3]. Those estimates also require the setting of a threshold (or scale parameter) to discard gross errors (outliers) from the relevant ones (inliers).

This paper proposes a solution for the estimation of the scale

parameter that can be used iteratively along with location parameter estimation [18, 4, 25] when there are many outliers, by assuming a mixture model. This method is based on the definition of two random variables Y and Z computed from the samples of variable X using a non-linear transformation. The distributions of those new variables have properties that allow us to define new estimates for the scale parameter. The computation of those new estimates requires us to detect a particular maximum over the distribution of the new variables. This is achieved using nonparametric kernel-based modelling of probability density functions. The resulting method is then both robust and unsupervised.

Section 2 presents related works. Section 3 presents our new estimates. Section 4 presents the computation. Paragraph 5 proposes to use this procedure for scale estimation iteratively with a location parameter estimation. Section 6 presents experimental results showing the robustness of the approach.

2 Related Works

Several works have been carried out on robust regression in the vision community [23], offering complementary views to statistics [14, 12]. Wang has recently proposed a clear overview in both domains [24], underlying several differences. In particular, in the vision community, the breakdown point is usually expected far below 50% of outliers to deal with real world applications, and proofs of robustness are usually inferred through experimental results. In statistics, formal proofs of robust estimator properties usually prevail (even if those are only valid under strict assumptions that may be scarcely encountered in practice) as it provides insights on the approaches [12].

We consider the problem of robust regression:

$$v = f(\mathbf{u}, \theta) + \epsilon$$

The mapping function f is assumed to be described by the vector θ (location parameter). From a set of observations $\{(v_i, \mathbf{u}_i)\}$, the goal of regression is to estimate θ .

Maximum likelihood is a popular estimation method. It relies on the modelling of the probability density function of the residual ϵ that expresses the error between each observation and its prediction by the mapping. Standard parametric modellings for the pdf of the residuals include Gaussian and Laplacian density functions [13]. Those models fail when gross errors or outliers occur in the observations. In this case, the pdf of the residuals can be expressed as a mixture:

$$\mathcal{P}(\epsilon|\theta, \sigma) = \mathcal{P}(\epsilon|\mathcal{C}, \theta, \sigma) \cdot \mathcal{P}(\mathcal{C}) + \mathcal{P}(\epsilon|\bar{\mathcal{C}}) \cdot \mathcal{P}(\bar{\mathcal{C}}) \quad (1)$$

where \mathcal{C} is the inlier class for the model designed by the location parameter θ and σ the scale parameter. The distribution of the inliers $\mathcal{P}(\epsilon|\mathcal{C}, \theta, \sigma)$ is modelled by a parametric expression, e.g. centered Laplacian [13] or Gaussian [25], depending on the location parameter θ to be estimated and the scale parameter σ also usually unknown. Those parametric models usually offers good description in real world applications.

Two main strategies have been proposed to overcome the lack of robustness. The first weights the residuals in the objective function to lessen the influence of gross error over the estimate. Weights are then a function of the residuals. The second approach consists in using sampling strategies and estimated from several randomly selected subsets of observations. The final selection is then performed by comparing each estimate and keeping the optimal one. In a way, it can be seen as a similar approach to weighting the residuals in taking null weights on left out data at each round. However, this approach is different as weights are set randomly on data and do not depend on the values of the residuals. M-estimators or R-estimators are examples of objective functions that used a weighting strategy [14, 12]. By their efficiency and their rather low computation cost, M-estimators in particular, have been widely applied to many computer vision applications such as camera motion estimation [20], object class learning [15], object detection [7] and recognition [2]. Using sampling has been made popular by Fischler and Bolles with the RANSAC estimator [10]. It has been successfully applied to camera calibration and image matching [10].

Both the M-estimator and RANSAC depend on the scale parameter σ that needs to be robustly assessed along with θ . Statistical solutions for the scale estimate include in particular the Median Absolute Deviation [12] and Least Median Square [18]. Those methods have been used for online estimation from current observations. It can also be set by hand [10], or learned a priori off-line, and then set once and for all in the context of specific applications [6, 13]. For the modelling of the distribution of outliers, $\mathcal{P}(\epsilon|\bar{\mathcal{C}})$, nonparametric representation using histograms have also been proposed [13] to define a more dedicated M-estimator for image matching.

However, in most applications, the scale parameter cannot be learned, and need to be estimated in parallel to θ with a

much lower breakdown than 50%. Recently several strategies involving nonparametric representations of the distribution of the residuals $\mathcal{P}(\epsilon|\theta, \sigma)$ have been proven efficient. For instance, Chen et al. [4] proposed to re-express the M-estimator objective function on the residuals, as a kernel density like estimator where the scale parameter is substituted by a bandwidth. After localising the main central peak, surrounding basins of attraction are detected heuristically and are used as thresholds to separate inliers from outliers. The segmented inliers are then processed further to reach an estimate of θ . Similarly in [25], the distribution of the residuals is modelled using a nonparametric kernel modelling. The central peak is assumed to be corresponding to the inliers, and is isolated from the rest of the pdf using a Meanshift based algorithm searching for surrounding valleys. From this rough classification, the scale parameter is then robustly estimated by computing the Median. This two step scale estimation (TSSE) is coupled with a location parameter θ estimation similar fashion to RANSAC [10].

