Local Occlusion Detection Under Deformations Using Topological Invariants [*]

Edgar Lobaton[1], Ram Vasudevan[2], Ruzena Bajcsy[2], and Ron Alterovitz[1]

[1] Department of Computer Science
University of North Carolina at Chapel Hill, NC 27599
{lobaton,ron}@cs.unc.edu
[2] Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720
{ramv,bajcsy}@eecs.berkeley.edu

Abstract. Occlusions provide critical cues about the 3D structure of man-made and natural scenes. We present a mathematical framework and algorithm to detect and localize occlusions in image sequences of scenes that include deforming objects. Our occlusion detector works under far weaker assumptions than other detectors. We prove that occlusions in deforming scenes occur when certain well-defined local topological invariants are not preserved. Our framework employs these invariants to detect occlusions with a zero false positive rate under assumptions of bounded deformations and color variation. The novelty and strength of this methodology is that it does not rely on spatio-temporal derivatives or matching, which can be problematic in scenes including deforming objects, but is instead based on a mathematical representation of the underlying cause of occlusions in a deforming 3D scene. We demonstrate the effectiveness of the occlusion detector using image sequences of natural scenes, including deforming cloth and hand motions.

1 Introduction

Inherent in the exhaustive work done on edge detection is the belief that discontinuities in image intensity provide valuable clues about scene structure. Edges resulting from occlusions are of special interest since they correspond to locations in an image where one surface is closer to the camera than another, which can provide critical cues about the 3D structure of a scene. Occlusion detection is used in numerous applications including shape extraction, figure-ground separation, and motion segmentation, e.g. [1–6]. The purpose of this paper is to present a completely local, bottom-up approach to detect and localize occlusions in order to provide this powerful low-level information to higher-level reasoning methods. Our approach is applicable to image sequences including deforming objects, which can present difficulties to classical methods.
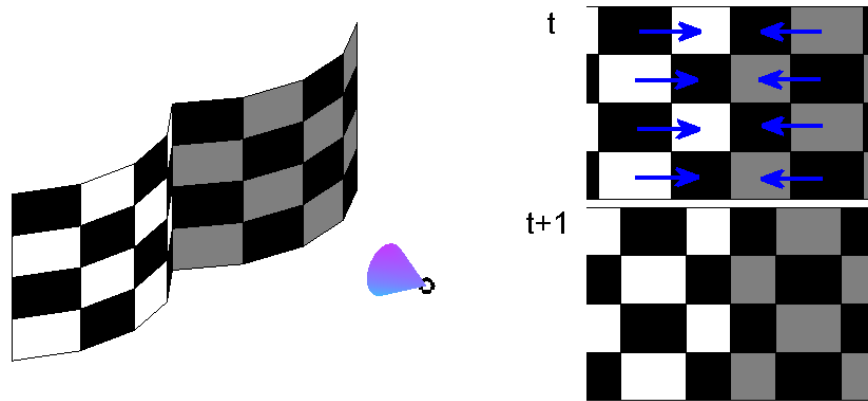
Fig. 1. The figure illustrates the inability of motion flow inconsistency to correctly identify occlusions in scenes with deforming objects. Consider two images (right column) obtained at times t and $t + 1$ after a paper in 3D is folded along a crease (left). Observe that this transformation is in fact not rigid and that there is no occlusion in either image. Assuming perfect motion estimation from the images on the right we would find that the gray colored checkerboard portion is moving to the left and the white colored checkerboard portion is moving to the right, which would seem to indicate the presence of a non-existent occlusion.

Traditional occlusion detectors rely almost entirely on spatiotemporal derivatives or matching to detect the artifacts of occlusions. These artifacts fall into two categories: motion flow inconsistency across an edge and the T-junction. Unfortunately, both methods are effective only under restrictive assumptions about the scene. The motion flow inconsistency approach implicitly assumes that only rigid transformations take place, such as a foreground and background layer moving in distinct directions. Due to this implicit assumption, motion inconsistencies do not necessarily imply an occlusion, as illustrated in Figure 1. Numerous methods are available to find T-junctions, but they all make assumptions about the orientations of the occluding contour. Moreover, even after a T-junction has been detected, an occlusion may not be present. Our method to detect occlusions works under far weaker assumptions than other methods. In particular, we only assume a weak bound on the magnitude of deformation on objects viewed by a camera and a bound on the color variations between frames.

In contrast to prior methods, we model the cause of occlusion, under a local deformation model, and show that the proper measurement of certain topological invariants serves as a definitive indicator to the presence of an occlusion. Prior approaches do not give any analytical guarantee on the validity of their detections, only experimental results. Our approach, in contrast to existing methods, is proven to yield a zero false positive rate as long as the required motion and color variation bounds are satisfied. The strength of our framework is that it is able to operate at different scales providing information that may otherwise be unavailable while not relying on noisy derivatives, not making strict assumptions
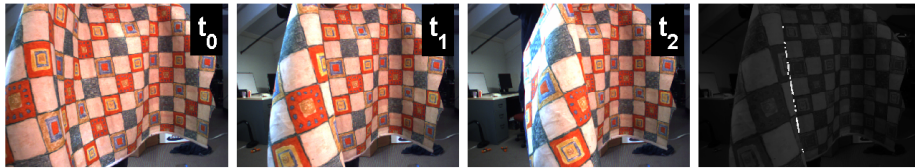
Fig. 2. The results of the occlusion detector presented in this paper applied to a sequence of images that depict a piece of cloth being folded. Observe the existence of a self-occlusion. Frames of a sequence are shown (first three columns) and the corresponding occlusion detections centered at time t_1 (right).

about the orientation of occluding contours, not building complex appearance models, and not performing any matching. In Figure 2, we apply our method to an image sequence of a cloth deforming in 3D, creating a self occlusion. Our method successfully detects the occlusion while producing no false positives. Note that these local detections can be fed into a global process such as graph cut in order to segment objects as is done in [2].

