

Weakly Convex Coupling Continuous Cuts and Shape Priors

Bernhard Schmitzer, Christoph Schnörr

University of Heidelberg

Abstract. We introduce a novel approach to variational image segmentation with shape priors. Key properties are convexity of the joint energy functional and weak coupling of convex models from different domains by mapping corresponding solutions to a common space. Specifically, we combine total variation based continuous cuts for image segmentation and convex relaxations of Markov Random Field based shape priors learned from shape databases. A convergent algorithm amenable to large-scale convex programming is presented. Numerical experiments demonstrate promising synergistic performance of convex continuous cuts and convex variational shape priors under image distortions related to noise, occlusions and clutter.

1 Introduction

1.1 Overview, Related Work

Various continuous variational approaches to image labeling and segmentation have been presented in the recent literature [4, 14, 13] based on tight convex relaxations of the underlying combinatorial problem. While the relaxation of the binary two-class case can be shown to compute the global combinatorial optimum after thresholding [4], the non-binary case of multiple labels [14, 13] also returns high-quality combinatorial solutions in practice, as numerical experiments based on primal-dual iterations show. Unlike algorithms for fully discrete graph-cut approaches [2] that may get stuck in a poor local minimum in the nonbinary case, their continuous convex counterparts do not suffer from such problems. Moreover, a broad range of robust first-order minimization algorithms from sparse convex programming are available for efficiently solving such large-scale problems [9, 3].

Variational approaches comprise a data term and a regularization term. In connection with image labeling, the data term is a linear form that does not impose any restriction on the type of image features to be processed. Concerning the regularization term, a large class of alternatives to the standard Potts prior has been suggested in [12], all of which do not compromise convexity of the variational approach.

While features and regularization terms can be handled quite flexibly within a convex variational framework, this is not the case for another major clue to

reliable segmentations: shape. Substantial research work has been done on variational representations of shape statistics as additional penalty terms, ranging from sampled contours and kernel techniques from machine learning to sophisticated manifolds of invariant shape representations [6, 18, 16]. Analogous ideas have been applied to embedding functions in connection with level set based approaches to shape representation and segmentation [5, 7, 11]. In this connection, we point out two properties of prior work that motivated the work presented in this paper:

- shape representations may not conform to the representation of image segmentations (contour spaces vs. regions or set of pixels after discretization);
- shape penalty functionals are nonconvex (except for the less attractive case of Gaussian shape statistics based on sampled contours);
- *nonconvex* level set based functionals have been used for image segmentation;
- adding a shape penalty functional compromises convexity of the overall approach.

1.2 Contribution, Organization

The approach introduced in this paper comprises

- (i) *separate convex modeling* of variational approaches to segmentation and shape priors, respectively, and
- (ii) *weak convex coupling* of these models in terms of a convex, but possibly indefinite, quadratic form

$$\|Ax - B\mu\|^2, \tag{1}$$

that measures similarity of segmentations x and shapes μ , respectively, *mapped to a common space* by linear mappings A and B .

As a consequence, our corresponding approach avoids all deficiencies of previous approaches, from the viewpoint of optimization.

Concerning (i) and segmentation, we use convex models of continuous cuts as introduced in [4, 14, 13]. Concerning (i) and shape priors, we apply strengthened local polytope convex relaxations of binary Markov Random Fields (MRFs) [17, 8, 15] whose structure and parameters are learned offline from a shape database using large-scale convex optimization in a preprocessing step [10]. Concerning (ii), we adopt the framework presented in [1] for coupling two proper convex and lower semicontinuous functionals defined on *different* spaces.

