Joint Priors for Variational Shape and Appearance Modeling

Jeremy D. Jackson and Anthony J. Yezzi
Georgia Institute of Technology
School of Electrical and Computer Engineering
777 Atlantic Av. Atlanta, GA 30332
gtg120d,ayezzi@ece.gatech.edu
http://users.ece.gatech.edu/~gtg120d

Stefano Soatto
University of California at Los Angeles
Computer Science Department
4732 Boelter Hall Los Angeles, CA 90095
soatto@ucla.edu
http://www.vision.cs.ucla.edu

Abstract

We are interested in modeling the variability of different images of the same scene, or class of objects, obtained by changing the imaging conditions, for instance the viewpoint or the illumination. Understanding of such a variability is key to reconstruction of objects despite changes in their appearance (e.g. due to non-Lambertian reflection), or to recognizing classes of objects (e.g. cars), or individual objects seen from different vantage points. We propose a model that can account for changes in shape or viewpoint, appearance, and also occlusions of line of sight. We learn a prior model of each factor (shape, motion and appearance) from a collection of samples using principal component analysis, akin a generalization of "active appearance models" to dense domains affected by occlusions. The ultimate goal of this work is stereo reconstruction in 3D, but first we have developed the first stage in this approach by addressing the simpler case of 2D shape/radiance detection in single images. We illustrate our model on a collection of images of different cars and show how the learned prior can be used to improve segmentation and 3D stereo reconstruction.

1. Introduction

An image can be thought of as a function from a compact domain (the "image plane") to the positive reals (the "intensity" range). Changes in the imaging conditions, for instance due to changes in viewpoint and illumination, cause changes in both the domain and range of such a function. For instance, a change of view of a Lambertian scene in ambient light can be modeled, away from occlusions, by a diffeomorphic deformation of the image domain [12], whereas changes of illumination on a static scene can be modeled as structured changes in intensity (for instance described by a low-dimensional linear variety, known as "illumination cone"). Unfortunately, however, changes in the domain and range of the image play overlapping roles: One can always explain classes of images of the same scene or "object" with changes in its domain (intensity values) or, modulo contrast functions [1], by deformations of the image domain, as in a "deformable template" [6] (transitive actions of infinite-dimensional groups of diffeomorphisms). Therefore, inferring domain deformations and changes in intensity of a sequence of images obtained with different viewpoints and/or illumination is an ill-posed problem, and suitable regularizers have to be imposed in order to arrive at a meaningful model.

A common regularizer for changes in intensity is obtained by assuming that such changes cause the images to move on or close to a low-dimensional linear variety. The most common approach is leads to principal component analysis (PCA), and has been used extensively in modeling and recognition of scenes when there are no changes of viewpoint [11]. Changes of viewpoint at a finite number of landmarks can also be modeled in a similar fashion, using procrustes analysis, which can be implemented using singular value decompositions in a manner similar to PCA [5]. Combined changes in intensity and shape can be modeled in a conditionally linear fashion, by assuming that intensity in a normalized frame is linear, and that normalization is achieved by procrustes analysis, leading to so-called "active appearance models" [4]. These have proven effective in modeling classes of objects, such as faces, with modest changes of appearance and shape, an free of occlusions. Learning the principal components of shape and appearance from a collection of images of an object provide a powerful prior model that can be used to detect a new instance, or to recognize the belonging of an object to the target class.

The problem becomes significantly more complex in the presence of occlusions. In this case, domain deformations are not only not diffeomorphic, but they are not even regular functions, since occlusions cause portions of the scene to disappear, and other portions to appear. In [8] the problem of modeling changes in motion and appearance of occlu-

1-4244-1180-7/07/$25.00 ©2007 IEEE

sion layers using the variational framework of Deformotion was addressed by the authors, introduced in [13], exploiting generic regularizers. Using those methods in this paper, we introduce a learning-based regularization approach that extends the active appearance model to scenes with occlusions, all in a principled variational framework. Since we model the shape, motion and appearance of each layer, we can also fill in missing portions of layers (that are not visible in all images) realizing a multi-view version of "image inpainting" [2]. For the case of just one layer, our model simplifies to standard active appearance models, but represented in a continuous domain rather than at a finite number of landmarks.

