Pattern Recognition 46 (2013) 788–794

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Generalised relaxed Radon transform (GR^{2}T) for robust inference

Rozenn Dahyot*, Jonathan Ruttle

School of Computer Science and Statistics, Trinity College Dublin, Ireland

ARTICLE INFO

Article history:
Received 27 January 2012
Received in revised form
17 August 2012
Accepted 30 September 2012
Available online 11 October 2012

Keywords:
Generalised Radon transform
Hough transform
Robust inference
M-estimator
Generalised projection based M-estimator

ABSTRACT

This paper introduces the generalised relaxed Radon transform (GR^{2}T) as an extension to the generalised radon transform (GRT) [1]. This new modelling allows us to define a new framework for robust inference. The resulting objective functions are probability density functions that can be chosen differentiable and that can be optimised using gradient methods. One of this cost function is already widely used in the forms of the Hough transform and generalised projection based M-estimator, and it is interpreted as a conditional density function on the latent variables of interest. In addition the joint density function of the latent variables is also proposed as a cost function and it has the advantage of including a prior about the latent variable. Several applications, including lines detection in images and volume reconstruction from silhouettes captured from multiple views, are presented to underline the versatility of this framework.

1. Introduction

Robust inference is one major focus in pattern recognition applications. The possibility to find several occurrences of the same object (pseudo-outliers) when the observations are contaminated with data that have no relation to the information of interest (outliers) is a major challenge in image and video processing. In the past decades, several innovative approaches have been proposed to deal with both outliers and pseudo-outliers (see Section 2 for a brief review). The aim of this paper is to explain and generalise in the same consistent non-parametric Bayesian framework several one-to-many and many-to-one robust strategies that have been proposed for inference in Computer vision. This new framework is derived from the generalised Radon transform (GRT) and is coined generalised relaxed Radon transform (GR^{2}T) (see Section 3). Breaking away from past interpretations, the first contribution of this paper is to be able to interpret the robust objective functions as probability density functions of the latent variables of interest. In this new framework, the addition of prior information is straightforward allowing the modelling to be fully Bayesian.

Secondly, we show how this framework can be used for a wide range of inferential problems by using an auxiliary variable. The versatility of our new framework is illustrated for various applications in Section 4. Depending on the experimenter's assumptions, this methodology can define discrete or differentiable smooth cost functions suitable for parameter estimation. Section 3.3 discusses the benefits and inconveniences of discrete versus differentiable cost functions for performing stochastic exploration.

* Corresponding author. Tel.: +353 1896 1760; fax: +353 1 677 0711.
E-mail addresses: Rozenn.Dahyot@scss.tcd.ie, Rozenn.Dahyot@tcd.ie (R. Dahyot).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.09.026

2. State of the art

We consider $\{\mathbf{x}^{(i)}\}_{i=1,\ldots,N}$, N independent observations of the random variable $\mathbf{x}$ that is related to a latent quantity of interest (λ,Θ) by

$$\lambda + F(\Theta,\mathbf{x}) = \epsilon \sim p_\epsilon(\epsilon) \tag{1}$$

where F is a known function. Using the observations, we note the residuals:

$$\epsilon^{(i)} = \lambda + F(\Theta,\mathbf{x}^{(i)}), \quad \forall i = 1,\ldots,N$$

This section presents several objective functions that have been proposed in the literature to estimate (λ,Θ). These objective functions originate from different frameworks namely: the standard likelihood approach used along with its extension to robust M-estimation (Section 2.1), the Hough transform (Section 2.2) and its extension (Section 2.3). More recent robust approaches in computer vision are reported in Section 2.4.

2.1. Maximum likelihood estimation

The likelihood cost function corresponds to the joint probability of the residuals (assumed independent):

$$\mathcal{L}(\lambda,\Theta) = p(\epsilon^{(1)},\ldots,\epsilon^{(N)}) = \prod_{i=1}^{N} p_\epsilon(\epsilon^{(i)}) \tag{2}$$

Using the negative log transformation, an estimate of (λ,Θ) can be computed by

$$(\hat{\lambda},\hat{\Theta}) = \arg\min_{\lambda,\Theta}\left\{-\log \mathcal{L}(\lambda,\Theta) = \sum_{i=1}^{N} -\log(p_\epsilon(\epsilon^{(i)}))\right\} \tag{3}$$

A popular choice for the density function p_ϵ is the normal distribution. However this choice is not robust to the presence of outliers, and other distributions such as the Cauchy distribution can be used. The maximum likelihood cost function has been extended to the generalised maximum likelihood estimators as follows:

$$(\hat{\lambda},\widehat{\Theta}) = \arg\min_{\lambda,\Theta}\left\{\mathcal{M}(\lambda,\Theta) = \sum_{i=1}^{N}\rho(\epsilon^{(i)})\right\} \tag{4}$$

where the function ρ is designed to penalise high residuals and is not limited to the choice $\rho = -\log p_\epsilon$ [2].

2.2. Hough transform

The Hough transform increments a histogram in the parameter space of the shape of interest by taking each observations independently, and incrementing all possible bins corresponding to the models that could have generated this observation [3]. This can be understood as a voting scheme and the resulting discrete objective function to optimise is a multidimensional histogram where each significant local maxima is associated with a likely occurrence of the object. This sort of strategy is referred as one-to-many mapping [4] since one observation votes for several possible models.

Mathematically, the Hough transform is defined [5,6] as

$$\mathcal{H}((\lambda,\Theta)\in\Omega) = \int p(\mathbf{x},\Omega)\, I(\mathbf{x})\, d\mathbf{x} = \sum_{i=1}^{N} p(\mathbf{x}^{(i)},\Omega) \tag{5}$$

where Ω defines a finite cell in the parameter space such that

$$p(\mathbf{x}^{(i)},\Omega) = \begin{cases} 1 & \text{if any curve of parameters} \\ & (\lambda,\Theta)\in\Omega \text{ passes through the point } \mathbf{x}^{(i)} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

and I is proportional to the empirical density function of $\mathbf{x}$:

$$I(\mathbf{x}) = \sum_{i=1}^{N}\delta(\mathbf{x}-\mathbf{x}^{(i)}) \tag{7}$$

The objective function $\mathcal{H}((\lambda,\Theta)\in\Omega) = \mathcal{H}(\Omega)$ is then computed on all cells Ω such that the whole parameter space is covered. The estimation for the global maximum is then performed by

$$\hat{\Omega} = \arg\max_{\Omega}\left\{\mathcal{H}(\Omega) = \sum_{i=1}^{N} p(\mathbf{x}^{(i)},\Omega)\right\} \tag{8}$$

The discrete nature of the voting kernel p (Eq. (6)) implies that the objective function $\mathcal{H}$ is discrete but this can be relaxed to smooth formulation [5]. Most applications of the Hough transform deal with the image data and there has been many efforts for modelling the noise in the image (observation) space and its consequence in the latent (Hough) parameter space [7,8,5,9]. For line detection, the function F is defined as $F(\mathbf{x}=(x_1,x_2),\Theta) = x_1\cos\Theta + x_2\sin\Theta$. In this case, the Hough transform can be understood as the Radon transform [10,11]. Another popular applications is for circle detection in images [7,5].