The main objective of this paper is to *introduce the general framework*. Therefore, we make no attempt to present a complete list of fully-fledged model components, but rather confine ourselves to demonstrating the key aspect – coupling of convex models – using preliminary versions of individual models. Specifically, concerning segmentation, we merely employ *binary* continuous cuts [4] and deliberately *do not elaborate the issue of feature extraction*, having in mind that, as discussed in the previous section, any image features computed in a preprocessing step could be used. We apply two different MRF models as shape priors in order

to indicate the potential of this research direction: a naive MRF directly defined on the pixel grid, and a hierarchically defined MRF that compares favourably from the viewpoint of learning and automatically extracts a part-based probabilistic representation of object shapes taken from different viewpoints. Under strong noise levels simulating feature imperfections, we numerically demonstrate promising synergistic performance of convex continuous cuts and convex variational shape priors.

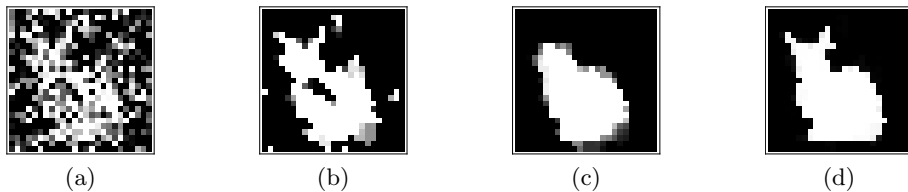


Fig. 1. Binary image segmentation: (a) noisy input, (b) only continuous cuts with low regularization, (c) only continuous cuts with strong regularization, (d) convex coupling of continuous cuts with low regularization and shape prior.

Our approach is general and can be extended in various directions. We indicate this below as we go along.

2 Variational Models

2.1 Segmentation by Continuous Cuts

We adopt the approach [4] for globally optimal foreground-background separation by convex optimization.

Let Ω denote the uniform pixel grid of size $|\Omega| = N$ corresponding to the domain $[0, 1]^2 \subset \mathbb{R}^2$, and $G \in \mathbb{R}^{2N \times N}$ a discrete gradient matrix corresponding to functions $x: \Omega \rightarrow \mathcal{C}$, $\mathcal{C} = [0, 1]^N$. Given any similarity values $s: \Omega \rightarrow \mathbb{R}^N$ extracted from image data beforehand, that locally indicate fore- or background in terms of its components $s_i, i = 1, \dots, N$, we look for a minimizer $x_{\min} \in \mathcal{C}$ of the functional

$$E_{\text{TV}}(x) = \alpha \text{TV}(x) + \langle x, s \rangle, \quad (2)$$

where $\alpha > 0$ is a regularization parameter, and the (discretized) total variation measure as regularizer is given by

$$\begin{aligned} \text{TV}(x) &= \sum_{i=1}^N \|(Gx)_i\| = \sum_{i=1}^N ((Gx)_{i,1}^2 + (Gx)_{i,2}^2)^{1/2} \\ &= \sigma_{\mathcal{D}}(Gx) = \sup_z \{ \langle Gx, z \rangle - \delta_{\mathcal{D}}(z) : z \in \mathbb{R}^{2N} \}, \\ \mathcal{D} &= \{ z \in \mathbb{R}^{2N} : (z_{i,1})^2 + (z_{i,2})^2 \leq 1, 1 \leq i \leq N \}. \end{aligned}$$

In [4] it is pointed out how these minimizers are related to finding an optimal solution to the two-level Mumford-Shah energy functional by thresholding.

2.2 MRF Based Shape Priors

General Variational Formulation Shape priors consist of feature vectors y of binary variables $y_i, 1 \leq i \leq M$ and are defined statistically in terms of a joint distribution

$$\begin{aligned} P(y) &= \frac{1}{Z(\theta)} \exp(\langle \theta, \phi(y) \rangle) \\ Z(\theta) &= \sum_{y \in \{0,1\}^M} \exp(\langle \theta, \phi(y) \rangle), \end{aligned} \quad (3)$$

where plausible choices corresponding to familiar shapes are assigned a higher probability. The form of $\phi(y)$ is given by an associated undirected graph G with vertices $V = \{1, \dots, M\}$ and edges $E \subseteq \{(i, j) : 1 \leq i < j \leq M\}$. Then $\phi(y) = (y_1, y_2, \dots, y_M, \dots, y_i y_j, \dots), (ij) \in E$.

