GRASP Recurring Patterns from a Single View

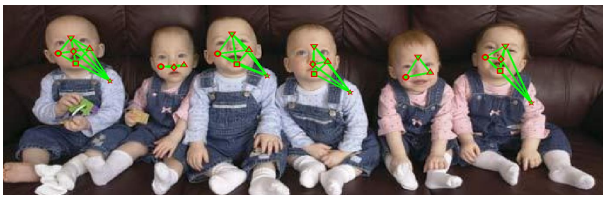
Jingchen Liu[1] Yanxi Liu[1,2]
[1] Computer Science and Engineering, [2] Electrical Engineering
The Pennsylvania State University
University Park, PA 16802, USA
{jingchen, yanxi}@cse.psu.edu

Abstract

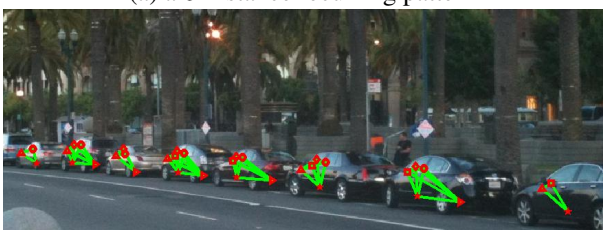
We propose a novel unsupervised method for discovering recurring patterns from a single view. A key contribution of our approach is the formulation and validation of a joint assignment optimization problem where multiple visual words and object instances of a potential recurring pattern are considered simultaneously. The optimization is achieved by a greedy randomized adaptive search procedure (GRASP) with moves specifically designed for fast convergence. We have quantified systematically the performance of our approach under stressed conditions of the input (missing features, geometric distortions). We demonstrate that our proposed algorithm outperforms state of the art methods for recurring pattern discovery on a diverse set of 400+ real world and synthesized test images.

1. Introduction

Similar yet non-identical objects, such as animals in a herd, cars on the street, faces in a crowd or goods on a supermarket shelf, are ubiquitous. There has been a surge of interest in unsupervised visual perception of such near-identical objects [1, 2, 3, 4, 5, 6, 7, 8], echoing an observation that *much of our understanding of the world is based on the perception and recognition of shared or repeated structures* [9]. To capture the **recurrence** nature within such patterns, we use the term *recurring pattern* to refer to the ensemble of multiple instances of a common *visual object* or object for short, which may or may not correspond to a complete physical object. As shown in Figure 1, each object of a recurring pattern is a geometric composition (green arcs) of *visual words* (distinct red iconic shapes), where partial matching among the objects is permitted. The recognition of recurring patterns has applications in effective image segmentation [4], compression and super-resolution [2], retrieval [5] and organization of unlabeled data [7]. More fundamentally, a recurring pattern is a domain independent representation for semantically meaningful mid-level grouping and scene interpretation [10, 11].



(a) a 6-instance recurring pattern



(b) an 8-instance recurring pattern

Figure 1. Unsupervised discovery of recurring patterns in real images by our proposed algorithm, where partial matching and low visual word recall rates (75% for (a), 71% for (b)) are allowed.

Two classic approaches for recurring pattern detection are: (A) *pairwise visual-word-matching* which matches pairs of visual words across all objects [7]; and (B) *pairwise object-matching* which matches feature point correspondences between a pair of objects [12, 5, 4]. Both of these methods are limited in that (1) Pairwise matching, though relatively simple, does not fully utilize all available information for optimal matching. (2) Visual word-pair matching also suffers from missing feature points (low visual word recall rate), as shown by our quantitative evaluations (Section 4). (3) Whether it is better to match object-pairs or visual word-pairs is unknown in advance, and due to the lack of a global decision mechanism, current pairwise-matching systems do not afford flexible and adaptive switching between the two.

We are thus motivated to propose an alternative joint-optimization framework for recurring pattern discovery by matching along both visual word and object dimensions *si-*

multaneously (Fig.2). Given the combinatoric nature of the problem, we further propose to use a *Greedy Randomized Adaptive Search Procedure* (GRASP)[13] for optimization. Our major contributions are: (1) a novel object-visual word joint optimization framework; (2) an effective adaptation of *GRASP* for this joint optimization problem using stochastic 'moves' specifically designed for fast convergence; (3) a formal and explicit treatment of *recurring patterns* with potential missing/spurious feature points in real images;

2. Related Work

Recurring pattern discovery has been referred to in the literature as common visual pattern discovery [14, 5], co-recognition/segmentation of objects [15, 16, 4], and high-order structural semantics learning [7]. [15, 17, 18] achieve unsupervised detection/segmentation of two objects in two separate images. Yuan and Wu [14] use spatial random partitioning to detect object pair(s) from one or a pair of images; Cho et al. formulate the same problem as correspondence association solved by *MCMC* exploration [16] and graph matching [3], respectively. [5] adopts graph matching to detect multiple recurring patterns between two images. To detect more than 2 recurring instances, Cho et al. generalize feature correspondence association under a many-to-many constraint and perform multiple object matching using agglomerative clustering [19] and *MCMC* association [4]. Both approaches are *pairwise-object matching* based methods. Gau et al. [7] use the approach of *pairwise visual word-matching*, while assuming that visual words can be detected on all recurring instances (i.e. a 100% feature recall rate is required).

