Matching Images Under Unstable Segmentations

Varsha Hedau[1], Himanshu Arora[2], Narendra Ahuja[3]
[1,3] *ECE Department, University of Illinois, Urbana, IL-61801, USA.* {vhedau2,n-ahuja}@uiuc.edu
[2] *Objectvideo Inc., Reston, VA-20191, USA.* harora@objectvideo.com

Abstract

Region based features are getting popular due to their higher descriptive power relative to other features. However, real world images exhibit changes in image segments capturing the same scene part taken at different time, under different lighting conditions, from different viewpoints, etc. Segmentation algorithms reflect these changes, and thus segmentations exhibit poor repeatability. In this paper we address the problem of matching regions of similar objects under unstable segmentations. Merging and splitting of regions makes it difficult to find such correspondences using one-to-one matching algorithms. We present partial region matching as a solution to this problem. We assume that the high contrast, dominant contours of an object are fairly repeatable, and use them to compute partial matching cost (PMC) between regions. Region correspondences are obtained under region adjacency constraints encoded by Region Adjacency Graph (RAG). We integrate PMC in a many-to-one label assignment framework for matching RAGs, and solve it using belief propagation. We show that our algorithm can match images of similar objects across unstable image segmentations. We also compare the performance of our algorithm with that of the standard one-to-one matching algorithm on three motion sequences. We conclude that our partial region matching approach is robust under segmentation irrepeatabilities.

1. Introduction

The regions of an object obtained by many segmentation methods in images taken under slightly different imaging conditions (*e.g*. lighting, viewpoint, *etc*.) differ significantly. This is because the existing segmentation algorithms do not always produce perceptually meaningful regions. This results in splitting and mergers of adjacent regions, even if the contrast changes very slightly. On the other hand, in conditions where stable regions can be reliably obtained, for instance, by using additional information such as, shape cues as in [20], the efficacy of region based matching has been successfully demonstrated for high level vision tasks such as object recognition. Despite their descriptive power, regions have not been a very popular choice for computer vision tasks requiring image matching, partly because of the lack of segmentation algorithms which produce repeatable regions.

Most of the existing algorithms for region correspondences employ both the absolute region properties *e.g*. shape, area, mean intensity *etc*., as well as relational constraints *e.g*. adjacency (regions sharing boundary) and hierarchy, encoded in graph structures *e.g*. *Region Adjacency Graphs* (RAGs) [17, 20] (encoding adjacency), or segmentation trees [1, 2, 13] (encoding hierarchy). The splits and mergers of regions corresponding to the same scene considerably affect both *(i)* the regions' absolute properties, and *(ii)* their relationships with other regions in the image. Consider for instance two example segmentations of the same object, shown in Fig. 1. The corresponding RAGs are overlaid on the top of segmented images (middle column). Here region b in the top image is a merger of regions $b1$ and $b2$ in the bottom image. In the ideal case, b should be matched to both $b1$ and $b2$. However, absolute properties of $b1$ and $b2$ differ significantly as compared to their merger b therefore the unary match cost of b is high with both b_1 and b_2. Also note that the adjacency relationships between regions change considerably, *e.g*., $b2$ in the bottom image is not adjacent to a, but b in the top image, which is the merger containing $b2$, is adjacent to a. This causes irrepeatability in the structure of (RAGs) of the two images, commonly referred to as structural noise. Matching b with its fragments $b1$ and $b2$ is thus difficult by i) any algorithm that outputs a one-to-one matching between regions, or ii) any algorithm that uses only the unary match cost based on regions' absolute properties. This paper is aimed at providing a solution for both of the above problems. It proposes a novel cost measure between regions, one that depends on the properties of the portion of overlap between regions after transforming them appropriately and not the properties of entire regions. We denote this measure by *Partial Match Cost* (PMC). Also, the paper integrates this cost measure into a many-to-one region matching framework.

Prior Work. One solution to the segmentation instabil-

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

ity problem is allowing access to region segmentations at multiple scales, e.g., contrast levels, [1, 2, 13]. However, even searching all possible scales of segmentation does not guarantee perfect repeatability. Similar observation is reported in [11], wherein the authors use regions obtained from multiple segmentations, at multiple scales, and determine if these regions reliably correspond to the spatial support of natural objects. They conclude that even with large number of regions available per image, explicit mergers of regions are still required to obtain desirable levels of support. Since mergers lead to combinatorial complexity, using regions under such conditions becomes computationally prohibitive quickly. An approach that would allow partial matching of regions, so that the split portions could match the whole, is therefore desirable.

The problem of structural noise in graphs has been addressed before by relaxing the enforcement of relational constraints between regions. For instance, instead of using the direct parent-child constraints between regions while matching segmentation trees, Torsello *et al*. [16] suggest using transitive closures in trees which connect every node in the tree to every descendant. Thus any ancestor-descendant pair in one tree can match with any such pair in the other, even if their levels are different. As another example, the technique in [17] uses a fixed number of regions with nearest centroids, instead of directly adjacent regions while matching RAGs. However, even with the relaxed constraints, the assumption of one-to-one correspondence is often found to be too restrictive for real images.

