From TV-L^1 Model to Convex & Fast Optimization Models for Image & Geometry Processing

Tony Chan, UCLA and NSF

Joint work with Xavier Bresson, UCLA

Papers: www.math.ucla.edu/applied/cam/index.html
Research group: www.math.ucla.edu/∼imagers

May 28, 2009

Outline

- The "classical" TV-L^2 model introduced by Rudin, Osher and Fatemi[1] is an efficient image denoising model. However, TV-L^2 does not preserve image contrast unlike the TV-L^1 model introduced by Chan and Esedoglu[2].
- The TV-L^1 model also leads to the convexification of several non-convex image processing models.
- Standard models are defined by non-convex energy minimization problems, which make them sensitive to initial conditions and slow to minimize. We show how to convexify standard image processing models and how to define fast optimization algorithms.
- Applications to Image Processing and Computer Vision: Image Segmentation, Multiview 3D Reconstruction, Stereo Evaluation, Optical Flow & Object Tracking, Multi-phase Segmentation, etc.

[1] Rudin-Osher-Fatemi, Nonlinear Total Variation Based Noise Removal Algorithms, 1992

[2] Chan-Esedoglu, Aspects of Total Variation Regularized L1 Function Approximation, 2005

Image Denoising: TV-L^2 Model (Rudin-Osher-Fatemi/ROF 92)

- The Total Variation (TV) norm has been successful in Image Processing. The TV-based ROF model defined as an energy minimization/variational model:

$$\min_{u, u-u_0 \in BV(\Omega)\times L^2(\Omega)} TV(u) + \frac{\lambda}{2}||u-u_0||_2^2,$$

$$\text{where } TV(u) = \int_\Omega |\nabla u| dx$$

removes the noise in u_0 while preserving discontinuities (edges).

- TV is important because:

• TV controls the size of jumps in signal since for u monotonic in $[a, b]$, then $TV(u) = |u(b) - u(a)|$, regardless of whether u is discontinuous or not. TV can handle image discontinuities.

• TV also controls the geometry of boundary since for the characteristic function 1_Σ of region $\Sigma \subset \Omega$ since we have: $TV(u = 1_\Sigma) = \int_{\partial\Sigma} ds = |\partial\Sigma| = Per(\partial\Sigma)$.

TV-L^1 Model: Contrast & Geometry Preservation (Alliney 97, Chan-Esedoglu 05)

- The ROF image denoising model preserves the geometry in the presence of noise. However, ROF loses the image contrast (Strong-Chan 96, Bellettini-Caselles-Novaga 02).
  Theorem: If $u_0 = 1_D$, D is convex, $\partial D \in C^{1,1}$ and for every $p \in \partial D$, $curv_{\partial D}(p) \leq \frac{|\partial D|}{|D|}$, then

$$u = \Big(1 - \underbrace{\frac{|\partial D|}{2\lambda|D|}}_{\text{Contrast Lost}}\Big)1_D$$

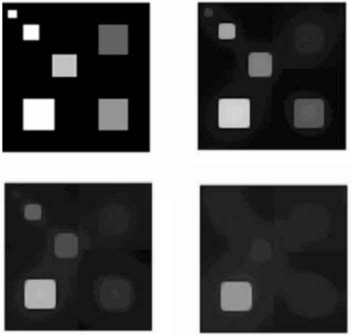
- The TV-L^1 energy minimization model:

$$\min_{u, u-u_0 \in BV(\Omega)\times L^1(\Omega)} TV(u) + \lambda||u - u_0||_1,$$

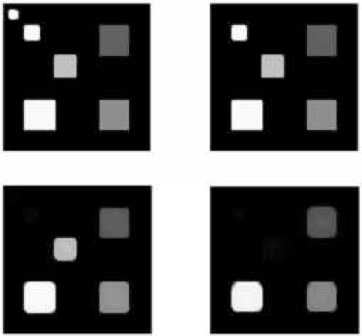
  - is robust to contrast and geometry perturbation in the presence of noise
  - does not perturb a clean image in the absence of noise

- Other properties of TV-L^1 Model:
  - Cleaner image multiscale decomposition than ROF
  - Data driven scale selection (detection of meaningful objects in images)
  - Shape denoising/Geometry regularization model