However looking for minima in pdf where few data occur [4, 25], is a more difficult operation than looking for maxima. Searching to maximize a pdf is also a more natural way for parameter estimation. We then propose another nonparametric approach in the next section that estimates the unknown scale parameter by searching for particular maxima in specific distribution. Another difference from [4, 25] is that only a maximum localisation is necessary to infer the scale parameter. It is similar to the Hough transform approach in that peaks in the estimated distribution are directly related to the unknown parameters [11].

3 New Estimates for σ

Another approach for scale parameter estimation has been proposed in [8] based on some properties of independent samples from a Gaussian distribution. Indeed assuming a Gaussian distribution for the inliers, the distribution of square root of the sum of squares of several independent samples is a χ distribution that shows a maximum directly related to the scale parameter. This approach is further studied in this article.

3.1 Assumptions

Let us define the random variable X that follows a two class mixture model. We call *inliers* the data that belong to the class of interest: $x \in \mathcal{C}$, and *outliers* the other data $x \in \bar{\mathcal{C}}$. It is possible to make some relatively loose assumptions about the distribution of the class \mathcal{C} of interest, to allow its statistics to be estimated from an observed mixture distribution $\mathcal{P}_X(x)$ [13]. In this work, we assume the distribution of the inliers to be a

Generalized centered Gaussian [1]:

$$\mathcal{P}_X(x|\mathcal{C}, \sigma) = \frac{1}{2\Gamma(\alpha) \cdot \alpha \cdot \beta^\alpha} \exp\left[\frac{-|x|^{1/\alpha}}{\beta}\right] \quad (2)$$

with $\beta = \sigma^{1/\alpha} \cdot \left[\frac{\Gamma(\alpha)}{\Gamma(3\alpha)}\right]^{1/(2\alpha)}$

Setting $\alpha = 1$ (Laplacian law) and $\alpha = 1/2$ (Gaussian law) in equation (2), are two popular hypotheses [13, 8]. Assuming the shape parameter α known, we focus on the estimation of the scale σ .

3.2 Definition of new variables

The variables $Z = \sum_{n=1}^n |X_n|^{1/\alpha}$ and $Y = Z^\alpha$ are defined with independent random variables X_n that follow the same probability density function (2).

For $\mathbf{n} = 1$ and $Z = |X|^{1/\alpha}$, the pdf $\mathcal{P}_Z(z|\mathcal{C})$ corresponds to the gamma distribution:

$$\mathcal{G}_{Z|(\alpha, \beta)}(z) = \frac{z^{\alpha-1}}{\Gamma(\alpha) \cdot \beta^\alpha} \exp\left[-\frac{z}{\beta}\right], \quad z \geq 0 \quad (3)$$

When $\mathbf{n} > 1$, the variable Z is then defined by the sum of i.i.d. variables that follow a gamma distribution. Using the characteristic function of the gamma distribution $\Phi(t) = (1 - \imath\beta t)^{-\alpha}$, the characteristic function of $\mathcal{P}_Z(z|\mathcal{C}, \sigma)$ is:

$$\Phi_Z(t) = \prod_{n=1}^{\mathbf{n}} \Phi(t) = (1 - \imath\beta t)^{-\mathbf{n}\alpha} \quad (4)$$

By Inverse Fourier transform, the p.d.f $\mathcal{P}_Z(z|\mathcal{C}, \sigma)$ is the gamma function $\mathcal{G}_{Z|(\mathbf{n}\alpha, \beta)}(z)$. From the distribution of Z , it is easy to infer the pdf of Y :

$$\mathcal{P}_Y(y|\mathcal{C}, \sigma) = \frac{y^{(\mathbf{n}-1)}}{\alpha \cdot \Gamma(\mathbf{n}\alpha) \cdot \beta^{\mathbf{n}\alpha}} \exp\left[\frac{-y^{1/\alpha}}{\beta}\right], \quad y \geq 0 \quad (5)$$

The maximum of the distributions $\mathcal{P}_Z(z|\mathcal{C}, \sigma)$ and $\mathcal{P}_Y(y|\mathcal{C}, \sigma)$ can be then computed (ignoring the solution $z = 0$ as a special case):

$$\begin{cases} Z_{\max \mathcal{C}} = \beta \cdot (\mathbf{n}\alpha - 1), & \mathbf{n}\alpha > 1 \\ Y_{\max \mathcal{C}} = [(\mathbf{n}-1) \alpha \beta]^\alpha, & \mathbf{n} > 1 \end{cases} \quad (6)$$

Those maxima depend on the parameter σ by definition of β (cf. eq. (2)).