The shape of a layer that represents an object in the scene is transformed by a diffeomorphism to model small-scale changes in the object and is transformed by a finite-dimensional group that describes coarse global motion. The intensity function associated with a layer is computed by minimizing a Mumford-Shah type energy that allows for occlusions and "in-paints" based on other images where there is no occlusion or if no information is available then the solution of Laplace's equation is used. In this work we take the two most complex pieces, the intensity function for a layer and the diffeomorphism representing variations in shape, and reduce these to a smaller, more reasonable space using principal components analysis.

The model we propose describes changes in motion, appearance, depth ordering, and shape of a number depth layers. In addition, we have to learn a number of bases for the appearance space, motion space, and shape space. This model is very powerful, but the notation tends to get heavy when all factors are taken into account, and the computational cost of inference can be significant. For this reason, we mostly restrict our attention to the important case of two layers (foreground and background) and refer the reader to a forthcoming technical report where the full model is described.

The ultimate application here is detecting 3D shape and radiance from multiple images. The type of joint priors developed here with would be used with stereoscopic segmentation [9]. This would involve using a suitable 3D shape/radiance prior whose perspective projections would bear resemblance to the types of 2D shape/radiance priors presented in this work. So this work is the first step in developing joint priors for stereo shape/radiance reconstruction.

2. Layered Deformotion

Here we will give a description of the variational method of layered deformotion. To describe a scene there are L layers indexed by $k = 1, 2, ..., L$ that may occlude each other in the order that layer 3 overlaps layer 2 and layer 2 overlaps layer 1, etc. Each layer has a shape and a radiance function. The shape of a layer is denoted by $\Omega^k \subset \mathbb{R}^2$ and the layer's radiance is $\rho^k : \Omega^k \to \mathbb{R}^+$. Each layer's global motion is represented by some affine or rigid group action g^k. The local deformations of the shape Ω^k of the layer k are given by the diffeomorphism $w^k : \Omega^k \to \mathbb{R}^2$. The shape Ω^k of a layer k is transformed to model an image I_t at time t by a diffeomorphism w_t^k and a finite group action g_t^k. The background layer is denoted by $\Omega^0 = \mathbb{R}^2$. A model image $\hat{I}_t$ is produced by the following:

$$\boxed{\begin{cases} \hat{I}_t(x_t^l) = \rho^l(x), & x \in \Omega^l \\ x_t^l = g_t^l \circ w_t^l(x), & l = \max\{k \mid x \in \Omega^k\}. \end{cases}} \tag{1}$$

So the energy to be minimized to produce $\hat{I}_t$ is

$$\begin{aligned} E = \sum_{t=1}^{N} \int_{\Omega^0} \left(I_t(x_t) - \rho^l(w_t^{l^{-1}} \circ g_t^{l^{-1}}(x_t)) \right)^2 dx_t \, + \\ + \beta \sum_{k=1}^{L} \int_{\Omega^k} \|\nabla \rho^k(x)\|^2 dx + \gamma \sum_{k,t=1}^{L,N} \int_{\Omega^k} r(w_t^k(x)) dx \end{aligned} \tag{2}$$

subject to $l = \max\{k \mid x \in \Omega^k\}$ where $\beta, \gamma \in \mathbb{R}^+$.

A typical regularizer $r(w)$ would be the typical one used in optical flow [7]. We can represent $w(x) : \mathbb{R}^2 \to \mathbb{R}^2$ as a vector field with $w(x) = [x + u(x), y + v(x)]$. Then

$$r(w(x)) = \int_{\Omega} \langle \nabla u(x), \nabla u(x) \rangle + \langle \nabla v(x), \nabla v(x) \rangle \, dx \tag{3}$$

The unknown quantities to be solved for are the radiances ρ^k, the shapes for each layer Ω^k, the global motions g_t^k from a layer k to an image t, and the deformations w_t^k from a layer k to an image t. These are solved for using gradient descent techniques. In the layered deformotion paper, the authors reduce the complexity of (2) to a moving, deforming foreground layer Ω^1,a fixed background layer Ω^0, and one image I in order to easily show the descent equations. We will keep this method for the sake of simplicity here as well.