Another class of approaches perform explicit region mergers, to deal with problem of high unary match cost between absolute region properties caused by region merging and splitting [7, 8, 13, 17]. These are guided by some heuristically set thresholds, e.g., those depending on intensity contrast between two regions. However, such mergers may not produce perfect one-to-one correspondences. Keselman *et al*. [9] obtain a model RAG, referred to as Least Common Abstraction (LCA), using RAG of exemplar images of a particular object category. Finding LCA requires search for isomorphic RAGs among those obtained by all possible region mergers in the original RAGs. Matching explicit splits and mergers of regions as performed by the above techniques can quickly become computationally prohibitive due to their combinatorial nature. The partial region matching approach proposed in this paper can reduce this combinatorial search space drastically. Another class of previous approaches is that of inexact graph matching [18], wherein graph edit operations such as node (or edge) substitution, deletion and insertions are defined to handle the inconsistencies in graph structure. Inexact graph matching is an n-p complete problem and is computationally intractable for real images, since the space of all possible edit operations is huge. A number of approximate solutions have been proposed in literature; however, the convergence in these cases depends heavily upon initialization.

Overview of our approach. We provide a solution to deal with the irrepeatibility of regions caused by merging and splitting, and noise in graph structure, while matching region adjacency graphs. We define the match score between two regions in a robust fashion, which allows fragments of the same region to have low unary match cost with the original region. We denote this cost as the Partial Match Cost (PMC). Towards computing PMC, we assume that the dominant contours of object are repeatable under different imaging conditions. In this paper, we consider adjacency relationships between regions contained in RAG. However note that PMC is generic enough to be integrated with other types of relations, *e.g*. containment, as observed in segmentation hierarchies (trees). To handle the structural noise in RAGs, we formulate the problem of RAG matching as a many-to-one label assignment problem, *i.e*., allowing for multiple regions from one RAG to match with a single region in the other RAG. We solve this assignment problem using the loopy belief propagation technique. Our algorithm can thus match fragments with their merger region, while respecting the adjacency constraints. In Fig. 1, third column, we show example results of our algorithm. Notice that our algorithm can handle both of the problems due to segmentation noise, mentioned above. This paper has the following **contributions**: 1) It uses partial region matching which is a robust similarity measure between regions. This eliminates the need for performing a combinatorially large number of explicit region mergers, which is otherwise necessary to deal with irrepeatibility of regions observed under unstable segmentations. 2) This paper also proposes an effective solution of many-to-one label assignments for the problem of structural noise in RAGs, which is possible due to our definition of partial match cost.

The rest of the paper is organized as follows: In Sec. 2, we explain our partial region match cost. In Sec. 3, we explain our label assignment formulation for RAG matching. Experiments and results are described in Sec. 4. Finally, we conclude the paper in Sec. 6.

2. Partial Region Matching

Since the global properties of regions change drastically among segmentations of the same scene under slightly different conditions, (*e.g*., lighting, viewpoint, *etc*.), they can not be used to obtain a reliable match between the segmentations. The properties of regions should instead be compared in the portions common to the two regions. Consider for instance the regions b and $b1$ shown in Fig. 1, corresponding to the torso and shirt of the man. Instead of comparing their global properties, if we compare the properties only in the common physical portion of these regions, *i.e*., shirt, we will be able to assign low cost to a match between

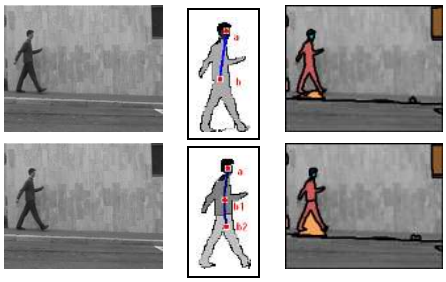


Figure 1. Images in the first column show adjacent frames of a video sequence. Second column shows the *region splitting* phenomenon and the changed RAG structure due to it. We show only a part of RAG for better visibilty. Third column shows some of the matched regions obtained by our algorithm. Matched regions are shown with the same color. Notice that *shirt* and *pant* fragments in the bottom image are successfully matched to the *torso* fragment in the top image.

these regions. We denote the match cost obtained in such a manner as Partial Match Cost (PMC). As we shall next explain, we hypothesize that deduction about such a common portion between the regions in different images can be obtained as a transformation $\mathbb{T}$, by robustly registering their contours. For two regions, $\boldsymbol{r}_1$ in the first image and $\boldsymbol{r}_2$ in the second image, and the transformation $\mathbb{T}$ between them, we can then define overlapping portions as $\boldsymbol{o}_1 = \boldsymbol{r}_1 \cap \mathbb{T}(\boldsymbol{r}_2)$ and $\boldsymbol{o}_2 = \boldsymbol{r}_2 \cap \mathbb{T}^{-1}(\boldsymbol{r}_1)$, in the domains of the first and second images, respectively. We define the PMC between $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ as

$$PMC(\boldsymbol{r}_1, \boldsymbol{r}_2) = ||f(\boldsymbol{o}_1) - f(\boldsymbol{o}_2))|| \qquad (1)$$

where $f(r)$ denotes the properties of region r and $||\cdot||$ is the euclidian norm. Since Eq. (1) computes the cost of matching overlapping portions of registered regions, assuming that the transformation between regions has been computed correctly, it will always return low cost for matching a region with one of its fragments (e.g. b with $b1$ in Fig. 1). Below we explain our methodology for estimating the transformation $\mathbb{T}$.