3.3 New estimates for σ

From equation (6), we propose to estimate σ using:

$$\begin{cases} \sigma_Z = \left(\frac{Z_{\max \mathcal{C}}}{\mathbf{n}\alpha - 1}\right)^\alpha \left[\frac{\Gamma(3\alpha)}{\Gamma(\alpha)}\right]^{1/2}, & \mathbf{n}\alpha > 1 \\ \sigma_Y = \frac{Y_{\max \mathcal{C}}}{(\mathbf{n}-1)^\alpha \cdot \alpha^\alpha} \left[\frac{\Gamma(3\alpha)}{\Gamma(\alpha)}\right]^{1/2}, & \mathbf{n} > 1 \end{cases} \quad (7)$$

The maximum of the distributions of Y and Z has first to be located. This can be made difficult by outliers occurring in

the observations such that the observed distribution for X is a mixture (as for Y and Z):

$$\mathcal{P}_X(x|\sigma) = \mathcal{P}_X(x|\sigma, \mathcal{C}) \cdot \mathcal{P}_X(\mathcal{C}) + \mathcal{P}_X(x|\bar{\mathcal{C}}) \cdot \mathcal{P}_X(\bar{\mathcal{C}}) \quad (8)$$

Depending on the proportion and the values of the outliers, the localisation of the maximum needed in the estimation gets more difficult. We assume that the relevant maximum for the estimation is the closest peak to zero in the distributions $\mathcal{P}_Y(y|\sigma)$ and $\mathcal{P}_Z(z|\sigma)$. Note that robust estimation using M-estimator for Gamma distribution has been proposed in the literature [17]. But this nonparametric method is shown to be more robust in section 6.

3.4 Remarks

Figure 1 presents the pdfs of the inlier class for Y and Z . Instinctively, the higher the maximum is, the better its localisation should be in the mixture of observations. A priori, considering the density of variable Y for the estimation of the scale parameter should then perform better than the variable Z . However on the other hand, the transformation from X to Z is spreading the range of variations of the observations, and consequently decreasing locally the density of the outliers. Expressions of

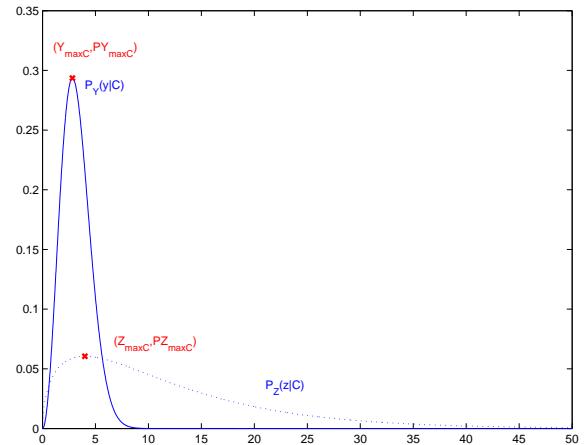


Figure 1: Probability density functions $\mathcal{P}_Y(y|\mathcal{C})$ and $\mathcal{P}_Z(z|\mathcal{C})$ ($\mathbf{n} = 3$ and $\alpha = 0.5$).

the maxima are given by:

$$\mathcal{P}_Z(Z_{\max \mathcal{C}}|\mathcal{C}, \sigma) = \frac{(\mathbf{n}\alpha - 1)^{\mathbf{n}\alpha - 1}}{\Gamma(\mathbf{n}\alpha) \cdot \beta} \exp[-(\mathbf{n}\alpha - 1)] \quad (9)$$

$$\mathcal{P}_Y(Y_{\max \mathcal{C}}|\mathcal{C}, \sigma) = \frac{((\mathbf{n}-1)\alpha)^{(\mathbf{n}-1)\alpha}}{\alpha \cdot \Gamma(\mathbf{n}\alpha) \beta^\alpha} \exp[-(\mathbf{n}-1)\alpha]$$

4 Robust Non-Parametric estimation

Section 4.1 gives details on the computation of samples of the variable Y and Z from the samples of X . Paragraph 4.2

presents the nonparametric approach taken to perform the computation of the relevant maxima in the estimation of the scale parameter.

4.1 Computing samples

From a set of observed independent samples $\mathcal{B}_x = \{x_i\}_{i \in \{1 \dots i\}}$, we need first to compute the samples $\mathcal{B}_z = \{z_j\}_{j \in \{1 \dots j\}}$ and $\mathcal{B}_y = \{y_j\}_{j \in \{1 \dots j\}}$ of Z and Y . It is performed by randomly selecting \mathbf{n} samples from \mathcal{B}_x to compute samples z_j and y_j [9]. This should be done without replacement to insure independence of samples of X . However if \mathbf{i} is large, it can also be performed with replacement.

It is assumed that the observed samples $\{z_j\}$ (or $\{y_j\}$) are generated from a mixture $\mathcal{P}_Z(z|\sigma)$ of $\mathcal{P}_Z(z|\mathcal{C}, \sigma)$ (inliers) and $\mathcal{P}_Z(z|\bar{\mathcal{C}})$ (outliers). A priori, the proportion of the inliers $\mathcal{P}_Y(\mathcal{C})$ in \mathcal{B}_y and $\mathcal{P}_Z(\mathcal{C})$ in \mathcal{B}_z (i.e. when z_j and y_j are computed using $x_n \in \mathcal{C}$, $\forall n \in [1; \mathbf{n}]$) is equal to $(\mathcal{P}_X(\mathcal{C}))^{\mathbf{n}}$. However this proportion can be increased using properties of the data. More precisely, audio and video data present strong spatio-temporal correlations that allow us to assume that data from a neighbourhood belong to the same class (\mathcal{C} or $\bar{\mathcal{C}}$) [8]. Using this assumption, samples for Y and Z are carefully generated in order to limit the proportion of outliers by mis-coupling samples of X . As a consequence, it is possible to compute samples of Y and Z such that the proportion of inliers in \mathcal{B}_x , $\mathcal{P}_X(\mathcal{C})$, is rather close to the proportions of inliers in \mathcal{B}_y and \mathcal{B}_z : $\mathcal{P}_X(\mathcal{C}) \sim \mathcal{P}_Y(\mathcal{C}) \sim \mathcal{P}_Z(\mathcal{C})$.