This corresponds to a minimally represented binary graphical model [17] with strictly convex and essentially smooth log-partition function $Z(\theta)$, for any choice of the model parameters $\theta \in \mathbb{R}^{(|V|+|E|)}$. These parameters are learned from shape databases, as described in Section 3.

The most probable configuration y_{\max} is given as solution to the problem

$$\operatorname{argmax}_y \{ \langle \theta, \phi(y) \rangle : y \in \{0, 1\}^M \}.$$

This discrete combinatorial problem can be reformulated as linear problem on the marginal polytope $\mathcal{M}(G)$ of the graph G [17]:

$$\max_y \{ \langle \theta, \phi(y) \rangle : y \in \{0, 1\}^M \} = \sup_{\mu} \{ \langle \theta, \mu \rangle : \mu \in \mathcal{M}(G) \}.$$

As the number of constraints that define $\mathcal{M}(G)$ grows exponentially with the size of the graph, this problem is in general unfeasible. Thus one is forced to relax the optimization set to the local polytope $\mathcal{L}(G)$ and to check carefully the global optimum of this convex relaxation. Methods have been proposed to tighten the standard local polytope relaxation by identifying and reintroducing violated constraints of the original optimization set, see for example [15].

Whatever set of constraints one chooses, they can be written as N_{con} affine inequalities leading to the variational problem formulation

$$\inf_{\mu} \{ E_{\text{MRF}}(\mu) : K\mu \leq k, K \in \mathbb{R}^{N_{\text{con}} \times (|V|+|E|)}, k \in \mathbb{R}^{N_{\text{con}}} \} \quad (4)$$

with

$$E_{\text{MRF}}(\mu) = -\langle \theta, \mu \rangle.$$

3 Variational Shape Priors

We describe two specific instances of the general framework (3).

3.1 Ising Shape Prior

As a baseline, we investigate a direct application of the two dimensional Ising Model: Every pixel is treated as binary feature, hence $M = N$ in (3).

The set E of edges defining the model is determined by the model parameters θ . Vanishing parameter values indicate missing edges. Parameter values are determined using an approach proposed by Hoefling and Tibshirani [10]. The corresponding algorithm maximizes the pseudo-likelihood of a set of training samples \mathbf{T} as a function of the parameter vector θ , constrained by a ℓ_1 -norm penalty $\rho\|\theta\|_1$. This penalty enforces sparse connectivity of the graph which is desirable for various reasons: Learning dense graphs tends to overfit the training data; dense graphs lead to a larger linear problem, and the local polytope relaxation tends to become weak. In our simulations about 3% of all possible edges were actually existent (see Fig. 2a).

In the training data some pixels were either always black or always white. As they do not contain any correlation information about other pixels they have been removed from the learning procedure and their unary θ -weight has been set to $\pm\infty$ by hand. The corresponding vertices have no connections to other vertices.

3.2 Hierarchical Part Based Shape Prior

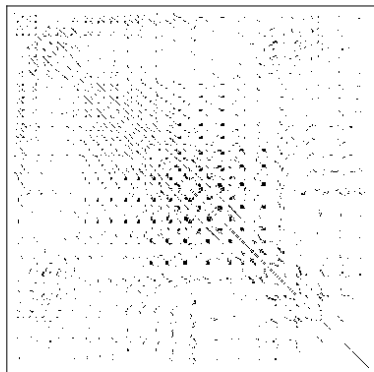
Besides the “flat” Ising Model discussed above, we also study a simple hierarchical model that can demonstrate the benefits of articulated object descriptions and support nonlocal interactions without compromising sparsity. Shapes will be decomposed into a torso and successive limbs. Limbs may depend on another and on the torso: E.g. a lower arm is only expected if the upper arm is present which, in turn, can only be present if the torso is there. Our model learns such statistical relationships from data and penalizes implausible constellations as follows.

