Co-recognition of Image Pairs by Data-Driven Monte Carlo Image Exploration

Minsu Cho, Young Min Shin, and Kyoung Mu Lee

Department of EECS, ASRI, Seoul National University, 151-742, Seoul, Korea
minsucho@diehard.snu.ac.kr, shinyoungmin@gmail.com, kyoungmu@snu.ac.kr

Abstract. We introduce a new concept of 'co-recognition' for object-level image matching between an arbitrary image pair. Our method augments putative local region matches to reliable object-level correspondences without any supervision or prior knowledge on common objects. It provides the number of reliable common objects and the dense correspondences between the image pair. In this paper, generative model for co-recognition is presented. For inference, we propose data-driven Monte Carlo image exploration which clusters and propagates local region matches by Markov chain dynamics. The global optimum is achieved by a guiding force of our data-driven sampling and posterior probability model. In the experiments, we demonstrate the power and utility on image retrieval and unsupervised recognition and segmentation of multiple common objects.

1 Introduction

Establishing correspondences between image pairs is one of the fundamental and crucial issues for many vision problems. Although the development of various kinds of local invariant features [1,2,3] have brought about notable progress in this area, their local ambiguities remain hard to be solved. Thus, domain specific knowledge or human supervision has been generally required for accurate matching. Obviously, the best promising strategy to eliminate the ambiguities from local feature correspondences is to go beyond locality [4,5,6,7]. The larger image regions we exploit, the more reliable correspondences we can obtain. In this work we propose a novel data-driven Monte Carlo framework to augment naive local region correspondences to reliable object-level correspondences in an arbitrary image pair. Our method establishes multiple coherent clusters of dense correspondences to achieve recognition and segmentation of multiple common objects without any prior knowledge of specific objects.

For the purpose, we introduce a perceptually meaningful entity, which can be interpreted as a common object or visual pattern. We will refer to the entity in an image pair as a Maximal Common Saliency (MCS) and define it as follows: (1) An MCS is a semi-global region pair, composed of local region matches between the image pair. (2) The region pair should be mutually consistent in geometry and photometry. (3) Each region of the pair should be maximal in size. Now, the goal of our work is defined to obtain the set of MCSs from an image pair. According to the naming conventions of some related works [5,8], we term it *co-recognition*.

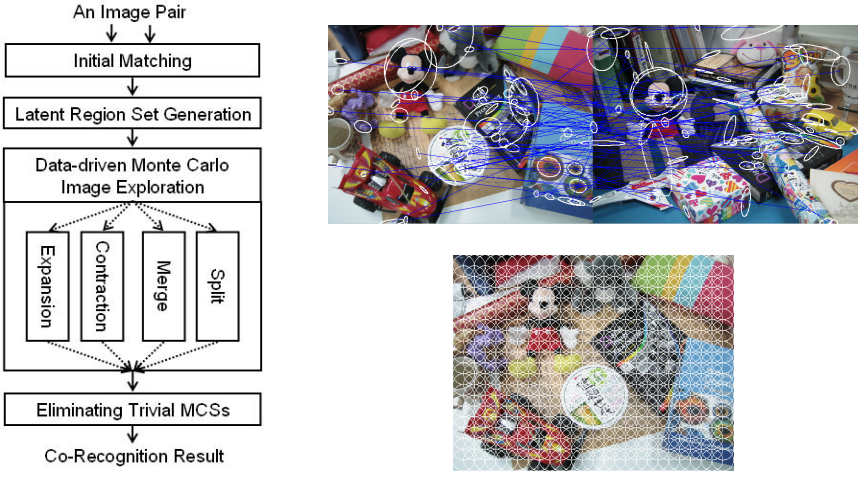
D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part IV, LNCS 5305, pp. 144–157, 2008.

Fig. 1. Result of co-recognition on our dataset *Mickey's*. Given an image pair, co-recognition detects all Maximal Common Saliencies without any supervision or prior knowledge. Each color represents each identity of the MCS, which means an object in this case. Note that the book (blue) is separated by occlusion but identified as one object. See the text for details.

As shown in Fig. 1, co-recognition is equivalent to recognizing and segmenting multiple common objects in a given image pair under two conditions: (1) All the common object appears mutually distinctive in geometry. (2) Each common object lies on different backgrounds in photometry.[1] Note that it can detect separated regions by occlusion as a single object without any prior knowledge or supervision. In this problem, local region correspondences can be established more easily if reliable poses of common objects are known in advance, and the converse is also true. We pose this chicken-and-egg problem in terms of data-driven Monte Carlo sampling with reversible jump dynamics [9,10] over the intra- and inter-image domain. A main advantage of our formulation is to combine bottom-up and top-down processes in a integrated and principled way. Thus, global MCS correspondences and their local region correspondences reinforce each other simultaneously so as to reach global optimum.

