

# Dynamic Shape and Appearance Modeling Via Moving and Deforming Layers

Jeremy D. Jackson<sup>1</sup>, Anthony Yezzi<sup>1</sup>, and Stefano Soatto<sup>2</sup>

<sup>1</sup> School of Electrical and Computer Engineering,  
Georgia Institute of Technology, Atlanta, GA 30332  
gtg120d@mail.gatech.edu, ayezzi@ece.gatech.edu

<sup>2</sup> UCLA Computer Science Department, 4732 Boelter Hall, Los Angeles, CA 90095-1596  
soatto@ucla.edu

**Abstract.** We propose a model of the shape, motion and appearance of a sequence of images that captures occlusions, scene deformations, arbitrary viewpoint variations and changes in irradiance. This model is based on a collection of overlapping layers that can move and deform, each supporting an intensity function that can change over time. We discuss the generality and limitations of this model in relations to existing ones such as traditional optical flow or motion segmentation, layers, deformable templates and deformation. We then illustrate how this model can be used for inference of shape, motion, deformation and appearance of the scene from a collection of images. The layering structure allows for automatic inpainting of partially occluded regions. We illustrate the model on synthetic and real sequences where existing schemes fail; we implement our gradient-based infinite-dimensional optimization using level set methods.

## 1 Introduction

We are interested in modeling video sequences where changes occur over time due to viewer motion, motion or deformation of objects in the scene – including occlusions – and appearance variations due to the motion of objects relative to the illumination. A suitable model will trade off generality, by allowing variations of shape, motion and appearance, with tractability, by being amenable to inference and analysis. The goal of modeling is to support inference, and depending on the application one may be more interested in recovering shape (e.g. in shape analysis, classification, recognition, registration), or recovering motion (e.g. tracking, optical flow), or appearance variations (e.g. segmentation). Traditionally, therefore, the modeling task has been approached by making strict assumptions on some of the unknowns in order to recover the others, for instance the brightness-constancy assumption in optical flow, or the affine warping in shape analysis and registration. This is partly justified because in any image-formation model there is ambiguity between the three factors – shape, motion and appearance – and therefore the most general inference problem is ill-posed. In some applications, for instance video compression, the ambiguity is moot since all that matters is for the model to capture the sequence as faithfully and parsimoniously as possible. Nevertheless, since all three factors affect the generation of the image, a more germane approach

would call for modeling all three jointly, then letting the analysis dictate the responsibility of each factor, and the application dictate the choice of suitable regularizers to make the inference algorithms well posed. We therefore concentrate our attention on modeling, not on any particular application. So, this is not yet another paper on tracking, nor on motion segmentation, nor on optical flow, nor on shape registration. It is a little bit of each.

We propose a model of image formation that is general enough to capture shape, motion and appearance variations (Sect. 2), and simple enough to allow inference (Sect. 3). We want to be able to capture *occlusion phenomena*, hence our model will entail a notion of *hierarchy* or *layering*; we want to capture image variability due to arbitrary *changes in viewpoint* for non-planar objects, hence our model will entail *infinite-dimensional deformations* of the image domain. Such deformations can be due to changes in viewpoint for a rigid scene, or changes of shape of the scene seen from a static viewpoint, or any combination thereof. Our model will not attempt to resolve this ambiguity, since that requires higher-level knowledge. Furthermore, we want to capture large-scale motion of objects in the scene, as opposed to deformations, hence we will allow for a choice of a finite-dimensional group, e.g. Euclidean or affine. Finally, we want to capture changes in appearance, hence scene radiance will be one of the unknowns in our model. Changes in radiance can come from changes in reflectance or changes in illumination, including changes in the mutual position between the light sources and the scene; again we do not attempt to resolve this ambiguity, since that requires higher-level knowledge. The image-formation model we propose is not the most general that one can conceive; far from it. Indeed, it is far less general than the simplest models considered acceptable in Computer Graphics, and we illustrate the lack of generality in Sect. 2.1. Nevertheless, it is more general than any other model used so far for motion analysis in Computer Vision, as we discuss also in Sect. 2.1, and is complex enough to be barely tractable with the analytical and computational tools at our disposal today. We pose the inference problem within a variational framework, involving partial differential equations, integrated numerically in the level set framework [15], although any other computational scheme of choice would do, including stochastic gradients or Markov-chain Monte Carlo. The point of this paper is to propose a model and show that it can be inferred with at least one particular computational scheme, not to advocate a particular optimization technique.