2.1. Region transformation

We assume that dominant contours corresponding to physical boundaries of different parts of an object are fairly repeatable. Due to the presence of a few spurious pixels on the boundaries, the originally closed regions of the object leak out to form big regions. It is this leakage that is mostly responsible for irrepeatability of regions across images of the same object. We therefore want to exploit the parts of the region boundaries which are undisturbed by the image noise, to estimate the transformation $\mathbb{T}$ between regions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. Let us denote the boundaries of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ by $\boldsymbol{bo}_1 = \{\boldsymbol{x}_1^1, \ldots, \boldsymbol{x}_k^1\}$ and $\boldsymbol{bo}_2 = \{\boldsymbol{x}_1^2, \ldots, \boldsymbol{x}_m^2\}$ where

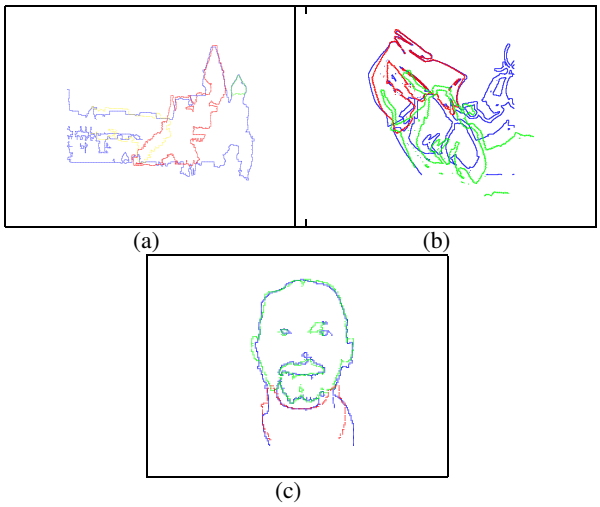


Figure 2. **KDC Registration.** Boundaries of regions from the first image are shown in blue and the transformed boundaries of corresponding fragments in the second image are overlaid on the top in different colors. **(a)** Regions from image pair in Fig. 3(c) are accurately registered despite huge scale variations. **(b)** Regions from image pair in Fig. 3(b) are accurately registered under affine transformations. **(c)** The *face-neck* region is successfully registered to *face* and *neck* regions (see Fig. 3(a)).

$\boldsymbol{x}_i^1$, $i \in \{1, \ldots, k\}$ and $\boldsymbol{x}_j^2$, $j \in \{1, \ldots, m\}$ are boundary pixels of regions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ respectively. Our goal is to align pieces of boundaries which are common in the two regions. Moreover, in order to successfully align fragments of a region with the whole region, we must do this alignment in a manner such that it is robust to outliers. There has been a much towards robustly estimating a transformation between point clouds in such a correspondenceless setting [6, 14, 19]. We adopt robust kernel density correlation (KDC) machinery, described in [15], for solving this problem. According to [15], the problem of robustly matching point clouds is formulated as that of matching the respective kernel density (also known as parzen window) estimates [5] of the samples in each of these point clouds. KDC is used as a measure of match between these kernel density estimates. The equivalence between kernel density estimators and robust M-estimators was shown in [3], and it is from here that KDC inherits its robustness to outliers.

Following the formulation in [15], we assume that the point sets $\boldsymbol{bo}_1$ and appropriately transformed $\boldsymbol{bo}_2$ are generated from the same underlying pdf. The quality of match between $\boldsymbol{bo}_1$ and $\mathbb{T}(\boldsymbol{bo}_2)$ is then given by the correlation of the corresponding estimates of the underlying pdf, given by the following equation, as described in [15].

$$KDC(\boldsymbol{bo}_1, \mathbb{T}(\boldsymbol{bo}_2)) = \sum_{\substack{\boldsymbol{x}_i \in \boldsymbol{bo}_1 \\ \boldsymbol{x}_j \in \boldsymbol{bo}_2}} K(H^{-1}(\boldsymbol{x}_i - \mathbb{T}(\boldsymbol{x}_j))) \quad (2)$$

where, K is the kernel function with zero mean, unit area

and identity covariance matrix, and H is a non singular bandwidth matrix. We use Gaussian kernel in our experiments. Maximizing KDC measure in Eq. (2) gives us the $\mathbb{T}$ between two regions. Given an appropriate initialization, the maximizer of Eq. (2) can be estimated efficiently in an iterative fashion by using variational optimization machinery, as shown in [15]. In our implementation, we restrict $\mathbb{T}$ to be an affine transformation. We discuss the initialization in Sec 4.

Fig. 2 shows some example results of our region transformation estimation using KDC. Boundary of the region in first image is shown in blue. The corresponding region of the object was broken into fragments in the segmentation of the second image. See segmentation images in second row in Fig. 3. We have shown the boundaries of the fragment regions overlaid on the top of the original region after transforming them using the computed transformation, with different colors. Notice how each fragment from second image is robustly registered on the single merger region in the first image.