Scale-Space Generated by the ROF Model and the TV-L^1 Model



(a) ROF



(b) TV-L^1

Multiscale Decomposition of TV-L^1 (Related: Tadmor-Nezzar-Vese 03; Kunisch-Scherzer 03)

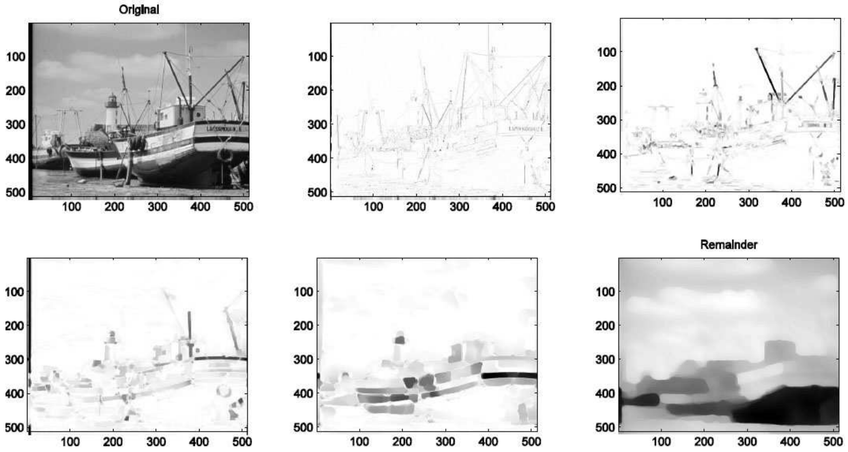


Figure: TV-L^1 decomposition gives well separated & contrast preserving features at different scales. E.g. boat masts, foreground boat appear mostly in only 1 scale.

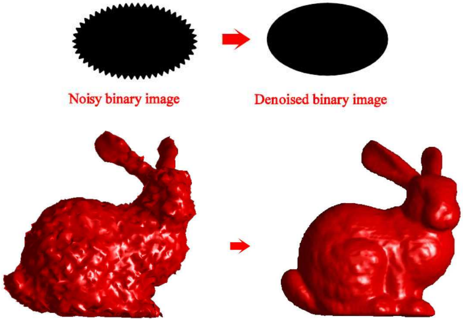
Shape Denoising/Geometry Regularization

- We will show that the convex TVL^1 model is equivalent to the non-convex shape denoising model defined as:

$$\min_{\Sigma} \; Per(\Sigma) + \lambda|\Sigma \Delta S|,$$

  where Per is the perimeter, $\lambda > 0$, and S is a given noisy shape.

- Examples of shape denoising (1D and 2D shapes):



Equivalence between TV-L^1 Model and Shape Denoising

- In the case of Shape Denoising, function u_0 is a binary function of the (noisy) shape D:

$$u_0(x) = 1_D(x) = \begin{cases} 1 & \text{if } x \in D \\ 0 & \text{otherwise} \end{cases}$$

- The co-area formula decomposes TV into the sum of Perimeters of level sets of u:

$$TV(u) = \int_\Omega |\nabla u| dx = \int_\mu Per(\underbrace{\{x : u(x) > \mu\}}_{\Sigma(\mu)}) d\mu.$$

- The "Layer Cake" formula decomposes the L^1-based data term as follows:

$$\int_\Omega |u - u_0| dx = \int_\mu |\Sigma(\mu) \Delta \{x : u_0(x) > \mu\}| d\mu.$$

- With $u_0 = 1_D$, we have

$$TVL^1(u) = \int_\mu Per(\Sigma(\mu)) + \lambda |\Sigma(\mu) \Delta D| d\mu.$$

which means that, for each upper level set of $u(x)$, we have the same geometry problem:

$$\min_\Sigma \; Per(\Sigma) + \lambda |\Sigma \Delta D|$$

TVL^1 provides a Global Minimizer of the Shape Regularization Problem (Chan-Esedoglu-Nikolova 05[3])

- Theorem: If the observed image $u_0(x)$ is the (noisy) binary function of a set D and if $u^\star$ is any minimizer of TVL^1, then for almost every $\mu \in (0, 1)$, the binary function $1_{\{x:u^\star(x)>\mu\}}(x)$ is also a minimizer of TVL^1.

- TVL^1 has "convexified" the shape regularization problem! In other words, we have established the equivalence of a convex problem (minimizing over all functions) to a non-convex problem (minimizing over geometric sets).