Letting $g = g^1$, $w = w^1$, $\hat{x} = g(w(x))$ and $\hat{\Omega}^1 = g(w(\Omega^1))$, the energy is as follows:

$$\begin{aligned} E &= \int_{\hat{\Omega}^1} \left(I(\hat{x}) - \rho^1(x) \right)^2 d\hat{x} + \int_{\Omega^0 \setminus \hat{\Omega}^1} \left(I(\hat{x}) - \rho^0(\hat{x}) \right)^2 d\hat{x} \\ &+ \beta \sum_{k=0}^{1} \int_{\Omega^k} \langle \nabla \rho^k, \nabla \rho^k \rangle \, dx \\ &+ \gamma \int_{\Omega^1} \langle \nabla u(x), \nabla u(x) \rangle + \langle \nabla v(x), \nabla v(x) \rangle \, dx \end{aligned} \tag{4}$$

The gradient descent equation for a parameter λ (such as x and y translation, scale, or rotation) of g is:

$$\begin{aligned} \frac{\partial E}{\partial \lambda} = \int_{\partial \hat{\Omega}^1} \left\langle \frac{\partial \hat{x}}{\partial \lambda}, \hat{N} \right\rangle \left(\left(I(\hat{x}) - \rho^1(x) \right)^2 - \left(I(\hat{x}) - \rho^0(\hat{x}) \right)^2 \right) d\hat{s} \, + \\ + 2 \int_{\hat{\Omega}^1} \left(I(\hat{x}) - \rho^1(x) \right) \left\langle \nabla \rho(x), Dw \frac{\partial g(\hat{x})}{\partial \lambda} \right\rangle d\hat{x} \end{aligned} \tag{5}$$

$\hat{N}$ is the outward unit normal and $d\hat{s}$ is the arclength element of $\partial\hat{\Omega}^1$. The solution of w is similar to the solution for g except there is included the laplacian terms for the regularizer.

The curve evolution is also similar to the boundary-based term for the evolution of g:

$$\frac{\partial C}{\partial t} = -\Big(\big(I(\hat{x})-\rho^1(x)\big)^2-\big(I(\hat{x})-\rho^0(\hat{x})\big)^2\Big)\hat{N} \qquad (6)$$

The solution of ρ^k is the solution of the usual Mumford-Shah problem for the radiance portion with Poisson-type equations.

$$\Delta\rho^1(x) = \frac{1}{\beta}\big(\rho^1(x) - I(\hat{x})\big), \qquad x \in \Omega^1 \qquad (7)$$

$$\Delta\rho^0(x) = \begin{cases} 0, & x \in \hat{\Omega}^1 \\ \frac{1}{\beta}\big(\rho^0(x) - I(x)\big), & x \in \Omega^0 \setminus \hat{\Omega}^1 \end{cases} \qquad (8)$$

3. Layered Deformotion with Joint Prior

After seeing all the gradient descent equations from the previous section, it becomes obvious that a solution would take a while to acquire. By fixing each layer Ω^k to an "average shape", it becomes possible to look at the radiances ρ^k and the diffeomorphisms w^k and build a prior on them. We take a database of images of an object of interest and run the "layered deformotion" algorithm on them to "learn" the radiances and to "learn" the diffeomorphisms that object has in the database. Then we reduce that space of radiances and diffeomorphisms using principal components analysis. The modeling of appearance and shape of any new object of that trained class becomes much more accurate and computationally efficient.

3.1. Derivation of w PCA flow with one constant

If we build $w(x)$ and $\rho^1(x)$ jointly out of principal components we get

$$\begin{bmatrix} \rho^1(x) \\ w(x) \end{bmatrix} = \begin{bmatrix} \rho^1(x) \\ x + u(x) \\ y + v(x) \end{bmatrix} \qquad (9)$$

$$= \begin{bmatrix} \bar{\rho}^1(x) + \sum_{i=1}^N c_i \rho_i^1(x) \\ x + \bar{u}(x) + \sum_{i=1}^N c_i u_i(x) \\ y + \bar{v}(x) + \sum_{i=1}^N c_i v_i(x) \end{bmatrix} \qquad (10)$$

where a bar over a variable denotes the mean. Let us recall the energy E:

$$\begin{aligned} E &= \int_{\hat{\Omega}^1} \big(I(\hat{x}) - \rho^1(x)\big)^2 - \big(I(\hat{x}) - \rho^0(\hat{x})\big)^2 \, d\hat{x} \\ &+ \beta \sum_{k=0}^{1} \int_{\Omega^k} \langle \nabla\rho^k, \nabla\rho^k \rangle \, dx \qquad (11) \\ &+ \gamma \int_{\Omega^1} \langle \nabla u(x), \nabla u(x) \rangle + \langle \nabla v(x), \nabla v(x) \rangle \, dx \end{aligned}$$