3. Matching Region Adjacency Graphs

We abstract the segmentation of each image as a Region Adjacency Graph (RAG) constructed in the following fashion. For an image I, we define an RAG $G = (\mathcal{P}, \mathcal{N})$ consisting of a set of nodes $\mathcal{P}$ and a set of edges $\mathcal{N} \subset \mathcal{P} \times \mathcal{P}$. Each node $p \in \mathcal{P}$ represents a region in the image. An edge $(p, q) \in \mathcal{N}$ exists if regions p and q share a common boundary (are adjacent), i.e., $p \sim q$. Towards matching RAGs of two images, $\mathcal{G}_1 = (\mathcal{P}, \mathcal{N}_1)$ and $\mathcal{G}_2 = (\mathcal{L}, \mathcal{N}_2)$, we want a correspondence between their nodes which has a low partial match cost, and preserves adjacency relations. We also want this correspondence space to include many-to-one matches since a node in one RAG might exist as multiple fragments in the other RAG. We thus define RAG matching as the minimum unary cost labeling or mapping $l : \mathcal{P} \rightarrow \mathcal{L}$ satisfying the constraints that $\forall (p, q) \in \mathcal{N}_1$, we have either $(l(p), l(q)) \in \mathcal{N}_2$ or $l(p) = l(q)$. For convenience, we use the notation $l_p = l(p)$. Note that the constraints on the labeling preserve adjacency by allowing adjacent nodes in $\mathcal{G}_1$ to take labels corresponding to adjacent nodes in $\mathcal{G}_2$ and can produce many-to-one matches by allowing adjacent nodes to take the same label. Following this discussion, we can write the energy associated with a labeling l as

$$E(l) = \sum_{p \in \mathcal{P}} D_p(l_p) + \sum_{p,q \in \mathcal{N}_1} V(l_p, l_q) \tag{3}$$

Here, $D_p(l_p)$ is the unary cost of matching nodes $p \in \mathcal{P}$ and $l_p \in \mathcal{L}$ computed as partial match cost using Eq. (1), *i.e.*, $D_p(l_p) = PMC(p, l_p)$. The term $V(l_p, l_q)$ is the cost of violating the constraints, defined as

$$V(l_p, l_q) = \begin{cases} 0 & \text{if} \quad l_p, l_q \in \mathcal{N}_2 \text{ or } l_p = l_q \\ K & \text{if} \quad l_p, l_q \notin \mathcal{N}_2 \end{cases} \tag{4}$$

Here, K is the constraint violation penalty and is set to a high value. Thus the minimizer of $E(l)$ gives an optimum assignment in terms of both the unary cost and adjacency relations.

Estimating a global minimum of the energy function in Eq. (3) requires a search over the discrete space of all possible labelings. However, efficient iterative techniques based on message passing algorithms [12] and graph cuts [10] have been proposed in literature to estimate strong local minimum of such an energy function. We use loopy belief propagation (LBP), which is a message passing algorithm, to solve the minimization in Eq. (3). In LBP, each node iteratively receives messages from its neighbors in the graph, and updates its own confidence of having a certain label. Once the messages stop changing beyond a threshold, the iterations are terminated, and each node computes the resulting belief over all the labels. The per node label is then obtained by maximization over these individual beliefs. If convergent, LBP has proved to be a good approximation for many applications [12]. In our experiments, LBP converges in approximately 50 iterations. For our formulation, we can write the message passed by node p to its neighbor q for label l_q at iteration t as

$$m^t_{p \rightarrow q}(l_q) = \min_{l_p} \left(V(l_p, l_q) + D_p(l_p) + \sum_{\substack{s \neq q \\ (s,p) \in \mathcal{N}_1}} m^{t-1}_{s \rightarrow p}(l_p) \right)$$

After T iterations of message passing, the belief vector at node q for label l_q is computed as

$$b_q(l_q) = D_q(l_q) + \sum_{\{(p,q) \in \mathcal{N}_1\}} m^T_{p \rightarrow q}(l_q) \tag{5}$$

Finally, the label l^*_q that minimizes $b_q(l_q)$ at node q is selected.

4. Experiments and Results

We test the performance of the proposed PMC and many-to-one matching approach with two different experiments: (a) matching images of similar objects taken under different imaging conditions, and (b) matching images of a motion sequence. For image matching, we present qualitative results on image pairs with varying degrees of complexities, validating the ability of our algorithm towards matching a region with its fragments, and many-to-one region matching (see Sec. 4.1). For matching images of motion sequences, we present both qualitative and quantitative results. On these sequences, we compare the performance of

our algorithm against its one-to-one counterpart which does not use PMC (explained in Sec. 4.2).