## 1.1 Relation to Existing Work

This work relates to a wide body of literature in scene modeling, motion estimation, shape analysis, segmentation, and registration which cannot be properly reviewed in detail in the limited space available. In Sect. 2.1 we illustrate the specific relationship between the model we propose and existing models. These include Layers [20, 12], which only model affine deformations of the domain and can therefore only capture planar scenes under small viewer motion or small aperture, and where there is no explicit spatial consistency within each layer and the appearance of each layer is fixed. As we will illustrate, our model allows deformations that can model arbitrary viewpoint variation, layer deformation and enforce spatial coherence within each layer. One could think of our work as a generalization of existing work on Layers to arbitrary view-

point changes, or arbitrary scene shape, and to changes in radiance (texture), all cast within a principled variational framework. Our work relates to a plethora of variational algorithms for optical flow computation, for instance [18, 1, 8] and references therein, except that we partition the domain and allow arbitrary smooth deformations as well as changes in appearance (that would violate the brightness constancy constraints that most work on optical flow is based on, with a few exceptions, e.g. [10]). It also relates to various approaches to motion segmentation, where the domain is also partitioned and allowed to move with a simple motion, e.g. Euclidean or affine, see for instance [7] and references therein. Such approaches do not allow deformations of the region boundaries, or changes in the intensity within each region. Furthermore, they realize a partition, rather than a hierarchy, of domain deformations, so our model can be thought of as motion segmentation with moving and deforming layers with changes in intensity and inpainting [3]. In this, our work relates to [19], except that we allow layers to overlap. So, our work can be thought of as a layered version of Deformation with changes in region intensities. Also very related to our work is work done by [17] in registering one distance function to another using a rigid and non-rigid transformation. Our work relates to deformable templates [14, 9], in the sense that each of our layers will be a deformable template. However, we do not know the shape and intensity profile of the template, so we estimate that along with the layering structure. Our work is also related to active appearance models [6, 2], in that we seek the same goal, although rather than imposing regularization of shape and appearance by projection onto suitably inferred linear subspaces we employ generic regularizers. One can therefore think of our work as a generalization of active appearance models to smooth shape and intensity deformations, cast in a variational framework. Of course this work relates more generically to active contours, e.g. [4, 13, 5, 16] and references therein. In the next section we introduce our model, and in Sect. 3 we illustrate our approach to infer its (infinite-dimensional) constitutive elements.

## 2 Modeling

We represent a scene as a collection of  $L$  overlapping *layers*. Each layer, labeled by an index  $l = 1, \dots, L$ , is a function that has associated with it a domain, or *shape*  $\Omega^l \subset \mathbb{R}^2$ , and a range, or *radiance*  $\rho^l : \Omega^l \rightarrow \mathbb{R}^+$ . Layer boundaries model the occlusion process, and each layer  $l$  undergoes a *motion*, described by a (finite-dimensional) group action  $g^l$ , for instance  $g^l \in SE(2)$  or  $A(2)$ , and a *deformation*, or *warping*, described by a diffeomorphism  $w^l : \Omega^l \rightarrow \mathbb{R}^2$ , in order to generate an image  $I$  at a given time  $t$ . Warping models changes of viewpoint for non-planar scenes, or actual changes in the shape of objects in the scene. Since each image is obtained from the given scene after a different motion and deformation, we index each of them by  $t$ :  $g_t^l$ ,  $w_t^l$ , and  $I_t$ . Finally, since layers occlude each other, there is a natural ordering in  $l$  which, without loss of generality, we will assume to coincide with the integers: Layer  $l = 1$  is occluded by layer  $l = 2$  and so on. So, the only layer that will contribute to the intensity at a pixel in a given image is the frontmost that intersects the warped domain. For simplicity we assume that  $\Omega^0 = \mathbb{R}^2$  (the backmost layer, or “the background”). With this notation, the model of how the value of the generic image  $I_t : \Omega^0 \rightarrow \mathbb{R}^+$  at the location  $x \in$

