Total Variation Theory and Its Applications

2nd UCC Annual Research Conference, Kingston, Jamaica

Peter Ndajah

University of the Commonwealth Caribbean, Kingston, Jamaica

September 27, 2018

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Introduction

The total variation of a real-valued function f defined on an interval $[a, b]$ is a measure of the $1 - D$ arc length from a to b. Mathematically we can define the total variation of f as

$$V_a^b(f) = \sup_p \sum_{i=1}^{n} |f(x_i) - f(x_{i-1}| \tag{1}$$

where the supremum runs over the set of all partitions $P = \{a = x_0, x_1, ..., x_n = b\}$. (1) is useful if we are measuring the arc length of a one-dimensional function f. to extend the notion to more general curves, we may begin by observing that equation (1) can be multiplied and divided by Δx to obtain

$$V_a^b(f) = \sup_p \sum_{i=1}^{n} \left| \frac{f(x_i) - f(x_{i-1}}{\Delta x} \right| \Delta x \tag{2}$$

Equation (2) can be transformed into the continuous case by taking limits as $\Delta x \to 0$ and $n \to \infty$.

Derivations of The Total Variation Functional: Method I

This method may be called the method of first principles.
A function f defined on an interval $[a, b]$ is said to be of bounded variation if there is a constant $C > 0$ such that

$$\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \leq C$$

for every partition

$$a = x_0 < x_1 < ... < x_n = b$$

of $[a, b]$ by points of subdivision $x_0, x_1, ..., x_n$. If f is of bounded variation, then by the total variation of f is meant the quantity

$$V_a^b(f) = \sup \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|$$

where the least upper bound is taken over all partitions of the interval $[a, b]$. The above definitions work quite well for Riemann integrable functions

Derivations of The Total Variation Functional: Method II

This method uses measure theory to describe the total variation functional. It is useful only when describing the total variation qualitatively. Let λ be a charge on X and let P, N be a Hahn decomposition for λ. The positive and negative variations of λ are the finite measuresλ^+, λ^- defined for E in X by

$$\lambda^+(E) = \lambda(E \cap P)$$

$$\lambda^-(E) = -\lambda(E \cap N)$$

The total variation of λ is the measure $|\lambda|$ defined for E in X by

$$|\lambda| = \lambda^+(E) + \lambda^-(E)$$

If f is Lebesgue-integrable and f belongs to $L(X, \mathbf{X}, \mu)$ with respect to a measure μ on X, and if λ is defined for E in X by

$$\lambda(E) = \int_E f d\mu,$$

Derivations of The Total Variation Functional: Method II

If f is Lebesgue-integrable and f belongs to $L(X, \mathbf{X}, \mu)$ with respect to a measure μ on X, and if λ is defined for E in X by

$$\lambda(E) = \int_E f d\mu,$$

then λ is a charge and

$$\lambda^+(E) = \int_E f^+ d\mu$$
$$\lambda^-(E) = \int_E f^- d\mu$$

so that

$$|\lambda|(E) = \int_E |f| d\mu = \int_E f^+ d\mu + \int_E f^- d\mu \tag{4}$$

Here, f is any Lebesgue-integrable function. The sum of the positive and negative variations of the function f gives its total variation.

Derivations of The Total Variation Functional: Method III

A function $\mu \in L^1(\Omega)$ whose partial derivatives in the sense of distributions are measures with finite total variation in Ω is called a function of bounded variation. This class of functions is usually denoted by $BV(\Omega)$. Thus, $u \in BV(\Omega)$ if and only if there are signed measures $\mu_1, \mu_2, ..., \mu_n$ defined in Ω such that for $i = 1, 2, ..., n$,

$$|Du|(\Omega) < \infty$$

and

$$\int uD_i\varphi dx = -\int \varphi d\mu_i \tag{5}$$

for all $\varphi \in C_0^\infty(\Omega)$. The gradient of μ will therefore be a vector valued measure with finite total variation:

$$\|Du\| =$$

$$\sup\{\int_\Omega u\mathrm{div}vdx : v = (v_1, ..., v_n) \in C_0^\infty(\Omega; R^n),$$