The contributions of this work are three-fold. First, in Section 3, we prove that under a deformation model occlusions occur when there does not exist homeomorphisms between pairs of images in an image sequences. Second, in Section 4, we define local topological invariants to detect an occlusion within an image. Finally, in Section 5, we demonstrate the applicability of our occlusion detector including some preliminary results on foreground/background segmentation.

2 Related Work

As described earlier, traditional approaches to occlusion detection can be divided into two categories: those that attempt to detect motion inconsistencies and those that detect T-junctions. Detecting motion inconsistencies is inspired by the classic work of Horn and Schunck [7] and the observation that the motion between the two sides of an occluding edge are generally dissimilar. This argument implicitly assumes that the objects being imaged undergo rigid transformations. This argument is inapplicable if the projection is instead allowed to transform in a more general fashion, as illustrated in Figure 1, and can result in false positives. The algorithms in this domain can be classified by the varying level of assumptions used in order to make the motion estimate robust. The presence of T-junctions in a contour has been shown to be a strong local cue for occlusion [8]. Unfortunately, not all T-junctions are occlusions, which can introduce false positives. Most algorithms in the T-junction domain can be classified according to the methodology they employ to detect and classify them.

At one extreme of motion estimation is the class of layered motion segmentation algorithms which employ a parametric model that is restricted to near-planar, rigidly-moving regions for each layer to segment regions based on the consistency of motion [9–12]. Incorporating a variety of techniques to estimate

these models, these algorithms assume a fixed number of layers in the scene, which does not scale well as the number of layers increases. Instead of relying on this requirement, we argue that the low-level reasoning done by an occlusion detector with a deformation model provides appropriate cues to high-level reasoning algorithms like those performing layered motion segmentation.

At the other extreme of motion estimators are those that make the motion estimate robust by smoothing the velocity field spatially or temporally [13, 14]. Regrettably, this has the unintended consequence of making the motion estimate inaccurate at boundaries where occlusions occur. An alternative to this smoothing approach is the use of an implicit model, either learned from local motion cues estimated from training data or based on some fixed model of the distribution of motion cues in the vicinity of occluding boundaries [15–18]. Though these approaches are appealing because they rely on well-defined statistical models, they remain sensitive to deviations of the actual data from the trained model.

T-junction detection has a rich history. Until recently, there have been two predominant approaches to T-junction detection: gradient or filter-based approaches [19–22] and model-based template matching [23]. These approaches work singularly to detect the T-junctions rather than distinguish an occluding T-junction from a non-occluding T-junction. More recently, others define what they call a proper T-junction as a T-junction at which an occlusion takes place [3]. They detect these proper T-junctions by exploiting a rank constraint on a data matrix of feature tracks that would normally be classified as outliers in a multiple-view geometry problem. Although mathematically correct, the method can be overly sensitive to even slight deviations from the given rank condition. Inspired by this work, other alternatives have exploited a discriminative framework to classify these proper T-junctions [1, 2]. Unfortunately, these methods utilize 2D spatiotemporal slices instead of volumes which mean that detections can only be made in fixed orientations.

In contrast to prior work, we prove that, under a deformation model, occlusions occur when pairs of images are not equivalent via deformation. We construct local topological invariants which exploit this result to localize occlusions in an image. Our method applies under weaker assumptions than the aforementioned detectors.

3   Modeling Scenes and Images

In this section, we describe our scene model. We let objects in $\mathbb{R}^3$ correspond to sets in the space. Each point on the surface of an object at a given time is assigned a color. We initially assume that the color at a given point on the surface of an object does not change over time, but we allow the object to deform via a homeomorphism. For simplicity, this model ignores lighting, shadows, and specularities while extensions to account for such effects are discussed in Section 4.2.

The motion of an object is determined via a continuous family of homeomorphisms: $F(x,t) : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$, where $F(\cdot, t)$ is a homeomorphism from $\mathbb{R}^3$ to

$\mathbb{R}^3$ for each t. The camera is located at the origin of our coordinate system. The **image domain**, Ω, is defined to be a sphere of radius 1 centered at the origin, $\mathbb{S}^2$. We employ an omni-directional camera model in order to avoid occlusions at the boundaries of the field of view. We consider the effect of boundaries and the case of directional cameras in the next section. Throughout our analysis, we assume for convenience that the camera is static and the world is moving. Fortunately, our analysis applies to situations where the camera and the environment are moving and changing simultaneously. We also assume objects in the scene remain outside the unit sphere at all times.