First the set of training samples is divided into groups of characteristic views, i.e. viewpoints of the object that yield similar shapes. Then to each group the following procedure is applied: pixels in the image are combined into families depending on the sample subsets in which they are active ($=1$) or not ($=0$). For example, usually there will be a “torso” family of pixels that are active in every sample of a specific view, and a “background” family of pixels that are always black in that group.

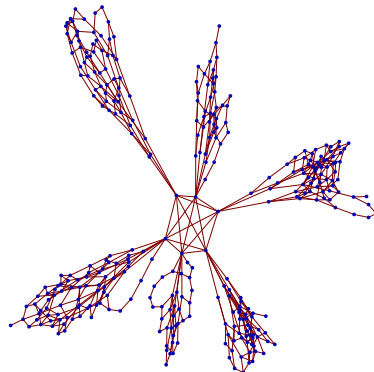
Correspondingly, a hierarchical dependency graph of families is constructed for every group: family i of pixels is considered dependent on family j if family j is always active if i is active.

This relation is transitive, and the dependency graph only represents the transitive reduction of this relation to keep the representation minimal. As to be expected, in all these graphs the torso-families will be the root vertices.

Using the groups of characteristic views and their dependency graphs a joint prior-graph with parameter vector θ is constructed in the following way: The core of the graph is constituted by the torso-families of all characteristic view



(a) Plot of the adjacency matrix of the graph G of the Ising prior learned from a set of training samples. The nodes corresponding to pixels of constant value in all training samples are not shown. Density: approx. 3%.



(b) Illustration of the graph structure of the hierarchical part based prior: The nodes of the fully connected subgraph in the center represent the different possible torsos.

Fig. 2. Illustrating the two MRF based priors.

groups. They are fully connected amongst each other. Then to each torso node the dependency graph of the associated characteristic view is appended (compare Fig. 2b). The parameter vector is initialized with $\theta_i = 0, 1 \leq i \leq M, \theta_{ij} = 0, (ij) \in E$. Then all pairs of torso nodes are considered: Plausibility dictates that at most one torso can be present at a time. Thus the simultaneous presence of two torsos has to be penalized. Therefore all θ_{ij} where i and j are two different torso nodes, are set to $-p$ where p is the positive penalty parameter. Then all dependency graphs of the torso nodes are processed: Whenever a node i is dependent on a node j then θ_i will be decreased by p and θ_{ij} will be increased by p . Thus the unlikely case where y_i is 1 while y_j is 0 will be penalized by $-p$ relative to the other possible combinations.

We are aware that this preliminary version of our learning procedure is somewhat heuristic, but fully data driven and controlled just by a single user parameter p .

The grouping into pixel families and the introduction of a partial ordering relation on the vertices capture some of the most striking features of the sample data. A learning procedure that maximizes the likelihood would deduce similar relations. Thus this step is only an approximation where the explicit forming of families each corresponding to one vertex helps reducing the problem size and simplifies the graph structure.

The clustering into characteristic views subsequently allows multiple vertices to be associated with the same pixel. This simplifies differentiating “for what reason” (i.e. which characteristic view) a pixel is white. Hence, no further edges

between different views are required which drastically reduces the number of cycles in the graph and thus supports the applied polytope relaxations.

4 Coupling Convex Models

In this section an approach by Attouch et al.[1] is briefly introduced. We will subsequently use its potential for combining TV-based segmentation with shape prior knowledge in a variational framework.