For all the experiments, image segmentation is performed using the mean-shift clustering approach [4]. For each image pair, we use the same scale parameters for the mean-shift segmentation. However, in order for our assumption of having repeatable contours in both images being valid, we pick a scale which preserves dominant contours in segmentation output of the two images. As we shall see in the results, there still are numerous examples of region irrepeatabilites making the region correspondence task really difficult. This selection of the scale parameter can be completely avoided by taking regions from segmentation outputs obtained at multiple scales. Note that we do not separately tune the scale parameters for each image in the pair to get similar regions. We set the value of constraint violation cost K in Eq. (4) to 100 and the region property function f in Eq. (1) is chosen as gray scale absolute intensity difference at each pixel. Further, we assume that the transformation to be estimated using KDC, $\mathbb{T}$, is an affine transformation. KDC being an iterative approach requires a good initialization to search for the global optimal. For image matching experiments, we use multiple initializations, obtained by matching local appearances at corner points on the image pairs. For matching images of a motion sequence, we initialize the transformation as an identity transformation.

4.1. Matching image pairs

We show the results of our RAG matching algorithm on a set of image pairs consisting of similar objects, captured under different conditions. Each image is represented as an RAG, and PMC is computed between every region pair consisting of a region from the first image and a region from the second image. Region correspondences are then obtained by matching RAGs using the algorithm described in Sec 3. Note that our original formulation for label assignment is many-to-one, however, we re-run the algorithm by reversing the order of images to obtain many-to-many correspondences. For each set of images in Fig. 3, the original images are shown in the top row, the segmentations in the middle row, and, the matched regions are shown in the same colors in the bottom row. The nodes which are not matched are shown in black.

Fig. 3(a) shows the matching of face images of the same person on different backgrounds. Here, the regions corresponding to face, neck and right collar in the left image fuse together to form a single *face-neck-collar* merger in the right image. Note that our formulation is able to find correspondences between this merger and its individual fragments. This shows the robustness of PMC based unary cost and the ability of our RAG matching algorithm towards handling many-to-one matches. Also note that the regions neighboring the *face-neck-collar* portion in the images still get matched appropriately despite their significantly different neighborhoods. Thus, our algorithm can also handle structural noise in RAGs. Fig. 3(b) shows matching results on images of a graffiti which have complex adjacency structure. Matching these images using any global region match measure or one-to-one RAG matching methodology would be very difficult, due to huge differences in segmentations in terms of mergers and splits of regions, as well as the adjacency structure. However, our matching framework is able to handle this successfully. Note the partial matching for the regions shown in blue, where the corresponding region in the second image occupies less than half of the area of the region in the first image (see the corresponding registrations in Fig. 2). Also note that despite of instances of structural noise in the graphs, we are able to match them correctly. For instance, the regions colored in blue and orange are adjacent in the first image, but not in the second, and yet they are matched correctly.

Fig. 3(c) and (d) show matching results on images of the Kremlin and the Statue of Liberty, respectively. Note that the Kremlin appears as a single region in one image and is fragmented into multiple regions in the second image. Our algorithm is able to match these fragments with the entire region. Many of these fragments, as shown in Fig. 2(a), are successfully registered to the single Kremlin region, providing a low match cost between these corresponding regions. A similar effect can be seen for the Statue of Liberty image pair.

Fig. 3(e) and (f) show matching results for rhino and horse images pairs. Note that the many regions corresponding to the object are matched correctly, but background regions are not. This is because the background regions and their adjacency relationships are different in these images. Also, some regions inside the object are not matched correctly. due to lack of sufficient boundary overlap for reliable transformation estimation using KDC.

4.2. Matching Motion Sequences

We compare the results of our algorithm, both qualitatively and quantitatively, against a one-to-one region matching algorithm, which uses absolute region properties for computing match cost between regions instead of PMC, implemented in the following manner. The regions are described using their area, centroid, eccentricity, solidity, perimeter, and, mean and variance of intensities. The cost of matching two regions is computed as sum-of-squared-differences between these absolute properties. The one-to-one RAG matching algorithm implemented is the same as the one described in Sec. 3, except that we change the constraint cost in Eq. (4) so that $V(l_p, l_q) = 0$ only when $l_p, l_q \in \mathcal{N}_2$. Thus, there is a high cost for neighboring regions taking the same label, *i.e.*, penalizing many-to-one