$\Omega^0 \subset \mathbb{R}^2$  is generated can be summarized as  $I_t(g_t^l \circ w_t^l(x)) = \rho^l(x)$ , with  $x \in \Omega^l$ ,  $l = \max\{k \mid x \in \Omega^k\}$ . To simplify the notation, we call  $x_t^l \doteq g_t^l \circ w_t^l(x)$ , which sometimes we indicate, for simplicity, as  $x_t$ , so that

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

Our goal in this work is to infer the radiance family  $\{\rho^l\}_{l=1,\dots,L}$ , the shape family  $\{\Omega^l\}_{l=1,\dots,L}$ , the motions  $\{g_t^l\}_{l=1,\dots,L;t=1,\dots,N}$  and the deformations  $\{w_t^l\}_{l=1,\dots,L;t=1,\dots,N}$  that minimize the discrepancy of the measured images from the ideal model (1), subject to generic regularity constraints. Such a discrepancy is measured by a cost functional  $\phi(\Omega^k, \rho^k, w_t^k, g_t^k)$  to be minimized as

$$\begin{aligned} \phi \doteq & \sum_{t=1}^N \int_{\Omega^0} \left( I_t(x_t) - \rho^l(w_t^{l-1} \circ g_t^{l-1}(x_t)) \right)^2 dx_t + \\ & + \lambda \sum_{k=1}^L \int_{\Omega^k} \|\nabla \rho^k(x)\|^2 dx + \mu \sum_{k,t=1}^{L,N} \int_{\Omega^k} r(w_t^k(x)) dx \end{aligned} \quad (2)$$

subject to  $l = \max\{k \mid x \in \Omega^k\}$ .

Here  $r$  is a regularizing functional, for instance  $r(w) \doteq |w| + \frac{1}{|w|}$ , and  $\lambda, \mu$  are positive constants. Note that  $l$  is a function, specifically  $l : \Omega^0 \rightarrow \mathbb{Z}^+$ .

## 2.1 Generality of the Model

It can be easily shown that eq. (1) models images of 3-D scenes with piecewise smooth geometry exhibiting Lambertian reflection with piecewise smooth albedo<sup>1</sup> viewed under diffuse illumination from an arbitrarily changing viewpoint. It does not capture global or indirect illumination effects, such as cast shadows or inter-reflections, complex reflectance, such as specularities, anisotropies or sub-surface scattering. These are treated as modeling errors and are responsible for the discrepancy between the model and the images, which is measured by  $\phi$  in eq. (2). We lump these discrepancies together with sensor errors and improperly call them “noise.” Although far from general, (1) is nevertheless a more ambitious model than has ever been used in the context of motion estimation and tracking. In fact, the reader can easily verify that many existing models are special cases of (1). In particular,  $L = 0$ ,  $g = Id$ ,  $\lambda = 0$  yields traditional **optical flow**, where  $\rho = I_{t+1}$ . There are too many variants of this model to review here, depending on the choice of norm and regularizers, stemming from the ancestor [11]. Choosing  $L = 1$ ,  $w = Id$ ,  $\lambda = 0$  yields **motion segmentation**, that has also been addressed by many, see for instance [7] and references therein for the case of affine motion  $g \in \mathbb{A}(2)$ .  $L = 1, \rho = \text{const}$ ,  $r(w) = \|w\|$  yields a model called **Deformation**

<sup>1</sup> The model can be further generalized by allowing  $\rho^l$  to be vector-valued to capture a set of radiance statistics such as the coefficients of a filter bank or other texture descriptors, but this is beyond the scope of this paper.

in [19], and has also been extended to grayscale images  $L = 1$ ,  $r(w) = \|w\|$ . Choosing  $L > 1$ ,  $w = Id$ ,  $\Omega^k$  unconstrained and  $g \in \mathbb{A}(2)$  would yield a variational version of the **Layers** model [20], that to the best of our knowledge has never been attempted. Note that this is different than simpler variational multi-phase motion segmentation, since in that case the motion of a phase affects the shape of neighboring phases, whereas in the model (1) layers can overlap without distorting underlying domains. One can think of the Layer model as a multi-phase motion segmentation with inpainting [3] of occluded layers and shape constraints. The model also relates to **deformable templates**, where  $\rho = \text{const}$  in the traditional model [9] and  $\rho = \text{smooth}$  in the more general version [14]. Another relevant approach is **Active Appearance Models** where the regions, warping and radiance are modeled as points in a linear space. The model (1) does not impose such restrictions, and render the problem well-posed by generic regularization instead.