4.1 Variational Approach

The approach [1] is based on the following functional jointly defined on two real Hilbert spaces \mathcal{U} and \mathcal{V} :

$$E(x, \mu) = f(x) + g(\mu) + \frac{\lambda}{2}Q(x, \mu), \quad x \in \mathcal{U}, \mu \in \mathcal{V}, \quad (5)$$

where

$$\begin{aligned} &\mathcal{U}, \mathcal{V} : \text{real Hilbert spaces,} \\ &\begin{aligned} f : \mathcal{U} &\rightarrow \mathbb{R} \cup \{+\infty\} \\ g : \mathcal{V} &\rightarrow \mathbb{R} \cup \{+\infty\} \end{aligned} : \text{closed convex proper functions,} \\ &Q : (x, \mu) \in \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}^+ : \text{nonnegative, quadratic form.} \end{aligned}$$

A minimizing pair of vectors (x, μ) is given by the limit of the series generated by the following alternating proximal update steps:

$$\begin{aligned} x_{n+1} &= \operatorname{argmin}_{\xi} \left\{ f(\xi) + \frac{\lambda}{2}Q(\xi, \mu_n) + \frac{\beta_f}{2} \|\xi - x_n\|^2 : \xi \in \mathcal{U} \right\}, \\ \mu_{n+1} &= \operatorname{argmin}_{\eta} \left\{ g(\eta) + \frac{\lambda}{2}Q(x_{n+1}, \eta) + \frac{\beta_g}{2} \|\eta - \mu_n\|^2 : \eta \in \mathcal{V} \right\}. \end{aligned} \quad (6)$$

Here $\lambda \geq 0$ is a coupling constant that regulates the strength of the interaction. $\beta_f > 0$ and $\beta_g > 0$ are damping constants that affect the convergence rate of the algorithm but not convergence itself. Indeed, as we work in finite-dimensional Euclidean spaces, [1, Thm. 2.1] immediately yields

Theorem 1. *The sequence (x_n, μ_n) generated by algorithm (6) converges to a minimum (x_∞, μ_∞) of the functional E in (5). Moreover, $f(x_n) \rightarrow f(x_\infty)$, $g(\mu_n) \rightarrow g(\mu_\infty)$, $Q(x_n, \mu_n) \rightarrow Q(x_\infty, \mu_\infty)$, $\|x_{n+1} - x_n\| \rightarrow 0$ and $\|\mu_{n+1} - \mu_n\| \rightarrow 0$ as $n \rightarrow \infty$.*

We point out that this result is non-trivial because functionals f and g are required to be convex and closed, but may be *non-smooth*, as in our applications. As a consequence, heuristic alternating minimization without complementing proper proximal point mappings, as is sometimes done in the field of computer vision, might not yield a minimizing sequence.

From the viewpoint of modeling, the result is appealing as well, because the quadratic form $Q(\cdot, \cdot)$ enables to couple and to compare points from “two different worlds” in various mathematically sound ways in a common third space.

4.2 Shape Constrained Cuts

We present a preliminary application of the idea of coupling two convex functionals to the image segmentation problem. We set $\mathcal{U} = \mathbb{R}^N$, $\mathcal{V} = \mathbb{R}^{|V|+|E|}$, and investigate the following joint energy:

$$E(x, \mu) = \lambda_1 E_{\text{TV}}(x) + \lambda_2 E_{\text{MRF}}(\mu) + \frac{\lambda}{2} \|x - P\mu\|^2 + \delta_{\mathcal{C}}(x) + \sup_{\omega} \left\{ \langle \omega, K\mu \rangle - \left(\delta_{\mathbb{R}_+^{N_{\text{con}}}}(\omega) + \langle \omega, k \rangle \right) \right\}, \quad (7)$$

where P is a linear mapping to be specified. Note that the terms in the second line constrain solutions $x \in \mathcal{U}$, $\mu \in \mathcal{V}$, to the optimization problem to the convex sets $x \in \mathcal{C}$ and $\{\mu: K\mu \leq k\}$, respectively.

Comparison to Eq. (5) yields

$$f(x) = \lambda_1 E_{\text{TV}}(x) + \delta_{\mathcal{C}}(x) \quad (8)$$

$$g(\mu) = \lambda_2 E_{\text{MRF}}(\mu) + \sup_{\omega} \left\{ \langle \omega, K\mu \rangle - \left(\delta_{\mathbb{R}_+^{N_{\text{con}}}}(\omega) + \langle \omega, k \rangle \right) \right\} \quad (9)$$

$$Q(x, \mu) = \|x - P\mu\|^2. \quad (10)$$

It is easy to verify that these choices satisfy the criteria that are required for the sequence generated by (6) to converge.