A **color image**, $\mathcal{I} : \Omega \to \mathbb{R}$, and **depth image**, $\mathcal{D} : \Omega \to \mathbb{R}^3$, are defined for every point $s \in \Omega$ via a ray drawn from the origin passing through s. We consider 1D color images (i.e. grayscale) for simplicity. The depth value assigned to s is obtained by finding the point in $\mathbb{R}^3$ at which the ray beginning at the origin through s first intersects. The color value at s is defined similarly. The set of **occluding contours** in an image is the set of points at which the depth image is discontinuous. The following result connects the homeomorphisms in $\mathbb{R}^3$ to homeomorphisms in Ω.

Proposition 1. *If there are no occluding contours for an ordered set of depth images indexed by t in $[0,1]$, then $\mathcal{D}_t(s)$ provides a homeomorphism between Ω and $\mathcal{D}_t(\Omega) \subset \mathbb{R}^3$ which implies that $f(s,t) := \mathcal{D}_t^{-1}(F(\mathcal{D}_t(s), t))$ is a continuous family of image homeomorphisms for which $\mathcal{I}_{t_1}(f(s,t_1)) = \mathcal{I}_{t_2}(f(s,t_2))$ for all t_1 and $t_2 \in [0,1]$.*

We refer to the existence of a family of continuous image homeomorphisms for which $\mathcal{I}_{t1}(f(s,t_1)) = \mathcal{I}_{t2}(f(s,t_2))$ as the **Image Homeomorphism Criterion**. If this criterion is violated, then using the previous theorem we conclude that an occluding contour exists. Though this argument guarantees the existence of an occlusion, it does not help us localize the occluding contour either temporally or spatially. The reader may wonder if the converse of Proposition 1 is valid. The following observation provides an important partial converse to the proposition:

Proposition 2. *If the Image Homeomorphism Criterion is satisfied by a set of color images, then there exists a realization of an object in $\mathbb{R}^3$ that generates the same color images with no occluding contours.*

One such realization corresponds to forming a sphere of radius 2 centered at the origin and coloring the sphere according to the color image. The motion homeomorphism $F(x,t)$ for $\mathbb{R}^3$ is then just the extension of the color image homeomorphism $f(s,t)$. This result verifies that the Image Homeomorphism Criterion is in fact the best achievable result to guarantee the existence of occlusions without making additional assumptions.

4   Localizing Occlusions

In this section, we introduce an approach to locally detect occlusions in image sequences over discrete time by extending the ideas in Section 3. This is done

by introducing an additional constraint on the size of deformations in $\mathbb{R}^3$. Then, we generalize the concepts for image sequences in which color information also varies. Finally, we focus on the case in which the image homeomorphisms can be decomposed into a translational and deformation component.

4.1   Local Detections Without Color Variation

To begin, we introduce a constraint on the size of deformations in $\mathbb{R}^3$:

Definition 1. *A family of homeomorphisms* $F(x,t) : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$ *is said to be* **Lipschitz** *if for all* $x \in \mathbb{R}^3$, t_1 *and* $t_2 \in \mathbb{R}$

$$||F(x,t_1) - F(x,t_2)|| \leq K|t_1 - t_2|$$

for some constant K *that is independent of* x, t_1 *and* t_2. *The smallest such* K *is called the* **Lipschitz constant**.

From now on, we require that the continuous family of homeomorphisms, F, that $\mathbb{R}^3$ transforms under be Lipschitz, and the Lipschitz constant, K, gives an upper bound on the size of these deformations. In practice, this requirement demands bounding the speed of objects in $\mathbb{R}^3$ based on the rate at which the camera captures images. If no occluding contours are present, then the induced image homeomorphism is also Lipschitz with the same constant, K, since all objects are required to remain outside of $\mathbb{S}^2$.

In order to verify if the Image Homeomorphism Criterion has been violated, we study changes to topological invariants of the set $\mathcal{I}^{-1}([a,b])$, where $[a,b] \subset \mathbb{R}$. In particular, we focus on the number of connected components. To illustrate the problem with naïvely comparing the number of connected components to detect local occlusions, consider the sets in Figure 3(a) and corresponding neighborhoods drawn in the rest of the figure. The first neighborhood, E_r, is a square with a side of length $2r$ drawn in Figure 3(b). The second neighborhood E_{r+K}
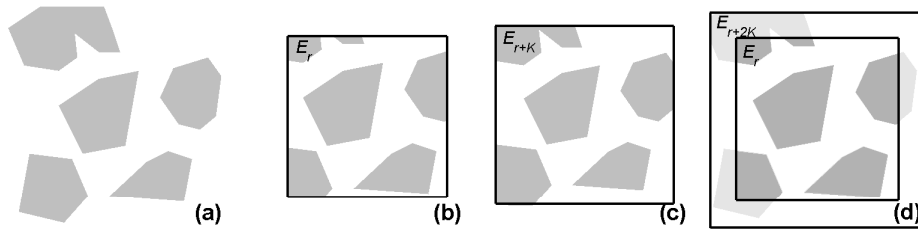
Fig. 3. Illustration of how to count connected components for neighborhoods E_r and E_{r+K}. (a) The original image before any window is applied. Counts of connected components: (b) 6 in E_r; (c) 5 in E_{r+K}; and (d) 5 in E_r after identification using E_{r+2K}. Without identification we would erroneously conclude that the Image Homeomorphism Criterion is violated and that there is an occlusion.

has a side of length $2(r + K)$ drawn in Figure 3(c). E_r has 6 connected components while E_{r+K} has 5 connected components suggesting that some set has disappeared and that the Image Homeomorphism Criterion has been violated. However, there are no occlusions in this instance. The problem arises because we count the same set twice in the smaller neighborhood. We remedy this problem by identifying sets in E_r that correspond to the same connected component in E_{r+2K}. This solution inspires the construction of a sort of local topological invariant.