Among recent works related to ours are co-segmentation [8], co-saliency [5], and common visual pattern discovery [11]. Rother et al. [8] defined co-segmentation as segmenting common regions simultaneously in two images. They exploited a generative MRF-based graph model and the color histogram similarity measure. Toshev et al. [5] defined co-saliency matching as searching for regions which have strong intra-image coherency and high inter-image similarity. The method takes advantage of the segmentation cue to address the ambiguity of local feature matching. Yuan and Wu [11] used spatial random partition to discover common visual patterns from a collection of images. The common pattern is localized by aggregating the matched set of partitioned images. However, none of these methods recognize multiple common objects as distinct entities. Moreover, [8] and [11] do not consider geometrical consistency in the detected region.

[1] With first condition unsatisfied, several distinct common objects can be recognized as one. With second unsatisfied, objects can include a portion of similar background.



(a) Overview of our approach (b) Initial matching and latent regions

Fig. 2. (a) Given two images, data-driven Monte Carlo image exploration solves co-recognition problem of the image pair. See the text for details. (b) Top: Several different types of local features can be used for initial matches. Bottom: Overlapping circular regions are generated covering the whole reference image for latent regions.

Our method has been inspired by the image exploration method for object recognition and segmentation, proposed by Ferrari et al [6]. The method is based on propagating initial local matches to neighboring regions by their affine homography. Even with few true initial matches, their iterative algorithm expands inliers and contracts outliers so that the recognition can be highly improved.[2] The similar correspondence growing approaches were proposed also in [4,7] for non-rigid image registration. Our new exploration scheme leads the image exploration strategy of [6] to unsupervised multi-object image matching by the Bayesian formulation and the DDMCMC framework [9]. Therefore, the co-recognition problem addressed by this paper can be viewed as a generalization of several other problems reported in the literature [5,6,8,11].

2 Overview of the Approach

Given an image pair, the goal of co-recognition is to recognize and segment all the MCSs and infer their dense correspondences in the pair simultaneously. Figure 2(a) illustrates the overview of our algorithm. First, we obtain initial affine region matches using several different types of local affine invariant features [3,2]. Then, each initial match forms an initial cluster by itself, which is a seed for an MCS. Second, one of the pair is set to be the reference image, and we generate

[2] Although recent object recognition and segmentation methods [12,13] based on local region features demonstrate more accurate results in segmentation, they require enough inliers to localize the object in their initialization step.

a grid of overlapping circular regions covering the whole reference image. All the overlapping regions are placed into a latent region set Λ, in which each element region waits to be included in one of the existing clusters (Fig. 2(b)). After these initialization steps, our data-driven Monte Carlo image exploration algorithm starts to search for the set of MCSs by two pairs of reversible moves; expansion/contraction and merge/split. In the expansion/contraction moves, a cluster obtains a new match or lose one. In merge/split moves, two clusters are combined into one cluster, or one cluster is divided into two clusters. Utilizing all these moves in a stochastic manner, our algorithm traverses the solution space efficiently to find the set of MCSs. The final solution is obtained by eliminating trivial MCSs from the result.

3 Generative Model of Co-recognition

We formulate co-recognition as follows. The set of MCSs is denoted by a vector of unknown variables θ which consists of clusters of matches:

$$\theta = (K, \{\Gamma_i \; ; \; i = 1, ..., K\}), \tag{1}$$

where Γ_i represents a cluster of matches, K means the number of clusters. Γ_i consists of local region matches across the image pair, expressed as follows:

$$\Gamma_i = \{(R_j, T_j); j = 1, ..., L_i\}, \tag{2}$$

where R_j denotes a small local region of the reference image, T_j indicates an affine transformation that maps the region R_j to the other image.[3] L_i denotes the number of local regions included in the cluster Γ_i.

In the Bayesian framework, we denote the posterior probability $p(\theta|I)$ as the probability of θ being the set of MCSs given an image pair I, which is proportional to the product of the prior $p(\theta)$ and the likelihood $p(I|\theta)$. Therefore, co-recognition is to find θ^* that maximizes this posterior as follows.

$$\theta^* = \arg\max_{\theta} p(\theta|I) = \arg\max_{\theta} p(I|\theta)p(\theta). \tag{3}$$

3.1 The Prior $p(\theta)$

The prior $p(\theta)$ models the geometric consistency and the maximality of MCSs.

Geometric Consistency of MCSs. To formulate the geometric constraint of a cluster Γ_i, we used the sidedness constraint of [6], and reinforced it with orientation consistency. Consider a triple (R_j, R_k, R_l) of local regions in the reference image and their corresponding regions $(R_j{}', R_k{}', R_l{}')$[4] in the other image. Let $c_j, c_j{}'$ be the centers of regions $R_j, R_j{}'$, respectively. Then, the sidedness constraint,

$$\text{sign}((c_k \times c_l)c_j) = \text{sign}((c_k{}' \times c_l{}')c_j{}') \tag{4}$$

[3] Registration of non-planar 3-d surfaces is approximated by a set of linear transformations of small local regions.