2.3. Mapping subsets of observations into the parameter space

Olson proposed to extend the Hough transform to map subsets of observations in the parameter space [6]. Considering all subsets $s = (\mathbf{x}^{(i_1)},\ldots,\mathbf{x}^{(i_K)})$ of K observations selected from $\{\mathbf{x}^{(i)}\}_{i=1,\ldots,N}$, the Hough transform is generalised to [6]

$$\mathcal{H}^K(\Omega) = \sum_{\forall s}\prod_{k=1}^{K} p(\mathbf{x}^{(i_k)},\Omega) \tag{9}$$

If the sets s are computed without repetitions (a.k.a. if the sets correspond to delete-d jackknife subsamples [12]), then there are $\binom{N}{K}$ possible different subsets s [6]. To simplify computation of $\mathcal{H}^K$, one can consider to randomly select a few to compute $\mathcal{H}^K$ [6]. Note that $\mathcal{H}^1$ corresponds to the standard Hough transform as defined in Eq. (5). In addition constraints can be added for selecting the best subsets [6].

This strategy can be classified as a many-to-one method [4] where several observations are grouped in a subset to map one solution in the parameter space (when the size of the subsets is equal or superior to the dimension of the parameter space $K \geq \dim(\Theta)+\dim(\lambda)$). RANSAC [13] is another popular example of a many-to-one method. Estimation of the latent variables can be computed by optimisation of the objective function $\mathcal{H}^K$.

2.4. Kernel-based cost functions for robust computer vision

Mittal et al. propose to use the generalised projection based M-estimator as a cost function [14,15]. Their cost function is defined like a kernel density estimate:

$$\mathcal{G}(\lambda,\Theta) = \frac{1}{N}\sum_{i=1}^{N} k(\lambda + F(\Theta,\mathbf{x}^{(i)})) \tag{10}$$

Their formulation originates from the M-estimation (cf. Section 2.1) where the kernel function is defined with respect to the M-estimator loss function ρ such that $k(u) = 1-\rho(u)$ [14]. Note that this kernel interpretation of M-estimators is only applicable to hard redescending M-estimator functions ρ [16] (ρ is non-negative, symmetric and non-decreasing with $|u|$, has a unique minimum $\rho(0)=0$ and a maximum of one for $|u|>1$ [14]). This trick allows the equivalence with the M-estimation approach:

$$(\hat{\lambda},\widehat{\Theta}) = \arg\min_{\lambda,\Theta}\left\{\frac{1}{N}\sum_{i=1}^{N}\rho(\epsilon^{(i)})\right\} = \arg\max_{\lambda,\Theta}\left\{\mathcal{G}(\lambda,\Theta) = \frac{1}{N}\sum_{i=1}^{N}k(\epsilon^{(i)})\right\} \tag{11}$$

M-estimation requires the estimation of a scale parameter along with the location parameter (λ,Θ) [2]. Mittal et al.'s framework also tackles the estimation of the nuisance adaptive scale parameter associated with each observation $\mathbf{x}^{(i)}$. The latent variables (λ,Θ) are then iteratively estimated along with the scale using gradient ascent methods to stochastically explore the $\mathcal{G}$ cost function (Eq. (10)) [15].

2.5. Remarks and contributions

The form of the cost functions $\mathcal{H}$ Eq. (5) and $\mathcal{G}$ Eq. (10) are well summarised by the following cost function $\mathcal{C}$ defined with a discrete or continuous function c that models the vote of each observation:

$$\mathcal{C}(\lambda,\Theta) \propto \sum_{i=1}^{N} c(\epsilon^{(i)})$$

However the origin of these cost functions is not clear when using statistics: both $\mathcal{H}$ and $\mathcal{G}$ have somewhat been related to a log probability (e.g. via M-estimation) [5,17–19,14]. This interpretation limits the definition of generalised projection based M-estimator to using specific kernels for instance [14].

Alternatively Dahyot interprets the Hough transform as a density probability function of the latent variables (λ,Θ) [20]. In this paper, we introduce next explicitly two objective functions, noted $\hat{p}_{\lambda|\Theta}$ and $\hat{p}_{\lambda,\Theta}$. We will show that the objective functions $\mathcal{H}$ and $\mathcal{G}$ presented in this section can be better understood as the

conditional density function $\hat{p}_{\lambda|\Theta}$. The origin of both density functions $\hat{p}_{\lambda|\Theta}$ and $\hat{p}_{\lambda\Theta}$ is loosely derived from the generalised Radon transform. Moreover, the joint density function $\hat{p}_{\lambda\Theta}$ encapsulates naturally prior information about the latent variables making the framework fully Bayesian. Finally we show that this framework also explains the cost function $\mathcal{H}^K$ and the likelihood $\mathcal{L}$.

3. Generalised relaxed Radon transform

Section 3.1 presents the definition of the generalised Radon transform (GRT) and proposes a statistical interpretation when the transform is applied to density functions. We relax the hypotheses about the noise ϵ and propose new cost functions in Section 3.2. Section 3.3 comments on the advantages of defining smooth differentiable cost functions. Finally, Section 3.4 generalised the cost functions to use subsets of observations like $\mathcal{H}^K$ and show how the likelihood $\mathcal{L}$ can also be understood in the same general approach.