Our method differs from previous work in two significant ways: (1) it solves a simultaneous visual word-object assignment problem; and (2) it explicitly and effectively deals with missing/spurious feature points in recurring patterns (feature recall rate from an image can be lower than 100%).

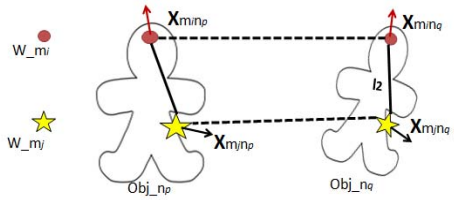
Another line of related work is unsupervised category discovery, e.g. [1, 20, 21]. Categories are found by clustering, typically, 40-1000 input images, and each image usually contains one (sometimes manually cropped) object. None of these methods can handle a single image input without predefined categories.

3. Our approach

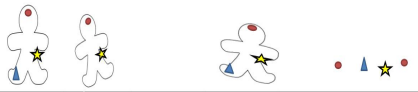
We start with a formalization of the concept of a recurring pattern and its components (Fig. 2), followed by a step-by-step overview of our proposed computational framework (Fig. 3). The key technical steps are the selection and grouping of representative feature points into key visual words and the exploration of the structural consistency among their topology/geometry by using *GRASP* optimization to discover recurring patterns.

3.1. Formalization of Recurring Patterns

We define a *recurring pattern* to have at least two visual objects. Likewise each *object* of a recurring pattern is required to have at least two distinct visual words. Thus, the smallest recurring pattern is conceptually a 4-tuple structure satisfying certain affinity constraints (Figure 2 (a)). The visual word distinctiveness requirement forces each object of a recurring pattern to have a compact representation (no nested recurrence of visual words within each object), thus qualifying it to serve as a structural-primitive for recurring pattern discovery. More importantly, this definition ensures the uniqueness of each recurring pattern while maximizing number of object instances. Mathematically, we construct a recurring pattern $\mathbf{\Omega}$ as a 2D feature-assignment matrix where each row corresponds to a visual word and each column corresponds to a visual object (Figure 2 (b)), that is, $\mathbf{\Omega}_{M,N}(m,n) = f_i$, where f_i corresponds to a feature point, $m = 1 \ldots M, n = 1 \ldots N$, and M and N are the number of visual words and objects, respectively. $\mathbf{\Omega}_{M,N}(m,n) = 0$ is used to indicate a corresponding feature point is missing.



(a)The 4-tuple structure of the smallest recurring pattern



	O-1	O-2	...	O-N	Remaining feature points
W-1	Ω(1,1)	Ω(1,2)		Ω(1,N)	f_i ...
W-2	Ω(2,1)	0	...	Ω(2,N)	f_j ...
...	...	...		...	
W-M	Ω(M,1)	Ω(M,2)	...	Ω(M,N)	f_k ...
Remaining Visual Words	....				

(b)A recurring pattern in assignment matrix form

Figure 2. (a) two potential objects of a smallest recurring pattern, n_1, n_2, each of which contains two visual words m_1, m_2; (b) The 2D feature assignment matrix, where each row corresponds to a visual word and each column to a visual object.

3.2. Visual Word Extraction

Given a set of feature points $\mathbf{F} = \{f_i | i = 1, \ldots, K\}$ (*e.g.*, *SIFT*), a *visual word* $\mathbf{W}$ is a subset of $\mathbf{F}$ such that all feature points in $\mathbf{W}$ share strong appearance similarity. Let v_i be the normalized descriptor of f_i, such that $\|v_i\|_2 =$

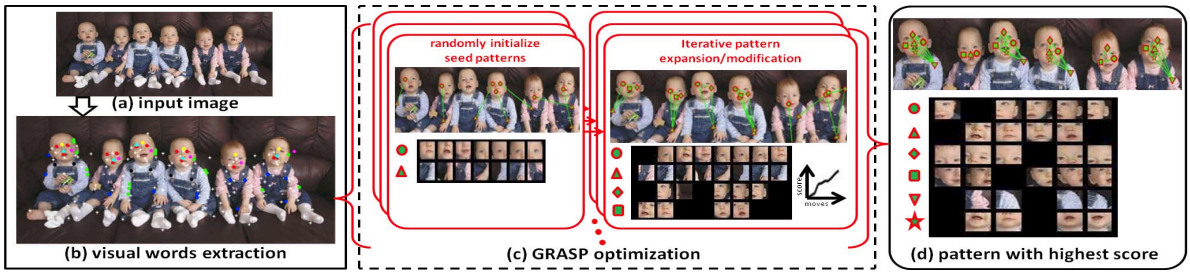


Figure 3. An overview of the proposed method: (a) input image; (b) extracted and clustered feature points with top 20 clusters color coded; (c) *GRASP* optimization framework; (d) automatically discovered recurring pattern after a joint optimization process.