### 3 Inference

Minimizing the cost functional in (2) is a tall order. It depends upon each domain boundary (a closed planar contour)  $\Omega^k$ , its deformation (a flow of planar diffeomorphisms)  $w_t^k$ , the radiance (a piecewise smooth function)  $\rho^k$ , all of which are infinite-dimensional unknowns. In addition, it depends on a group action per layer per instant,  $g_t^k$ , and on the occlusion model, which is represented by the discrete-valued function  $l(x) = \max\{k \mid x \in \Omega^k\}$ , and all of this for each layer  $k = 1, \dots, K$ . The first simplification is to notice that, as long as each layer is a compact region bounded by a simple smooth curve, there is no loss of generality in assuming that  $\Omega^k$  are fixed. This is because each diffeomorphism  $w_t^k$  will act transitively on it. Therefore, we assume that each region  $\Omega^k$  is a circle in most of the examples. While there is no loss of generality, there is a loss of energy, in that if we were allowed to also optimize with respect to the initial regions we would be able to reach each deforming layer with less energy. This, however, does not enhance the generality of the model, hence we will forgo it (see Fig. 1 for an illustration of this effect).

Apart from this simplification, we proceed by minimizing the functional (2) using simultaneous gradient flows with respect to the groups (motion), the radiance (appearance) and the diffeomorphisms (deformation). The detailed evolution equations are a bit complicated depending upon the number of layers and the occlusion structure between layers. To help avoid excessive subscripting and superscripting and multiple-case definitions according to occlusion relationships, we will outline some of the key properties of the various gradient flows for the case of a background layer  $\Omega^0$ , a single image  $I$ , and a single foreground layer  $\Omega^1$ . We will also, to help keep the illustration simple, assume that the group action  $g^0$  and the warp  $w^0$  for the background layer are simply the identity transforms. This is the simplest possible scenario that will allow us to still show the key properties of the gradient flows.

Let  $\hat{x} = g^1(w^1(x))$  and  $\hat{\Omega}^1 = g^1(w^1(\Omega^1))$ . With this notation, we may write image-dependent terms in our energy functional as follows.

$$E = \int_{\hat{\Omega}^1} (I(\hat{x}) - \rho^1(x))^2 d\hat{x} + \int_{\Omega^0 \setminus \hat{\Omega}^1} (I(x) - \rho^0(x))^2 dx \tag{3}$$

If  $g$  denotes any single parameter (e.g. horizontal translation) of the group  $g^1$ , then differentiating yields

$$\begin{aligned} \frac{\partial E}{\partial g} &= \int_{\partial\hat{\Omega}^1} \left\langle \frac{\partial\hat{x}}{\partial g}, \hat{N} \right\rangle \left( (I(\hat{x}) - \rho^1(x))^2 - (I(\hat{x}) - \rho^0(\hat{x}))^2 \right) d\hat{s} \\ &+ 2 \int_{\hat{\Omega}^1} (I(\hat{x}) - \rho^1(x)) \left\langle \nabla\rho(x), \text{inv} \left[ (w^1)' \right] \frac{\partial}{\partial g} \text{inv}[g^1](\hat{x}) \right\rangle d\hat{x} \end{aligned} \tag{4}$$

where  $\hat{N}$  and  $d\hat{s}$  denote the outward unit normal and the arclength element of  $\partial\hat{\Omega}^1$  respectively. We are able to note two things. First, the update equations for the group involve measurements both along the boundary of its corresponding layer (first integral) as well as measurements within the layer’s interior (second integral). Notice that this later integral vanishes if a constant radiance  $\rho$  is utilized for the layer. We also see that it is not necessary to differentiate the image data  $I$ . Derivatives land on the estimated smooth radiance  $\rho$  instead, which is a significant computational perk of our model that results in considerable robustness to image noise.

A similar gradient structure arises for the case of the infinite dimensional warp  $w$  (boundary based terms and region based terms for each layer similar to previous integrals). However, additional terms arise in the gradient flow equations for  $w$  depending upon the choice of regularization terms in the energy functional (smoothness penalties, magnitude penalties etc.).