4.2 Non-parametric estimation

4.2.1 Estimating distributions

From the collection \mathcal{B}_y of samples of Y , a kernel estimate of the density function $\mathcal{P}_Y(y)$ can be computed:

$$\hat{\mathcal{P}}_Y(y) = \frac{1}{\mathbf{i}} \sum_{i=1}^{\mathbf{i}} \frac{1}{h_{y_i}} \cdot k\left(\left|\frac{y - y_i}{h_{y_i}}\right|\right)$$

with $k(\cdot)$ chosen as a gaussian $\mathcal{N}(0, 1)$. The variable bandwidths $\{h_{y_i}\}$ are selected automatically selection following Comanicciu et al. scheme [5]. The only change concerns the initial setting of the center point bandwidth: instead of the rule of thumb plug-in, the more efficient Sheather-Jones plug-in bandwidth h_{SJ} is computed [22].

4.2.2 Relation between bandwidths

Using the relation $z_i = y_i^2$ between samples in \mathcal{B}_y and \mathcal{B}_z , variable bandwidths are automatically computed for samples in \mathcal{B}_y , and then inferred for samples in \mathcal{B}_z by:

$$h_{z_i} = h_{y_i} \cdot \sqrt{4y_i^2 + 2h_{y_i}} \quad (10)$$

This relation between h_{z_i} and h_{y_i} is derived by assuming a gaussian variable $y \sim \mathcal{N}(y_i, h_{y_i}^2)$ (from the kernel gaussian assumption) and then by inferring the variance $h_{z_i}^2$ of the variable $z = y^2$.

4.2.3 MeanShift

The closest mode to zero can be computed using mean shift from the minimum value of the samples [8, 5]:

$$\begin{aligned} &\text{Init } y^{(0)} = \min_{y_i} \mathcal{B}_y \\ &\left| \begin{aligned} y^{(m+1)} &= \frac{\sum_{i=1}^{\mathbf{i}} \frac{y_i}{h_{y_i}^3} g\left(\left|\frac{y^{(m)} - y_i}{h_{y_i}}\right|^2\right)}{\sum_{i=1}^{\mathbf{i}} \frac{1}{h_{y_i}^3} g\left(\left|\frac{y^{(m)} - y_i}{h_{y_i}}\right|^2\right)} \end{aligned} \right. \quad (11) \\ &\text{till convergence } Y_{\max \mathcal{C}} = y^{(\mathbf{m})} \end{aligned}$$

with $g(t) = \frac{-k'(\sqrt{t})}{2\sqrt{t}}$. The same procedure is used to estimate $Z_{\max \mathcal{C}}$. Using equation (7), estimates for the scale parameter can be inferred.

5 Robust Regression

For simplicity, we consider linear regression where observations $\mathcal{B}_{uv} = \{(v_i, u_i)\}_{i=1 \dots i}$ follow the linear mapping $v_i = u_i^T \theta + \epsilon_i$, $\forall i$. The joint estimation of θ and σ is performed by using the scale estimate introduced in the previous paragraphs, iteratively with a least squares¹ estimation of θ performed on a subset \mathcal{S}_{uv} of p observations (p is chosen superior or equal to the dimension of θ). This is similar to the RANSAC approach with an added scale estimate. The algorithm can be described as:

- Repeat B times (see [10]) for the choice of B
 1. Select a subset $\mathcal{S}_{uv}^{(b)}$ of p points randomly selected from \mathcal{B}_{uv} ,
 2. Least Squares estimation of the location parameter $\theta^{(b)}$ on $\mathcal{S}_{uv}^{(b)}$.
 3. Compute the residuals $\{\epsilon_i = v_i - u_i^T \theta^{(b)}\}_{i=1 \dots i}$, and samples of Y and Z .
 4. Compute the bandwidths $\{h_{y_i}\}$ and $\{h_{z_i}\}$ as explained in section 4.2.
 5. Estimate the scale parameter $\sigma^{(b)}$ using procedure (11) and relations (7).
 6. Compute the objective function $\mathcal{J}^{(b)}(\theta^{(b)}, \sigma^{(b)})$.
- Infer $(\hat{\theta}, \hat{\sigma})$ from $\arg \max \mathcal{J}^{(b)}$ (ou $\arg \min$ depending on the chosen objective function).

¹The inliers are assumed gaussian: $\alpha = 0.5$.

In this article, a similar objective function as Wang et. al has been chosen [25].

$$\mathcal{J}(\theta, \sigma) = \frac{\sum_{i=1}^{\mathbf{i}} \mathbf{1}\{|\epsilon_i| < 2.5\sigma\}}{\mathbf{i} \cdot \sigma} \quad (12)$$

Note that, with our algorithm, a fixed θ leads to an estimate of σ , hence the space of all possibilities (θ, σ) is not fully explored but only the space (θ, σ_θ) . As a consequence the objective function \mathcal{J} has only θ as a variable. Some representations of the objective function are presented in the experimental results in section 6.2. Similarly to the Hough transform [11], the problem of recovering several models (or multi-lines when θ is of dimension $p = 2$) is to find all maxima (or minima) in \mathcal{J} .