3.1. Generalised Radon transform (GRT)

The generalised Radon transform is defined by [1]

$$\mathcal{R}[f(\mathbf{x})](\lambda,\Theta)=\int_{\mathbb{R}^{d_{\mathbf{x}}}} A(\mathbf{x},\Theta)\, f(\mathbf{x})\, \delta(\lambda-F(\mathbf{x},\Theta))\, d\mathbf{x}$$
$$\text{with } \mathbf{x}\in\mathbb{R}^{d_{\mathbf{x}}},\Theta\in\mathbb{R}^{d_\Theta}, \quad \text{and} \quad \lambda\in\mathbb{R} \tag{12}$$

The functions f, A and F are defined such that the integral (12) can be computed at (λ,Θ) [1]. The integral (12) can be equivalently rewritten as

$$\mathcal{R}[f(\mathbf{x})](\lambda,\Theta)=\int_{\mathcal{D}} A(\mathbf{x},\Theta)\, f(\mathbf{x})\, d\mathbf{x} \tag{13}$$

where the domain $\mathcal{D}\subset\mathbb{R}^{d_{\mathbf{x}}}$ is defined as

$$\mathcal{D}=\{\mathbf{x}\in\mathbb{R}^{d_{\mathbf{x}}}\,|\,\lambda-F(\mathbf{x},\Theta)=0\} \tag{14}$$

This domain of integration can be rewritten as follows

$$\mathcal{D}=\{\mathbf{x}\in\mathbb{R}^{d_{\mathbf{x}}}\,|\,\lambda-F(\mathbf{x},\Theta)=\epsilon\sim\delta(\epsilon)\} \tag{15}$$

where the random variable ϵ is introduced explicitly and its distribution is the Dirac function centred on 0. Indeed the Dirac function $\delta(\lambda-F(\mathbf{x},\Theta))$ in integral (12) can be understood as the conditional density function $p_{\lambda|\mathbf{x}\Theta}(\lambda|\mathbf{x},\Theta)$ given the equation $\lambda-F(\mathbf{x},\Theta)=\epsilon\sim\delta(\epsilon)$.

Lets apply now the Generalised Radon Transform to functions A and f that are density functions such that $A(\mathbf{x},\Theta)$ is the conditional $p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x})$ and $f(\mathbf{x})$ is the density function of $\mathbf{x}$, noted $p_{\mathbf{x}}(\mathbf{x})$. In this case, integral (12) becomes (using Bayes theorem):

$$\mathcal{R}[p_{\mathbf{x}}(\mathbf{x})](\lambda,\Theta)=\int p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x})\, p_{\mathbf{x}}(\mathbf{x})\, p_{\lambda|\mathbf{x}\Theta}(\lambda|\mathbf{x},\Theta)\, d\mathbf{x}=p_{\lambda\Theta}(\lambda,\Theta) \tag{16}$$

where $p_{\lambda\Theta}(\lambda,\Theta)$ is the joint density function of the random variables λ and Θ. The Generalised Radon Transform in Eq. (16) can also be understood as an expectation w.r.t. the random variable $\mathbf{x}$:

$$\begin{aligned}\mathcal{R}[p_{\mathbf{x}}(\mathbf{x})](\lambda,\Theta)&=\mathbb{E}_{\mathbf{x}}[p_{\lambda|\mathbf{x}\Theta}(\lambda|\mathbf{x},\Theta)\; p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x})]\\&=\mathbb{E}_{\mathbf{x}}[\delta(\lambda-F(\mathbf{x},\Theta))\, p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x})]\end{aligned} \tag{17}$$

Having collected independent observations $\{\mathbf{x}^{(i)}\}_{i=1,\ldots,N}$ of the random variable $\mathbf{x}$, the density function of $\mathbf{x}$, $p_{\mathbf{x}}(\mathbf{x})$, can be approximated by its empirical density function:

$$\hat{p}_{\mathbf{x}}(\mathbf{x})=\sum_{i=1}^{N}\delta(\mathbf{x}-\mathbf{x}^{(i)})\,\pi_i \tag{18}$$

where δ is the Dirac density function, and π_i is the prior associated with the observation $\mathbf{x}^{(i)}$. In general, the observations are equiprobable and $\pi_i=1/N,\forall i$. By replacing $p_{\mathbf{x}}(\mathbf{x})$ by its empirical estimate in integral (16), the joint density $p_{\Theta\lambda}(\Theta,\lambda)$ can then be estimated by the empirical average [21]:

$$\begin{aligned}\hat{p}_{\lambda\Theta}(\lambda,\Theta)&=\sum_{i=1}^{N} p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x}^{(i)})\, p_{\lambda|\mathbf{x}\Theta}(\lambda|\mathbf{x}^{(i)},\Theta)\,\pi_i\\&=\sum_{i=1}^{N} p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x}^{(i)})\,\delta(\lambda-F(\mathbf{x}^{(i)},\Theta))\,\pi_i\end{aligned} \tag{19}$$

The term $\delta(\lambda-F(\mathbf{x}^{(i)},\Theta))$ corresponds to the vote of data point $\mathbf{x}^{(i)}$ and the term $p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x}^{(i)})$ allows the experimenter the possibility to insert a prior on Θ that is depending (or not) on the observation $\mathbf{x}^{(i)}$. When no prior knowledge is available then one can assume independence between the variables Θ and $\mathbf{x}$ (hence $p_{\Theta|\mathbf{x}}=p_\Theta$) leading to

$$\hat{p}_{\lambda\Theta}(\lambda,\Theta)=p_\Theta(\Theta)\underbrace{\sum_{i=1}^{N}\delta(\lambda-F(\mathbf{x}^{(i)},\Theta))\,\pi_i}_{\hat{p}_{\lambda|\Theta}(\lambda|\Theta)} \tag{20}$$

Using the joint density estimate $\hat{p}_{\Theta\lambda}$ for inference of the latent variables (λ,Θ) gives the user the opportunity to add prior information about Θ and perform the inference in a Bayesian framework. Alternatively the conditional estimate $\hat{p}_{\lambda|\Theta}$ can also be used as a cost function to find estimates of the latent variables.

In practice though, the Dirac kernel $\delta(\lambda-F(\mathbf{x}^{(i)},\Theta))$ cannot be computed, and both density estimates $\hat{p}_{\lambda\Theta}$ and $\hat{p}_{\lambda|\Theta}$ needs to be approximated by a discrete histogram. By relaxing the hypothesis about ϵ (Eq. (15)), we will propose in Section 3.2 a smooth differentiable estimate $\hat{p}_{\lambda\Theta}(\lambda,\Theta)$ more suitable for computation and optimisation.

Applications of GRT: The generalised Radon transform has been widely applied to pattern detection. In its simplest form, the Radon transform also known as the Hough transform, is very much used for line detection in images [3,22,10,5,18]. The generalised Radon transform is approximated by a discrete multidimensional histogram and its maxima provides candidates (λ,Θ) of shapes. For instance, Toft proposed to use the discrete GRT for detection of curve in images, and presented a way of setting a threshold for this histogram to decide if a curve (lines) occurs or not, depending on the noise level in the image [23]. Hansen and Toft [24] used a discrete formulation of GRT to analyse seismic data and find hyperbolas. They also point out the relation of GRT to the Hough transform.