Definition 2. *Given a color image $\mathcal{I}$ over the image domain Ω, two collection of intervals, $\mathcal{B} = \{[a_k, b_k]\}_{k=1}^{N_B}$ and $\mathcal{B}' = \{[a_k', b_k']\}_{k=1}^{N_B}$, **the histogram of connected components in a neighborhood** $E \subset \Omega$ **given the bins** $\mathcal{B}$ is defined as the vector $\alpha_E(\mathcal{I}|\mathcal{B}) = (\alpha_k)_{k=1}^{N_B}$, where*

$$\alpha_k = cc\left(\mathcal{I}^{-1}([a_k, b_k]) \cap E\right),$$

*and $cc(A)$ is the number of connected components in the set A. Under the assumption that $E \subset E'$ and $[a_k, b_k] \subset [a_k', b_k']$ for all k, **the histogram of connected components in E identified with the neighborhood E' given the bins $\mathcal{B}$ and $\mathcal{B}'$**, denoted $\alpha_{E|E'}(\mathcal{I}|\mathcal{B}, \mathcal{B}')$, is computed in the same way except the connected components in $\mathcal{I}^{-1}([a_k, b_k]) \cap E$ are identified (i.e. treated as the same connected component) if they correspond to the same component in $\mathcal{I}^{-1}[a_k', b_k']) \cap E'$.*

Definition 3. *The **color support in a neighborhood** E **given the bins** $\mathcal{B}$ is the vector $\sigma_E(\mathcal{I}|\mathcal{B}) = (\sigma_k)_{k=1}^{N_B}$, where σ_k is 0 if $\mathcal{I}^{-1}([a_k, b_k]) \cap E = \emptyset$ and 1 otherwise.*

Employing this new method to calculate connected components, guarantees that under Lipschitz image homeomorphisms, the histogram of connected components in a neighborhood E_r identified with the neighborhood E_{r+2K} is always less than the histogram of connected components in a neighborhood E_{r+K}. If we apply this procedure to the example in Figure 3 and compare the number of connected components, we find that the Image Homeomorphism Criterion is not violated (the number of connected components in E_r after identification is 5 which is the number of components in E_{r+K}). The following result proves that this argument can be used to identify local violations of the Image Homeomorphism Criterion, which allows us to define a local occlusion detector.

Theorem 1. *Given bins $\mathcal{B}$, neighborhoods E_r, E_{r+K}, and E_{r+2K} centered around a common point s, and color images $\mathcal{I}_{t_1}$ and $\mathcal{I}_{t_2}$ where $|t_1 - t_2| = 1$, if either*

$$\alpha_{E_r|E_{r+2K}}(\mathcal{I}_{t1}|\mathcal{B}, \mathcal{B}) \leq \alpha_{E_{r+K}}(\mathcal{I}_{t2}|\mathcal{B}) \tag{1}$$

or

$$\sigma_{E_{r+K}}(\mathcal{I}_{t2}|\mathcal{B}) \leq \sigma_{E_{r+2K}}(\mathcal{I}_{t1}|\mathcal{B}) \tag{2}$$

is violated, where the inequality is checked element wise, then the Image Homeomorphism Criterion is not satisfied between $\mathcal{I}_{t1} \cap E_r$ and $\mathcal{I}_{t2} \cap E_{r+K}$.

The argument for the proof proceeds as follows: For a scene deforming under a Lipschitz homeomorphism, if a set is in the interior of the neighborhood E_r, then it is in the interior of E_{r+K}. If two sets in E_r are connected in E_{r+K} then the path connecting them is in E_{r+2K} which explains why identification using E_{r+2K} guarantees the condition in Equation 1. The second condition just specifies that the colors observed in E_{r+K} are present in E_{r+2K}.

At this point, we make two additional remarks. First, the previous theorem allows us to consider occlusion detection in the case of directional cameras in a straightforward manner. Second, observe that Theorem 1 works both forward and backward in time so that appearance and disappearance events, which are each just types of occlusions, can be identified locally.

4.2 Generalizing to Color Variation

In this section, we generalize our model to include color variations in the images. These variations may include soft shadows and slow lighting variations. However, we do not claim to solve the problem for strong shadows and specularities which are a challenge for all occlusion detection algorithms. In order to quantify the amount of uncertainty allowed, we consider color variations that are Lipschitz over time with constant K_c. That is, if $f(s,t)$ is a family of image homeomorphisms, we must have

$$|\mathcal{I}_{t1}(f(s,t_1)) - \mathcal{I}_{t2}(f(s,t_2))| \le K_c|t_1 - t_2|$$

instead of $\mathcal{I}(f(s,t_1)) = \mathcal{I}(f(s,t_2))$ (i.e. $K_c = 0$) as was assumed in the previous section. From now on, we require that the color variations be Lipschitz with constant K_c.