6 Experimental results

Assessment of the robustness of the proposed method is done using simulations. Those simulations are performed under certain conditions explained below. That reflect real situations of interest encountered in data analysis. In section 6.1, two scenarios are tried : outliers with uniform distribution and pseudo outliers with a Gaussian distribution. Results for robust regression are reported in section 6.2.

6.1 Scale Estimate

In the following experiments, we chose $\alpha = \frac{1}{2}$, or a Gaussian distribution for the inliers, $\mathbf{i} = 1000$ (cardinal of \mathcal{B}_Y and \mathcal{B}_Z), the groundtruth scale parameter $\sigma = 2$ and the degree of freedom $\mathbf{n} = 3$:

- $\forall n = 1, \dots, \mathbf{n}$, sets \mathcal{B}_{x_n} are used to compute sets \mathcal{B}_y and \mathcal{B}_z such that the proportion of inliers $\forall n$, $\mathcal{P}_{X_n}(\mathcal{C}) = \mathcal{P}_Y(\mathcal{C}) = \mathcal{P}_Z(\mathcal{C})$. Outliers follow a uniform distribution $\forall n$, $\mathcal{P}_{X_n}(x|\bar{\mathcal{C}}) = \mathcal{U}([-50; 50])$.
- Pseudo-outliers follows a Gaussian distribution with the same variance as the inliers and a mean μ . The proportion of the inliers is fixed such that $\forall n$, $\mathcal{P}_{X_n}(\mathcal{C}) = \mathcal{P}_Y(\mathcal{C}) = \mathcal{P}_Z(\mathcal{C}) = 0.1$.

6.1.1 Robustness to outliers

Figure 2 presents the mean of the estimates depending on the proportion of the inliers computed over 50 tries. As the proportion of inliers increases, the accuracy improves. Although σ_Z is less accurate than σ_Y , it shows a better robustness to the proportion of outliers. It is understood that the peak localisation for the estimation is easier performed on the distribution of Z than Y . In fact, when too many outliers occur, the inlier peak is not anymore distinguishable in the pdf of Y .

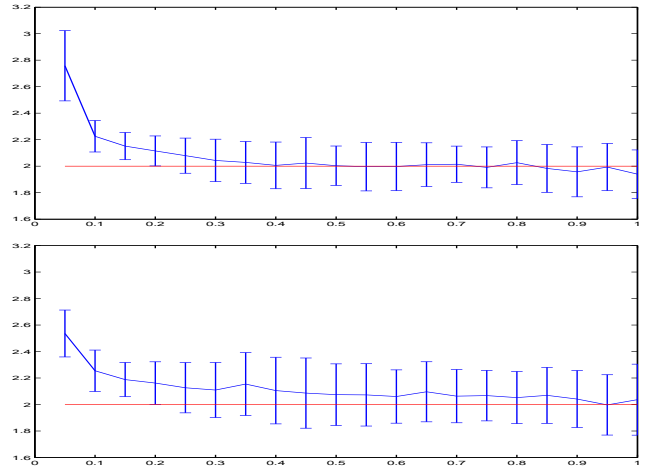


Figure 2: **Robustness to outliers:** Estimates σ_Y (top) and σ_Z (bottom) with standard deviation, w.r.t. the proportion of inliers $\mathcal{P}_Y(\mathcal{C}) = \mathcal{P}_Z(\mathcal{C})$. The red line corresponds to the groundtruth scale parameter.

6.1.2 Robustness to pseudo-outliers

Figure 3 presents the mean of the scale estimates σ_Y and σ_Z with standard deviation, depending on the mean of the pseudo-outliers, computed over 50 tries. The estimate σ_Y is shown to be accurate when the mean μ of the pseudo-outliers is above 3σ . The estimate σ_Z is rather accurate already when $\mu > 2.5\sigma$. Figure 4 presents the distributions $\mathcal{P}_X(x|\sigma)$, $\mathcal{P}_Y(y|\sigma)$ and

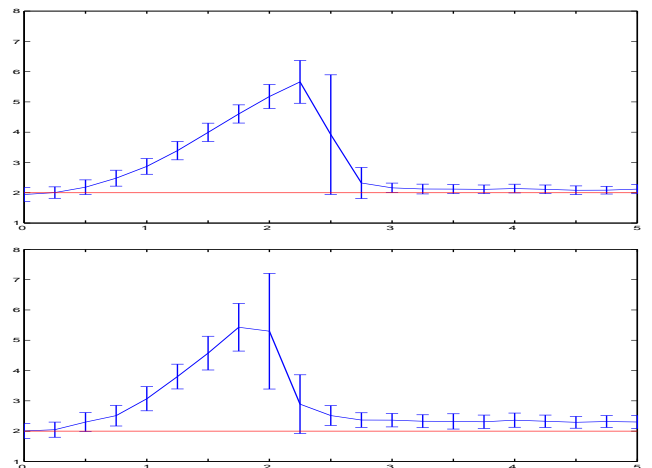


Figure 3: **Robustness to pseudo-outliers:** Estimates σ_Y (top) and σ_Z (bottom) with standard deviation, w.r.t. the mean of the gaussian of pseudo-outliers expressed as a multiplicative of σ (i.e. abscissa equal to 2 means $\mu = 2\sigma$).