1, we define a normalized affinity metric between features f_i, f_j as

$$A(i,j) = \frac{v_i^T v_j - \mathbf{avg}\{v_p^T v_q | p,q = 1,2,\ldots,K\}}{\mathbf{std}\{v_p^T v_q | p,q = 1,2,\ldots,K\}}. \quad (1)$$

and evaluate the intra-visual-word similarity of $\mathbf{W}$ by

$$s_{\mathbf{W}} = \frac{1}{|\mathbf{W}|} \sum_{i,j \in \mathbf{W}} A(i,j). \quad (2)$$

Starting with an initial assignment of $\mathbf{W} = \{i,j\}$ where $A(i,j)$ is maximum among all feature pairs in $\mathbf{F}$, we use a *forward selection* scheme where new feature points f_i are sequentially included into $\mathbf{W}$ that maximizes Eqn. 2. Once Eqn. 2 can no longer be increased, the growing of the current $\mathbf{W}$ stops and the extraction process then continues on $\mathbf{F} - \mathbf{W}$ to find the next visual word. Our *visual word forward-selection* method differs significantly from K-means, in that we only extract inlier subsets of $\mathbf{F}$ to form a vocabulary of key visual-words for recurring patterns, while ignoring a considerable amount of background noise or outliers.

For efficiency, the affinity matrix A can be made sparse by setting $A(i,j) = 0$ for $A(i,j) < \tau$. In our experiments, we set $\tau = 2$ to remove feature pairs with distance that exceeds 'two-sigma'. Given the sparsity of A, typically $30 \sim 200$ valid visual words can be extracted from a single image depending on the image content and resolution.

Ideally, different feature points from the same visual-word $\mathbf{W}$ should be present in the corresponding relative locations of all N objects of a recurring pattern, i.e. $N = |\mathbf{W}|$. Due to image noise and distortion, we may only obtain $N^* < N$ inlier feature points for word W while getting $|W| - N^*$ outliers. To quantify the levels of such difficulty for recurring pattern discovery algorithms we define the **visual-word recall** R_{VW} **and precision** P_{VW} **rates** respectively as: $\mathbf{R_{VW}} = \mathbf{N^*/N}, \mathbf{P_{VW}} = \mathbf{N^*/|W|}$.

3.3. Object Geometric Affinity and a Joint Optimization Problem

Objects of a potential *recurring pattern* need to be supported not only by the appearance similarity of their matched feature-pairs from distinct visual words, but also by the geometric consistency of the spatial layouts of their corresponding visual words across objects. The geometric configuration of an entry (feature point) in $\mathbf{\Omega}(m,n)$ is defined by $(x_{m,n}, s_{m,n}, \theta_{m,n})$, denoting the centroid, scale and rotation of the corresponding local image patch. The geometric affinity metric for a 4-tuple feature structure, the smallest recurring pattern (Fig. 2(a)), is defined over two distinct visual words w_{m_i}, w_{m_j} and 2 objects o_{n_p}, o_{n_q} as

$$g(m_i, m_j, n_p, n_q) = \exp\left(-\frac{\Delta_\theta^2}{2\sigma_\theta^2} - \frac{\Delta_s^2}{2\sigma_s^2}\right), \quad (3)$$

where Δ_s and Δ_θ are normalized scale and angular distances measured by

$$\Delta_s = \frac{1}{2} \sum_{m=m_i,m_j} \frac{s_{m,n_q} - r \cdot s_{m,n_p}}{\sqrt{r \cdot s_{m,n_p} s_{m,n_q}}}, \quad (4)$$

$$\Delta_\theta = \frac{1}{2} \sum_{m=m_i,m_j} \angle(\theta_{m,n_q} - \theta_{m,n_p} - \theta_0), \quad (5)$$

$r = d(x_{m_i n_q} - x_{m_j n_q})/d(x_{m_i n_p} - x_{m_j n_p})$, $\angle(\cdot)$ is absolute angular distance, and θ_0 is the angle between line segment $\overline{x_{m_i n_p} x_{m_j n_p}}$ and $\overline{x_{m_i n_q} x_{m_j n_q}}$ (Fig. 2(a)). Parameters σ_θ, σ_s control the tolerance for shape deformation; both are set to 0.2 in the experiments. The overall geometric affinity of a candidate recurring pattern $\mathbf{\Omega}_{M,N}$ is then given by

$$G(\mathbf{\Omega}_{M,N}) = \frac{1}{M \cdot N - N_0} \sum_{\substack{m_i,m_j=1..M \\ n_p,n_q=1..N}} g(m_i, m_j, n_p, n_q), \quad (6)$$

where N_0 is the number of missing features. Finally, we can formalize recurring pattern detection as a *joint optimization problem*:

$$\mathbf{\Omega}^* = \arg \max_{\mathbf{\Omega},M,N} \{G(\Omega_{M,N})\}. \quad (7)$$

Tolerance to missing features:
We set $g(m_i, m_j, n_p, n_q) = 0$ in case any of the features