Hendricks et al. uses a discrete GRT to find circles and hypersphere in images and their study focuses on the memory consumption of such approach [25]. Indeed the multidimensional histogram approximating $p_{\lambda\Theta}(\lambda,\Theta)$ uses a lot of memory depending on the resolution of its bins and the dimension of the latent space (λ,Θ) to explore. For circle detection, they propose to store the best radius for each centre location instead of storing the full histogram. One limitation is that information about concentric circles is lost (only one circle would be detected). Daras et al. used GRT to compute invariants suitable for indexing and retrieval of 3D meshes [26]. They considered two link functions F corresponding to a sphere and a line in a 3D space [26].

3.2. Generalised relaxed Radon transform (GR^2T)

We relax the definition of the GRT by changing the assumption for the perturbation ϵ (Eq. (15)) as follows:

$$\lambda-F(\mathbf{x},\Theta)=\epsilon\sim p_\epsilon(\epsilon) \tag{21}$$

ϵ is now a random vector modelling the perturbation or noise and has a given distribution $p_\epsilon(\epsilon)$. This reformulation allows to include explicitly uncertainty. The conditional density function $p_{\lambda|\mathbf{x}\Theta}(\lambda|\mathbf{x},\Theta)$ is now

$$p_{\lambda|\mathbf{x}\Theta}(\lambda|\mathbf{x},\Theta) = p_\epsilon(\lambda - F(\mathbf{x},\Theta)) \tag{22}$$

and consequently the joint density function of Θ and λ can be computed by

$$\begin{aligned} p_{\lambda\Theta}(\lambda,\Theta) &= \int p_{\lambda|\mathbf{x}\Theta}(\lambda|\mathbf{x},\Theta)\, p_{\mathbf{x}\Theta}(\mathbf{x},\Theta)\, d\mathbf{x} \\ &= \int p_\epsilon(\lambda - F(\mathbf{x},\Theta))\, p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x})\, p_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \\ &= \mathbb{E}_{\mathbf{x}}[p_\epsilon(\lambda - F(\mathbf{x},\Theta))\, p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x})] \end{aligned} \tag{23}$$

Note that GRT is a special case of GR^{2}T when the density p_ϵ is the Dirac distribution $\delta(\epsilon)$. Having N independent observations of x, $\{\mathbf{x}^{(i)}\}_{i=1,\dots,N}$, the density $p_{\Theta\lambda}(\Theta,\lambda)$ can be estimated by the empirical average:

$$\hat{p}_{\lambda\Theta}(\lambda,\Theta) = \sum_{i=1}^{N} p_\epsilon(\lambda - F(\mathbf{x}^{(i)},\Theta))\, p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x}^{(i)})\, \pi_i \tag{24}$$

While GRT is defined for $\lambda \in \mathbb{R}$ (Eq. (12), GR^{2}T generalises to $\lambda \in \mathbb{R}^{d_\lambda}$. In this case, Eq. (21) is now a system of stochastic equations with the error $\epsilon \in \mathbb{R}^{d_\lambda}$ and the link function F is defined from the domain of definition of $(\mathbf{x},\Theta)$ to $\mathbb{R}^{d_\lambda}$.

Assuming independence between Θ and $\mathbf{x}$ (i.e. $p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x}) = p_\Theta(\Theta)$), the joint density estimate (Eq. (24)) becomes

$$\hat{p}_{\lambda\Theta}(\lambda,\Theta) = p_\Theta(\Theta) \sum_{i=1}^{N} p_\epsilon(\lambda - F(\mathbf{x}^{(i)},\Theta))\, \pi_i \tag{25}$$

and an estimate of the conditional density function $p_{\lambda|\Theta}$ can easily be computed as

$$\hat{p}_{\lambda|\Theta}(\lambda|\Theta) = \sum_{i=1}^{N} p_\epsilon(\lambda - F(\mathbf{x}^{(i)},\Theta))\, \pi_i \tag{26}$$

Hence depending on the chosen assumptions and knowledge available, both density estimates $\hat{p}_{\lambda\Theta}$ (assuming p_ϵ and $p_{\Theta|\mathbf{x}}$ are known) and $\hat{p}_{\lambda|\Theta}$ (given p_ϵ and, assuming Θ and $\mathbf{x}$ are independent) can be used for inference. In this later case, it is interesting to compare the estimate $\hat{p}_{\lambda|\Theta}$ with the likelihood formulation (Eq. (2)): while the likelihood cost function multiplies the contribution of each observation, the conditional $\hat{p}_{\lambda|\Theta}$ modelled thanks to GR^{2}T, adds the contribution of each observation. The addition of the contribution of each observation as opposed to their multiplication allows the modelling to be more robust to outliers. Indeed an outlier $\mathbf{x}^{(i)}$ (for the structure parameterised by (λ,Θ)) would more likely create a value $p_\epsilon(\lambda - F(\mathbf{x}^{(i)},\Theta))$ close to zero which would not affect much the overall score of $\hat{p}_{\lambda|\Theta}$.

By the strong law of large numbers, the density estimate $\hat{p}_{\Theta\lambda}(\Theta,\lambda)$ converges almost surely to the exact density function $p_{\Theta\lambda}(\Theta,\lambda)$ [21]:

$$\lim_{N\to\infty} \hat{p}_{\lambda\Theta}(\lambda,\Theta) = p_{\lambda\Theta}(\lambda,\Theta) \quad a.s. \tag{27}$$

and similarly for the conditional density function:

$$\lim_{N\to\infty} \hat{p}_{\lambda|\Theta}(\lambda|\Theta) = p_{\lambda|\Theta}(\lambda|\Theta) \quad a.s. \tag{28}$$

In other words, as more independent observations are collected about the random variable $\mathbf{x}$, the more accurate the estimate $\hat{p}_{\lambda\Theta}(\lambda,\Theta)$ (resp. $\hat{p}_{\lambda|\Theta}(\lambda|\Theta)$) is to approximate the true density function $p_{\lambda\Theta}(\lambda,\Theta)$ (resp. $p_{\lambda|\Theta}(\lambda|\Theta)$).