Derivations of The Total Variation Functional: Method III

The divergence of a vector field is denoted by $\mathrm{div} v$ and is defined by

$$\mathrm{div} v = \sum_{i=1}^{n} D_i v_i = \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i}$$

If $u \in BV(\Omega)$, the total variation $\|Du\|$ may be regarded as a measure, for if f is a non-negative real-valued continuous function with compact support in Ω, we may define

$$\|Du\|(f) = \sup\{\int_\Omega u \mathrm{div} v dx : v = (v_1, ..., v_n) \in C_0^\infty(\Omega; R^n), |v(x)| \leq f(x) \forall x \in \Omega\} \quad (8)$$

$||Du||$ is additive, continuous under monotone convergence and a non-negative Radon measure on Ω.

Derivations of The Total Variation Functional: Method III

The space of absolutely continuous u with $u' \in L^1(R^1)$ is contained in $BV(R^1)$. In the same manner in R^n, a Sobolev function is also BV. That is, $W^{1,1}(\Omega) \subset BV(\Omega)$. If $u \in W^{1,1}(\Omega)$ then,

$$\int_\Omega u \operatorname{div} v dx = -\int_\Omega \sum_{i=1}^{n} D_i u v dx \tag{9}$$

and the gradient of u has finite total variation with

$$||Du||(\Omega) = \int_\Omega |Du| dx. \tag{10}$$

Derivations of The Total Variation Functional: Method III

This measures both the positive and negative variations implicitly rather than explicitly. If $u \in C^1(\Omega)$, then

$$\int_\Omega |Du| dx = \int_\Omega |\nabla u(x)| dx \tag{11}$$

Let $\Omega \in R^2$ be an image surface. Then we have

$$\int_\Omega |\nabla u(x)| dx = \int \int \sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2} dxdy \tag{12}$$

which is the total variation of an image u with two independent variables x, y.

Ndajah and Kikuchi Derivation: Method IV

We derive the total variation of an image using the vector gradients. Images are two dimensional spatial functions. In general, to find the total variation of an n-dimensional mathematical object, we consider the directional derivative of a scalar function $f(\vec{x}) = f(x_1, x_2, ..., x_n)$ along a unit vector $\vec{u} = (u_1, ..., u_n)$. The directional derivative is defined to be the limit

$$\nabla_{\vec{u}} f(\vec{x}) = \lim_{h \to 0^+} \frac{f(\vec{x} + h\vec{u}) - f(\vec{x})}{h} \tag{13}$$

This means that the directional derivatives exist along any unit vector $\vec{u}$, and one has

$$\nabla_{\vec{u}} f(\vec{x}) = \nabla f(\vec{x}) \cdot \vec{u} \tag{14}$$

Ndajah and Kikuchi Derivation: Method IV

For an image we represent the directional derivatives by **i**, **j** so that

$$\begin{cases} \nabla_x f(\vec{x}) = \nabla f(\vec{x}) \cdot \mathbf{i} & = \frac{\partial f}{\partial x} \\ \nabla_y f(\vec{x}) = \nabla f(\vec{x}) \cdot \mathbf{j} & = \frac{\partial f}{\partial y} \end{cases} \tag{15}$$

The components $\nabla_x f(\vec{x})$ and $\nabla_y f(\vec{x})$ are orthogonal so that

$$\nabla f(\vec{x}) = \mathbf{i}\frac{\partial f}{\partial x} + \mathbf{j}\frac{\partial f}{\partial y} \tag{16}$$

and the inner product is

$$\langle \nabla f(\vec{x}), \nabla f(\vec{x}) \rangle = \|\nabla f\|^2 = \int\int \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} dxdy \tag{17}$$