Alg.1: Greedy randomized adaptive search procedure

Repeat: randomly initialize the feature matrix $\mathbf{\Omega}_{MN}$
  Repeat: sequentially apply *moves*[1] 1-5
  Until: no valid moves increase G in Eqn. 7
Until: maximum number of re-initializations reached

[1]Only moves that improve Eqn. 7 are valid.

in the 4-tuple structure is missing ($\mathbf{\Omega}(m,n) = 0$). This is later compensated for by a smaller normalization term $(M \cdot N - N_0)$ (Eqn.6). Our empirical results demonstrate that our framework can tolerate missing features as long as the recall rate $R_{VW} \geq 50\%$ and every object contains at least 50% of the valid visual words. It can be proven that the first missing entry from a full M-by-N matrix will on average decrease the G score by $3/(MN-1)$.

3.4. GRASP Optimization

The optimization problem specified in Eqn.7 is NP-hard. We hereby adopt a *Greedy Randomized Adaptive Search Procedure*(GRASP) [13, 22]. This approach is commonly applied to solve difficult combinatorial optimization problems, athough it has rarely been applied to computer vision problems in the past.

3.4.1 *GRASP* Framework

GRASP is a multi-start metaheuristic algorithm for solving combinatorial optimization problems. Each iteration consists of two phases: (1) random initialization of a feasible solution: since *GRASP* seeks a local optimum for each random initialization, it is important that a variety of initial states be generated to fully explore the solution space. We randomly select 2 visual words for initialization in our experiments.(2) local greedy optimization: We define 5 basic local moves to construct a neighborhood-traversal system in the solution space: (a)*add a visual word*, (b)*add an object*, (c)*modify a single feature point*, (d)*remove a visual word* and (e)*remove an object*. These correspond to adding/removing a row/column or modifying an entry in the assignment matrix $\mathbf{\Omega}_{MN}$ (Fig. 2(b)). We apply stochastic greedy moves, which means that all candidate moves are evaluated and we randomly select among the top 3 moves that improve the objective function. This approach lets us explore a variety of local optima through different randomized paths during the expansion of $\mathbf{\Omega}$.

3.4.2 Local Moves

For all moves described below, only moves that improve Eqn. 7 are valid (see Alg. 1).

1. Add-a-visual-word: $\mathbf{\Omega}_{M,N} \rightarrow \Omega_{M+1,N}$. Let a new word candidate contain $n' = 1,\ldots,N'$ feature points ($|\mathbf{W}_{M+1}| = N'$). The assignment of N' points to N objects can be solved using graph matching by defining an $N'N$-by-$N'N$ affinity matrix U, with each entry $U(i,j)$ indicating the co-assignment affinity of n'_i to n_i and n'_j to n_j, evaluated by:

$$U(i,j) = \sum_{m=1}^{M} g(m, M+1, n_i, n_j). \quad (8)$$

The assignment of features in the new word $\mathbf{W}_{M+1}$ to each object can be determined from the optimal binary indicator vector: $x^* = \arg\max(x^T U x)$, subject to an additional 1-to-1 feature allocation constraint.

Eqn.8 reflects one distinctive characteristic of our approach: we are looking for a consistent matching between the new candidate $\mathbf{W}_{M+1}$ and *all* existing visual words $\mathbf{W}_m, m = 1,\ldots,M$, instead of only isolated pairwise matches. Although the graph matching problem with 1-to-1 constraints is NP-hard, our subproblem is of small scale, and can be handled by state-of-the-art empirical graph matching methods (e.g.[3]).

2. Add-an-object: $\mathbf{\Omega}_{M,N} \rightarrow \mathbf{\Omega}_{M,N+1}$. Any remaining feature points f_i in $\mathbf{W}_{m'}, m' = 1,\ldots,M$ can start a new column $\mathbf{\Omega}_{m',N+1}$ to propose a new candidate object. We then fill in the rest of column $N+1$ by independently examining leftover feature points in W_m as well as missing features $\mathbf{\Omega}(m, N+1) = 0$, for $m = \{1,\ldots,M\} \setminus m'$, and optimizing the affinity between the new object $N+1$ and all existing objects: $\mathbf{\Omega}(m, N+1)^* = arg\max \sum_{n=1}^{N} g(m, m', n, N+1)$.

3. Modify-feature-entries enumerates and replaces all entries in $\mathbf{\Omega}_{M,N}(m,n)$ with the remaining feature points in the same words $\mathbf{W}_m$ or $\mathbf{\Omega}_{m,n} = 0$.

4/5. Remove-a-visual-word/object removes a row or column from the feature matrix $\mathbf{\Omega}_{M,N}$.

4. Experimental Validation

We validate the effectiveness of our algorithm and compare to existing algorithms on four different datasets (422 images total): (1) a synthesized image set (262 images) generated under controlled visual words precision/recall rates and geometric deformations (Fig. 5); (2) a subset of public domain supermarket image set [23] (100 images); (3) a public domain face dataset [24] (30); and (4) our own collection of various real world recurring patterns (30 images containing diverse photos, paintings and texture synthesized images).