3.3. Discrete vs smooth objective functions

Depending of the chosen models for the densities functions p_ϵ and $p_{\Theta|\mathbf{x}}$, then the objective functions, $\hat{p}_{\Theta\lambda}(\Theta,\lambda)$ and $\hat{p}_{\lambda|\Theta}(\lambda|\Theta)$, can be discrete or smooth differentiable functions. When defining an estimator of the latent variables (Θ,λ) by

$$(\hat{\lambda},\hat{\Theta}) = \arg\max \hat{p}_{\Theta\lambda}(\Theta,\lambda) \quad \text{or} \quad (\hat{\lambda},\hat{\Theta}) = \arg\max \hat{p}_{\lambda|\Theta}(\lambda|\Theta)$$

then using a smooth objective function improves the convergence rate of that estimator compared to using a discrete objective function [18,27,5].

Moreover, when the densities functions p_ϵ and $p_{\Theta|\mathbf{x}}$ are chosen differentiable, then the objective functions can be optimised with standard stochastic exploration techniques and in particular with gradient ascent algorithms [21]. Using gradient ascent techniques to find local maxima requires selecting initial guesses in the latent space that iteratively climb towards their nearest maxima. These strategies for stochastic exploration are well suited for parallel architectures with GPU for instance [28] and are less memory demanding than computing the objective function on a grid spanning the latent space. Indeed memory consumption of the discrete GRT is one major problem that has limited it application for inference in high dimensional latent space [25]. However when the latent space is of small dimension, inference with an exhaustive search on a discrete grid is not sensitive to initial guesses required by gradient ascent algorithms [17].

If the global maximum of the cost function is the only one of interest, a simulated annealing approach can be used [29,30]. As opposed to searching local maxima with gradient ascent that is sensitive to the starting guesses selected in the latent space to initialise the algorithm, the simulated annealing approach to find the global maximum is not sensitive to the choice of the initial guess in the latent space.

3.4. GR^{2}T as a general framework for estimation

λ as an auxiliary variable: Eq. (21) implies that the relationship between the observed and latent random variables includes one additive latent variable λ. So what should we do if the problem to solve is written as $F(\mathbf{x},\Theta) = \epsilon$ without additive latent random variable? In this case, the problem $F(\mathbf{x},\Theta) = \epsilon$ can be augmented by adding an auxiliary variable λ to formulate the problem as in Eq. (21). Inference of Θ can then be done using GR^{2}T objective functions $\hat{p}_{\lambda\Theta}(\lambda,\Theta)$ or $\hat{p}_{\lambda|\Theta}(\lambda|\Theta)$ computed at $\lambda = 0$. Note that choosing a modelling such that λ is an auxiliary variable (of non-interest) allows to choose a prior p_Θ or $p_{\Theta|\mathbf{x}}$ related to the latent variable of interest in the cost function $\hat{p}_{\lambda\Theta}$.

Mapping subsets of observations into the parameter space: We can extend our framework to consider subsets of size K of observations as Olson did with the Hough transform [6] (see Section 2.3). To do that a system of Eqs. (33) is defined to link the latent random variables with the observed one (see Table 1). Note that this requires to define K additive random variables $\{\lambda_k\}_{k=1,\dots,K}$ to estimate the cost functions $\hat{p}_{\overrightarrow{\lambda}_K\Theta}$ and $\hat{p}_{\overrightarrow{\lambda}_K|\Theta}$. The joint density estimate has the following expression:

$$\hat{p}_{\overrightarrow{\lambda}_K\Theta}(\overrightarrow{\lambda}_K,\Theta) = \sum_{i=1}^{N^*} \pi_i\, p_{\Theta|\overrightarrow{\mathbf{x}}}(\Theta|\overrightarrow{\mathbf{x}}^{(i)}) \prod_{k=1}^{K} p_\epsilon(\lambda_k + F(\mathbf{x}^{(i_k)},\Theta)) \tag{29}$$

where $\overrightarrow{\mathbf{x}}^{(i)} = (\mathbf{x}^{(i_1)},\dots,\mathbf{x}^{(i_K)})$ is a subset of K observations taken from $\{\mathbf{x}^{(i)}\}_{i=1,\dots,N}$ without repetition. The weight π_i is the prior associated with the subset $\overrightarrow{\mathbf{x}}^{(i)}$ and can be chosen equiprobable $\pi_i = 1/N^*$.

Table 1
GR^{2}T for inference extended to mapping subsets of K observations in the parameter space.

Observations	N independent observations are collected: $\{\mathbf{x}^{(i)}\}_{i=1,\dots,N}$	
User design	Choose relation F and subset size K	
	Case $K=1$: $\lambda+F(\mathbf{x},\Theta)=\epsilon$	$\vec{\lambda}_K+\vec{F}(\vec{\mathbf{x}},\Theta)=\vec{\epsilon}$ equivalent to: $\begin{cases}\lambda_1+F(\mathbf{x}_1,\Theta)=\epsilon_1\\ \vdots\\ \lambda_K+F(\mathbf{x}_K,\Theta)=\epsilon_K\end{cases}$ (33)
Randomvariables:		
Latent:	Θ	Θ
Observed:	$\mathbf{x}$, with the associated N observations $\{\mathbf{x}^{(i)}\}_{i=1,\dots,N}$	$\vec{\mathbf{x}}=(\mathbf{x}_1,\dots,\mathbf{x}_K)$, with the associated $N^*=\binom{N}{K}$ observations $\{\vec{\mathbf{x}}^{(i)}\}_{i=1,\dots,N^*}$ with $\vec{\mathbf{x}}^{(i)}$ subset of size K selected without repetition from $\{\mathbf{x}^{(i)}\}_{i=1,\dots,N}$
Hypotheses		
	Error ϵ with distribution: $\epsilon\sim p_\epsilon(\epsilon)$ Model $p_{\Theta\|\mathbf{x}}$ available, or assuming independence $p_{\Theta\|\mathbf{x}}=p_\Theta$	Error $\vec{\epsilon}=(\epsilon_1,\dots,\epsilon_K)$ with distribution: $\vec{\epsilon}\sim\prod_{k=1}^K p_\epsilon(\epsilon_k)$ Model $p_{\Theta\|\vec{\mathbf{x}}}$ available or assuming independence $p_{\Theta\|\vec{\mathbf{x}}}=p_\Theta$
Obj. functions	$\hat{p}_{\lambda\Theta}(\lambda,\Theta)$ or $\hat{p}_{\lambda\|\Theta}(\lambda\|\Theta)$	$\hat{p}_{\vec{\lambda}_K\Theta}(\vec{\lambda}_K,\Theta)$ or $\hat{p}_{\vec{\lambda}_K\|\Theta}(\vec{\lambda}_K\|\Theta)$