[3] Chan-Esedoglu-Nikolova, Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models, 2006

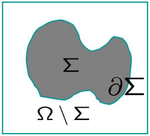
Image Segmentation: Active Contour Models

- Image segmentation consists in partitioning an image into multiple regions. Image segmentation locates meaningful objects in images. Applications are in video surveillance, medical imaging, etc.

- A well-posed mathematical model is the active contour model (Kass-Witkin-Terzopoulos 88). Objective: Find the set $\Sigma \subset \Omega \subset \mathbb{R}^N$ which provides the global minimum of the shape optimization problem:

$$\min_{\Sigma} \quad F_{AC}(\Sigma) = \underbrace{\int_{\partial\Sigma} w_b ds}_{Per_{w_b}(\Sigma)} + \lambda \underbrace{\int_{\Sigma} w_r^{in} dx}_{Area_{w_r^{in}}(\Sigma)} + \lambda \underbrace{\int_{\Omega\backslash\Sigma} w_r^{out} dx}_{Area_{w_r^{out}}(\Omega\backslash\Sigma)} . \quad (1)$$

- The shape model (1) is not convex because the set of $\{\Sigma\}$ and the energy F_{AC} are not convex.



- The convex TV-$\langle , \rangle$ energy defined as:

$$F_{TV\langle , \rangle}(u) = \underbrace{\int_\Omega w_b |\nabla u| dx}_{TV_{w_b}(u)} + \lambda \underbrace{\int_\Omega w_r^{in} u dx}_{<w_r^{in}, u>} + \lambda \underbrace{\int_\Omega w_r^{out}(1-u) dx}_{<w_r^{out}, 1-u>}$$

can be decomposed into a sum of upper level set energies (using weighted co-area formula and layer-cake formula):

$$F_{TV\langle , \rangle}(u) = \int_\mu Per_{w_b}(\Sigma(\mu)) + \lambda Area_{w_r^{in}}(\Sigma(\mu)) + \lambda Area_{w_r^{out}}(\Omega \setminus \Sigma(\mu)) \; d\mu$$

which means that, for each upper level set of $u(x)$, we have the same geometry problem:

$$\min_\Sigma \; Per_{w_b}(\Sigma) + \lambda Area_{w_r}(\Sigma) + \lambda Area_{w_r^{out}}(\Omega \setminus \Sigma) = F_{AC}(\Sigma)$$

which corresponds to the non-convex active contour minimization problem.

TV-$\langle , \rangle$ provides a Global Minimizer of the Active Contour Problem (Chan-etal 06, Bresson-etal 07)

- Theorem: Suppose that $w_b(x) \in \mathbb{R}_+$, for any fixed $w_r^{in}(x), w_r^{out}(x) \in \mathbb{R}$ and $\lambda \in \mathbb{R}_+$, if $u^\star$ is any minimizer of $\min_{0 \le u \le 1} \; F_{TV\langle , \rangle}(u)$, then for almost every $\mu \in (0,1)$, the binary function $1_{\{x:u^\star(x)>\mu\}}(x)$ is also a global minimizer of $F_{TV\langle , \rangle}$ and the active contour energy F_{AC}.

  TV-$\langle , \rangle$ has convexified the image segmentation problem!

- The non-convex Chan-Vese model:

$$\min_{\Sigma} \quad F_{CV}(\Sigma) = Per(\Sigma) + \lambda \int_{\Sigma} (c^{in} - u_0)^2 dx + \lambda \int_{\Omega \setminus \Sigma} (c^{out} - u_0)^2 dx,$$

  can be convexified as follows:

$$\min_{0 \le u \le 1} \; F_{CV}^c(u) = \int_{\Omega} |\nabla u| dx + \lambda \int_{\Omega} (c^{in} - u_0)^2 u dx + \lambda \int_{\Omega} (c^{out} - u_0)^2 (1-u) dx$$

  If $u^\star$ is any minimizer of F_{CV}^c, then the binary function $1_{\{x:u^\star(x)>\mu\}}(x)$ is a global minimizer of F_{CV}.

- What optimization algorithm for TV-$\langle , \rangle$? Standard algorithms for continuous energy minimization problems were usually slow to converge because they are non-convex (and non-differentiable).
  Since TV-$\langle , \rangle$ is convex, we can use efficient optimization techniques (that can lead to real-time applications).