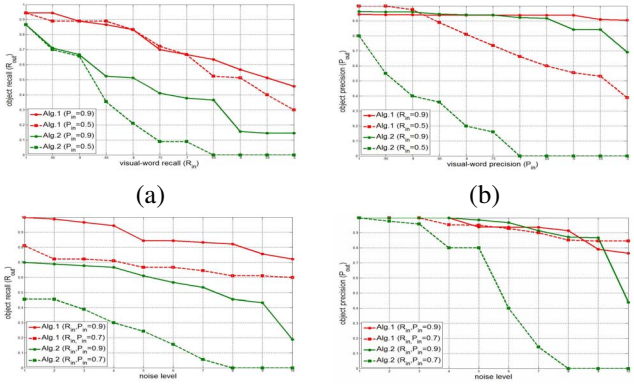
Generalization to multiple patterns/images: Although the description of our approach (Section 3) has focused on discovering a single recurring pattern in a single image, we can generalize it to multiple patterns/images. We treat multiple images as one huge image and impose an extra con-

straint that forbids feature points across different images to be associated to form an 'object'. Multiple recurring patterns are discovered with a recursive greedy approach: each time a pattern is discovered, all its associated feature points are removed and the discovery process restarts.

4.1. Controlled Stress-Test and Comparison

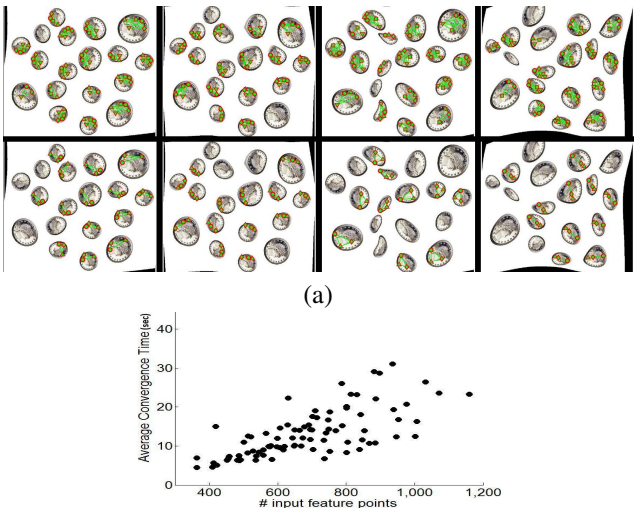
We carefully divide computational challenges at the input feature level for recurring pattern discovery into 3 categories: (1) **low visual word recall rate** R_{in} due to imperfect feature extraction, (2)**low visual word precision rate** P_{in} due to background clutter, and (3) **noisy feature locations** due to geometric deformation. To control the level of difficulty, we simulate these challenges on an image of coins by randomly eliminating feature points, generating outlier features at random locations, and adding Gaussian noise to feature point locations (Fig. 5)(a). The performance of the algorithms is then evaluated at the object-level by the **object recall** (R_{out}) **and precision**(P_{out}) **rates** (with known ground-truth). We compare our method with the pairwise visual word association method in [7], at different levels of visual-word recall/precision rate as well as geometric deformation as plotted in Fig. 4. It is interesting to note that under high input feature recall rate R_{in}, performance difference is minimal; it is when the feature recall R_{in} is low that the algorithms diverge in performance. Compared with [7], our proposed approach illustrates robustness against low R_{in} (red curves in Fig. 4) and has a low false alarm rate in all cases. Object recall/precision rates under varying geometric deformation levels are shown in Fig. 4(c,d).

Computational time and Repeatability We provide empirical time estimates for *GRASP* optimization on the coin



(a) (b) (c) (d)

Figure 4. Quantitative evaluation under controlled conditions. Alg.1: our approach (Red); Alg.2: [7] (Green). We show the object recall and precision rates under two levels of feature recall and precision rates, 0.9 and 0.7, respectively, while fixing the geometric deformation at level 5. See the full evaluation and movie in our supplemental material.



(a) (b)

Figure 5. (a) Synthesized coins with increased geometric deformation (left to right). top: results of our approach; bottom: results of [7]. (b) Average convergence time (seconds) of GRASP optimization versus the number of input feature points

images under different R_{in}, P_{in} and deformation levels as plotted in Fig.5(b). The average convergence time after each random initialization appears to be linear with the number of input feature points (total input visual words). Although the *GRASP* optimization requires repeated calculation and evaluation of candidate moves, most calculation involves the geometric affinity estimation of Eqn. 3. These affinities can be pre-computed on all 4-tuple combinations, and the main computation during optimization is only the summation and indexing from Eqn. 7. The algorithm is implemented using Matlab with no compiler optimization and was run on a 2.6GHz, i7 CPU. To evaluate the repeatability and variance of *GRASP* initialization, we performed 30 random initializations for *GRASP*, and observed on average that 29% of the attempts end up with equivalently high R_{out}/P_{out} performances.