Therefore, total variation in two dimensions can be written as

$$\int\int |\nabla f| \, dxdy \tag{18}$$

Total Variation Diminishing

In numerical methods, total variation diminishing (TVD) is a property of certain discretization schemes used to solve hyperbolic partial differential equations. It is mostly applied in computational fluid dynamics. This concept was first introduced by Ami Harten.
In systems described partial differential equations, such as

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \tag{19}$$

$$TV = \int \left| \frac{\partial u}{\partial x} \right| dx \tag{20}$$

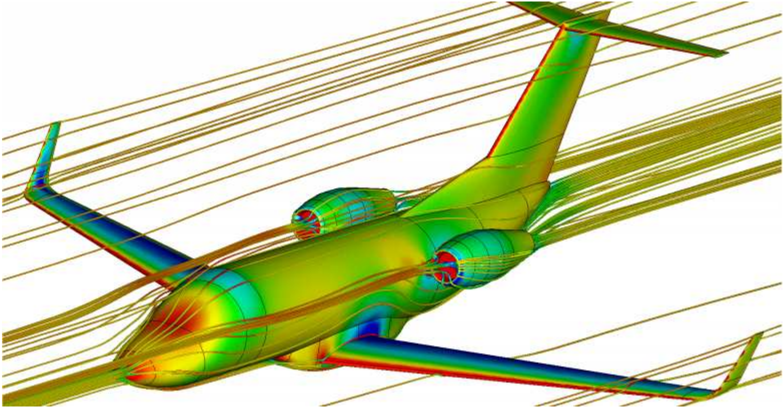
A numerical method is said to be total variation diminishing (TVD) if,

$$TV(u^{n+1}) \leq TV(u^n)$$

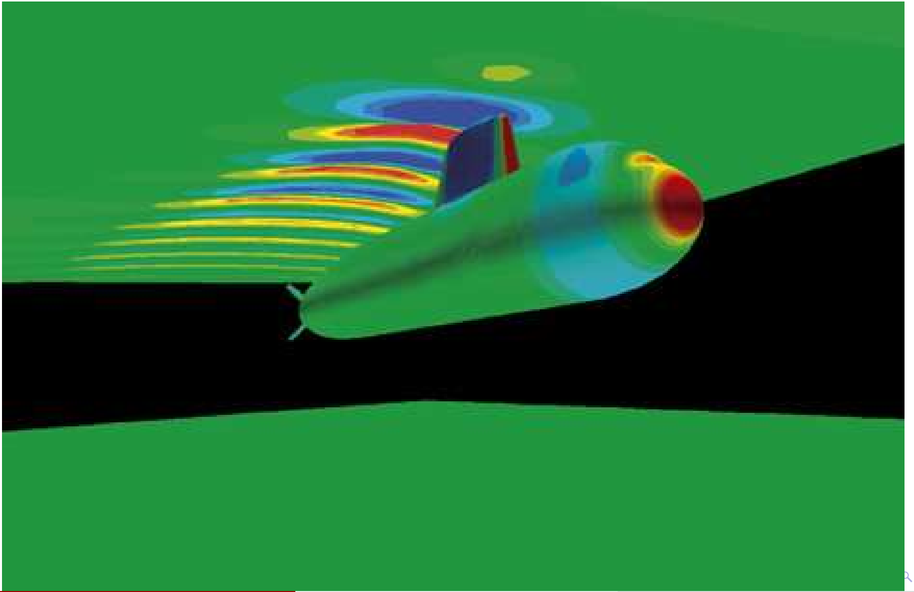
Total Variation Diminishing

In Computational Fluid Dynamics, TVD scheme is used to capture sharper shock predictions without leaving any misleading oscillations when variation of field variable is discontinuous. On the other hand, to capture the variation, fine grids ($\Delta x =$ very small) are needed and the computations can become heavy and uneconomic. The use of coarse grids with central difference scheme, upwind scheme, hybrid difference scheme, and power law scheme gives false shock predictions. The TVD scheme enables sharper shock predictions on coarse grids saving computation time and as the scheme preserves monotonicity there are no spurious oscillations in the solution.