- Rich literature on continuous convex optimization algorithms: Rockafellar 76 (Proximal Point Algorithm), Hestenes 69, Powell 69 (Method of Multipliers/ Augmented Lagrangian), Passty 79, Gabay 83, Tseng 88 (Forward-Backward), Lions 1978, Passty 79 (Double-Backward), Lions and Mercier 79 (Peaceman-Rachford), Lions and Mercier 79 (Douglas-Rachford).
  There has been a new interest for these optimization models to solve the problem of compressive sensing of Candes-Romberg-Tao 06, for examples Yin-Osher-Goldfarb-Darbon 08, Goldstein-Osher 08, Zhang-Burger-Bresson-Osher 09.

- We define an efficient algorithm based on Bregman Iteration (which is a special case of Augmented Lagrangian method and Douglas-Rachford algorithm) to minimize TV-$\langle , \rangle$.

The TV-$\langle,\rangle$ Segmentation Model is Non-Differentiable

- The TV-$\langle,\rangle$ minimization problem:

$$\min_{0\leq u\leq 1} F_{TV\langle,\rangle}(u) = \int_\Omega w_b|\nabla u|dx + \underbrace{\lambda < w_r^{in}, u > +\lambda < w_r^{out}, (1-u) >}_{\lambda < w_r, u > + Cte,\ w_r := w_r^{in} - w_r^{out}}$$

is difficult to solve because TV is non-differentiable. Wang-Yin-Zhang 07 and Goldstein-Osher 08 have recently proposed a splitting approach to overcome this issue in the context of denoising and Compressed Sensing.

- The original unconstrained minimization problem is replaced by the constrained minimization problem:

$$\min_{u\in[0,1],\ d\in\mathbb{R}^2} \underbrace{\int_\Omega w_b|d|dx}_{|d|_{w_b}} + \lambda < w_r, u > \quad \text{such that } d = \nabla u,$$

and relaxed to this unconstrained minimization problem:

$$\min_{u,d} \underbrace{|d|_{w_b} + \lambda < w_r, u >}_{F(u,d)} + \frac{\rho}{2}\underbrace{\int_\Omega |d - \nabla u|^2}_{||d-\nabla u||^2}.$$

- How to enforce the constraint? Wang-Yin-Zhang used the continuation principle, i.e. $\rho_k \to \infty$.

Split-Bregman Iteration: An Alternative of the Continuation Principle (Goldstein-Osher 08, Goldstein-Bresson-Osher 09)

- The Split-Bregman iteration method allows to exactly enforce the constraint $d = \nabla u$ during the minimization process.
- The core of this method is the "Bregman Distance" of the (convex) functional $F(u, d)$:

$$D_F(u, d, u^k, d^k, p_u^k, p_d^k) = F(u, d) - < p_u^k, u - u^k > - < p_d^k, d - d^k >,$$

where p_u, p_d are the subgradient of F w.r.t. u, d.

- It can be shown that the following iterative process:

$$\begin{cases} (u^{k+1}, d^{k+1}) &= \arg\min_{u\in[0,1],d} \; D_F(u, d, u^k, d^k, p_u^k, p_d^k) + \frac{\rho}{2}||d - \nabla u||^2 \\ p_u^{k+1} &= p_u^k - \rho div(d^{k+1} - \nabla u^{k+1}) \\ p_d^{k+1} &= p_d^k - \rho(d^{k+1} - \nabla u^{k+1}) \end{cases}$$

holds the following properties:
- $||d^k - \nabla u^k|| \to 0$ as $k \to \infty$
- The limit $u^\star := \lim_{k\to\infty} u^k$ satisfies the original constrained problem:

$$u^\star := \arg \min_{0\le u\le 1} \; F_{TV\langle,\rangle}(u)$$

Split-Bregman = Augmented Lagrangian (Glowinski-Le Tallec 89, Setzer 09)

- The Split-Bregman iterative process can be re-written as an Augmented Lagrangian algorithm:

$$\begin{cases} (u^{k+1}, d^{k+1}) & = \arg\min_{0\le u\le 1, d} \; F(u,d) + < b^k, \nabla u - d > + \frac{\rho}{2} ||\nabla u - d||^2 \\ b^{k+1} & = b^k + \nabla u^{k+1} - d^{k+1} \end{cases}$$