The curve evolution is also similar to the boundary based term for the evolution of the  $g$ ’s:

$$\frac{\partial C}{\partial t} = - \left( (I(\hat{x}) - \rho^1(x))^2 - (I(\hat{x}) - \rho^0(\hat{x}))^2 \right) \hat{N} \tag{5}$$

Finally, the optimality conditions for the smooth radiance functions  $\rho^0$  and  $\rho^1$  are given by the following Poisson-type equations.

$$\Delta\rho^1(x) = \lambda(\rho^1(x) - I(\hat{x})), \quad x \in \Omega^1 \tag{6}$$

$$\Delta\rho^0(x) = \begin{cases} 0, & x \in \hat{\Omega}^1 \\ \lambda(\rho^0(x) - I(x)), & x \in \Omega^0 \setminus \hat{\Omega}^1 \end{cases} \tag{7}$$

Notice that the background radiance  $\rho^0$  is “inpainted” in regions occluded by the foreground layer  $\Omega^1$  by harmonic interpolation (as it satisfies Laplace’s equation  $\Delta\rho^0 = 0$ ) of values along the boundary of  $\hat{\Omega}^1$ .

## 4 Experiments

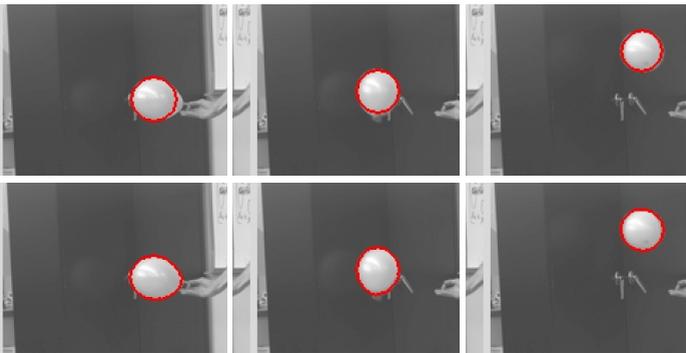
The first simple experiment is meant to illustrate that there is no lack of generality in the model by assuming that the shape of the initial regions  $\Omega$  is fixed. This is because the transformations  $g$  and  $w$  act transitively to obtain regions bounded by general simple, closed, smooth planar contours. In Fig. 1 we illustrate this point by allowing a circular region to capture the motion and deformation of a rectangle. This simulation involves one background layer, one foreground layer, and one set of transformations  $g$  and  $w$ . We



**Fig. 1.** Rectangle Captured By a Circle: The shape of the region is fixed to be a circle (left), and the appearance of the images (in this case a simple binary image of a rectangular domain) is captured by its motion  $g$  (middle) and deformation  $w$  (right) without loss of generality

can simultaneously find  $g$  and  $w$ , but in this experiment the similarity transformation  $g$  is allowed to reach steady state and then the warp  $w$  is found. The data fidelity term used is a Mumford-Shah term, so the radiances representing each layer are smooth functions. Figure 1 shows the initial circle placed over the image and then the image with the similarity transformation at steady state; Finally, the warp  $w$  is applied. Adding the regions  $\Omega$  to the model, therefore, would only simplify the optimization procedure, and allow us to reach various shapes with less energy, but not add to the generality of the model.

In the next experiment we illustrate the capability of our model to track deforming layers. In Figure 2 we show three sequences of an image where a deflating balloon is undergoing a rather errating motion while deforming from an initial waterdrop shape to a circular one, finally to a spermatozoidal shape. On the top row of Fig. 2 we show the layer boundaries for a model that only allow for affine deformations of the initial contour (a circle). This is essentially a variational implementation of the model of [20]. As it can be seen, it captures the gross motion of the balloon, but it cannot capture the subtler