and write it with the portion that have principal components:

$$\begin{aligned} E &= \int_{\hat{\Omega}^1} (I(\hat{x}) - \rho^1(x^1, y^1))^2 - (I(\hat{x}) - \rho^0(\hat{x}))^2 d\hat{x} \\ &+ \beta \int_{\Omega^1} \left\langle \nabla\rho^1, \nabla\left(\bar{\rho}^1(x) + \sum_{i=1}^N c_i \rho_i^1(x)\right)\right\rangle dx (12) \\ &+ \gamma \int_{\Omega^1} \left\langle \nabla u(x), \nabla\left(\bar{u}(x) + \sum_{i=1}^N c_i u_i(x)\right)\right\rangle dx \\ &+ \gamma \int_{\Omega^1} \left\langle \nabla v(x), \nabla\left(\bar{v}(x) + \sum_{i=1}^N c_i v_i(x)\right)\right\rangle dx \end{aligned}$$

where

$$x^1 = g^{-1}(\hat{x}) - \bar{u}(x) - \sum_{i=1}^N c_i u_i(x) \qquad (13)$$

$$y^1 = g^{-1}(\hat{y}) - \bar{v}(x) - \sum_{i=1}^N c_i v_i(x) \qquad (14)$$

By differentiating we get:

$$\begin{aligned}\frac{\partial E}{\partial c_j} &= \int_C \left\langle RS([u_j(x), v_j(x)]^T), J(g'w'T)\right\rangle f(x)ds \\ &+ 2\int_{\Omega^1} |g'||w'|(I(g(w(x)))–\rho^1(x)) \left\langle \nabla\rho^1(x), w_j(x)\right\rangle dx \\ &- 2\int_{\Omega^1} |g'||w'|(I(g(w(x))) - \rho^1(x))\rho^1_j(x)dx \\ &+ 2\gamma\int_{\Omega^1} \langle \nabla u_j(x), \nabla\bar{u}(x)\rangle\, dx \\ &+ 2\gamma\sum_{i=1}^{N} c_i \int_{\Omega^1} \langle \nabla u_j(x), \nabla u_i(x)\rangle\, dx \\ &+ 2\gamma\int_{\Omega^1} \langle \nabla v_j(x), \nabla\bar{v}(x)\rangle\, dx \\ &+ 2\gamma\sum_{i=1}^{N} c_i \int_{\Omega^1} \langle \nabla v_j(x), \nabla v_i(x)\rangle\, dx \\ &+ 2\beta\int_{\Omega^1} \left\langle \nabla\rho^1_j(x), \nabla\bar{\rho}^1(x)\right\rangle dx \\ &+ 2\beta\sum_{i=1}^{N} c_i \int_{\Omega^1} \left\langle \nabla\rho^1_j(x), \nabla\rho^1_i(x)\right\rangle dx \\ &= 0\end{aligned} \tag{15}$$

where

$$f(x) = [(I(g(w(x)))–\rho^1(x))^2–(I(g(w(x)))–\rho^0(g(w(x))))^2 \tag{16}$$

4. Experiments

The training set consisted of 300 cars from the dataset from [10] consisting of segmentations like that of Figure 1. An average shape was derived from all 300 examples and each segmentation was registered (with a rigid component and a non-rigid component) to the average shape. Then the image data could be mapped onto the average shape. This gives the training data for the radiances ρ^1 and the warps w^1. The modes for the appearance are in Figure 2. The modes for the diffeomorphisms are in Figure 3. Remeber for the warps that there is a rigid component that captures variablility in shape as well. The modes for the appearance and the diffeomorphism jointly are in Figure 4.

Using 10 principal components for the radiance and the warp jointly, we obtained the results in Figures 5 and 6. Figures 5 and 6 shows the initial placement of the contour along with the segmentation obtained by using pca for the radiances and the warps. The given example is not in the training database as we can see with the reconstruction of the radiance. The last pictures shown in both Figure 5 and 6 are just a segmentation not using principal components analysis, but using rigid registration of the average shape with the Chan-Vese Algorithm [3],[14].