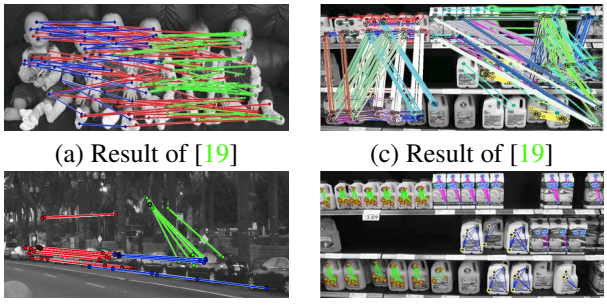
4.2. Supermarket and Face Datasets: Validation and Comparison

Supermarket scene images typically contain multiple recurring patterns. We evaluate dominant patterns detected by the algorithm in [7] and our approach. We also run our algorithm on the face dataset used in [7]. For each run we randomly form a collage of $6 \sim 24$ faces, and repeat 30 times. The result statistics on both datasets are given in Table 1. Sample results of our algorithm on the supermarket and face datasets are shown in Fig. 6.

Figure 6. TOP: Sample results of our algorithm on the supermarket scene (Table 1) where it captures multiple recurring patterns in single images. BOTTOM: Sample results on a collage from the face dataset with object-level precision rate 98.4% and recall rate 63.8% (Table 1 right column).

Dataset	supermarket	face
#Total recurring patterns	100	30
Avg. #obj/per pattern ±std	8.82 ± 4.47	22 ± 2.86
Recall rate (%) [ours]	**87.8**	**63.8**
Recall rate (%) [7]	71.5	39.4
Precision rate (%) [ours]	**95.4**	**98.4**
Precision rate (%) [7]	88.5	96.3

Table 1. Object-Level precision/recall rates on publicly available Supermarket [23] and Face [24] datasets



(a) Result of [19] (c) Result of [19]

(b) Result of [19] (d) Our Result finds 3 RPs

Figure 7. Qualitative comparison with [19]. The top 3 object pair clusters detected by [19] are shown in r,g,b in (a),(b),(c). Corresponding results of our algorithm for (a) and (b) can be found in Figure 1. In (c) 29 object pairs (out of 178) detected by [19] while our approach recognizes 3 dominant recurring patterns with 13-, 11-, 6-instances respectively, achieving an average object detection recall rate of 90%.

4.3. Qualitative Comparisons

State-of-the-art work in object-pair matching [19] uses agglomerative clustering (many-to-many) for recurring pattern detection. We use the publicly available code of [19], setting the parameter for number of initial matches to $4,000$ and maximum matches per feature to 10. The output of the algorithm is associated object pairs: feature point correspondences between the same object pair are associated together and shown in the same color. As can be seen in Fig.7, [19] recognizes 3 (overlapping) object pairs (red,green, blue) in (a) and (b) instead of the recurring patterns that contain 6 babies and 8 cars respectively. The algorithm thus fails to recognize the smallest recurring 'object' in these cases, which is an issue not addressed by previous work. In Fig.7(c), [19] recognizes 29 object pairs, while the image contains 3 recurring patterns with $N_1 = 14$, $N_2 = 12$ and $N_3 = 7$ instances, respectively, which result in a total of $\sum_i \binom{N_i}{2} = 178$ groundtruth object pairs. The main reason for this is that pairwise-independent object matching does not consider recurring patterns of multiple instances as a whole.

4.4. Real World Recurring Patterns: Validation

The recurring pattern dataset we have collected contains a variety of challenging real-world images including street views, architecture, faces, animals, hand-painting, texture-like patterns and objects with symmetries. The recurring patterns exhibit different levels of deformation, rotation, scaling, background clutter, image resolution and a wide range in the number of recurring instances (from 2 to 30). Sample results of our approach are shown in Figures 1 and 8. Our method achieves an object recall rate of 92% with an object precision of 96%.

4.5. Applications

As a byproduct of recurring pattern discovery, we can directly match and register (visual word by visual word) all instances found in a recurring pattern, which can be further used for regularity evaluation and categorization. Fig. 9(a) shows a detected recurring pattern with an *average pairwise normalized correlation* (APNC) score of 0.86 after pattern registration, suggesting high likelihood of a doctored photo of crowds, as compared with an APNC score of a small group of llamas at 0.64 (Fig. 9(b)). Using the average object geometry deformation and the APNC score of a recur-

Figure 9. (a) a doctored photo with found recurring pattern: APNC score = 0.86; (b) an example of registered recurring pattern of a group of llamas with an APNC scores of 0.64, indicating their inherent and statistically significant category differences.