Assuming independence of Θ and $\vec{\mathbf{x}}$, the conditional density estimate has a similar form as the cost function $\mathcal{H}^K$ defined by Olson [6] (cf. Eq. (9)):

$$\hat{p}_{\vec{\lambda}_K|\Theta}(\vec{\lambda}_K|\Theta)=\sum_{i=1}^{N^*}\pi_i\prod_{k=1}^{K}p_\epsilon(\lambda_k+F(\mathbf{x}^{(i_k)},\Theta)) \tag{30}$$

In the case $K=N$, the conditional density estimate corresponds to

$$\hat{p}_{\vec{\lambda}_N|\Theta}(\vec{\lambda}_N|\Theta)=\prod_{k=1}^{N}p_\epsilon(\lambda_k+F(\mathbf{x}^{(k)},\Theta)) \tag{31}$$

This conditional density estimate limited to the subspace $\lambda_k=\lambda,\forall k=1,\dots,N$ becomes

$$\hat{p}_{\vec{\lambda}_N|\Theta}(\lambda,\lambda,\dots,\lambda|\Theta)=\prod_{k=1}^{N}p_\epsilon(\lambda+F(\mathbf{x}^{(k)},\Theta)) \tag{32}$$

corresponding to the likelihood cost function (Eq. (2)). Imposing $\lambda_k=\lambda,\forall k=1,\dots,K$ in the cost functions $\hat{p}_{\vec{\lambda}_K\Theta}$ and $\hat{p}_{\vec{\lambda}_K|\Theta}$ (Eqs. (31) and (32)) is a similar trick as using λ as an auxiliary variable: the cost functions are modelled on a bigger latent space than needed but it is easy to limit the search to a particular subspace or manifold.

In conclusion the objective function $\hat{p}_{\vec{\lambda}_K|\Theta}$ explains the cost functions $\mathcal{L}$ (with $K=N$), $\mathcal{G}$ (with $K=1$), $\mathcal{H}$ (with $K=1$) and $\mathcal{H}^K$ reviewed in Section 2. Any probability density function p_ϵ can be chosen by the user and unlike the cost function $\mathcal{G}$ (Eq. (10)), the cost function $\hat{p}_{\vec{\lambda}_K|\Theta}$ is not limited to some specific kernels related to a robust M-estimator. The joint density estimate $\hat{p}_{\vec{\lambda}_K,\Theta}$ extends $\hat{p}_{\vec{\lambda}_K|\Theta}$ and provides a Bayesian framework where prior information about the latent variable of interest can be added. We summarise in Table 1 our approach for defining objective functions for inference with GR^{2}T.

4. Applications of GR^{2}T

Modelling with GR^{2}T is illustrated for line and circle detection, line tracking and 3D shape from silhouettes.

4.1. Line and circle detection in images

The most common application of the Hough transform is for detecting lines in images. In this case, the link function F is often chosen as

$$F(\mathbf{x},\Theta)=\mathbf{x}^T\begin{pmatrix}\cos\Theta\\ \sin\Theta\end{pmatrix}=x_1\cos\Theta+x_2\sin\Theta$$

where $\mathbf{x}\in\mathbb{R}^2$ is a spatial position for which N observations $\{\mathbf{x}^{(i)}=(x_1^{(i)},x_2^{(i)})\}_{i=1,\dots,N}$ have been collected. In image processing, these observations corresponds to the positions of the edge pixels. The latent variable Θ is an angle between $-\pi/2$ and $\pi/2$. Without any model to link Θ and $\mathbf{x}$, independence can be assumed and the cost function $\hat{p}_{\lambda|\Theta}$ can be used for inference. This can be generalised to infer hyperplanes in point clouds and by choosing the density p_ϵ differentiable (e.g. normal [20,30]), gradient ascent algorithms can be defined to find the local and global maxima for the latent variable (λ,Θ) [30].

Using the spatial derivatives of the image (noted (x_3,x_4)) computed at location (x_1,x_2) that gives information about the orientation of the contour passing through (x_1,x_2), Dahyot [20] proposed to model $p_{\Theta|\mathbf{x}}$ as follows:

$$p_{\Theta|\mathbf{x}}(\Theta|\mathbf{x})=\frac{1}{\sqrt{2\pi\frac{\sigma^2}{x_3^2+x_4^2}}}\exp\left(-\frac{\left(\Theta-\arctan\frac{x_3}{x_4}\right)^2}{2\frac{\sigma^2}{x_3^2+x_4^2}}\right) \tag{34}$$

where σ is a constant capturing the noise on the image derivatives, and the collected observations are now

$$\{\mathbf{x}^{(i)}=(x_1^{(i)},x_2^{(i)},x_3^{(i)},x_4^{(i)})\}_{i=1,\dots,N}$$

The cost function $\hat{p}_{\lambda\Theta}$ can be used efficiently to find the linear contour without discarding any pixel in the image [20].

This framework for finding lines can be extended to other parametric shapes such as circles using for instance the algebraic distance [31]:

$$\lambda+\underbrace{\begin{pmatrix}\cos\theta_1\cos\theta_2\\ \sin\theta_1\cos\theta_2\\ \sin\theta_2\end{pmatrix}^T\begin{pmatrix}x_1\\ x_2\\ x_1^2+x_2^2\end{pmatrix}}_{F(\mathbf{x}=(x_1,x_2),\Theta=(\theta_1,\theta_2))}=\epsilon$$

Gander et al. have studied several different functions F to infer circles with the standard likelihood function $\mathcal{L}$ [31]. They showed that the geometric distance is a better alternative to find circles with $\mathcal{L}$:

$$\lambda+\underbrace{\sqrt{(x_1-\theta_1)^2+(x_2-\theta_2)^2}}_{F(\mathbf{x}=(x_1,x_2),\Theta=(\theta_1,\theta_2))}=\epsilon$$

where the latent variable λ models the radius of the circles, and (θ_1,θ_2) are the spatial coordinates of their centres. Note that the likelihood $\mathcal{L}$ is a special case of $\hat{p}_{\underset{\lambda}{\rightarrow}\,N|\Theta}$ (see Section 3.4). Alternatively many papers on the discrete Hough transform have also focused on line and circle detection [7,5,17,6].