- It can also be shown that the Alternating Split-Bregman (ASB) iterative process:

$$\begin{cases} u^{k+1} & = \arg\min_{0\le u\le 1} \; F(u,d^k) + < b^k, \nabla u - d^k > + \frac{\rho}{2} ||\nabla u - d^k||^2 \\ d^{k+1} & = \arg\min_{d} \; F(u^{k+1},d) + < b^k, \nabla u^{k+1} - d > + \frac{\rho}{2} ||\nabla u^{k+1} - d||^2 \\ b^{k+1} & = b^k + \nabla u^{k+1} - d^{k+1} \end{cases}$$

  is equivalent to a Douglas-Rachford Splitting (DRS) Algorithm on the dual.

- Proof: Minimization problems like:

$$\min_u \; J(u) + H(u) \;\Leftrightarrow\; -\min_p \; J^\star(p) + H^\star(p)$$

  satisfies the Karush-Kuhn-Tucker conditions: $0 \in A(\hat{p}) + B(\hat{p})$. The solution $\hat{p}$ can be computed with the DRS algorithm:

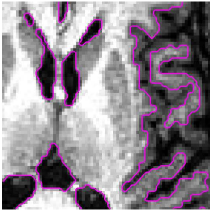
$$\begin{aligned} t^{k+1} &= Prox_{\eta A}(I - \eta B)p^k + \eta B p^k \\ p^{k+1} &= Prox_{\eta B}(t^{k+1}) \end{aligned}$$

  Finally, ASB is equivalent to DRS for $A = \partial(<,>^\star \; o \; (-div))$, $B = \partial(||)^\star$ and $\eta = \rho$.

Comparison with Graph-Cuts

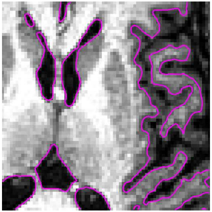
- The TV-$\langle , \rangle$ model can be optimized with Graph-Cuts (parametric max flow/min cut, Ford-Fulkerson 62). Graph-Cuts is a combinatorial technique that exactly optimizes binary energies (which fits well for our problem). This optimization method is fast (Boykov-Kolmogorov 04). Graph-Cuts have also some limitations. This technique:
  - uses anisotropic schemes to approximate the length (it cannot produce curved lines)
  - is pixel accurate
  - has memory limitations for 3D data
  - is not easily parallelized
  - and its speed depend on spatial connectivity.
- The proposed continuous optimization algorithm for TV-$\langle , \rangle$:
  - is faster than Graph-Cuts (much faster than Level Set Method).
  - uses isotropic schemes
  - is sub-pixel accurate
  - has low memory usage for 3D data
  - is easy to parallelize
  - requires a stopping criterion.

Comparison with Graph-Cuts



(a) Graph-Cuts. Final Contour.

(b) Graph-Cuts. Final u.



(c) Our Algorithm. Final Contour.

(d) Our Algorithm. Final Contour.

Figure: Image size is 256×256 (the presented results are zoomed in). The computational time for Graph-Cuts is 0.2 seconds and 0.07 seconds for our algorithm. Our algorithm is more accurate because it uses isotropic schemes and is sub-pixel accurate (smoothness of contours).

Convex Segmentation/Classification Models in High-Dimensional Spaces (Bresson-Chan 08[9])

- We extend the continuous model TV-$\langle , \rangle$ to work in high-dimensional spaces.
  Main advantage: segmentation/classification process for any data representation (image intensity, image patch, image, etc).
  Representation: the high-dimensional data (point in $\mathbb{R}^n$) are represented by the vertices of a graph.
  Applications: Interactive Segmentation, Machine Learning, etc.

- Related works: growing interest to develop PDEs/Wavelets on graph. For examples Coifman-Maggioni[4], Szlam-Maggioni-Coifman[5], Zhou-Scholkopf[6], Gilboa-Osher[7], Jones-Maggioni-Schul[8], etc.