In our example we chose the coupling space to which x and μ are mapped to be identical with the space x lives in. Thus the matrix A in Eq. (1) is the identity on \mathbb{R}^N . The matrix P in the coupling term maps the initially abstract feature vector μ to the pixel space. The coupling then encourages the segmentation proposal x to be close to the image represented by the features μ under this map. In our setup μ contains not only rows for unary marginals of variables y_i but also pairwise marginals of $y_i y_j$ corresponding to edges $(ij) \subset E$. These pairwise marginals are not considered by the map.

In the Ising-Model prior P takes the marginals of the pixel features y_i and maps them to the appropriate pixel positions i in the image space. It is given by

$$\left(\mathbb{I}_{|V| \times |V|} \ 0_{|V| \times |E|} \right)$$

where $\mathbb{I}_{n \times n}$ denotes the n -dimensional identity matrix and $0_{m \times n}$ a $m \times n$ matrix with all entries being 0.

For the hierarchical part based prior, in each column of P that corresponds to a unary marginal of a pixel family the entries of the corresponding pixels are 1. The subspace of pairwise marginals is again ignored. Despite the unusual generation of the parameter vector θ this model is still part of the family given by Eq. (3).

5 Experiments and Discussion

5.1 Setup

Training Data Shape priors are learned from 2D views of 3D object models. The training samples are views of the model from equidistant angles rotated

around its vertical axis. We chose a fixed window size that in applications may cover different region sizes of a given image, corresponding to different levels of a multiscale representation of the image.

In the first scenario the training data consist of 25×25 pixel views of a bunny (see Fig. 3). To demonstrate the generality of our approach a second scenario with 50×50 pixel views of a horse was also studied.



Fig. 3. Some of the training samples of scenario 1: 25×25 pixel b/w views of a 3D-model of a bunny.

In scenario 1 the Ising prior was trained with 50 equidistant views, the hierarchical part based model with 100 equidistant views.

Input Data As input data *new views with random angles* between the training views were taken and distorted to simulate *noise, occlusion and clutter*.

Gaussian noise of mean $\mu_{\text{noise}} = 0$ and some standard deviation σ_{noise} followed by projecting each pixel value onto the interval $[0, 1]$ simulates feature imperfections. Occlusion by another object in the foreground was simulated by drawing black circles on the input image. Clutter could be generated by other objects in the image with texture similar to the object in question and was simulated by drawing white circles on the input image. In the given setup the similarity vector s was obtained by $s = (1/2)_N - u$ where $(1/2)_N$ denotes the vector of length N with each entry being $1/2$ and $u: \Omega \rightarrow \mathcal{C}$ is the distorted input image.

Parameter Values Unless otherwise noted in the simulations the following parameter settings have been used: TV regularization parameter $\alpha = 0.5$ (eqn. (2)), part based MRF implausibility penalty $p = 100$ (cf. Section 3.2), coupling constants $\lambda_1 = \lambda_2 = 1, \lambda = 10$ (joint energy (7)), proximal damping parameters $\beta_f = \beta_g = 0.01$ (algorithm (6)). By default the noise level was set to $\sigma_{\text{noise}} = 0.75$.

The part based MRF penalty p is chosen to be on a higher energy scale than the continuous cuts segmentation to favour segmentations of familiar shapes. The coupling parameter λ was set to a high value to make the difference of x and $P\mu$ negligible. In plots that illustrate segmentation results only x is shown but $P\mu$ would not look much different.

5.2 Results

Figure 4 shows (a) several input images with distortions and each time the segmentations given by (b) the uncoupled continuous cuts approach, (c) the coupled Ising prior and (d) the coupled simple part based prior. Furthermore, distortions due to noise, occlusion and clutter was examined, as described in the previous section.