4.2. Line tracking in videos

Liu et al. uses the Hough transform for detecting and tracking lanes on road images, and having detected the parameters of a line in the image at time t, this information is used to reduce the domain for searching the updated parameter values at time $t+1$ [32]. To understand how one can limit the search in the space of the latent variable of interest in the GR^{2}T framework, lets consider the following link function:

$$F(\mathbf{x}=(x_1,x_2),\Theta=(\theta_1,\theta_2))=\theta_1+x_1\cos\theta_2+x_2\sin\theta_2 \quad (35)$$

where the latent variable Θ characterises entirely a line, then inference about Θ can be used with the density estimates $\hat{p}_{\lambda|\Theta}(\lambda=0|\Theta)$ and $\hat{p}_{\lambda\Theta}(\lambda=0,\Theta)$ (using λ as an auxiliary variable). The joint density function $\hat{p}_{\lambda\Theta}(\lambda=0,\Theta)$ allows to consider a prior about Θ given $\mathbf{x}$. For instance, choosing the prior $p_{\Theta|\mathbf{x}}$ as the rectangular distribution allows to limit the domain of the search for the best estimate in the Θ-space.

4.3. Inferring 3D shapes from silhouettes [33]

Three-dimensional reconstruction of an object that is seen by multiple image sensors has many applications such as 3D modelling [34,35]. Shape from silhouettes methods are very popular in computer vision because of their simplicity and their computational efficiency. Laurentini [36] has defined the visual hull as the best reconstruction that can be computed using an infinite number of silhouettes captured from all viewpoints outside the convex hull of the object. Volume-based approaches focus on the volume of the visual hull while surface-based approaches, less numerically stable, aim at estimating a surface representation of the visual hull [37]. Ruttle et al. infer 3D shape from silhouettes using the following link function [33]:

$$F(\mathbf{x},\Theta)=\begin{pmatrix} x_1-\frac{\theta_1x_3+\theta_2x_4+\theta_3x_5+x_6}{\theta_1x_{11}+\theta_2x_{12}+\theta_3x_{13}+x_{14}} \\ x_2-\frac{\theta_1x_7+\theta_2x_8+\theta_3x_9+x_{10}}{\theta_1x_{11}+\theta_2x_{12}+\theta_3x_{13}+x_{14}} \end{pmatrix}$$

where:

- $\mathbf{x}\in\mathbb{R}^{14}$ with (x_1,x_2) the variables modelling the pixel coordinates recorded by the camera with matrix of parameters:

$$\begin{bmatrix} x_3 & x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 & x_{10} \\ x_{11} & x_{12} & x_{13} & x_{14} \end{bmatrix} \quad (36)$$

  For real cameras, the relation between the pixel position and their corresponding 3D world coordinates is not linear [38]. The observations $\{\mathbf{x}^{(i)}\}_{i=1,\ldots,N}$ correspond to vectors concatenating the positions of the pixels in the silhouettes images with their corresponding camera parameters. Only the N pixels on the foreground of the image silhouettes are kept as observations.
- $\Theta=(\theta_1,\theta_2,\theta_3)$ is the 3D coordinates in the real world.
- $F(\mathbf{x},\Theta)$ is defined from $\mathbb{R}^{14}\times\mathbb{R}^3$ to $\mathbb{R}^2$ and models the projection of a position in the real world (Θ) on the pixel position in the images [38].
- $\epsilon=(\epsilon_1,\epsilon_2)\in\mathbb{R}^2$ is the error on the pixel position. The density p_ϵ is chosen normal centred on 0 with diagonal covariance matrix with standard deviations h_1 and h_2. We choose $h_1=h_2=1$ to model the uncertainty about pixel positions due to digitisation of the images.

Ruttle et al. infer the 3D shape by searching the maxima of the conditional $p_{\lambda|\Theta}(\lambda=0|\Theta)$ using λ as an auxiliary variable [33]. $\hat{p}_{\lambda|\Theta}(\lambda=0|\Theta)$ is a smooth density estimate of the visual hull of the object since it is inferred using silhouettes. The volume of the object can be inferred by thresholding $\hat{p}_{\lambda|\Theta}(\lambda=0|\Theta)$ or by using gradient ascent method [33]. The cost function $p_{\lambda|\Theta}(\lambda=0|\Theta)$ is a smooth extension to volume based approaches based on voxel occupancy [39–42] that rely on the optimisation of a discrete objective function.

5. Conclusion

This paper presented the generalised relaxed Radon transform (GR^{2}T) as a generic framework for statistical inference. Modelling with GR^{2}T explains well-known cost functions such as the Hough transform and the likelihood function as non-parametric probability density functions of the latent variables of interest. As an improvement to the Hough transform, GR^{2}T also allows to consider prior information about the latent variables. The proposed framework is versatile and the resulting cost function depends on the choices made by the experimenter: the selection of the link function F, using or not the additive latent variable λ as an auxiliary variable, considering subsets of size K of observations, density distribution chosen for modelling the noise ϵ, prior information about the latent variable.

GR^{2}T also explains robust cost functions such as the generalised projection based M-estimator [14,15]. One important problem in robust computer vision is not only to compute estimates for the latent variables but also simultaneously estimate the scale parameter [14,15,43]. Indeed the distribution p_ϵ also depends on nuisance parameters v (e.g. if p_ϵ is the normal distribution $\mathcal{N}(0,\Sigma)$, then the covariance matrix or scale parameter Σ is often unknown and is a nuisance parameter that needs to be estimated). The nuisance parameter v may vary with the latent and/or observed variables such that $p_\epsilon(\epsilon;v(\mathbf{x},\Theta))$, e.g. the noise may be different for each structures Θ, and even may vary in the observed space. Future efforts will focus on extending GR^{2}T to also allow for the estimation of these nuisance parameters for a broader range of applications.

References

[1] K. Denecker, J.V. Overloop, F. Sommen, The general quadratic radon transform, Inverse Problems (1998) 615–633.
[2] P. Huber, Robust Statistics, John Wiley and Sons, 1981.
[3] P. Hough, Methods of Means for Recognising Complex Patterns, US Patent 3069654, 1962.
[4] N. Kiryati, Y. Eldar, A.M. Bruckstein, A probabilistic Hough transform, Pattern Recognition 24 (1991) 303–316.
[5] J. Princen, J. Illingworth, J. Kittler, A formal definition of the Hough transform: properties and relationships, Journal of Mathematical Imaging and Vision 1 (1992) 153–168.
[6] C.F. Olson, Constrained Hough transform for curve detection, Computer Vision and Image Understanding 73 (1999).