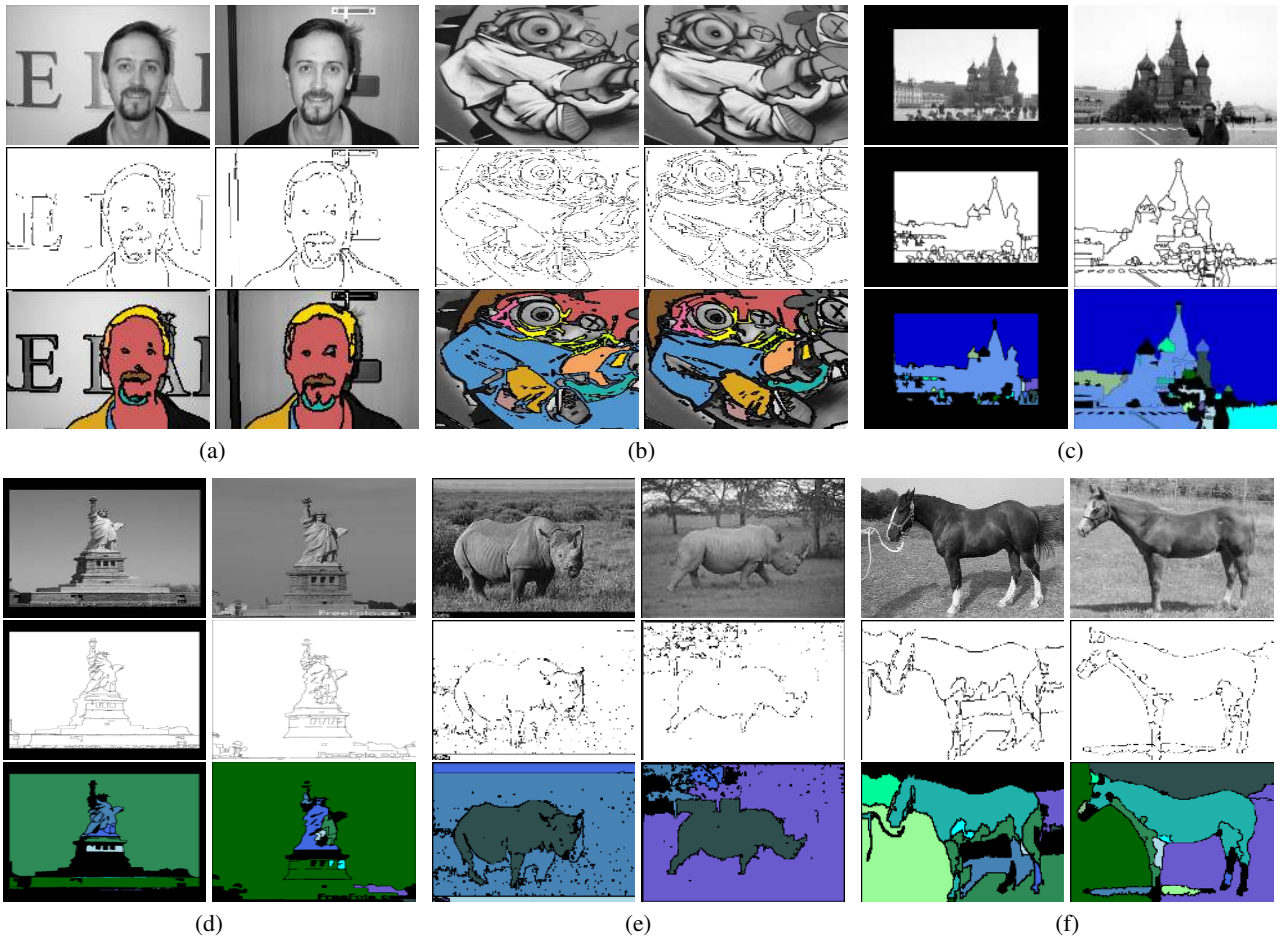
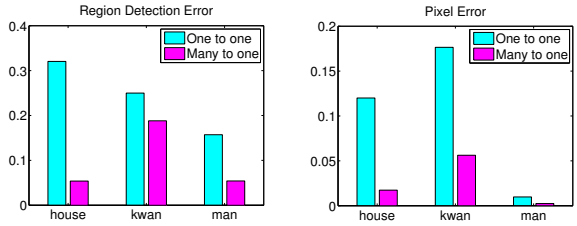


Figure 3. **Image matching results**. Each set of images shows the original images in the **top row**, the segmented images in the **middle row**, and matching results in the **bottom row**. Matched region pairs are shown with the same color in the two images. Images in **(a)** show results for matching face of a person on different backgrounds. Images in **(b)** show matching results for Graffiti images which have complex adjacency structure. Note that the instances of partial matching and many-to-many matchings are detected correctly. Images in **(c)** shows matching of the Kremlin images under scale variations, all the fragments of the structure are correctly matched to the corresponding merger region. Similarly **(d)**, **(e)**, **(f)** show results on instances of the Statue of Liberty, rhinos and horses. Notice how the different segments on the object are matched accurately. (This figure is best viewed in color.)

matching.

We choose motion sequences for this comparison. Note that the assumption of one-to-one correspondence is suitable under this setting. We have considered three different motion sequences here, namely, the *house* sequence from CMU motion database, the *man* sequence of a person walking on static background, and, the *Kwan* sequence of olympic ice-skater Michelle Kwan. Matching is performed on successive frames, returning matched region pairs, which are compared against ground truth region matches marked by us. A region pair is termed as an *error*, if either a) it is a ground truth pair, but not returned by the algorithm, or, b) it is returned by the algorithm, but not a ground truth match pair. The total *region mismatch error* per image is computed as the ratio of the total number of *erroneous* matched pairs returned by the algorithm, and the total number of ground truth region pairs. The *pixel mismatch error* for an erroneous pair is computed as the minimum of the areas of the two regions in the matched pair. We choose the minimum area to respect partial matching. The per image pixel mismatch error is computed as the ratio of sum of pixel mismatch errors of each erroneous pair, and the total image area. For each sequence, we report the mean per image region and pixel mismatch errors, computed as the average of region and pixel mismatch errors of all the frames of that sequence, shown in Fig. 4. Note that the many-to-one matching algorithm has lower errors as compared to the one-to-one matching algorithm, for each of the sequences. As can be seen in Fig. 5, the house sequence has the highest ambiguity due to multiple window regions which are very similar in appearance. This is not the case with the other sequences, and hence they have lower region matching errors.