$\mathcal{P}_Z(z|\sigma)$ for different values of the mean μ of the pseudo-outliers. The distributions are estimated using the kernel modelling as explained in section 4.2. When the inlier and pseudo-outlier Gaussians are too close, their respective maximum is not anymore distinguishable.

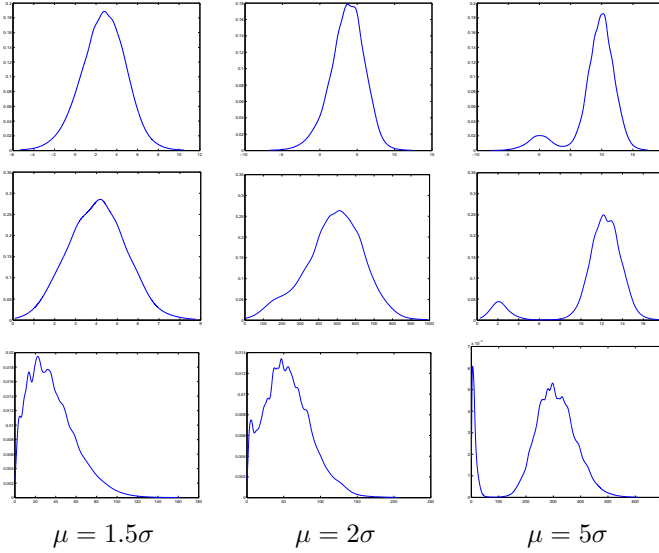


Figure 4: Distributions $\mathcal{P}_X(x)$ (top), $\mathcal{P}_Y(y)$ (middle), $\mathcal{P}_Z(z)$ (bottom). The relevant maximum for the estimation of the scale parameter becomes distinguishable after $\mu > 3\sigma$.

6.1.3 Remarks

Similar results have been obtained for various inliers distribution i.e. different values of α (in between 0.5 to 1) and $\mathbf{n} = 2, 3$. In practice, the choice of \mathbf{n} should be as low as possible to simplify the computation of samples of Y and Z .

6.2 Robust Regression

In a similar experience as in [25], line parameters are estimated iteratively with the scale parameter of the residuals [18, 25], following the procedure described in section 5. Figure 5 shows the result of the estimation when 90% of outliers, uniformly distributed, appear in the observations. This result has been obtained using the estimate σ_Z and it is identical to the one computed with σ_Y . Several simulations have been run on different randomly generated sets of observations. Ten estimates of the line are reported on the graph.

The previous experience is repeated with an added line of 50 points generated with the equation $u = v$. Figure 6 shows the observations where alignments are barely distinguishable (left). Both lines can however be recovered in analysing maxima of the objective function. Figure 7 presents the objective function \mathcal{J} in function of the 2-dimensional θ . Two peaks are clearly localised at the neighborhoods of $\theta = (1, 0)$ and $\theta = (-1, 100)$, corresponding to the two lines coefficients. Numerically the first two maxima computed correspond to $\hat{\theta} = (-0.9710, 98.4041)$ and $\hat{\theta} = (1.0135, -0.1504)$.

Figure 7 presents the computed objective function as it is performed in the robust estimation proposed in paragraph 5. For comparison, we computed the objective function for all values

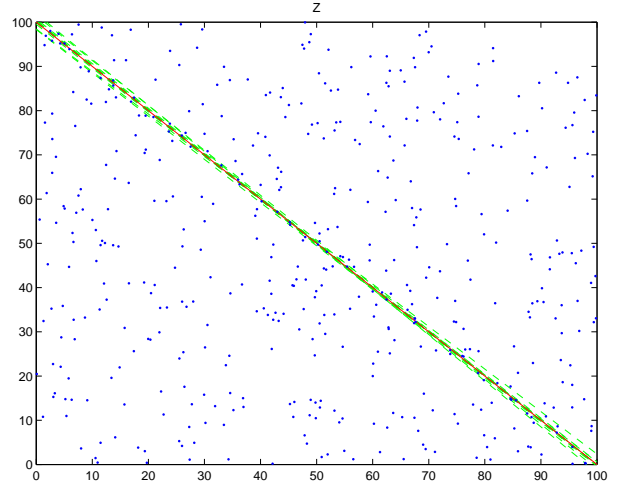


Figure 5: Robust line fitting. Inliers: 50 points on a line $x \in [0 - 100]$, $u = -v + 100$ and $\sigma = 1$; Outliers: 450 points uniformly distributed [25]. The red line represents the groundtruth, and the green lines represent the estimates performed on 10 trials.

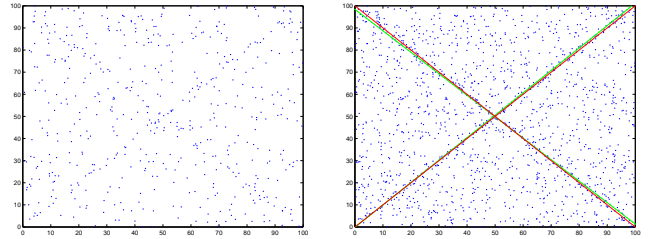


Figure 6: A two line fitting problem: Observations and estimated lines (green lines) with ground truth (red).