[4] That is, $R_i{}' = T_i R_i$.

(a) Sidedness constraint b) Reinforced with orientation consistency

Fig. 3. (a) 1 should be on the same side of the directed line from 2 to 3 in both images. (b) 4,5 and 1 satisfies sidedness constraint, while it does not lies on the same object. We can filter out this outlier triplet by checking if orientation(red arrow) changes in the triplet are mutually consistent.

means that the side of c_j w.r.t the directed line $(c_k \times c_l)$ should be just the same as the side of $c_j{}'$ w.r.t the directed line $(c_k{}' \times c_l{}')$ (Fig. 3(a)). This constraint holds for all correctly matching triplets of coplanar regions. Since the sidedness constraint is valid even for most non-planar regions, it is useful for sorting out triplets on a common surface. As illustrated in Fig. 3(b), we reinforce it with orientation consistency to deal with multiple common surfaces for our problem as follows:

$$\forall(m,n) \in \{(j,k),(j,l),(k,l)\}, \quad |\text{angle}(\text{angle}(o_m, o_m{}'), \text{angle}(o_n, o_n{}'))| < \delta_{\text{ori}} \tag{5}$$

where o_m means the dominant orientation of R_m in radian, while angle() denotes the function which calculates the clockwise angle diffrence in radian. Hence, the reinforced sidedness error with the orientation consistency is defined by

$$\text{err}_{\text{side}}(R_j, R_k, R_l) = \begin{cases} 0 & \text{if (4) and (5) hold} \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

A triple violating the reinforced sidedness constraint has higher chances of having one or more mismatches in it. The geometric error of $R_j (\in \Gamma_i)$ is defined by the share of violations in its own cluster such that

$$\text{err}_{\text{geo}}(R_j) = \frac{1}{v} \sum_{R_k, R_l \in \Gamma_i \setminus R_j, k>l} \text{err}_{\text{side}}(R_j, R_k, R_l), \tag{7}$$

where $v = (L_i - 1)(L_i - 2)/2$ is the normalization factor that counts the maximum number of violations. When $L_i < 3$, $\text{err}_{\text{geo}}(R_j)$ is defined as 1 if the cluster $\Gamma_i (\ni R_j)$ violates the orientation consistency, otherwise 0.

The geometric error of a cluster is then defined by the sum of errors for all members in the cluster as follows:

$$\text{err}_{\text{geo}}(\Gamma_i) = \sum_{j=1}^{L_i} \text{err}_{\text{geo}}(R_j). \tag{8}$$

Maximality of MCSs. To encode the degree of maximality of θ, the relative area of each cluster should be examined. We approximate it by the number of

matches in each cluster since all the latent regions have the same area and the number is constant after initialization. The maximality error is formulated as

$$\mathrm{err}_{\mathrm{maxi}}(\theta) = \sum_{i=1}^{K} \left((\frac{L_i}{N})^{0.8} - \frac{L_i}{N} \right), \tag{9}$$

where N is the initial number of the latent region set Λ. The first term encourages the clusters of θ to merge, and the second term makes each cluster of θ to expand.

3.2 Likelihood $p(I|\theta)$

Photometric Consistency of MCSs. The likelihood encodes the photometric consistency of θ using the observation of the given image pair. Let us define the dissimilarity of two regions by

$$\mathrm{dissim}(R_1, R_2) = 1 - \mathrm{NCC}(R_1, R_2) + \frac{\mathrm{dRGB}(R_1, R_2)}{100}, \tag{10}$$

where NCC is the normalized cross-correlation between the gray patterns, while dRGB is the average pixel-wise Euclidean distance in RGB color-space after independent normalization of the 3 colorbands for photometric invariance [6]. R_1 and R_2 are normalized to unit circles with the same orientation before computation. Since a cluster of matches should have low dissimilarity in each match, the overall photometric error of a cluster is defined as follows.

$$\mathrm{err}_{\mathrm{photo}}(\Gamma_i) = \sum_{j=1}^{L_i} \mathrm{dissim}(R_j, {R_j}')^2. \tag{11}$$

Visual patterns in each MCS are assumed to be mutually independent in our model. Hence, the likelihood is defined as follows.

$$p(I|\theta) \propto \exp\left(-\lambda_{\mathrm{photo}} \sum_{i=1}^{K} \mathrm{err}_{\mathrm{photo}}(\Gamma_i) \right). \tag{12}$$

3.3 Integrated Posterior $p(\theta|I)$

From (8), (9), and (12), MCSs in a given image pair I can be obtained by maximizing the following posterior probability:

$$p(\theta|I) \propto \exp\left(-\lambda_{\mathrm{geo}} \sum_{i=1}^{K} \mathrm{err}_{\mathrm{geo}}(\Gamma_i) - \lambda_{\mathrm{maxi}} \mathrm{err}_{\mathrm{maxi}}(\theta) - \lambda_{\mathrm{photo}} \sum_{i=1}^{K} \mathrm{err}_{\mathrm{photo}}(\Gamma_i) \right). \tag{13}$$

This posterior probability reflects how well the solution generates the set of MCSs from the given image pair.