Total Variation Diminishing



Total Variation Diminishing



Total Variation Diminishing

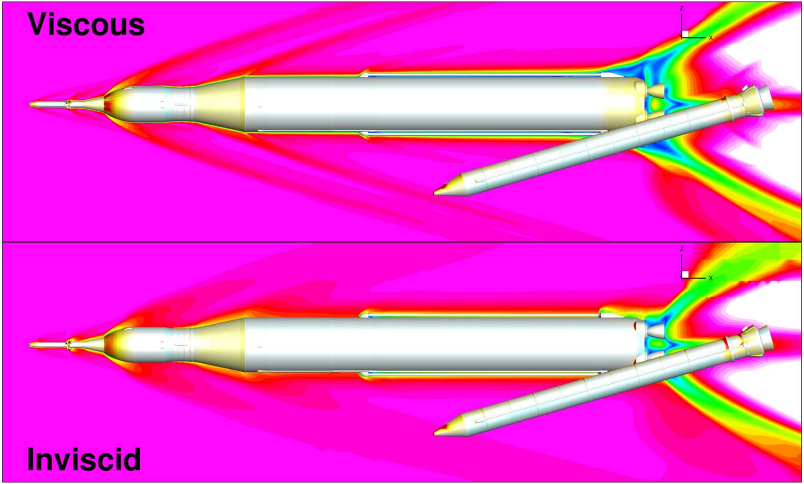


Image Denoising

Rudin, Osher and Fatemi (1990) developed finite difference schemes were developed there which were used to enhance mildly blurred images significantly while preserving the variation of the original image. They concluded that the space of BV functions is the proper class for many basic image processing tasks. Thus, the constrained minimization problem is

$$\min \int_{\Omega} \sqrt{u_x^2 + u_y^2}\, dxdy \tag{21}$$

subject to the constraints involving u_0. They took the two constraints as above:

Image Denoising

This constraint signifies the fact that the white noise $\eta(x, y)$ is of zero mean and

$$\int_{\Omega} \frac{1}{2}(u - u_0)^2 = \sigma^2$$

where $\sigma > 0$ is given. The second constraint uses a priori information that the standard deviation of the noise $\eta(x, y)$ is σ. Thus, there is a linear and nonlinear constraint.

They arrive at the Euler-Lagrange equations

$$0 = \frac{\partial}{\partial x}\left(\frac{u_x}{\sqrt{u_x^2 + u_y^2}}\right) + \frac{\partial}{\partial y}\left(\frac{u_y}{\sqrt{u_x^2 + u_y^2}}\right) - \lambda_1 - \lambda_2(u - u_0) \qquad (22)$$

in Ω, with $\frac{\partial u}{\partial n} = 0$ on the boundary of $\Omega = \partial\Omega$

Image Denoising

The solution uses a parabolic equation with time as an evolution parameter, or equivalently, the gradient descent method. This means that we solve

$$u_t = \frac{\partial}{\partial x}\left(\frac{u_x}{\sqrt{u_x^2 + u_y^2}}\right) + \frac{\partial}{\partial y}\left(\frac{u_y}{\sqrt{u_x^2 + u_y^2}}\right) - \lambda(u - u_0) \tag{23}$$

for $t > 0$, $x, y \in \Omega$, $u(x, y, 0)$ given $\frac{\partial u}{\partial n} = 0$ on $\partial\Omega$ As t increases, we approach a denoised version of the image. We must compute $\lambda(t)$. If steady state has been reached, the left hand side of equation (23) vanishes. We then have

$$\lambda = -\frac{1}{2\sigma^2}\int\left[\sqrt{u_x^2 + u_y^2} - \left(\frac{(u_0)_x u_x}{\sqrt{u_x^2 + u_y^2}} + \frac{(u_0)_y u_y}{\sqrt{u_x^2 + u_y^2}}\right)\right] dxdy \tag{24}$$