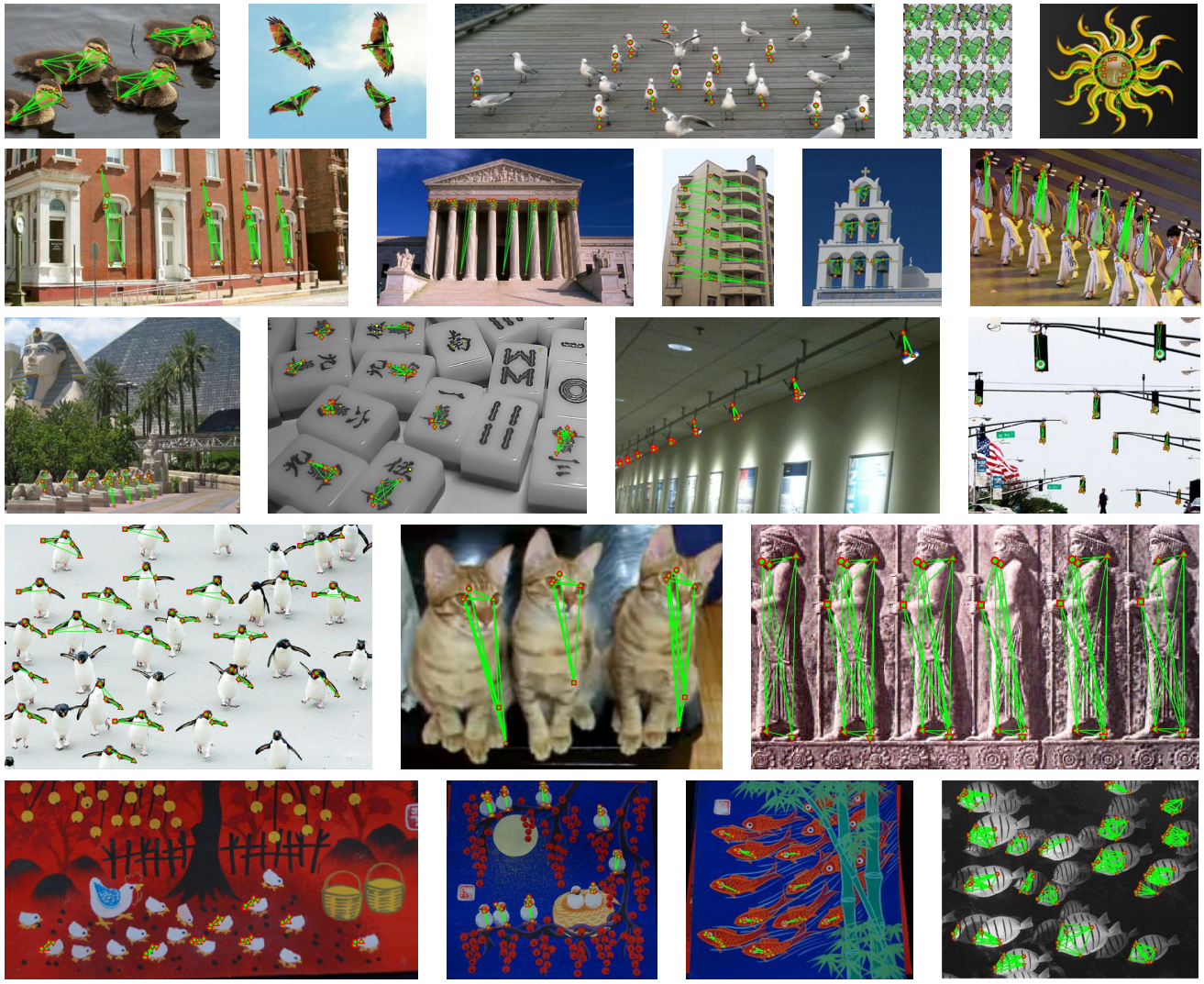


Figure 8. Sample output of our algorithm. Object recall and precision rates are 92% and 96% respectively. Average objects per recurring pattern is 10. Worth noting: bottom-left painting contains only all the chicks facing to the left while the rotated object instances in the top-right image are found completely. This reflects that at object-level the instances are shift- and rotation-invariant yet not reflection invariant.

ring pattern to approximate its quantified regularity in a 2D space (Figure 10), we observe a striking relation between recurring pattern categories and their geometry/photometry deviation from perfect regularity. We thus see much potential of our unsupervised recurring pattern discovery algorithm to contribute to object recognition and pattern categorization.

5. Discussion and Conclusion

Our work is the first to propose a joint object-visual word level optimization for recurring pattern discovery, extending beyond popular pairwise matching. We present a novel adaptation of *GRASP* and demonstrate its effectiveness through extensive evaluations on a variety of difficult synthetic and real-world images. Compared to state-of-the-art approaches, our method achieves superior object-level precision and recall rates under challenging stress-test conditions, in particular when feature-level recall rates are low. Although *GRASP* is not theoretically guaranteed to reach the global optimum compared to *MCMC*, practically it terminates a bad initialization quickly and thus explores the solution space more efficiently (Fig. 5(b)). The potential applications of an automated recurring pattern discovery tool are enormous, ranging from image registration, segmentation, people/product counting, surveillance to saliency perception. Our future work includes further exploration of alternative low-level feature descriptors to enrich the visual