4 Data-Driven Monte Carlo Image Exploration

The posterior probability $p(\theta|I)$ in (13) has a high-dimensional and complicated landscape with a large number of local maxima. Moreover, maximizing the posterior is a trans-dimensional problem because neither the number of MCSs nor the number of matches in each MCS are known. To pursue the global optimum of this complex trans-dimensional posterior $p(\theta|I)$, we propose a new image exploration algorithm based on the reversible jump MCMC [10] with data-driven techniques [9].

The basic idea of MCMC is to design a Markov chain to sample from a probability distribution $p(\theta|I)$. At each sampling step, we propose a candidate state θ' from a proposal distribution $q(\theta'|\theta)$. Through the Metropolis-Hastings rule, the candidate state is accepted with the following acceptance probability.

$$\alpha = \min\left(1, \frac{q(\theta|\theta')p(\theta'|I)}{q(\theta'|\theta)p(\theta|I)}\right). \tag{14}$$

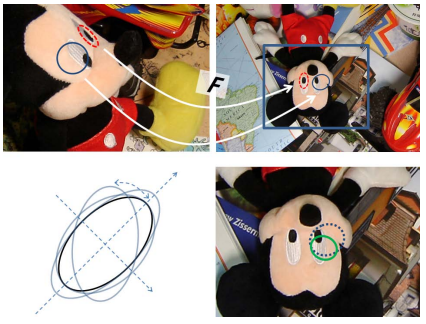
Theoretically, it is proven that the Markov chain constructed in this manner has its stationary distribution as $p(I|\theta)$ irrespective of the choice of the proposal $q(\theta'|\theta)$ and the initial state [10]. Nevertheless, in practice, the choice of the proposal significantly affects the efficiency of MCMC. Recently in computer vision area, data-driven MCMC [9] has been proposed and proven to improve the efficiency by incorporating domain knowledge in proposing new states of the Markov chain. In our algorithm, we adopt the data-driven techniques to guide our Markov chain using the current observation obtained by local region matches in the image pair. Our Markov chain kernel consists of two pairs of reversible jump dynamics which perform expansion/contraction and merge/split, respectively. At each sampling step, a move $m \in \{\text{expand, contract, split, merge}\}$ is selected with the constant probability $q(m)$.

4.1 Expansion/Contraction Moves

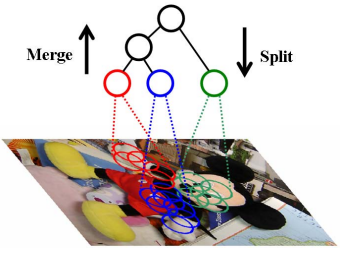
Expansion is to increase the size of an existing cluster by picking a region out of the latent region set Λ and propagating it with a support region in the cluster. Conversely, contraction functions to decrease the size by taking a region out of the members in the cluster and sending it back to Λ. Suppose, at a certain sampling step, that a cluster Γ_i is expanded to $\Gamma_i{}'$, or conversely that $\Gamma_i{}'$ is contracted to Γ_i, then this process can be expressed as the following form without loss of generality:

$$\theta = (K, \{\Gamma_i, ...\}) \leftrightarrow \left(K, \{\Gamma_i{}', ...\}\right) = \theta', \text{where } \Gamma_i \cup (R_k, T_k) = \Gamma_i{}'. \tag{15}$$

The Pathway to Propose Expansion. An expansion move is proposed by the following stochastic procedure with data-driven techniques. Firstly, a cluster is chosen among the current K clusters with the probability $q(\Gamma_i|\text{expand}) \propto \sqrt{L_i}$, which reflects a preference to larger clusters. Secondly, among the matches in the cluster, a support for propagation is selected with probability $q(R_j|\Gamma_i, \text{expand}) \propto$



(a) Propagation and refinement



(b) Mergence tree

Fig. 4. (a) At the top, a support match (red dotted) propagates one of the latent regions (blue solid) by affine homography F. At the bottom, by adjusting the parameter of the ellipse, the initially propagated region (blue dotted) is refined into the more accurate region (green solid). (b) Each of the present clusters has its own mergence tree, which stores hierarchical information of the preceding clusters of itself. It helps to propose a simple and reversible merge/split moves at low cost.