[4] Coifman-Maggioni, Diffusion wavelets, 2006

[5] Szlam-Maggioni-Coifman, Regularization on Graphs with Function Adapted Diffusion Processes, 2008

[6] Zhou-Scholkopf, A Regularization Framework for Learning from Graph Data, 2004

[7] Gilboa-Osher, Nonlocal linear image regularization and supervised segmentation, 2007

[8] Jones-Maggioni-Schul, Universal Local Parametrizations via Heat Kernels and Eigenfunctions of the Laplacian, 2007

[9] Bresson-Chan, Non-local Unsupervised Variational Image Segmentation Models, 2008

TV-$\langle , \rangle$ on Graph

- The extended TV-$\langle , \rangle$ on graph is:

$$\min_{0 \le u \le 1} \sum_{\Omega} w_b |\nabla_{\mathrm{G}} u| + \lambda < w_r, u >$$

where the domain Ω can be the standard image domain, but more generally the set of all vertices of the conidered graph (vertices can belong to high-dimensional space).

- TV on graph[10,11,12]:

$$TV_{\mathrm{G}}(u) \quad = \quad \sum_{\Omega} |\nabla_{\mathrm{G}} u|,$$

where the gradient operator on graph ∇_{G} is defined for a pair of points (x, y) in the domain Ω as:

$$\nabla_{\mathrm{G}} u(x, y) \quad := \quad (u(y) - u(x))\sqrt{w(x, y)} : \Omega \times \Omega \to \mathbb{R},$$

where $w(x, y)$ is the edge function of the graph G between vertices x and y.

[10] Chan-Osher-Shen, The digital TV filter and nonlinear denoising, 2001

[11] Zhou-Scholkopf, A Regularization Framework for Learning from Graph Data, 2004

[12] Gilboa-Osher, Nonlocal linear image regularization and supervised segmentation, 2007

Global Minimization Theorem

- Can we extend the global minimization theorem to graph? Not directly since the coarea formula does not exist on graph. However, if we assume that the set of points/vertices of the graph belong to a submanifold $\mathcal{M}$ of $\mathbb{R}^n$, where the dimension of $\mathcal{M}$ is much smaller that the ambient space $d < n$ [13], then we can show[14,15]:

$$\sum_{\Omega} \frac{1}{\epsilon^{\frac{d+2}{4}}} |\nabla_{\mathrm{G}_\epsilon} u| \underset{\epsilon \longrightarrow 0}{\longrightarrow} \int_{\mathcal{M}} |\nabla_{\mathcal{M}} u|,$$

$$\text{with the graph } \mathrm{G}_\epsilon \text{ defined with } w_\epsilon(x,y) = \exp\Big(- \frac{||x-y||^2}{\epsilon} \Big)$$

- Using the co-area formula on manifold (which exists), the global minimization theorem can be extended. In other words, to extract a global minimizer, we need to compute the minimizer of TV_{G}-$\langle , \rangle$ and threshold it.

[13] Belkin, Problems of Learning on Manifolds, 2003

[14] Bresson-Chan, Non-local Unsupervised Variational Image Segmentation Models, 2008

[15] Coifman-Lafon, Diffusion maps, 2006

The High-Dimensional Space of Image Patches

- Image patches used in Texture Synthesis (Efros-Leung 99) and Image Denoising (Buades-Coll-Morel 05 (Non-local Means), Gilboa-Osher 08).

- Each vertice of the graph corresponds to an image patch, which lives in a high-dimensional space (typically, the space has $n = 5 \times 5 = 25$ dimensions).

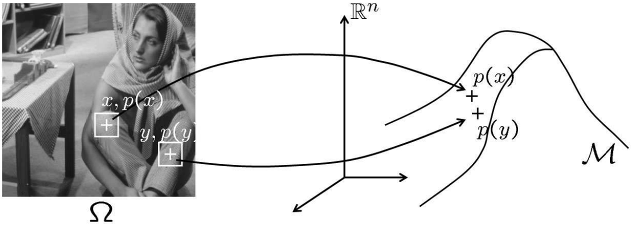


Figure: The set of image patches are samples from the manifold $\mathcal{M}$

Unsupervised Segmentation: Chan-Vese Model in the High-Dimensional Space of Image Patches (Bresson-Chan 08)

- Chan-Vese on the space/graph of image patches:

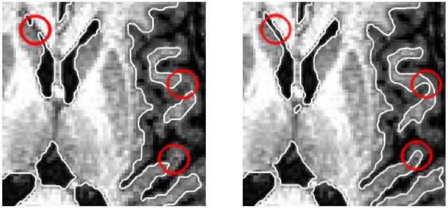
$$\min_{0\leq u\leq 1} \sum_{\Omega} |\nabla_{\mathrm{G}} u| + \lambda < w_r, u >$$

$$\text{with } w_r = (c^{in} - u_0)^2 - (c^{out} - u_0)^2$$

$$\text{and } w(x,y) = \exp\left(- \frac{||p(x) - p(y)||^2}{h} \right),$$

where $p(.)$ is the patch at x.