Image Denoising

This gives a dynamic value for $\lambda(t)$, which appears to converge as $t \to \infty$. This is merely the gradient projection method of Rosen in his 1961 paper titled "The Gradient Projection Method for Nonlinear Programming part II, nonlinear constraints" published by SIAM. The numerical method in two dimensions is:

$x_i = ih$, $y_j = jh$ $i, j = 0, 1, ..., N$ with $Nh = 1$
$t_n = n\Delta t$, $n = 0, 1, ...$
$u_{ij}^n = u(x_i, y_i, t_n)$,
$u_{ij}^0 = u(ih, jh) + \sigma\varphi(ih, ih)$

Image Denoising

So the numerical approximation to equations (23) and (24) is

$$
\begin{aligned}
u_{ij}^{n+1} = u_{ij}^{n} + & \\
& \frac{\Delta t}{h}\Bigg[\Delta_-^x\left(\frac{\Delta_+^x u_{ij}^n}{(\Delta_+^x u_{ij}^n)^2 + (m(\Delta_+^x u_{ij}^n, \Delta_-^y u_{ij}^n)^2)^{\frac{1}{2}}}\right) \\
& \quad + \Delta_-^y\left(\frac{\Delta_+^y u_{ij}^n}{\Delta_+^y u_{ij}^n + (m(\Delta_+^x u_{ij}^n, \Delta_-^x u_{ij}^n)^2)^{\frac{1}{2}}}\right) \\
& \qquad - \Delta t \lambda^n (u_{ij}^n - u_0(ih, ih)), \quad (25)
\end{aligned}
$$

for $i, j = 1, ..., N$ with boundary conditions
$u_{0j}^n = u_{1j}^n$, $u_{Nj}^n = u_{N-1,j}^n$, $u_{i0}^n = u_{iN}^n = u_{i,N-1}^n$
Here, $\Delta_{\mp}^x u_{ij} = \mp(u_{i\mp1,j} - u_{ij}$

Image Denoising

and similarly for $\Delta_{\mp}^{y} u_{ij}$.
λ^n is defined discretely as

$$\lambda^n = -\frac{h}{2\sigma^2}\Bigg[\sum\Bigg(\sqrt{(\Delta_+^x u_{ij}^n)^2 + (\Delta_+^y u_{ij}^n)^2} - \frac{(\Delta_+^x u_{ij}^0)(\Delta_+^x u_{ij}^n)}{\sqrt{(\Delta_+^x u_{ij}^n)^2 + (\Delta_+^y u_{ij}^n)^2}} - \frac{(\Delta_+^y u_{ij}^0)(\Delta_+^y k u_{ij}^n)}{\sqrt{(\Delta_+^x u_{ij}^n)^2 + (\Delta_+^y u_{ij}^n)^2}} \quad (26)$$

A step size restriction is imposed for stability: $\frac{\Delta t}{h^2} \leq c$

Image Denoising

Image Denoising

Reference Image: MSE = 255, SSIM = 1
(a)

Denoised Image: $\lambda = 60, \tau = 0.01$,
MSE = 255, SSIM = 0.653430
(b)

Denoised Image: $\lambda = 12$, $\tau = 0.01$,
MSE = 255, SSIM = 0.892388
(c)

Denoised Image: $\lambda = 2$, $\tau = 0.01$,
MSE = 255, SSIM = 0.748494
(d)

Denoised Image: $\lambda = 1$, $\tau = 0.01$,
MSE = 255, SSIM = 0.712412
(e)

Denoised Image: $\lambda = 0.5$, $\tau = 0.01$,
MSE = 255, SSIM = 0.685501
(f)

Image Edge Detection

We show edge detection using total variation and compare the result with the Marr-Hildreth method. The scaling factor σ enables us to tune the edge detection process.

$$G(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{27}$$

to smooth the image $u(x, y)$. This is accomplished by a convolution of both functions i.e. $G(x, y) \star u(x, y)$. σ is the standard deviation of $G(x, y)$ and it acts as a scaling factor, blurring out noise and structures with scales below σ. Therefore, we get no absolute definition of edges. We only talk about edges at a certain scale.