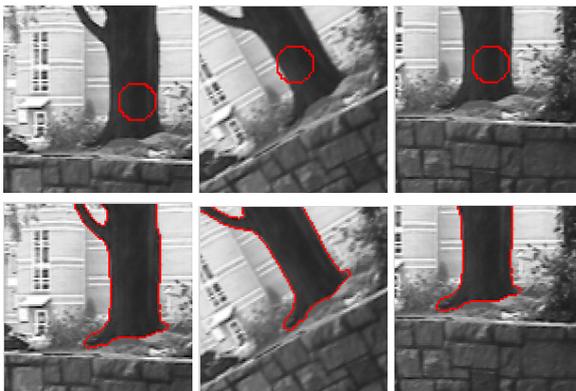
**Fig. 2.** Tracking a Balloon: Three sample views are shown from a sequence of a deflating balloon moving with an erratic motion while changing its shape from a drop-like shape to a circle. In the top row we show the boundary of the first layer as estimated by an affine layer model that does not allow for layer deformation, akin to a variational implementation of traditional layer models. As it can be seen, the model tracks the motion of the layer, but it fails to capture its deformation. On the bottom row we show the same three images with the first layer superimposed, where the layer is allowed to both move (affinely) and deform (diffeomorphically), yielding 12% lower RMS residual error, and capturing the subtler shape variations.



**Fig. 3.** Original Tree Background 300x300 image

shape variations. The second row shows the same three sample images with the boundary of the first layer superimposed, where the layer is allowed to deform according to the model we have introduced. Again the data fidelity term used is a Mumford-Shah term so the radiances representing each layer are smooth functions. As it can be seen, the layer changes shape to adapt to the deforming balloon, all while capturing its rather erratic motion. The average RMS error per image for the affine layer model is **30.87**, whereas the residual for the case of the deforming layers is **5.51**, corresponding to a 12% improvement. More importantly, the phenomenology of the scene, visible in the figure, has been correctly captured.

Figure 3 shows a 300x300 image of a tree that was used in an experiment that allows the foreground boundary (contour) to move and allows separate transformations for the foreground and background layers to be found. Figure 4 shows 100x100 images that have been cut out of Figure 3 that were used for this experiment. Figure 4 shows this example that has a rigid transformation for the foreground layer and has a separate rigid transformation for the background layer. The contour that bounds the



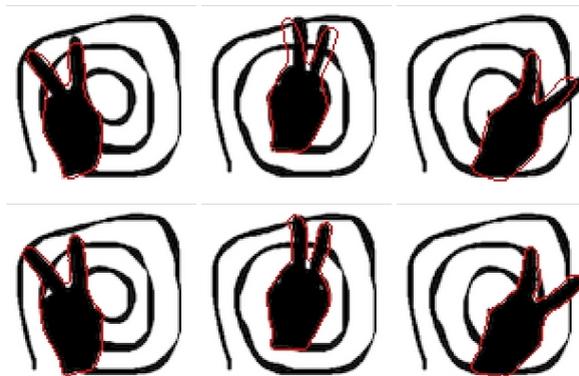
**Fig. 4.** Curve evolution with background and foreground transformations: The 100x100 images here are taken from Fig 3. The top row shows the initial curve, the bottom row shows the final segmentation and registration of the tree. A transform is computed for the background as well and gives rise to the next Figure 5.