5. Conclusion

By incorporating a joint prior on the appearance and warp for each layer, we are able to significantly improve the method of layered deformotion by exploiting the flexibility of that technique. By fixing the shape of each layer, it became possible to perform dimensionality reduction (via principal components analysis) on the most complex functions to solve for–the appearances and the warps. By using principal components analysis, we were able to obtain more useful and accurate segmentations. Since our goal is stereo reconstruction in 3D, we have shown the first stage in this approach by addressing the simpler case of 2D shape/radiance detection in single images. The results obtained so far show tremendous promise for 3D.

References

[1] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel. Axioms and fundamental equations of image processing. *Arch. Rational Mechanics*, 123, 1993. 1

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of SIGGRAPH 2000, New Orleans, USA, July 2000*. 2

[3] T. Chan and L. Vese. Active contour without edges. *IEEE Transactions on Image Processing*, 10:266–277, 2001. 4

[4] T. F. Cootes, C. J. Taylor, D. M. Cooper, and J. Graham. Active shape models – their training and applications. 61(1):38–59, 1995. 1

[5] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. Wiley, Chichester, 1998. 1

[6] U. Grenander. *Lectures in Pattern Theory*. Springer, Berlin, 1976. 1

[7] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981. 2

[8] J. D. Jackson, A. Yezzi, and S. Soatto. Dynamic shape and appearance modeling via moving and deforming layers. In *EMMCVPR 2005*, 2005. 1

[9] H. Jin, S. Soatto, and A. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *Intl. J. Computer Vision*, 63(3):175–189, July 2005. 2

[10] E. Jones and S. Soatto. Layered active appearance models. In *ICCV*, volume 2, pages 1097–1102, Oct 2005. 4

[11] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991. 1

[12] A. Vedaldi and S. Soatto. Features for recognition: viewpoint invariance for non-planar scenes. In *Proc. of the Intl. Conf. of Comp. Vision*, October 2005. 1

[13] A. Yezzi and S. Soatto. Deformotion: deforming motion, shape average and the joint segmentation and registration of images. *Intl. J. of Comp. Vis.*, 53(2):153–167, 2003. 2

[14] A. Yezzi, L. Zollei, and T. Kapur. A variational framework for joint segmentation and registration. In *Mathematical Methods in Biomedical Image Analysis*, pages 44–51, Dec 2001. 4