(a) Region Error (b) Pixel Error

Figure 4. **Performance comparison.** Detection error for our method (many-to-one) is shown in pink and the error for one-to-one matching method is shown in blue. Average errors for three motion sequences are shown. **(a)** We have obtained lower **region detection error** for all the three sequences. Performance gain is the highest for *house* sequence because of its complex RAG structure, due to which this sequence is more prone to structural noise. **(b) Pixel matching error** using one-to-one matching is high for the *kwan* sequence since it has large background clutter and a number of instances of non-repeating regions in adjacent frames. Our PMC based method gives very low error on this sequence. *Man* sequence was the least challenging amongst the three and hence both the methods perform almost equally well on it.

However, the area occupied by these ambiguous regions is very small, and hence the pixel matching errors are lower than the *Kwan* sequence. The *man* sequence is the easiest with not much complex motion, and high repeatability, and hence produces lowest errors.

In Fig. 5 we show example frames of the above motion sequences. Top row shows original images of consecutive frames in the motion sequences middle row shows matching results by our method and the bottom row shows matching results returned by the one-to-one method. Notice how our method can match shirt, pant with the torso region in the *man* sequence. Also hair and face regions are matched by our method but not by one-to-one. In *kwan* sequence similar effect can be observed. We can match the torso of *kwan* with both upper body and lower body whereas one-to-one matches it only with the upper body.

5. Computational Complexity

The most expensive step towards computing the partial match cost is the computation of robust transformation using KDC. From Eq. (2), the complexity of computing the KDC similarity measure between samples from boundaries of two regions, $\boldsymbol{bo}_1 = \{\boldsymbol{x}_1^1, \ldots, \boldsymbol{x}_k^1\}$ and $\boldsymbol{bo}_2 = \{\boldsymbol{x}_1^2, \ldots, \boldsymbol{x}_m^2\}$ for a given transformation $\mathbb{T}$ is $\mathcal{O}(km)$, where k and m are the number of boundary points in the respective boundaries. Let us assume that the KDC converges to a stable transformation in T iterations. Then the complexity of KDC computations for estimating $\mathbb{T}$ is given by $\mathcal{O}(mkT)$. Let us denote by nb_i^j, the number of boundary points in th i-th region of j-th image, and by $Nb^j = \sum_i nb_i^j$, the sum of boundary points of all regions in the jth image. Then, the total complexity of KDC stage can be computed as

$$C_{KDC} = \mathcal{O}(\sum_{i1,i2} Tnb_{i1}^1 nb_{i2}^2) = \mathcal{O}(TNb^1 Nb^2) \quad (6)$$

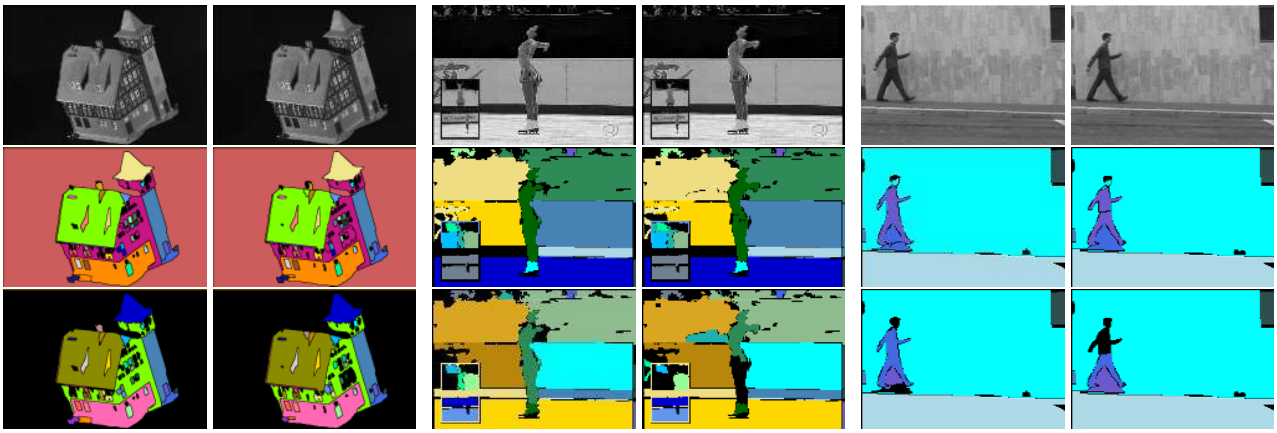
The loopy belief propagation for many-to-one RAG matching has the complexity, $C_{BP} = \mathcal{O}(N^1 N^2 N^2 t)$, with number of nodes (regions in first image) N^1 and number of labels (regions in second image) N^2 and t iterations. The one-to-one matching approach without PMC has computational complexity of $\mathcal{O}(N^1 N^2)$ for computing the match cost between N^1 regions in the first image and N^2 regions in the second image. The complexity of graph matching stage using label assignment is same as C_{BP} for both the methods. Thus PMC provides increased accuracy at the cost of increased computations. However, as demonstrated earlier, even in simplest settings assumption of one-to-one region correspondences is very restrictive. Note that, as compared to methods which perform explicit region mergers we save substantially in terms of computations. For instance, if we consider all possible mergers of N^1 nodes we get 2^{N^1} nodes and this thus makes the complexity of computing match cost exponential $\mathcal{O}(2^{N^1} 2^{N^2})$. The cost of graph matching stage is also increased exponentially.