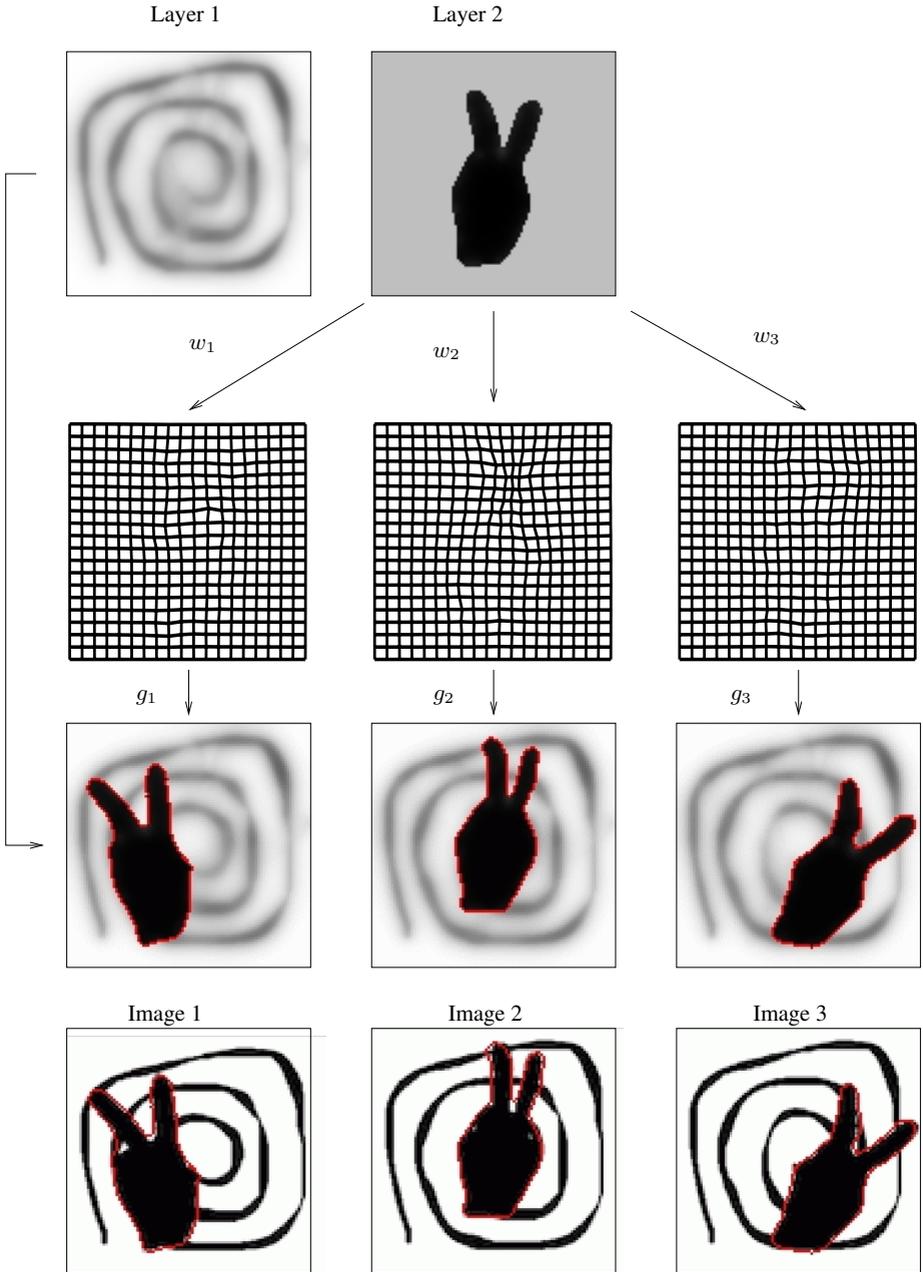
**Fig. 5.** Reconstructed Tree Background: This image is the smooth approximation of the background  $\rho^0$  given by the background regions from the three images in Figure 4

foreground layer is allowed to evolve to capture the tree. The reconstructed background function is shown in Figure 5.

In the next experiment we illustrate all the features of our model by showing how it allows recovering the background behind partially occluded layers while recovering their motion and deformation. In Figure 6 we show a few samples from this dataset where the silhouette of a moving hand forms a victory sign while moving the relative position between the fingers. The background, which is partially occluded, is a spiral. Here we use the average shape as the initial shape of the foreground layer to find its



**Fig. 6.** Victory sign, with deforming hand, moving in front of a partially occluded background portraying a spiral. The goal here is to recover the radiance of each layer (the spiral in the background and the constant black intensity of the hand), as well as the motion and deformation of the foreground layer. Note that current layer models based only on affine motion would fail to capture the phenomenology of this scene by over-segmenting the region into three regions, each moving with independent affine motion. Our model captures the overall motion of the layer with an affine group, and then the relative motion between the fingers as a deformation, as we illustrate in the next figure 7.



**Fig. 7.** Multiple Layers Mapping onto Multiple Images: The inference process returns an estimate of the albedo in each layer (top). Since we are assuming smooth albedo, the spiral is smoothed. The deformation of each layer is estimated (second row) together with its affine motion, to yield an approximation of the image (third row). This is used for comparison with the measured images (bottom row) that drives the optimization scheme.

affine motion, and then the diffeomorphic warp  $w_i$ . Again we assume smooth radiance within each layer, so when we recover the background layer we will only be able to show a slightly smoothed version of the spiral (of course we could further segment the black spiral from the background and thus obtain sharp boundaries, but this is standard and therefore we do not illustrate it here.)