We generalize the results of Theorem 1 to incorporate color variations.

Theorem 2. *Assume the same setup as in Theorem 1 with Lipschitz color images with Lipschitz constant K_c. Define $\mathcal{B}^c := \{[a_k - c, b_k + c]\}_{k=1}^{N_B}$ for $c > 0$. If either*

$$\alpha_{E_r|E_{r+2K}}(\mathcal{I}_{t1}|\mathcal{B}, \mathcal{B}^{2K_c}) \le \alpha_{E_{r+K}}(\mathcal{I}_{t2}|\mathcal{B}^{K_c}) \tag{3}$$

or

$$\sigma_{E_{r+K}}(\mathcal{I}_{t2}|\mathcal{B}^{K_c}) \le \sigma_{E_{r+2K}}(\mathcal{I}_{t1}|\mathcal{B}^{2K_c}) \tag{4}$$

is violated, then the Image Homeomorphism Criterion under color variation is not satisfied between $\mathcal{I}_{t1} \cap E_r$ and $\mathcal{I}_{t2} \cap E_{r+K}$.

Figure 4 illustrates this process where an object is moved behind a book resulting in an occlusion detection. The sets $\mathcal{I}_{t1}^{-1}([a_k, b_k]) \cap E_r$ are shown in white on the middle row for increasing k from left to right, the sets $\mathcal{I}_{t1}^{-1}([a_k - 2K_c, b_k + 2K_c]) \cap E_{r+2K}$ are also shown in gray on the middle row, and the sets $\mathcal{I}_{t2}^{-1}([a_k - K_c, b_k + K_c]) \cap E_{r+K}$ are shown on the bottom row. The images in this example are of size 240×240, $\mathcal{B} = \{[40(k-1), 40k]\}_{k=1}^{7}$, $r = 40$, $K = 10$, and $K_c = 5$. The outcome for this example is

$$\alpha_{E_r|E_{r+2K}}(\mathcal{I}_{t1}|\mathcal{B}, \mathcal{B}^{2K_c}) = [1\ 0\ 1\ 2\ 0\ 0\ 0]^\top$$
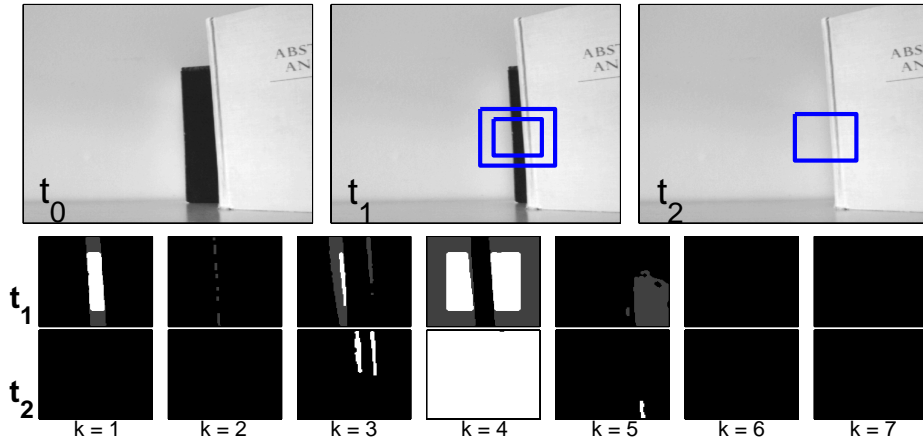
Fig. 4. Demonstration of our local topological occlusion detector. Three images corresponding to an object moving behind a book (top). The sets $\mathcal{I}_{t1}^{-1}([a_k, b_k]) \cap E_r$ and $\mathcal{I}_{t1}^{-1}([a_k - 2K_c, b_k + 2K_c]) \cap E_{r+2K}$ are shown in white and gray respectively (middle). The sets $\mathcal{I}_{t2}^{-1}([a_k - K_c, b_k + K_c]) \cap E_{r+K}$ are shown on the bottom row. Our framework detects an occlusion since the condition in Equation 3 is violated.

and

$$\alpha_{E_{r+K}}(\mathcal{I}_{t2}|\mathcal{B}^{K_c}) = [0\ 0\ 2\ 1\ 1\ 0\ 0]^\top.$$

Since the condition in Equation 3 is not satisfied, we conclude that an occlusion has occurred.

4.3   Estimating Translational Component

Theorem 2 gives a mechanism to detect occlusions in situations in which the motions and deformations of objects in 3D are unknown. In certain situations, it may be convenient to take advantage of this structure and decompose the homeomorphism into a translational and pure deformation component.