Clearly the coupling to a shape prior enhances the segmentation results: In row 1 the ears of the bunny are lost to regular segmentation while they are still restored by the coupled models. Similarly dealing with clutter and occlusion works much better with prior.

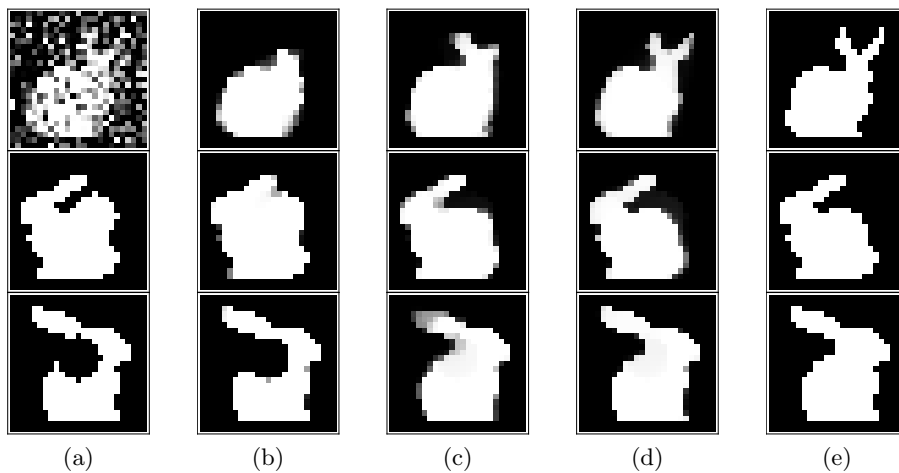


Fig. 4. Modelling different kinds of distortion: Row 1: noise $\sigma_{\text{noise}} = 0.50$, row 2: clutter, row 3: occlusion. Segmentation results: (a) input, (b) only continuous cuts, (c) Ising prior, (d) part based prior, (e) ground truth.

In Figure 5 it is shown how increased coupling strength forces the variables x and μ to correspond to the same image. In our simulations it turned out that most of the times a large coupling constant yields best results.

6 Conclusions and Further Work

We introduced a novel approach to variational image segmentation with shape priors. Key properties are convexity of the joint energy functional and weak coupling of convex models from different domains by mapping corresponding solutions to a common space. Specifically, we combined state-of-the-art variational approaches for TV-based image segmentation and for MRF-based learning of image classes from examples. A convergent algorithm amenable to large-scale con-

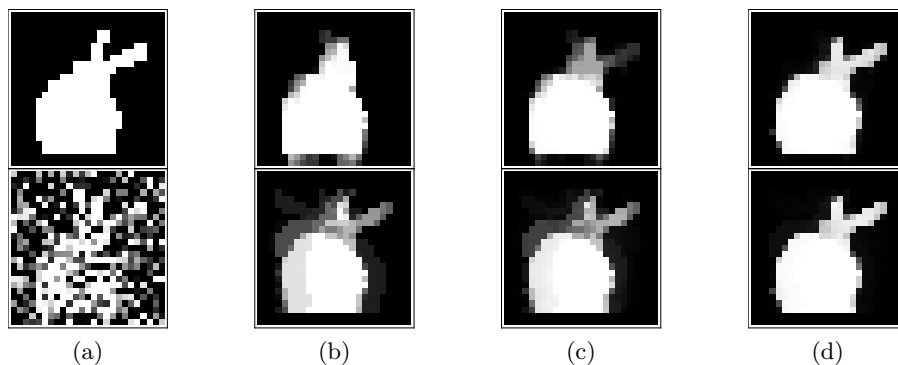


Fig. 5. Segmentation results of the coupled part based prior for various coupling strengths λ : (a) ground truth and noise-distorted input. The upper row shows x , the lower row $P\mu$: (b) $\lambda = 0.1$, (c) $\lambda = 1$, (d) $\lambda = 10$.