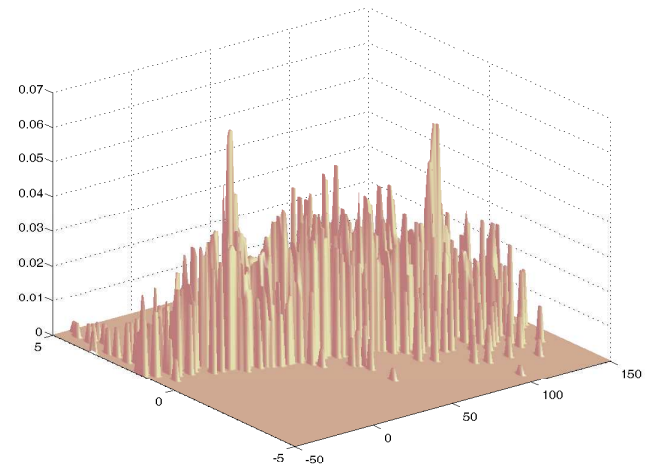


Figure 7: Objective function computed during the estimation. Two lines are present in the observations with parameter $\theta = (-1, 100)$ and $\theta = (1, 0)$. Both can be recovered by localising the two main peaks.

(θ, σ) on a finite domain and the graph is reported in figure 8.

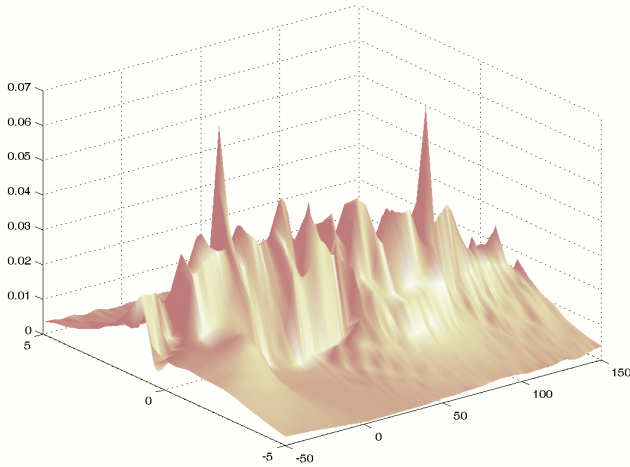


Figure 8: Simulated objective function.

6.3 Applications in Computer Vision

Application to edge detection in images of this scale parameter estimation is suggested [21]. Applications to silence detection in audio data streams, and region segmentation in colour images have been proposed in [8]. We consider here the problem of robust object recognition [16] in colour images [6].

A linear mapping is estimated by applying Principal Component Analysis [6] on 72 training colour images (image pixel values are arranged in a vector in lexicographic order) representative of an object class (cf. fig. 9). Eigenvectors associated with the highest eigenvalues summarize the informative visual content over the training set. The representative eigenspace is cho-



Figure 9: Example of training colour images for one object varying under different viewpoints [19].

sen of dimension 3 for this experiment. Figure 10 presents the eigenbasis including the mean image and the first three eigenvectors. The last image is the reconstruction of the template t_0 performed on this basis. Considering an unknown observa-

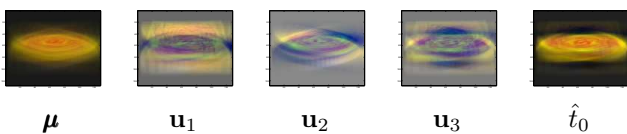


Figure 10: Mean and eigenvectors selected to represent the object class, and the reconstruction of one of the training template in this eigenspace.

tion (image), its recognition consists in two tasks. The first is to estimate its coordinate on the eigenspace (corresponding to the 3-dimensional location parameter θ). Then the recognition is completed by comparing this estimate with the coordinates indexing the training images. The estimation problem can be written as follow:

$$\begin{cases} R_i = \mu_i^R + (\mathbf{u}_i^R)^T \theta + \epsilon_i^R \\ G_i = \mu_i^G + (\mathbf{u}_i^G)^T \theta + \epsilon_i^G \\ B_i = \mu_i^B + (\mathbf{u}_i^B)^T \theta + \epsilon_i^B \end{cases} \quad (13)$$

where (R_i, G_i, B_i) are the colour values at the pixel i in the observed image, and $(\mu_i^R, \mu_i^G, \mu_i^B)$ and $(\mathbf{u}_i^R, \mathbf{u}_i^G, \mathbf{u}_i^B)$ are the mean and the eigenvector values at pixel i (from the learning). Residuals on each colour band are independent ($\mathbf{n} = 3$). Noise is assumed gaussian ($\alpha = 0.5$) and outliers typically occur because of a changing background or partial occlusions.

The estimation of θ is performed as described in section 5. A comparison with robust recognition using M-estimators [6] is proposed and several results are presented in figure 11. Observations present the object with added gaussian noise, different colour background and possible partial occlusions. The observations with a yellow background (same yellow as the main colour of the object itself) is tricky for the M-estimator based recognition method. In fact, M-estimation tries to match as many pixels as possible in the observation and consequently matches templates with the highest number of yellow pixels. Our method however, with its objective function taking into account both the scale and location parameter estimates from the current observation, gives a more accurate match for recognition.