In this section, we assume that the image homeomorphism can be locally decomposed as follows:

$$f(s,t) = f_t(s,t) + f_d(s,t),$$

where f_t is a translation and f_d is a deformation with Lipschitz constants K_t and K_d, respectively. Using the framework developed in the previous section, we would need to compare the connected components in E_r and $E_{r+K_t+K_d}$. If $K_t >> K_d$ then $E_{r+K_t+K_d}$ would be a large set which would decrease the utility of our algorithm to detect occlusions.

In order to take advantage of our knowledge about the translation components of the homeomorphism, we would like to split $E_{r+K_t+K_d}$ into N_t^2 evenly spaced regions that cover $E_{r+K_t+K_d}$. The diameter of the decomposed regions

needs to be sufficiently larger in order to guarantee that the conditions in Theorem 2 are fulfilled by at least one of the subregions whenever the Image Homeomorphism Criterion is satisfied. Figure 5 illustrates this situation. In order to guarantee that a deformed neighborhood E_r is found in the interior of at least one of the decomposed regions, an overlap of greater than $2(r + K_d)$ between the regions is required. If we let $2(r + D)$ be the length of the side of the decomposed regions, then the minimum length required corresponds to:

$$D = K_d + \frac{K_t}{N_t}. \tag{5}$$

Solving for the spacing d required between centers of the regions, we obtain

$$d = \frac{2}{N_t} K_t. \tag{6}$$

Hence, given that we have decomposed the neighborhood $E_{r+K_t+K_d}$ into N_t^2 regions of length $2(r + D)$ with centers spaced d units away, then the Image Homeomorphims Criterion between regions E_r at time t_1 and $E_{r+K_d+K_t}$ at time t_2 is violated if the conditions in Equations 3 and 4 are not satisfied by any of decomposed regions. This result is used in Section 5 to identify occlusions between objects with large translational components. Note that this decomposition approach can be used to estimate motion flow between regions without direct tracking or differential operators applied to the images.
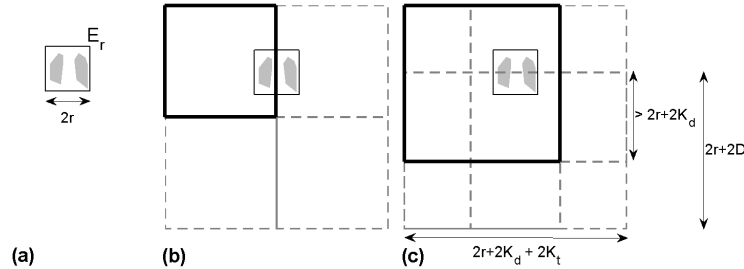
Fig. 5. Choosing an appropriate length for the coveraging sets of $E_{r+K_t+K_d}$ with $N_t = 2$. (a) Given a region E_r, and bounds K_t and K_d, our objective is to obtain a cover that guarantees that E_r can be found within one of the covering sets. (b) A coverage in which E_r is not contained within any of the covering sets since the coverage does not have enough overlap. The dark square corresponds to one of the regions in the coverage. (c) Given that the overlap between the region is greater than $2r + 2K_d$ then the neighborhood E_r can be found in at least one of the covering sets. The length of the sets needed to guarantee this fact are labeled $2r + 2D$.

5   Experiments

In this section, we present experimental results for occlusion detection applying Theorem 2 to make local detections. We also briefly consider how to utilize these results to perform foreground segmentation.

5.1   Implementation

The implementation of our approach takes the following image deformation parameters: the Lipschitz constants K_c, K_t and K_d. It also requires the following algorithmic parameters: the radius for the base regions r, the color bins $\mathcal{B}$, and the number of subsections N_t in which each local neighborhood is decomposed. For simplicity we take N_t to be odd. Note that any choice of the algorithmic parameters is appropriate (e.g. different choices of radius r may yield different detections, but still no false positives). The algorithm takes a base image $\mathcal{I}_{t1}$ and marks detections against image $\mathcal{I}_{t2}$.

We define a grid of points evenly spaced by a distance d (as defined by Equation 6). For each of the points in $\mathcal{I}_{t1}$ we compute $\alpha_{E_r|E_{r+2D}}(\mathcal{I}_{t1}|\mathcal{B}, \mathcal{B}^{2K_c})$ and $\sigma_{E_{r+2D}}(\mathcal{I}_{t1}|\mathcal{B}^{2K_c})$, where D is given by Equation 5. For each of the points in $\mathcal{I}_{t2}$ we compute $\alpha_{E_{r+D}}(\mathcal{I}_{t2}|\mathcal{B}^{K_c})$ and $\sigma_{E_{r+D}}(\mathcal{I}_{t2}|\mathcal{B}^{K_c})$.

For a fixed location x in the grid, let E_r be the neighborhood centered at x in image $\mathcal{I}_{t1}$ and let $E_{r+K_t+K_d}$ be the neighborhood centered at x in image $\mathcal{I}_{t2}$. We test the conditions in Theorem 2 by comparing the histogram of connected components of E_r against each of the histograms from the resulting N_t^2 regions in which $E_{r+K_t+K_d}$ is decomposed. If the conditions are not satisfied by any of the regions then position x at time t_1 is marked as an occluded location.