Image Edge Detection

The convolution

$$D(x,y) = G(x,y) \star u(x,y) \tag{28}$$

yields a smoothed image at a scale σ to which the Laplacian operator is applied i.e. $\nabla^2 \left(G(x,y) \star u(x,y)\right)$. This operation is commutative and gives the same result as

$$\left(\nabla^2 G(x,y)\right) \star u(x,y) = \nabla^2 D(x,y) \tag{29}$$

The expression $\nabla^2 G$ is called the Laplacian of a Gaussian (LoG) and is expressed as

$$\nabla^2 G(x,y) = \left[\frac{x^2+y^2-2\sigma^2}{\sigma^4}\right] e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{30}$$

Image Edge Detection

Likewise, for the total variation approach, we first smooth the image by convolving $u(x, y)$ and the Gaussian function $G(x, y)$ just as in the case of the Laplacian approach. We then apply the total variation operator to the convolved image $D(x, y)$ i.e. $\frac{\nabla^2 D(x,y)}{|\nabla D|}$. This operation is not commutative as in the case of the LoG. Below we show results of edge detection based on the LoG images and total variation filtered images. Edge detection for both kinds of images do not follow exactly the same procedure. The traditional method of detecting edges in LoG images is by zero crossings.

Image Filters

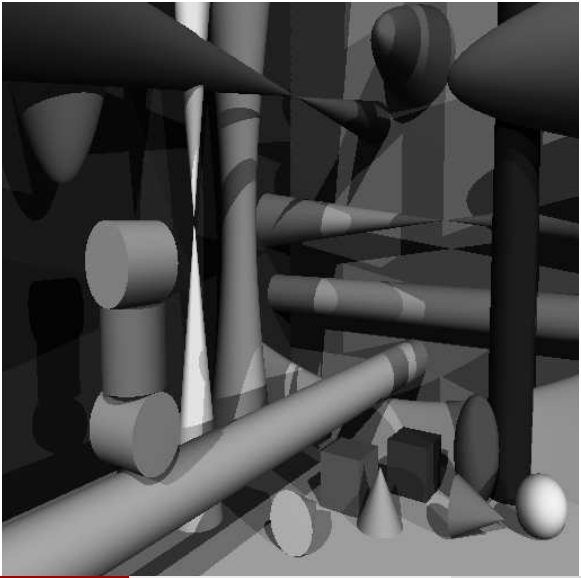


Image Filters

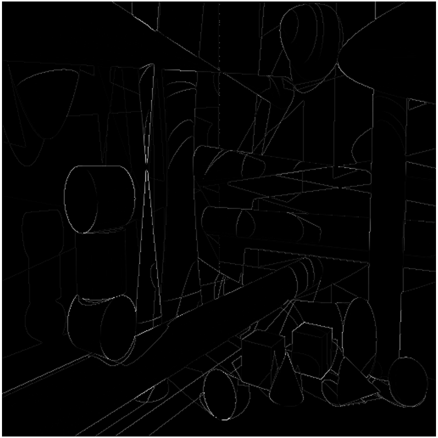


Figure: Laplacian Filtered Image Showing Edges

TV Image Filter

We obtained a minimization of the total variation functional by means of the Euler-Lagrange minimization method. We use the resulting nonlinear steady state partial differential equation given by

$$J_{\min} u = \frac{\nabla^2 u}{|\nabla u|} \tag{31}$$

with the Dirichlet boundary condition $\frac{\partial u}{\partial n} = 0$ to filter images in place of the Laplacian. The changes in the image texture are well captured in the filtered image. Also, the Laplacian forms double edges during image filteration process. The total variation filter overcomes this disadvantage by producing only single a edge.