[7] S. Shapiro, Generalization of the Hough transform for curve detection in noisy digital images, in: Joint Conference on Pattern Recognition, Kyoto, Japan, pp. 710–714.
[8] P. Thrift, S.M. Dunn, Approximating Point Set Images by Line Segments Using a Variation of the Hough Transform, vol. 21, pp. 383–394.
[9] N. Kiryati, A.M. Bruckstein, Heteroscedastic hough transform (htht): an efficient method for robust line fitting in the errors in the variables problem, Computer Vision and Image Understanding 78 (2000) 69–83.
[10] S.R. Deans, Hough transform from the Radon transform, IEEE Transactions on Pattern Analysis and Machine Intelligence 3 (1981).
[11] S.R. Deans, The Radon Transform and Some of Its Applications, John Wiley and Sons, 1983.
[12] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall/CRC, 1998.
[13] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of ACM 24 (1981) 381–395.
[14] S. Mittal, S. Anand, P. Meer, Generalized projection based m-estimator: theory and applications, in: Computer Vision and Pattern Recognition Conference, Colorado Springs, CO, pp. 2689–2696.
[15] S. Mittal, S. Anand, P. Meer, Generalized projection based m-estimator, IEEE Transactions on Pattern Analysis and Machine Intelligence, http://dx.doi.org/10.1109/TPAMI.2012.52, in press.
[16] C. Stewart, Bias in robust estimation caused by discontinuities and multiple structures, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 818–833.
[17] M. Soffer, N. Kiryati, Guaranteed convergence of the Hough transform, Computer Vision and Image Understanding 69 (1998) 119–134.
[18] A. Goldenshluger, A. Zeevi, The Hough transform estimator, The Annals of Statistics 32 (2004).
[19] N. Toronto, B.S. Morse, D. Ventura, K. Seppi, The Hough transform's implicit Bayesian foundations, in: International Conference on Image Processing (ICIP), San Antonio, Texas USA.
[20] R. Dahyot, Statistical Hough transform, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 1502–1509.
[21] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer Verlag, 1999.
[22] R.O. Duda, P.E. Hart, Use of the Hough transformation to detect lines and curves in pictures, Communications of the ACM 15 (1972) 11–15.
[23] P. Toft, Using the generalized radon transform for detection of curves in noisy images, in: 1996 Conference Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP-96, 1996, vol. 4, pp. 2219–2222.
[24] K. Hansen, P. Toft, Fast curve estimation using preconditioned generalized Radon transform, IEEE Transactions on Image Processing 5 (1996) 1651–1661.
[25] C.L.L. Hendricks, M. van Ginkel, P.W. Verbeek, L.J. van Vliet, The generalized Radon transform: sampling, accuracy and memory considerations, Pattern Recognition 38 (2005) 2494–2505.
[26] P. Daras, D. Zarpalas, D. Tzovaras, M. Strintzis, Efficient 3-d model search and retrieval using generalized 3-d Radon transforms, IEEE Transactions on Multimedia 8 (2006) 101–114.
[27] J. Kim, D. Pollard, Cube root asymptotics, Annals of Statistics (1990) 191–219.
[28] B.V. Srinivasan, H. Qi, R. Duraiswami, Gpuml: graphical processors for speeding up kernel machines, in: Workshop on High Performance Analytics—Algorithms, Implementations, and Applications, Siam Conference on Data Mining, 2010.
[29] C. Shen, M. Brooks, A. van den Hengel, Fast global kernel density mode seeking: applications to localization and tracking, IEEE Transactions on Image Processing 16 (2007).
[30] R. Dahyot, Mean-shift for Statistical Hough Transform, Technical Report 01/09, School of Computer Science and Statistics, Trinity College Dublin, 2009.
[31] W. Gander, G. Golub, R. Strebel, Least-square fitting of circles and ellipses, BIT Numerical Mathematics (1994) 558–578.
[32] G. Liu, F. Wörgötter, I. Markelić, Combining statistical hough transform and particle filter for robust lane detection and tracking, in: Intelligent Vehicles Symposium (IV), IEEE, 2010, pp. 993–997.
[33] J. Ruttle, M. Manzke, R. Dahyot, Smooth kernel density estimate for multiple view reconstruction, in: Proceedings of the 7th European Conference for Visual Media Production, CVMP 2010, pp. 74–81.
[34] C.R. Dyer, Volumetric scene reconstruction from multiple views, in: L.S. Davis (Ed.), Foundation of Image Understanding, Kluwer, Boston, 2001, pp. 469–489.
[35] J. Franco, E. Boyer, Efficient polyhedral modeling from silhouettes, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 414–427.
[36] A. Laurentini, How far 3d shapes can be understood from 2d silhouettes, IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (1995) 188–195.
[37] S. Lazebnik, E. Boyer, J. Ponce, On computing exact visual hulls of solids bounded by smooth surfaces, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 156–161.
[38] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.
[39] W.N. Martin, J.K. Aggarwal, Volumetric descriptions of objects from multiple views, IEEE Transactions on Pattern Analysis and Machine Intelligence 5 (1983) 150–158.
[40] M. Potmesil, Generating octree models of 3d objects from their silhouettes in a sequence of images, Computer Vision, Graphics, and Image Processing vol. 40:I-29, 1987.
[41] P. Srivasan, P. Liang, S. Hackwood, Computational geometric methods in volumetric intersections for 3d reconstruction, Pattern Recognition 23 (1990) 843–857.
[42] R. Szeliski, Rapid octree construction from image sequences, CVGIP: Image Understanding 58 (1993) 23–32.
[43] H. Wang, T.-J. Chin, D. Suter, Simultaneously fitting and segmenting multiple-structure data with outliers, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (2012) 1177–1192.

Rozenn Dahyot received Master degrees in Physics (1998, Ecole Nationale Superieure de Physique de Strasbourg) and Image Processing (1998, University Louis Pasteur Strasbourg) in France. She gained her Ph.D. in 2001 from the University Louis Pasteur Strasbourg. From 2002 to 2005, she was a research fellow in Trinity College Dublin, Ireland, and University of Cambridge, UK. Since 2005, she is an assistant professor in the School of Computer Science and Statistics, in Trinity College Dublin. Her research interest includes robust statistics, pattern detection and recognition, image and video analysis.

Jonathan Ruttle received Master degree in Computer Science (Interactive Entertainment Technologies) in 2008 from Trinity College Dublin, Ireland. Since 2008 Jonathan is a PhD candidate in Trinity College, and his research interests include multiple view 3D reconstruction and statistical inference from depth cameras.