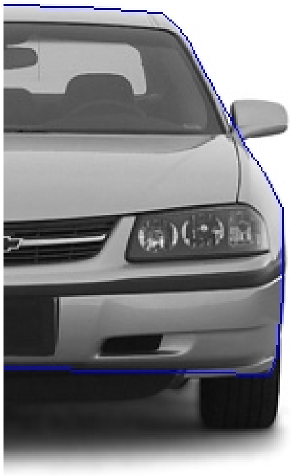

Figure 1. Training Example

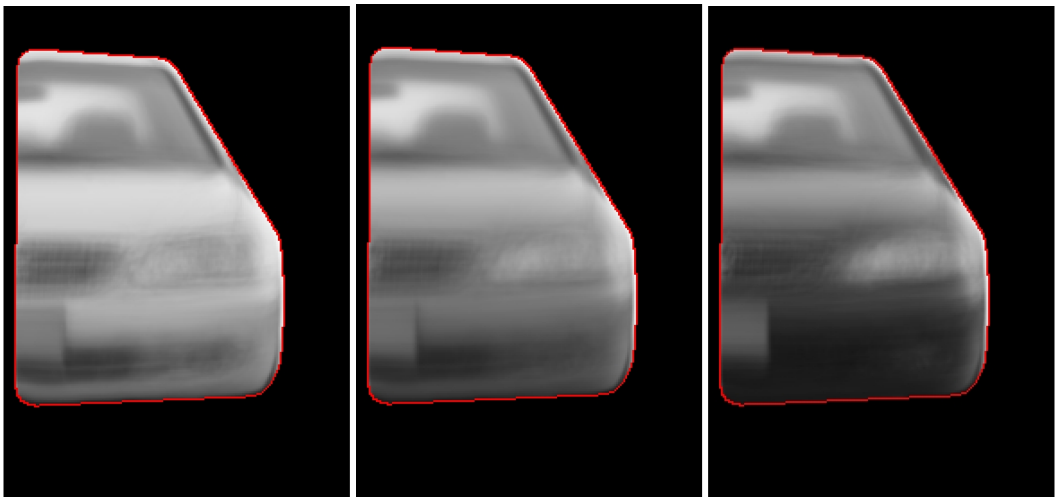

Figure 2. Appearance Modes:mean$-1\sigma*1$st, mean, mean$+1\sigma*1$st

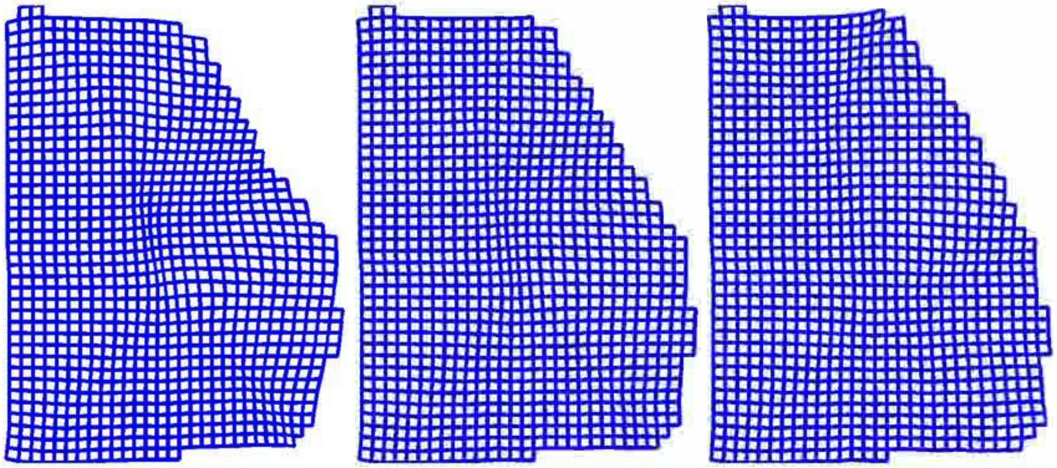

Figure 3. Warp Modes:varying 4th mode from the mean warp

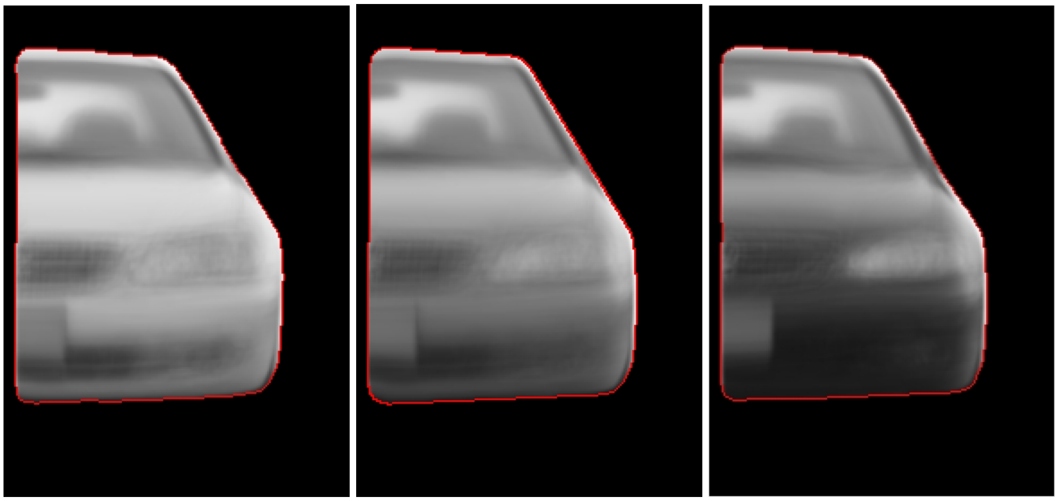

Figure 4. Warp and Appearance Modes:mean$-1\sigma*1$st, mean, mean$+1\sigma*1$st