$\sum_{R\in\Lambda} \exp\left(-\frac{\text{dist}(R_j,R)}{2\sigma^2_{\text{expand}}}\right)$, where dist() denotes the Euclidean distance between the region centers. In this stochastic selection, the supports that have more latent regions at nearer distance are favored. Finally, a latent region to propagate by the support is chosen with the probability $q(R_k|R_j,\Gamma_i,\text{expand}) \propto \exp\left(-\frac{\text{dist}(R_k,R_j)^2}{2\sigma^2_{\text{expand}}}\right)$, which means a preference to closer ones.

Propagation Attempt and Refinement. The building block of expansion is based on the propagation attempt and refinement in [6]. If an expansion move is proposed, we perform a propagation attempt followed by the refinement. As illustrated in Fig. 4(a), consider the case that a red dotted elliptical region R_1 in the reference image is already matched to R'_1 in the other image. Each R_1 and R'_1 has an affine transformation A and A', respectively, which transform the regions onto the orientation normalized unit circles. Thus, we can get the affine homography F between R_1 to R'_1 by $F = (A')^{-1}A$, satisfying $FR_1 = R'_1$. If a latent region R_2 is close enough to R_1 and lie on the same physical surface, we can approximate R'_2 in the other image by $R'_2 = FR_2$ as shown in Fig. 4(a). In that case, we state that the support match (R_1, R'_1) attempts to propagate the latent region R_2. Next, by locally searching the parameter space of the current affine homography F, the refiner adjusts it to find R'_2 with minimum dissimilarity such that $F_r = \arg\min_F \text{dissim}(R_2, FR_2)$ as shown at the bottom of Fig. 4(a).

The Pathway to Propose Contraction. An previously expanded region is proposed to contract by the following stochastic procedure with data-driven techniques. Firstly, a cluster is chosen among the current K clusters with the probability $q(\Gamma_i|\text{contract}) \propto \sqrt{L_i}$. Then, among matches supporting no other region

in the cluster, one match is selected with the probability $q(R_k|\Gamma_i, \text{contract}) \propto \exp\left(\frac{\text{err}_{\text{geo}}(R_k)^2+\text{err}_{\text{photo}}(R_k)^2}{2\sigma^2_{\text{contract}}}\right)$, favoring the matches with higher error in geometry and photometry.

4.2 Merge/Split Moves

This pair of moves is for merging two different clusters into a new one or splitting one into two clusters. Suppose, at a certain sampling step, that a cluster Γ_i is split into two cluster Γ_l and Γ_m, or conversely that that Γ_l and Γ_m is merged into a cluster Γ_i, then the processes can be represented as the following form without loss of generality.

$$\theta = (K, \{\Gamma_i, ...\}) \leftrightarrow (K+1, \{\Gamma_l, \Gamma_m, ...\}) = \theta', \text{where } \Gamma_i = \Gamma_l \cup \Gamma_m. \quad (16)$$

The Pathway to Propose Merge. We propose the merge of two clusters along the following stochastic procedure. Firstly, among the current K clusters, one cluster is chosen with the probability $q(\Gamma_l|\text{merge}) \propto 1/K$. Then, another cluster is selected with the probability $q(\Gamma_m|\Gamma_l, \text{merge}) \propto \exp\left(-\frac{\text{dist}(\Gamma_m,\Gamma_l)^2}{2\sigma^2_{\text{merge}}}\right)$, where dist() denotes the Euclidean distance between the cluster centroids. This represents the sampling from a Gaussian Parzen window centered at the centroid of the first cluster Γ_l.

Mergence Trees. Unlike merge, its reverse move, split, is complicated to propose since it involves classifying all the member regions of a cluster into two potential clusters. Moreover, to satisfy the detailed balance condition of MCMC [10], all the move sequences in dynamics should be reversible, which means that if a merge move can be proposed, then the exact reverse split move should be possible. To design efficient and reversible merge/split, we construct *mergence trees* for merge/split over all the process. Each cluster has its own mergence tree which stores the information of all the constituent clusters of itself in the tree structure (Fig. 4(b)). Utilizing the mergence trees, we can propose a simple but potential split move at low cost, that is the move to the state just before the latest merge move. Note that we always begin from the clusters with a single initial match, and the clusters are grown up gradually by the accepted moves among four types of proposals. Thus, one of the best split moves is simply tracing back to the past.

The Pathway to Propose Split. A previously merged cluster can be proposed to split into two as follows using the mergence tree. Firstly, a cluster among the current K clusters is chosen with the probability $q(\Gamma_i|\text{split}) \propto 1/K$. Then, the cluster is proposed to split into two clusters corresponding to child nodes in its mergence tree, with the probability $q(\Gamma_l, \Gamma_m|\Gamma_i, \text{split}, \text{mergence trees}) = 1$.