Fig. 6. Four pairs of noisy input, and part based prior supported segmentations of different views. All views were processed with the same prior.

vex programming was presented. Numerical experiments demonstrated promising synergistic performance of convex continuous cuts and convex variational shape priors under (simulated) noise, occlusion and clutter.

Our further work will further explore components of the coupling quadratic form Q in (5). For example, linear mappings that represent collections of linear shape features could be used to compare shapes after projecting them onto a low-dimensional feature space. Another line of research concerns elaborating the part-based variational shape prior so as to cope with several objects simultaneously and efficiently.

Acknowledgement This work was supported by the DFG, grant GRK 1653.

References

1. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Alternating proximal algorithms for weakly coupled convex minimization problems. applications to dynamical games and pde's. *Journal of Convex Analysis* 15(3), 485–506 (2008)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Patt. Anal. Mach. Intell.* 23(11), 1222–1239 (2001)
3. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision* (2010), published online



Fig. 7. Binary image segmentation of a horse: (a) noisy input, (b) only continuous cuts with high scale resolution $\alpha = 0.1$, (c) only continuous cuts with low scale resolution $\alpha = 0.3$, (d) coupled shape prior with high scale resolution $\alpha = 0.1$. Without prior a high resolution (b) can not filter out noise, with low resolution (c) details (legs) are lost. The prior can help to distinguish noise and details.

4. Chan, T., Esedoglu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* 66(5), 1632–1648 (2006)
5. Charpiat, G., Faugeras, O., Keriven, R.: Approximations of shape metrics and application to shape warping and empirical shape statistics. *Found. Comp. Math.* 5(1), 1–58 (2005)
6. Cremers, D., Kohlberger, T., Schnörr, C.: Shape statistics in kernel space for variational image segmentation. *Patt. Recognition* 36(9), 1929–1943 (2003)
7. Dambreville, S., Rathi, Y., Tannenbaum, A.: A framework for image segmentation using shape models and kernel space shape priors. *IEEE Trans. Patt. Anal. Mach. Intell.* 30(8), 1385 – 1399 (2008)
8. Deza, M., Laurent, M.: *Geometry of Cuts and Metrics*. Springer (1997)
9. Esser, E., Zhang, X., Chan, T.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imag. Sci.* 3(4), 1015–1046 (2010)
10. Hoefling, H., Tibshirani, R.: Estimation of sparse binary pairwise markov networks using pseudo-likelihoods. *Journal of Machine Learning Research* 10, 883–906 (2009)
11. Lecumberry, F., Pardo, A., Sapiro, G.: Simultaneous object classification and segmentation with high-order multiple shape models. *IEEE Trans. Image Proc.* 19(3), 625–635 (2010)
12. Lellmann, J., Becker, F., Schnörr, C.: Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In: *ICCV* (2009)
13. Lellmann, J., Kappes, J., Yuan, J., Becker, F., Schnörr, C.: Convex multi-class image labeling by simplex-constrained total variation. In: Tai, X.C., Mörken, K., Lysaker, M., Lie, K.A. (eds.) *Scale Space and Variational Methods in Computer Vision (SSVM 2009)*. LNCS, vol. 5567, pp. 150–162 (2009)
14. Pock, T., Chambolle, A., Cremers, D., Bischof, H.: A convex relaxation approach for computing minimal partitions. In: *CVPR*. pp. 810–817 (2009)
15. Sontag, D.A.: *Approximate Inference in Graphical Models using LP Relaxations*. Ph.D. thesis, Massachusetts Institute of Technology (2010)
16. Sundaramoorthi, G., Mennucci, A., Soatto, S., Yezzi, A.: A new geometric metric in the space of curves, and applications to tracking deforming objects by prediction and filtering (2010), <http://cvgmt.sns.it/papers/sunmensoa10/>, to appear
17. Wainwright, M.J., Jordan, M.I.: Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning* 1(1-2), 1–305 (2008)
18. Younes, L., Michor, P., Shah, J., Mumford, D.: A metric on shape space with explicit geodesics. *Rend. Lincei Mat. Appl.* 9, 25–57 (2008)