5.2   Detecting Occluding Contours

To begin, we consider results on real images. Figures 2 and 6 illustrate the results of applying our algorithm on a variety of image sequences: a deforming cloth, a walking person, a closing hand, and a folding colored Macbeth board. The first three columns correspond to frames from the sequence and the last column is the detection results corresponding to the frame at time t_1. Animations of the image sequences and the detected occlusions can be found at: http://www.cs.unc.edu/~ron/research/ECCV2010/

See the supplementary materials for animations of the image sequences and the detected occlusions.

Our method successfully detected occlusions without introducing any false positives. Note that several occluding contours were not highlighted in our detections due to our unconstrained assumptions about the scene (i.e. we made no prior assumptions about the environment and allowed for any type of deformations). To illustrate this point, consider the image sequence with the closing hand (second row in Figure 6). Though there are occluding contours along the edges near the palm of the hand, the hand's movement does not reveal the existence of any local occlusions here which means that the Image Homeomorphism
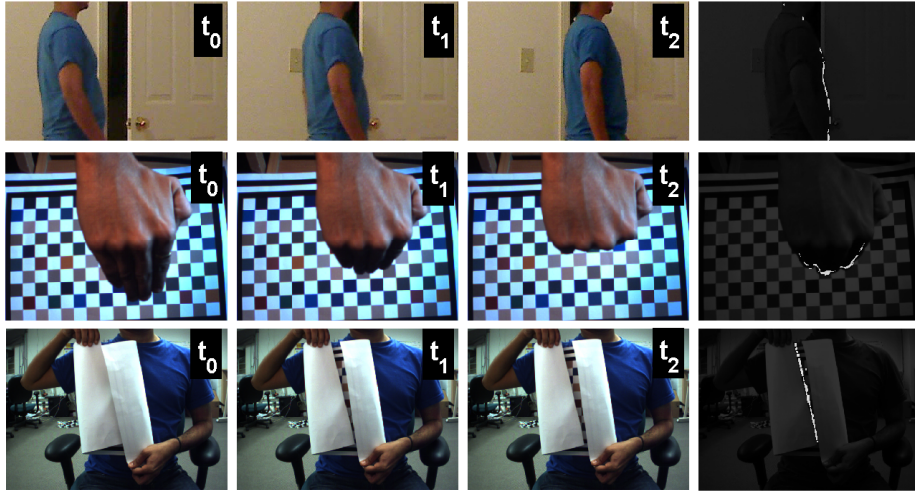
Fig. 6. Detection results for image sequences of a walking person, a closing hand, and a folding colored Macbeth board. Sample frames are displayed (first three columns) and the occluding contours corresponding to the frames at time t_1 (right column).

Criterion is never violated. Proposition 2 implies that there exists a 3D realization of each image in this sequence around this edge that does not contain an occlusion. By applying more global reasoning, one could hope to recover these type of detections, which is the focus of future research.

The method presented in this paper is not directly comparable to other approaches in the literature since our goal is to obtain local detections in unconstrained deforming scenes, an area that has not previously been explored. In future work, we plan to integrate local detections into consistent occluding contours for deforming scenes, which requires a new dataset of deforming scenes for evaluation and comparison of approaches.

5.3   Foreground/Background Segmentation

In this section, we briefly consider how one can employ the presented occlusion detector to do foreground versus background segmentation. We assume that there are two objects each with distinguishable color distributions, one performing the occluding (the foreground) and the other being occluded (the background). When an occlusion occurs, the neighborhood E_r contains samples of a set that becomes occluded and the neighborhood E_{r+K} contains samples of the set that perform the occlusion. We can use this elementary information to learn the color distribution of the foreground and background. After this distribution has been learned, we can test to which segment a given pixel belongs. Figure 7 illustrates this approach applied to a synthetic (top row) and real (bottom row) image sequences.
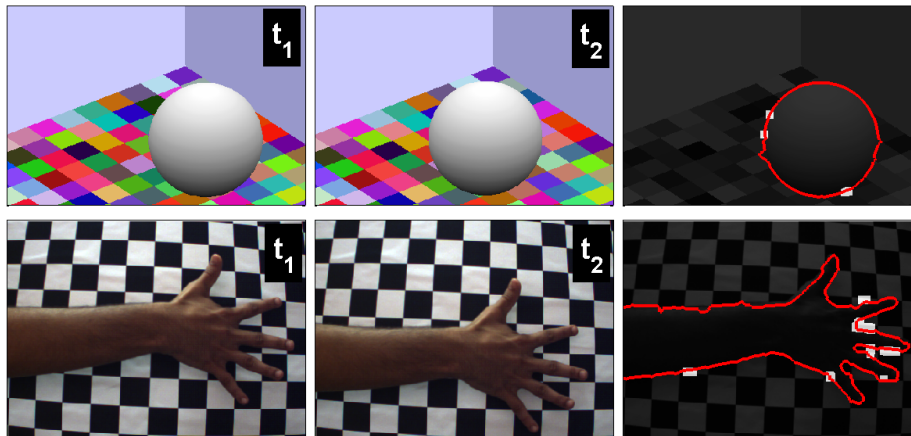
Fig. 7. Foreground/background segmentation results for a synthetic sequence of a ball moving through a room with multicolored tiles (top) and a real sequence of a hand moving in front of a checkerboard (bottom). Sample frames from the sequences (first two columns). Occlusion detections in white and segmentation by learning color distribution in red (right column) centered at time t_1.