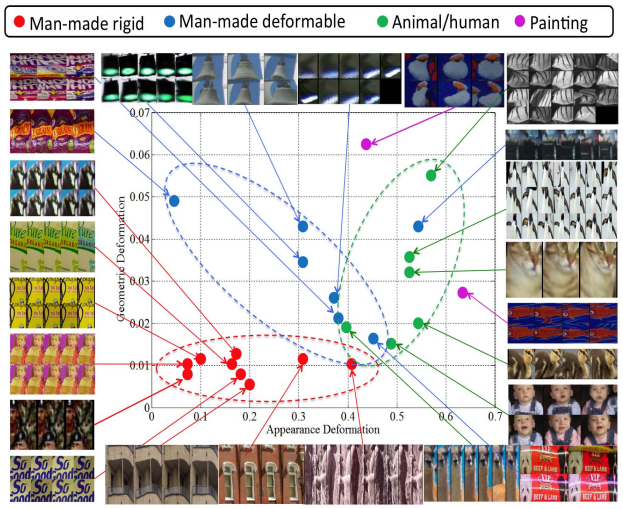


Figure 10. Distributions of automatically discovered recurring patterns by our algorithm arranged in a 2D geometry and appearance regularity space. Point (0,0) means no deviations from a perfect regular pattern. The general trend agrees with common sense: man-made rigid objects are more regular than man-made deformable objects, which are in turn more regular than herds of animals (cats, penguins, lamas, ducks) and human babies.

word vocabulary, and development of a self-enhancement strategy for recurring pattern discovery.

Acknowledgment This work is funded partially under NSF grants IIS-1248076 and IIS-1144938.

References

[1] Faktor, A., Irani, M.: "clustering by composition" - unsupervised discovery of image categories. In: ECCV. (2012) 1, 2

[2] Spinello, L., Triebel, R., Vasquez, D., Arras, K.O., Siegwart, R.: Exploiting repetitive object patterns for model compression and completion. In: ECCV. (2010) 1

[3] Cho, M., Lee, J., Lee, K.M.: Reweighted random walks for graph matching. In: ECCV. (2010) 1, 2, 4

[4] Cho, M., Shin, Y.M., Lee, K.M.: Unsupervised detection and segmentation of identical objects. In: CVPR. (2010) 1, 2

[5] Liu, H., Yan, S.: Common visual pattern discovery via spatially coherent correspondences. In: CVPR. (2010) 1, 2

[6] Bagon, S., Brostkovski, O., Galun, M., Irani, M.: Detecting and sketching the common. In: CVPR. (2010) 1

[7] Gao, J., Hu, Y., Liu, J., Yang, R.: Unsupervised learning of high-order structural semantics from images. In: ICCV. (2009) 1, 2, 5, 6

[8] Garg, R., Ramanan, D., Seitz, S.M., Snavely, N.: Where's waldo: Matching people in images of crowds. In: CVPR. (2011) 1

[9] Mitra, N.J., Guibas, L., Pauly, M.: Partial and approximate symmetry detection for 3d geometry. ACM Transactions on Graphics **25** (2006) 560–568 1

[10] Park, M., Brocklehurst, K., Collins, R., Liu, Y.: Translation-symmetry-based perceptual grouping with applications to urban scenes. In: ACCV. (2010) 1

[11] Wu, C., Frahm, J., Pollefeys, M.: Repetition-based dense single-view reconstruction. In: CVPR. (2011) 1

[12] Suh, Y., Cho, M., Lee, K.M.: Graph matching via sequential monte carlo. In: ECCV. (2012) 1

[13] Feo, T., Resende, M.: Greedy randomized adaptive search procedures. Global Optimization **6** (1995) 109–133 2, 4

[14] Yuan, J., Wu, Y.: Spatial random partition for common visual pattern discovery. In: ICCV. (2007) 2

[15] Rother, C., Minka, T., Blake, A., Kolmogorov, V.: Cosegmentation of image pairs by histogram matching - incorporating a global constraint into mrfs. In: CVPR. (2006) 2

[16] Cho, M., Shin, Y.M., Lee, K.M.: Corecognition of image pairs by data-driven monte carlo image exploration. In: ECCV. (2008) 2

[17] Toshev, A., Shi, J., Daniilidis, K.: Image matching via saliency region correspondences. In: CVPR. (2007) 2

[18] Kannala, J., Rahtu, E., Brandt, S., Heikkila, J.: Object recognition and segmentation by non-rigid quasi-dense matching. In: CVPR. (2008) 2

[19] Cho, M., Lee, J., Lee, K.M.: Feature correspondence and deformable object matching via agglomerative correspondence clustering. In: ICCV. (2009) 2, 6

[20] Sivic, J., Russell, B., Zisserman, A., Freeman, W., Efros, A.: Unsupervised discovery of visual object class hierarchies. In: CVPR. (2008) 2

[21] Todorovic, S., Ahuja, N.: Unsupervised category modeling, recognition, and segmentation in images. PAMI **30** (2008) 2158–2174 2

[22] Resende, M., Ribeiro, C.: Greedy randomized adaptive search procedures (evaluation). Handbook of Metaheuristics (2003) 219–249 4

[23] Zhang, Y., Wang, L., Hartley, R., Li, H.: Handling significant scale difference for object retrieval in a supermarket. In: DICTA. (2009) 4, 6

[24] F-F, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In: CVPR Workshop on Generative-Model Based Vision. (2004) 4, 6