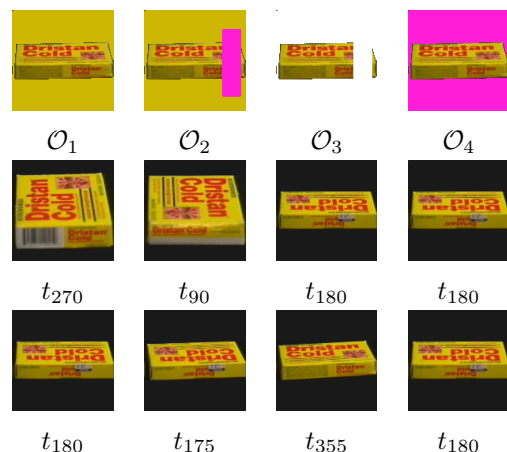


Figure 11: Observations (top), recognition performed with M-estimators as in [6] (middle), recognition with simultaneous robust estimation of the scale and location parameters (bottom).

7 Conclusion

The main idea of this article is to consider the generation of new variables whose distributions show relevant properties for the estimation of an unknown parameter of the original variable. In particular, modes or maxima related to the unknown parameter can be located using non-parametric modelling. Accuracy of this estimation relies on the accuracy of the estimated density function, here performed by nonparametric modelling such as kernels. The association with location parameter estimation performs very well in term of robustness to outliers. The proposed method is fully automatic, one drawback being the computation of samples of Y or Z that must be carefully done to limit the proportion of outliers.

References

- [1] B. Aiazzi, L. Alparone, and S. Baronti. Estimation based on entropy matching for generalized gaussian pdf modeling. *IEEE Signal Processing Letters*, 6(6), June 1999.
- [2] M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal on Computer Vision*, 26(1):63–84, January 1998.
- [3] P. Bouthémy, M. Gelgon, and F. Ganansia. A unified approach to shot change detection and camera motion characterization. *IEEE Transactions on Circuits and Systems for Video Technology*, 9:1030–1044, 1999.
- [4] H. Chen and P. Meer. Robust regression with projection based m-estimators. In *International Conference on Computer Vision*, pages 878–885, Nice, France, October 2003.
- [5] D. Comaniciu, V. Ramesh, and P. Meer. The variable bandwidth mean shift and data-driven scale selection. In *International Conference on Computer Vision*, pages 438–445, Vancouver, Canada, July 2001.
- [6] R. Dahyot, P. Charbonnier, and F. Heitz. Robust visual recognition of colour images. In *IEEE proceedings of the conference on Computer Vision and Pattern Recognition*, volume 1, pages 685–690, Hilton Head Island, USA, June 2000.
- [7] R. Dahyot, P. Charbonnier, and F. Heitz. A bayesian approach to object detection using probabilistic appearance-based models. *Pattern Analysis and Applications*, 7(3):317–332, December 2004.
- [8] R. Dahyot, N. Rea, A. Kokaram, and N. Kingsbury. Inlier modeling for multimedia data analysis. In *IEEE International Workshop on MultiMedia Signal Processing*, Siena Italy, September 2004.
- [9] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1998.
- [10] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.
- [11] A. Goldenshluger and A. Zeevi. The hough transform estimator. *The Annals of Statistics*, 32(5), October 2004.
- [12] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics : The Approach Based on Influence Functions*. John Wiley and Sons, 1986.
- [13] D. Hasler, L. Sbaiz, S. Süsstrunk, and M. Vetterli. Outlier modeling in image matching. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 25(3):301–315, March 2003.
- [14] P.J. Huber. *Robust Statistics*. John Wiley and Sons, 1981.
- [15] F. De la Torre and M. J. Black. Robust principal component analysis for computer vision. In *Proceedings of the International Conference on Computer Vision*, Vancouver, Canada, July 2001.
- [16] A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000.
- [17] Ruffieux C Marazzi A. Implementing m-estimators of the gamma distribution. *Robust Statistics, Data Analysis, and Computer Intensive Methods, in Honor of Peter Huber's 60th Birthday*, pages 277–297, 1996.
- [18] J. V. Miller and C. V. Stewart. Muse: Robust surface fitting using unbiased scale estimates. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 300–306, June 18-20 1996.
- [19] S. A. Nene, S. K. Nayar, and H. Murase. Columbia object image library (coil-100). Technical Report CUCS-006-96, Department of Computer Science, Columbia University, February 1996.
- [20] J-M. Odobez and P. Bouthémy. Robust multiresolution estimation of parametric motion models. *Journal of visual communication and image representation*, 6:348–365, 1995.
- [21] P. L. Rosin. Edges: saliency measures and automatic thresholding. *Machine Vision and Applications*, 9:139–159, 1997.
- [22] S. J. Sheather. Density estimation. *Statistical Science*, 19(4):588–597, 2004.
- [23] C.V. Stewart. Robust parameter estimation in computer vision. *SIAM Reviews*, 41(3):513–537, September 1999.
- [24] H. Wang. *Robust Statistics for Computer Vision: Model fitting, Image Segmentation and Visual Motion Analysis*. PhD thesis, Monash University, Clayton Victoria, Australia, February 2004.
- [25] H. Wang and D. Suter. Robust adaptive-scale parametric model estimation for computer vision. *Transactions on Pattern Analysis and Machine Intelligence*, 26(11), November 2004.