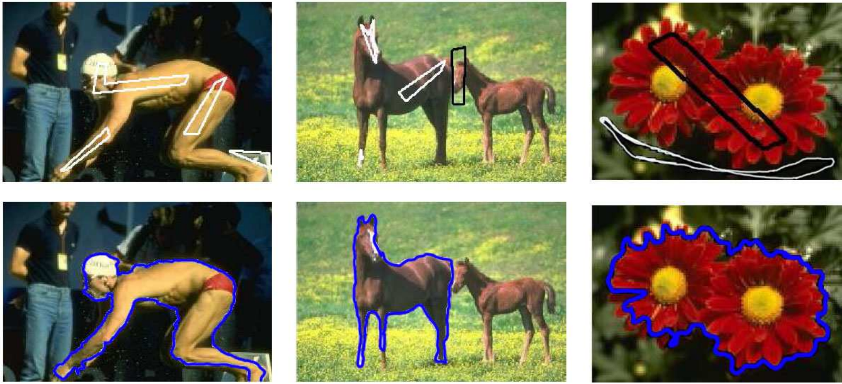
- The original CV model is limited to segment small structures because standard TV decreases the length of iso level sets. In the case of TV defined on the graph of image patches, the new CV model is able to denoise and preserve small structures which are repetitive (which is often the case in natural images).



Standard CV *NL-CV*

Semi-Supervised Segmentation in the High-Dimensional Space of Image Patches (Houhou-Bresson-Szlam-Chan-Thiran 09[18])

- The objective is to label a large amount of data from a small amount of labeled data. Application: Interactive Segmentation/Extraction of Objects (e.g. Li-Sun-Tang-Shum 04[16], Protiere-Sapiro 07[17]).



[16] Li-Sun-Tang-Shum, Lazy Snapping, 2004

[17] Protiere-Sapiro, Interactive image segmentation via adaptive weighted distances, 2007

[18] Houhou-Bresson-Szlam-Chan-Thiran, Semi-Supervised Segmentation based on Non-local Continuous Min-Cut, 2009

Semi-supervised classification/ Machine Learning in the High-Dimensional Space of Images (On-Going Work)

- Chan-Vese (CV) in the space of images for semi-supervised classification. Connection to K-means: The data term of CV corresponds to a 2-means algorithm. The main advantage of CV over K-means is the TV regularization term (CV is a better classification method).
- The extended CV model to K-means can be applied to Machine Learning. Application: classification of digit numbers using the MNIST database of handwritten digits (training set of 60,000 examples).

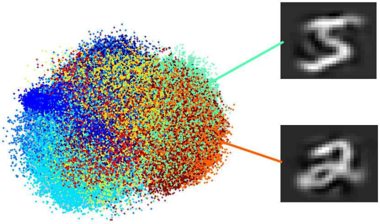


Figure: A cloud of points represents the numbers 0,...,9 projected on a 3D space with PCA.

- The convexification approach introduced with TV-L^1 and TV-$\langle , \rangle$ models have been used to convexify Image Processing and Computer Vision problems such as Multiview 3D Reconstruction[19], Stereo Evaluation[20], Optical Flow & Object Tracking[21], Multi-phase Segmentation[22], etc.
- Standard non-convex image processing problems have been reformulated as convex optimization problems, which are guaranteed to provide the global minimizing solution (independently of the initial condition). Besides, convex optimization problems can benefit from fast continuous optimization algorithms borrowed from Operator Splitting techniques.

[19] Kolev-Klodt-Brox-Esedoglu-Cremers, Continuous global optimization in multiview 3d reconstruction, 2007

[20] Pock-Schoenemann-Cremers-Bischof, A convex formulation of continuous multi-label problems, 2008

[21] Zach-Pock-Bischof, A Duality Based Approach for Realtime TV-L1 Optical Flow, 2007

[22] Chambolle-Pock-Cremers-Bischof, A Convex Relaxation Approach for Computing Minimal Partitions, 2009