4.3 Overall Markov Chains Dynamics and Criterion of Reliable MCSs

Our DDMC image exploration algorithm simulates a Markov chain consisting of two pairs of sub-kernels, which continuously reconfigures θ according to $p(\theta|I)$.

At each sampling step, the algorithm chooses a move m with probability $q(m)$, then the sub-kernel of the move m is performed. The proposed move along its pathway is accepted with the acceptance probability (14). If the move is accepted, the current state jumps from θ to θ'. Otherwise, the current state is retained. In the early stage of sampling, we perform only expansion/contration moves without merge/split moves because the unexpanded clusters in the early stage are prone to unreliable merge/split moves. After enough iterations, merge/split moves incorporate with expansion/contraction moves, helping the Markov chains to have better chances of proposing reliable expansion/contraction moves and estimating correct MCSs.

To evaluate the reliability of MCSs in the best sample θ^*, we define the expansion ratio of an MCS as the expanded area of the MCS divided by the entire image area. Since a reliable MCS is likely to expand enough, we determine the reliable MCSs as those expanded more than the threshold ratio ϵ in both of two images. This criterion of our method eliminates the trivial or false correspondences effectively.

4.4 Implementation Details

For initialization, we used Harris-Affine [3] and MSER [2] detectors with SIFT as a feature descriptor. After nearest neighbor matching, potential outliers are filtered out through the ratio test with threshold 0.8 [1]. In our experiments, the grid for the latent region set is composed of regions of radius $h/25$, spaced $h/25$, where h denotes the height of the reference image. The radius trades correspondence density and segmentation quality for computational cost. It can be selected based on the specific purpose. The parameters in the posterior model were fixed as follows: $\delta_{\text{ori}} = \pi/4, \lambda_{\text{geo}} = 3, \lambda_{\text{photo}} = 20, \lambda_{\text{maxi}} = 6$. In the sampling stage, we set the probability of selecting each sub-kernel as $q(\text{expand}) = q(\text{contr}) = 0.4, q(\text{split}) = q(\text{merge}) = 0.1$, and the parameters of sub-kernels are set to $\sigma_{\text{expand}} = l/100, \sigma_{\text{contract}} = 0.5, \sigma_{\text{merge}} = l/10$, where l means the diagonal length of the reference image. The results were obtained after 7000 iteration runs. Only the expansion/constration moves are performed in the first 1000 samplings. In most of our tests, the MAP θ^* was generated within about 5000 samplings. The expansion threshold ratio ϵ for reliable MCSs in all our experiments is set to 2% of each image.

5 Experiments

We have conducted two experiments: (i) unsupervised recognition and segmentation of multiple common objects and (ii) image retrieval for place recognition.

5.1 Unsupervised Recognition and Segmentation of Multiple Common Objects

Since there is no available public dataset for this problem yet, we built a new challenging dataset including multiple common objects with mutual occlusion

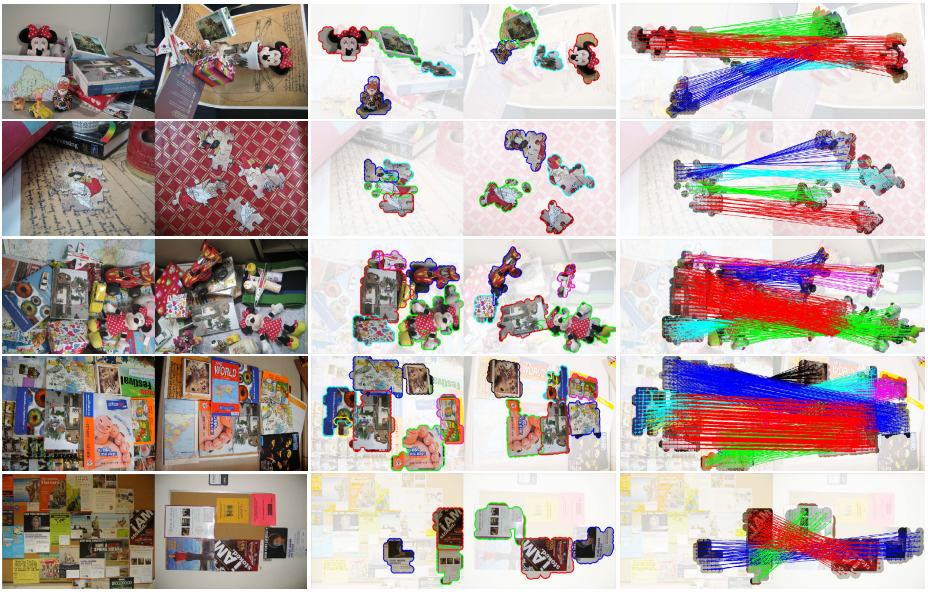


Fig. 5. Co-recognition results on *Minnie's*, *Jigsaws*, *Toys*, *Books*, and *Bulletins*. We built the datasets for evaluation of co-recognition except for *Bulletins*, which is borrowed from [11] for comparison.

