

Picture Collage

Jingdong Wang¹ Jian Sun² Long Quan¹ Xiaoou Tang² Heung-Yeung Shum²

¹The Department of Computer Science, HKUST, {welleast, quan}@cs.ust.hk

²Microsoft Research Asia, Beijing, China, {jiansun, xitang, hshum}@microsoft.com

Abstract

In this paper, we address a novel problem of automatically creating a picture collage from a group of images. Picture collage is a kind of visual image summary — to arrange all input images on a given canvas, allowing overlay, to maximize visible visual information. We formulate the picture collage creation problem in a Bayesian framework. The salient regions of each image are firstly extracted and represented as a set of weighted rectangles. Then, the image arrangement is formulated as a Maximum a Posterior (MAP) problem such that the output picture collage shows as many visible salient regions (without being overlaid by others) from all images as possible. Moreover, a very efficient Markov chain Monte Carlo (MCMC) method is designed for the optimization. Applications to desktop image browsing and image search result summarization demonstrate the effectiveness of our approach.

1 Introduction

With the rapid growth of digital image content, it becomes more and more of a challenge to browse through a huge amount of images, such as hundreds of vacation photos in a home directory or thousands of images returned for a query on the internet. To make browsing process more efficient, image summarization is often needed to address this problem. Most previous image summarization works mainly focus on content based techniques such as image clustering [3] and categorization [4] to provide a high-level description of a set of images. In this paper, we propose a visual image summarization approach — picture collage. Figure 1(c) shows an example of picture collage.

The top row of Figure 1 shows a group of images. A common summarization method is to select a smaller number of representative images and create an image mosaic as shown in Figure 1(a). However, the disadvantage of this approach is that the image mosaic will contain lots of uninformative regions (such as sky and grass in this case). An ideal image summary should contain as many informative regions as possible on a given space. Figure 1(b) shows a collage produced by a commercial image browsing software.

Images are randomly placed on a canvas allowing overlay. Although all images are displayed, more than half of the images are occluded. Additionally, each image is down-sampled, and cropped without considering image content. Figure 1(c) is a picture collage generated by the proposed approach in this paper. Compared with the previous two results, picture collage shows the most informative regions of all images on a single canvas without down-sampling and cropping. In another word, picture collage creates a visual image summarization of a group of images while maximizing visible visual information.

1.1 Related work

To create an ideal image summarization, the first step is to determine which regions of each image are informative or salient. One discriminative approach is the saliency-based visual attention model [6]. This model combines multi-scale image features (color, texture, orientation) into a single topographical saliency map. A dynamical neural network detects attended regions from the saliency map.

A simple technique for image arrangement is page layout [2][5]. It mainly aims to maximize page coverage without image overlap. However, page layout techniques share a common drawback of image mosaics — paying homogeneous attention over the entire image.

An interactive approach to combine multiple images is interactive digital photomontage [1]. The user manually specifies salient regions on each image and the system creates a single composite image. This technique works well only when all input images are roughly aligned. But in our application, i.e. picture collage, input images are usually considerably different.

One generative approach for selecting salient regions and generating a summary image is epitomic analysis [7]. The epitome of an input image is a condensed version of the image that contains all constitutive textural and shape primitives necessary for reconstructing the image. But the epitome image is originally designed for the purpose of reconstruction not viewing. Semantic structures and objects in the input images can not be preserved in the epitome image.

The most similar work to ours is digital tapestry [12].



Figure 1. Collage Comparison. Top: eight images from a consumer photo collection. Bottom: three different ways to create a collage: (a) mosaic collage of representative images; (b) picture collage created by Google Picasa; (c) picture collage created by the proposed approach.