6   Conclusion

In this paper, we present a mathematical framework to detect and localize occlusions in image sequences of scenes that can include deforming objects. The method works by measuring changes in a local topological invariant, which guarantees a zero false positive rate when certain motion and color variation bounds are satisfied. Our occlusion detector works under far weaker assumptions than other detectors. If the 3D scene transforms in a more restrictive fashion, the method presented in this paper can be viewed as complementary to traditional detectors. We also presented preliminary results on extending the detections to perform figure-ground separation when the model undergoes Lipschitz deformations. Most current such algorithms employ a fixed statistical model for the variation allowed in the background, but the framework presented here is more general and can work in tandem with a statistical model.

As future research, we plan to study how to integrate this local information in order to come up with global solutions to problems such as segmentation and matching. The descriptors that we used for identifying occlusions can also be thought of as topological features which are robust to deformations. Integrating the information from these local descriptors could lead to the development of new matching and recognition techniques.

The generality of our framework can benefit applications such as tracking in medical applications which involve soft, deformable tissues. For example, knowledge of occlusions could help in the reconstruction of surgical scenes and in performing foreground/background segmentation in scenes with soft tissue.

References

1. Apostoloff, N., Fitzgibbon, A.: Learning Spatiotemporal T-Junctions for Occlusion Detection. In: IEEE Conf. on Computer Vision and Pattern Recognition. (2005)
2. Apostoloff, N., Fitzgibbon, A.: Automatic video segmentation using spatiotemporal T-junctions. In: British Machine Vision Conf. Volume 3. (2006) 1089
3. Favaro, P., Duci, A., Ma, Y., Soatto, S.: On exploiting occlusions in multiple-view geometry. In: IEEE Intl. Conf. on Computer Vision. (2003) 479–486
4. Irani, M., Rousso, B., Peleg, S.: Computing occluding and transparent motions. Intl. J. of Computer Vision **12** (1994) 5–16
5. Niyogi, S., Adelson, E.: Analyzing and recognizing walking figures in XYT. In: IEEE Conf. on Computer Vision and Pattern Recognition. (1994) 469–474
6. Stein, A., Hebert, M.: Local detection of occlusion boundaries in video. Image and Vision Computing (2008)
7. Horn, B., Schunck, B.: Determining Optical Flow. Artificial Intelligence **17** (1981) 185–203
8. Biederman, I.: Recognition-by-components: A theory of human image understanding. Psychological Review **94** (1987) 115–147
9. Ogale, A., Fermuller, C., Aloimonos, Y.: Motion segmentation using occlusions. IEEE Trans. on Pattern Analysis and Machine Intelligence **27** (2005) 988–992
10. Smith, P., Drummond, T., Cipolla, R.: Layered motion segmentation and depth ordering by tracking edges. IEEE Trans. on Pattern Analysis and Machine Intelligence **26** (2004) 479–494
11. Xiao, J., Shah, M.: Accurate motion layer segmentation and matting. In: IEEE Conf. on Computer Vision and Pattern Recognition. Volume 2. (2005)
12. Yin, P., Criminisi, A., Winn, J., Essa, I.: Tree-based classifiers for bilayer video segmentation. In: IEEE Conf. on Computer Vision and Pattern Recognition. (2007)
13. Anandan, P.: A computational framework and an algorithm for the measurement of visual motion. Intl. J. of Computer Vision **2** (1989) 283–310
14. Black, M., Anandan, P.: Robust dynamic motion estimation over time. In: IEEE Conf. on Computer Vision and Pattern Recognition. (1991) 296–302
15. Black, M., Fleet, D.: Probabilistic detection and tracking of motion discontinuities. Intl. J. of Computer Vision **38** (2000) 231–245
16. Fleet, D., Black, M., Nestares, O.: Bayesian inference of visual motion boundaries. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA (2003)
17. Nestares, O., Fleet, D.: Probabilistic tracking of motion boundaries with spatiotemporal predictions. In: IEEE Conf. on Computer Vision and and Pattern Recognition. Volume 2. (2001) 358–365
18. Stein, A., Hoiem, D., Hebert, M.: Learning to Find Object Boundaries Using Motion Cues. In: IEEE Intl. Conf. on Computer Vision. (2007) 1–8
19. Beymer, D.: Finding Junctions Using the Image Gradient. Massachusetts Institute of Technology, Artificial Intelligence Laboratory (1991)
20. Freeman, W., Adelson, E.: The design and use of steerable filters. IEEE Trans. on Pattern Analysis and Machine Intelligence **13** (1991) 891–906
21. Li, D., Sullivan, G., Baker, K.: Edge detection at junctions. In: Alvey Vision Conf. Volume 2. (1989)
22. Perona, P.: Steerable-Scalable Kernels for Edge Detection and Junction Analysis. In: European Conf. on Computer Vision, Springer-Verlag London, UK (1992) 3–18
23. Parida, L., Geiger, D., Hummel, R.: Junctions: detection, classification, and reconstruction. IEEE Trans. on Pattern Analysis and Machine Intelligence **20** (1998) 687–698