Table 1. Performance evaluation of segmentation

Dataset	**Mickey's**	**Minnie's**	**Jigsaws**	**Toys**	**Books**	**Bulletins**	Average
Hit Ratio	80.7%	83.2%	80.0%	83.5%	94.6%	91.2%	85.5%
Bk Ratio	20.6%	37.4%	22.8%	25.2%	11.8%	16.8%	22.4%

and complex clutters. The ground truth segmentation of the common objects has been achieved manually[5]. Figure 5 and 1 show some of co-recognition results on them. Each color of the boundary represents identity of each MCS. The inferred MCSs, their segmentations (the 2nd column), and their dense correspondences (the 3rd column) are of good quality in all pairs of the dataset. On the average, the correct match ratio started from less than 5% in naive NN matches, growing to 42.2% after initial matching step, and finally reached to 92.8% in final reliable MCSs. The number of correct matches increased to 651%.

We evaluated segmentation accuracy by hit ratio h_r and background ratio b_r.[6] The results are summarized in Table 1. It also shows high accuracy in segmentation. For example, the dataset *Bulletins* is borrowed from [11], and our result of $h_r = 0.91, b_r = 0.17$ is much better than their result of $h_r = 0.76, b_r = 0.29$ in [11]. Moreover, note that our method provides object-level identities and

[5] The dataset with ground truth is available at http://cv.snu.ac.kr/∼corecognition.

[6] $h_r = \frac{|\text{GroundTruth} \cap \text{Result}|}{|\text{GroundTruth}|}$, $b_r = \frac{|\text{Result}| - |\text{Result} \cap \text{GroundTruth}|}{|\text{Result}|}$.

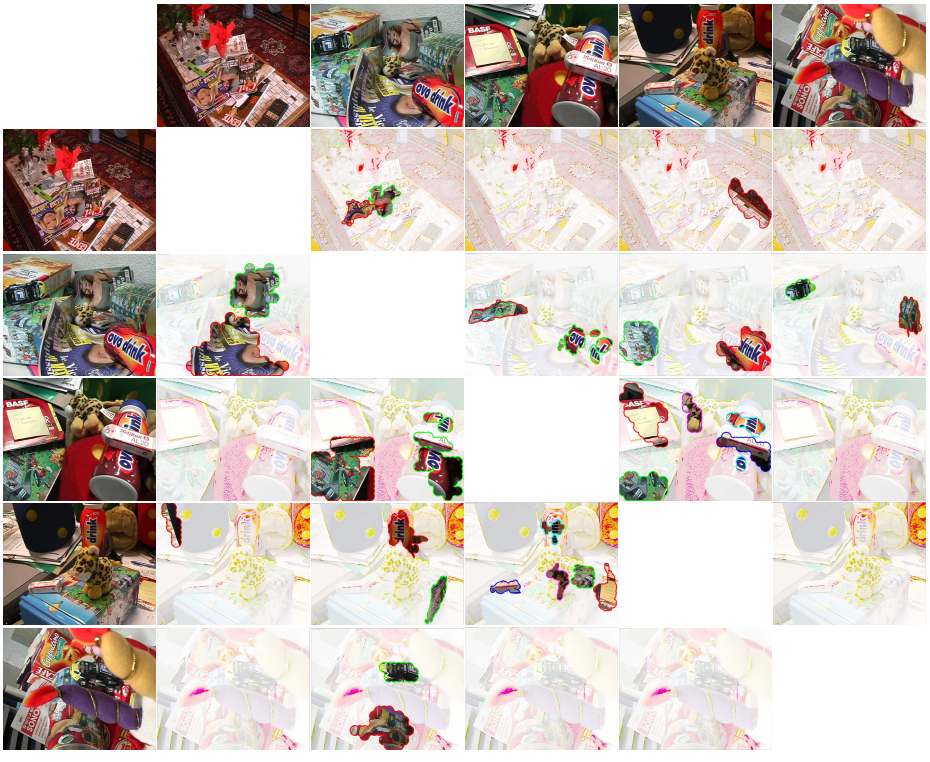


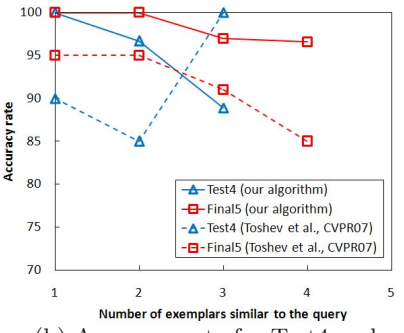
Fig. 6. Co-recognition on all combination pairs of 5 test images from the ETHZ Toys dataset. Both the detection rate and the precision are 93%.

dense correspondences, which are not provided by the method of [11]. Most of the over-expanded regions increasing the background ratio result from mutually similar background regions.