Digital tapestry formulates the selection of salient regions and their placement together as a Markov Random Field (MRF) problem. However, artifacts are also introduced along the boundaries of two neighboring salient regions coming from two different images in digital tapestry. Although some artifact removal methods can be used, there still exists obvious artifacts, which makes the final tapestry unnatural or unrealistic as pointed out in [12]. Contrarily, picture collage is different from digital tapestry in three aspects: 1) picture collage introduces an overlay style to avoid artifacts caused by digital tapestry. This collage style is more common in real life and can often be found in the album designed by artist; 2) the oriented placement and the layer ordering of the image are two unique features in picture collage. They substantially improve the visual impression of the results. It is not trivial to apply digital tapestry on our picture collage generation; 3) both formulation and optimization of picture collage are different from those in digital tapestry.

1.2 Our approach

We argue that a nice picture collage should have the following properties: 1) salience maximization. A picture collage should show as many visible salient regions (without being overlaid by others) as possible. 2) blank space minimization. A picture collage should make the best use of the canvas. 3) salience ratio balance. Each image in the collage has a similar salience ratio (the percentage of visible salient region). 4) orientation diversity. The orientations of the images are diverse. This property is used to imitate the collage style created by humans. After salient regions of each image are extracted by an image saliency analysis algorithm, we present a Bayesian formulation of the picture collage by incorporating the above properties.

The picture collage is also related to the rectangle packing problem, which is known to be NP-complete [9][10].

The picture collage is a more challenging problem because of the placement order. Therefore, we design a Markov chain Monte Carlo (MCMC) sampling algorithm for the optimization. To effectively and efficiently explore the posterior, we construct a mixture of three well-designed proposals: local proposal, pairwise proposal, and global proposal. Moreover, a parallel sampling method is proposed for a large number of images. Experimental results show that the proposed optimization approach is very efficient.

The remainder of this paper is organized as follows. Section 2 formulates picture collage as a Bayesian problem. In Section 3, an efficient MCMC scheme is presented. Section 4 gives the experimental results. The conclusion is given in Section 5.

2 Problem Formulation

Given N input images $\{I_i\}_{i=1}^N$ and their saliency maps $\{a_i\}_{i=1}^N$ (saliency map representation will be discussed in subsection 2.1), picture collage arranges all images on a canvas C . We denote all these inputs as observation $z = \{\{I_i\}_{i=1}^N, \{a_i\}_{i=1}^N, C\}$. In a picture collage, each image I_i has a set of state variables $x_i = \{s_i, o_i, l_i\}$, where s_i is the 2D spatial coordinate of the center of image I_i in C , and o_i is the orientation angle. Each image has a unique layer index $l_i \in \{1, 2, \dots, N\}$ such that we can determine the placement order of all the images.

To incorporate several desired properties, such as salience maximization, blank space minimization, salience ratio balance, and orientation diversity as mentioned before, we formulate the picture collage creation in a Bayesian framework, i.e. infer the Maximum a Posteriori (MAP) solution of the state variables $x = \{x_i\}_{i=1}^N$ given the observation z :

$$x^* = \arg \max_x p(z|x)p(x), \quad (1)$$

where $p(z|x)$ is the likelihood model and $p(x)$ is the prior

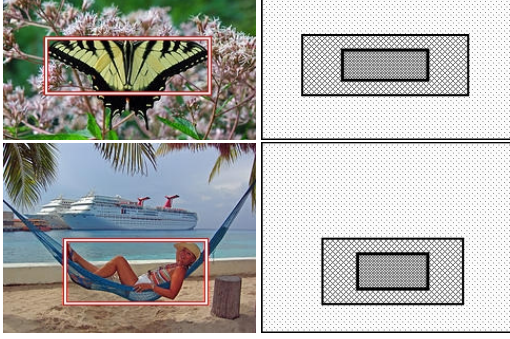


Figure 2. Saliency Representation. In the left column images, the attracted salient area is inside the rectangle. The right column shows the approximated saliency region by two rectangles.

model.