In Fig. 7 we illustrate the results of this experiments, arranged to summarize the modeling process. On the top row we show the recovered layers. Since we are assuming a smooth radiance within each layer, we can only recover a smoothed version of the spiral. These layers are deformed according to a diffeomorphism, one per layer, defined on the domain of the layer (second row) and then moved according to an affine motion. The third row shows the image generated by the model, which can therefore be thought of as a generative (although deterministic) model since it performs comparison at the image level, not via some intermediate feature. The corresponding images are displayed in the last row, with the layers superimposed for comparison.

## 5 Discussion

We have presented a generative model of the appearance (piecewise smooth albedo), motion (affine transformation) and deformation (diffeomorphism) of a sequence of images that include occlusions. We have used this model as a basis for a variational optimization algorithm that simultaneously tracks the motion of a number of overlapping layers, estimates their deformation, and estimates the albedo of each layer, including portions that were partially occluded. Where no information is available, the layers are implicitly inpainted by their regularizers.

This model generalizes existing layer models to the case of deforming layers. Alternatively, one can think of our algorithm as a layered version of deformable tracking algorithms, or as a generalized version of optical flow or motion segmentation where multiple layers are allowed to occlude each other without disturbing the estimate of adjacent and occluded ones.

Our numerical implementation of the flow-based algorithm uses level set methods, and is realized without taking derivatives of the image, a feature that yields significant robustness when compared with boundary-based or optical flow algorithms. We have illustrated our approach on simple but representative sequences where existing methods fail to capture the phenomenology of the scene by either over-segmenting it, or by failing to capture its deformation while only matching its affine motion.

## References

1. L. Alvarez, J. Weickert, and J. Sanchez. A scale-space approach to nonlocal optical flow calculations. In *In ScaleSpace '99*, pages 235–246, 1999.
2. S. Baker, I. Matthews, and J. Schneider. Image coding with active appearance models. Technical report, Carnegie Mellon University, The Robotics Institute, 2003.
3. M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 417–424. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

4. A. Blake and M. Isard. *Active contours*. Springer Verlag, 1998.
5. V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *Int. J. of Computer Vision*, 22(1):61–79, 1997.
6. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. of the Eur. Conf. on Comp. Vis.*, pages 484–496, 1998.
7. D. Cremers. Multiphase levelset framework for variational motion segmentation. In *Intl. Conf. on Scale-Space Theories in Computer Vision*, pages 599–614, June 2003.
8. R. Deriche, P. Kornprobst, and G. Aubert. Optical flow estimation while preserving its discontinuities: a variational approach. In *Proc. of ACCV*, volume 2, pages 290–295, 1995.
9. U. Grenander. *General Pattern Theory*. Oxford University Press, 1993.
10. H. W. Haussecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):661–673, 2001.
11. B. K. P. Horn and B. G. Schunk. Determining optical flow. *Artificial Intell.*, 17:185–203, 1981.
12. S. Hsu, P. Anandan, and S. Peleg. Accurate computation of optical flow by using layered motion representations. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages 1621–1626, 1992.
13. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *Int. J. of Computer Vision*, 1(4):321–331, 1987.
14. M. I. Miller and L. Younes. Group action, diffeomorphism and matching: a general framework. In *Proc. of SCTV*, 1999.
15. S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi equations. *J. of Comp. Physics*, 79:12–49, 1988.
16. N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3):266–280, 2000.
17. Nikos Paragios, Mikael Rousson, and Visvanathan Ramesh. Non-rigid registration using distance functions. *Comput. Vis. Image Underst.*, 89(2-3):142–165, 2003.
18. C. Schnörr. Computation of discontinuous optical flow by domain decomposition and shape optimization. *Int. J. of Computer Vision*, 8(2):153–165, 1992.
19. S. Soatto and A. Yezzi. Deformation: deforming motion, shape average and the joint segmentation and registration of images. In *Proc. of the Eur. Conf. on Computer Vision (ECCV)*, volume 3, pages 32–47, 2002.
20. J. Wang and E. Adelson. Representing moving images with layers. In *IEEE Trans. on Image Processing*, volume 3(5), pages 625–638, 1994.