TV-Filtered Image

TV filter

Comparing Marr-Hildreth and Total Variation

Figure: Gaussian Smoothed Image with $\sigma = 2$

Comparing Marr-Hildreth and Total Variation

Figure: Marr-Hildreth Edge Detection with $\sigma = 2$

Comparing Marr-Hildreth and Total Variation

Figure: Total Variation Edge Detection with $\sigma = 2$

Comparing Marr-Hildreth and Total Variation

Figure: Gaussian Smoothed Image with $\sigma = 3$

Comparing Marr-Hildreth and Total Variation

Figure: Marr-Hildreth Edge Detection with $\sigma = 3$

Comparing Marr-Hildreth and Total Variation

Figure: Total Variation Edge Detection at $\sigma = 3$

Comparing Marr-Hildreth and Total Variation

Figure: Gaussian Smoothed Image with $\sigma = 4$

Comparing Marr-Hildreth and Total Variation

Figure: Marr-Hildreth Edge Detection with $\sigma = 4$

Comparing Marr-Hildreth and Total Variation

Figure: Total Variation Edge Detection at $\sigma = 4$

Comparing Marr-Hildreth and Total Variation

Figure: Gaussian Smoothed Image with $\sigma = 5$

Comparing Marr-Hildreth and Total Variation

Figure: Marr-Hildreth Edge Detection with $\sigma = 5$

Comparing Marr-Hildreth and Total Variation

Figure: Total Variation Edge Detection at $\sigma = 5$

Object Recognition

Edge detection algorithms are at the heart of object detection and recognition in computer vision systems. Processes such as image segmentation and pattern recognition depend on image edge detection. These primary techniques combined with deep learning is what seems to give machines the ability to recognize objects and act accordingly in different situations. An example is self-navigating unmanned ground vehicles sometimes called driverless cars. For the car to successfully travel on a road, it must be able to recognize lane marks. The algorithm central to this ability is image edge detection algorithm. Earlier methods used for this purpose include the Canny and Marr-Hildreth algorithms. As we have seen in previous sections, these algorithms will fail after scale 4 (i.e. $\sigma = 4$). This means that if the road marks are not bright enough, the vehicle will not be able to recognize them.

Lane Recognition

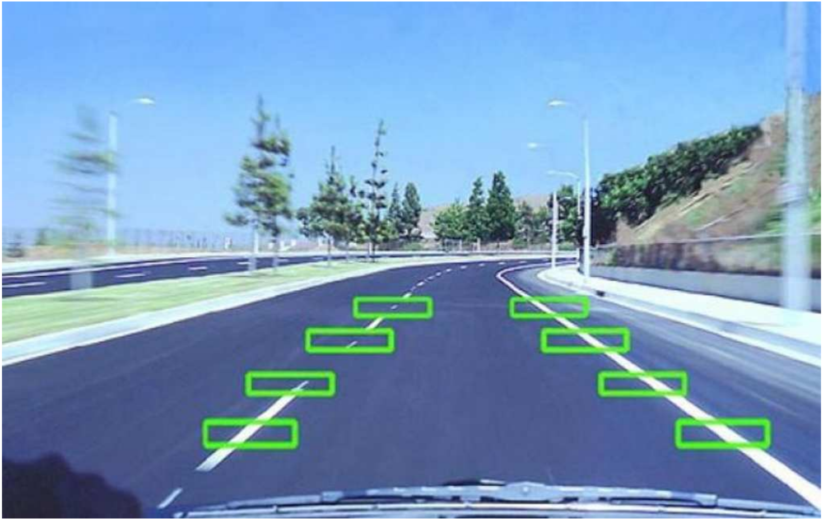


Figure: Driverless Car

Total Variation Book