6. Conclusion

In this paper, we propose a solution of partial region matching for matching a region with its component fragments. We further address the problem of structural noise in region adjacency graphs through our formulation of many-to-one assignment problem. The results of our algorithm for matching images of similar objects with different segmentations show that the concept of partial region match cost is promising for handling irrepeatability usually observed in the output of image segmentation algorithms. We also demonstrate performance enhancement with our many-to-one matching method based on partial match cost, over the one-to-one method which takes absolute properties of the regions into consideration. An immediate possible extension of our work is matching region adjacency graphs constructed using regions obtained from multiscale segmentation outputs, to ensure better repeatability in terms of image contours. Possibility of integrating hierarchical and adjacency constraints in such region adjacency graphs can also be explored.

7. Acknowledgements

The support of the National Science Foundation under grant NSF IBN 04-22073 is gratefully acknowledged.



(a) House Sequence (b) Kwan Sequence (c) Man Sequence

Figure 5. **Example frames from three motion sequences.** Original images are shown in the top row, matching results obtained by our method is shown in the middle row and one-to-one matching results are shown in the bottom row. Matched regions are shown with the same color in the left and right images, regions which are not matched are shown in black. **(a) House.** Note that some of the small regions corresponding to windows are not matched correctly by one-to-one method while we can match them. **(b) Kwan.** We can match the lower body and upper body of *kwan* in the right image with the whole body region in the left image whereas one-to-one matching can not achieve this. **(c) Man.** Notice how the shirt, face, and hair regions are left unmatched by one-to-one method which are matched accurately by our method.

References

[1] N. Ahuja and S. Todorovic. Learning the taxonomy and models of categories present in arbitrary images. In *Proc. of IEEE Intl. Conf. on Comp. Vis*, 2007.

[2] H. Arora and N. Ahuja. Modelling objects as region layouts and hierarchies. In *Proc. of IEEE Conf. on Comp. Vis. Patt. Rec.*, pages 927–934, 2006.

[3] H. Chen and P. Meer. Robust computer vision through kernel density estimation. In *Proc. of 7th Euro. Conf. on Comp. Vis.*, pages 236–250, London, UK, 2002. Springer-Verlag.

[4] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. 24:603–619, 2002.

[5] R. Duda, P. Hart, and D. Stork. Pattern classification. *2nd edR John Wiley Sons*.

[6] A. W. Fitzgibbon. Robust registration of 2d and 3d point sets. *Image Vision Comput.*, 21(13-14):1145–1153, 2003.

[7] C. Gomilla. Graph based object tracking. In *Proc. of IEEE Intl. Conf. on Image Proc.*, 2003.

[8] E. M. J Llados and J. Villanueva. Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. In *IEEE Trans. Patt. Anal. and Machine Intell.*, pages 1137–1143, 2001.

[9] Y. Keselman and S. J. Dickinson. Generic model abstraction from examples. In *Sensor Based Intelligent Robots*, pages 1–24, 2000.

[10] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? In *ECCV (3)*, pages 65–81, 2002.

[11] T. Malisiewicz and A. Efros. Improving spatial support for objects vis multiple segmentations. In *Proc. of British Machine Vis. Conf.*, 2007.

[12] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *Proc. of Uncertainty in AI'99*, pages 467–475, 2009.

[13] M. Pelillo, K. Siddiqi, and S. W. Zucker. Many-to-many matching of attributed trees using association graphs and game dynamics. In *IWVF*, pages 583–593, 2001.

[14] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. of 3rd Intl. Conf. on 3D Digital Imaging and Modeling (3DIM)*, June 2001.

[15] M. Singh, H. Arora, and N. Ahuja. Robust registration and tracking using kernel density correlation. In *Proc. of CVPRW'04*, page 174, 2004.

[16] A. Torsello and E. R. Hancock. Computing approximate tree edit distance using relaxation labeling. *Pattern Recognition Letters*, 24(8):1089–1097, 2003.

[17] P. Tu, T. Saxena, and R. Hartley. Recognizing objects using color-annotated adjacency graphs. In *Shape, Contour and Grouping in Computer Vision*, pages 246–263, 1999.

[18] C. Wang and K. Abe. Region correspondence by inexact attributed planar graph matching. In *Proc. of Intl. Conf. on Comp. Vis.*, pages 440–447, 1995.

[19] F. Wang, B. C. Vemuri, A. Rangarajan, I. M. Schmalfuss, and S. J. Eisenschenk. Simultaneous nonrigid registration of multiple point sets and atlas construction. In *ECCV (3)*, pages 551–563, 2006.

[20] P. L. Worthington and E. R. Hancock. Region-based object recognition using shape-from-shading. In *Proc. of 6th Euro. Conf. on Comp. Vis.*, pages 455–471, 2000.