To demonstrate the unsupervised detection performance of co-recognition in view changes or deformation, we tested on all combination pairs of 5 complex images from the ETHZ toys dataset[7]. None of the model images in the dataset are included in this experiment. As shown in Fig. 6, although this task is very challenging even for human eyes, our method detected 13 true ones and 1 false one among 14 common object correspondences in the combination pairs. The detection rate and the precision are all 93%. Note that our method can recognize the separate regions as one MCS if mutual geometry of the regions is consistent according to the reinforced sidedness constraint (6). Thus, it can deal with complex partial occlusion which separates the objects into fragments. This allows us to estimate the correct number of identical entities of separate regions as in result of Fig. 1 and Fig. 6.

[7] http://www.robots.ox.ac.uk/∼ferrari/datasets.html

(a) Co-recognition on ICCV2005 datasets



(b) Accuracy rate for Test4 and Final5.

Fig. 7. (a) Co-recognition deals with object-level correspondence, which is higher than segment-level correspondence. (b) Comparison with co-saliency matching [5] on ICCV2005 datasets.

5.2 Image Retrieval for Place Recognition

For image retrieval, we have conducted the experiment as in [5] on ICCV 2005 Computer Vision Contest datasets[8]. Each of two datasets (*Test4* and *Final5*) has been split into exemplar and query set. *Test4* has 19 query images and 9 exemplar images, while *Final5* has 22 query images and 16 exemplar images. Each of query images is compared with all exemplar images, and all the matched image pairs are ranked according to the total area of reliable MCSs. For every query image having at least k similar examplars, the accuracy rate is evaluated with how many of them are included in top k ranks. The result in Fig. 7(b) reveals that our co-recognition outperforms co-saliency matching [5] largely in this experiment. The reason can be explained by comparing our result of the top in Fig. 7(a) with the result of the same pair in [5]. Co-recognition deals with object-level correspondences, which are higher than segment-level correspondences as [5], our method generates larger, denser, and more accurate correspondence without segmentation cue.

6 Conclusion

We have presented a novel notion of *co-recognition* and the algorithm, which recognizes and segments all the common salient region pairs with their maximal sizes in an arbitrary image pair. The problem is formulated as a Bayesian MAP problem and the solution is obtained by our stochastic image exploration algorithm using DDMCMC paradigm. Experiments on challenging datasets show promising results on the problem, some of which even humans cannot achieve easily. The proposed co-recognition has various applications for high-level image matching such as object-driven image retrieval.

[8] http://research.microsoft.com/iccv2005/Contest/

Acknowledgements

This research was supported in part by the Defense Acquisition Program Administration and Agency for Defense Development, Korea, through the Image Information Research Center under the contract UD070007AD, and in part by the MKE (Ministry of Knowledge Economy), Korea under the ITRC (Information Technolgy Research Center) Support program supervised by the IITA (Institute of Information Technology Advancement) (IITA-2008-C1090-0801-0018).

References

1. Lowe, D.G.: Object recognition from local scale-invariant features. In: ICCV, pp. 1150–1157 (1999)
2. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: BMVC (2002)
3. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 128–142. Springer, Heidelberg (2002)
4. Vedaldi, A., Soatto, S.: Local features, all grown up. In: CVPR, pp. 1753–1760 (2006)
5. Toshev, A., Shi, J., Daniilidis, K.: Image matching via saliency region correspondences. In: CVPR (2007)
6. Ferrari, V., Tuytelaars, T., Gool, L.: Simultaneous object recognition and segmentation from single or multiple model views. IJCV 67(2), 159–188 (2006)
7. Yang, G., Stewart, C.V., Michal Sofka, C.L.T.: Registration of challenging image pairs:initialization, estimation, and decision. PAMI 29(11), 1973–1989 (2007)
8. Rother, C., Minka, T.P., Blake, A., Kolmogorov, V.: Cosegmentation of image pairs by histogram matching - incorporating a global constraint into MRFs. In: CVPR, pp. 993–1000 (2006)
9. Tu, Z., Chen, X., Yuille, A.L., Zhu, S.C.: Image parsing: unifying segmentation, detection, and recognition. In: ICCV, vol. 1, pp. 18–25 (2003)
10. Green, P.: Reversible jump markov chain monte carlo computation and bayesian model determination. Biometrica 82, 711–732 (1995)
11. Yuan, J., Wu, Y.: Spatial random partition for common visual pattern discovery. In: ICCV, pp. 1–8 (2007)
12. Simon, I., Seitz, S.M.: A probabilistic model for object recognition, segmentation, and non-rigid correspondence. In: CVPR (2007)
13. Cho, M., Lee, K.M.: Partially occluded object-specific segmentation in view-based recognition. In: CVPR (2007)