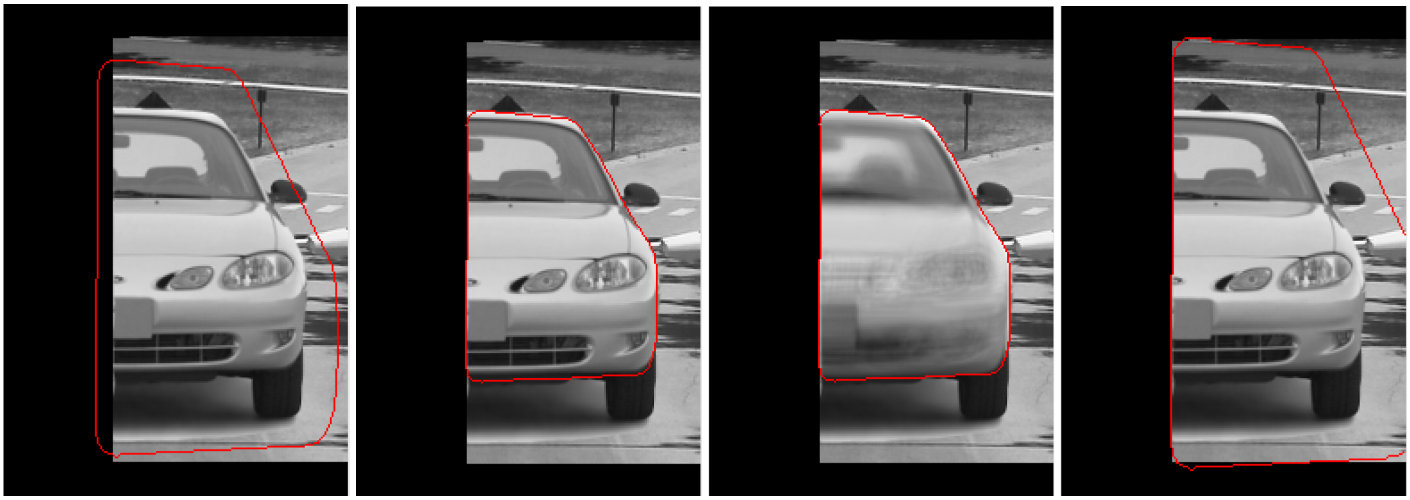

Figure 5. Initialization, Using our joint prior, Its pca reconstruction, Segmentation using Chan-Vese rigid registration with no prior

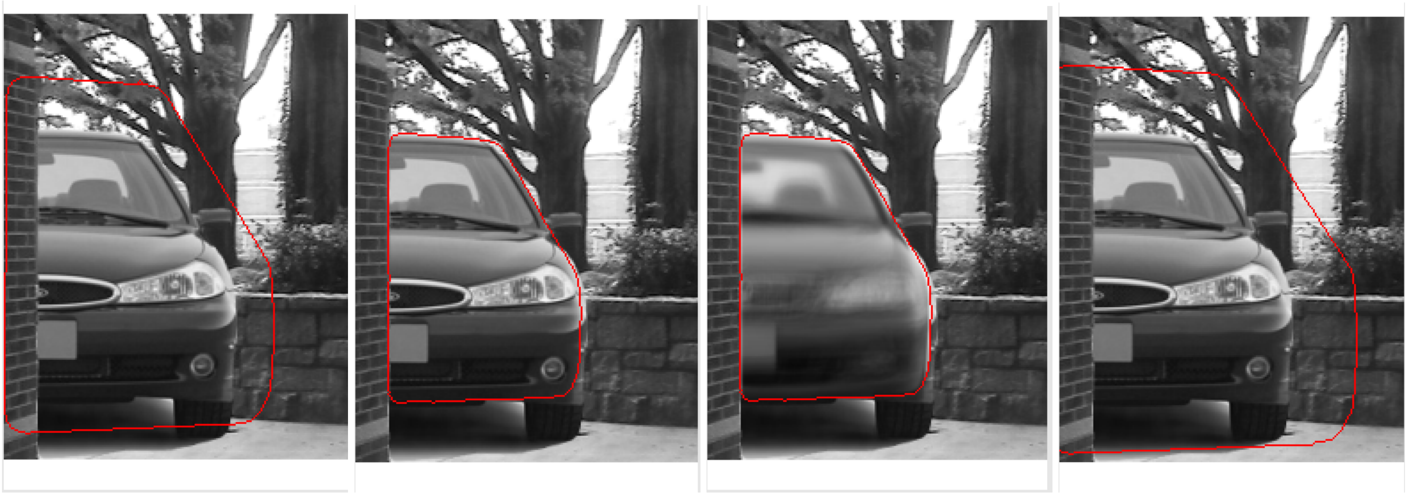

Figure 6. Initialization, Using our joint prior, Its pca reconstruction, Segmentation using Chan-Vese rigid registration with no prior