2.1 Saliency Representation

Before creating photo collage, it is necessary to extract salient regions from the input images. In digital tapestry, a high contrast image block is assumed to be salient. In our approach, we extract the biggest salient region for each input image by adopting a visual attention model [6] which combines multi-scale image features (color, texture, orientation) in a discriminative framework.

Then the salient map is approximated by several weighted rectangles. An example is shown in Figure 2. The concentric rectangles in the right column are used to approximate the salient region in the left column. The innermost rectangle has the largest saliency density, where the saliency map is represented using a combination of several (particularly two in Figure 2) rectangles, and each one has a homogeneous saliency density. With this approximated representation, the likelihood evaluation can be efficiently performed by very simple polygon boolean operations. A fast implementation can be obtained from <http://www.cs.man.ac.uk/~toby/alan/software/gpc.html>.

2.2 Likelihood

The likelihood measures the quality of a picture collage given the state variables x . It is modeled as an exponential distribution:

$$p(z|x) = \frac{1}{Z} \exp(-(\bar{A}_{occ} + \lambda_B \bar{B} + \lambda_V V)). \quad (2)$$

where Z is a normalization constant, \bar{A}_{occ} is the normalized sum of occluded saliency regions, \bar{B} is the normalized sum of uncovered regions on the canvas, and V is the variance of saliency ratios. Three terms encode salience maximization, blank space minimization, and salience ratio balance of a picture collage.

Salience Maximization This property aims to maximize the total amount of visible saliency $A_{vis} = \sum_i a_i^{vis}$, where

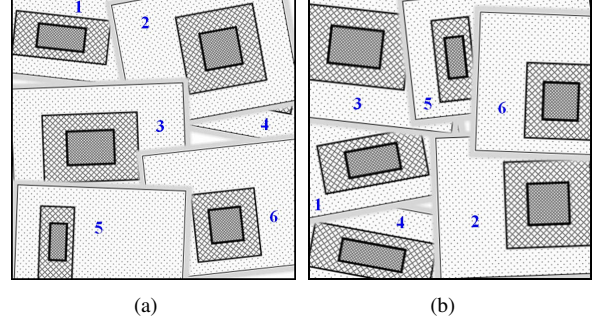


Figure 3. Saliency Ratio Balance. Without saliency ratio balance, some images may be heavily occluded. In (a) most of image 4 is invisible. With saliency ratio balance, (b) is a better balanced collage for the same input images.

a_i^{vis} is the visible part of the saliency region a_i . It is equivalent to minimizing the sum of occluded saliency regions $A_{occ} = A_{max} - A_{vis}$, where $A_{max} = \sum_i a_i$. We further normalize this measure into the range $[0, 1]$:

$$\bar{A}_{occ} = A_{occ}/A_{max}. \quad (3)$$

Blank Space Minimization The blank space is the space in the canvas which is not covered by any image. The blank space is the difference of the canvas bounding rectangle R_C from the union of all the images: $B = R_C - \bigcup_{i=1}^N R_i$, where R_i is the bounding rectangle of image I_i . B should be minimized to make the best use of canvas space. We also compute the normalized term

$$\bar{B} = Area(B)/Area(R_C). \quad (4)$$

Saliency Ratio Balance Due to the canvas size limitation, the visible part of a saliency region may be very small as image I_4 in Figure 3(a). To avoid such a result, a visible saliency ratio balance can be introduced to obtain a well-balanced collage as shown in Figure 3(b). The visible saliency ratio of one image is calculated as $r_i = a_i^{vis}/a_i$. The variance of all saliency ratios

$$V = \frac{1}{N} \sum_{i=1}^N (r_i - \bar{r})^2 \quad (5)$$

is used to evaluate the balance, where $\bar{r} = \frac{\sum_{i=1}^N r_i}{N}$. A better balanced picture collage should have a smaller variance V .

To balance the relative importance of the above three factors, we set the weights λ_B and λ_V in Equation (2) as 0.1 and 0.2.

2.3 Prior

We factorize the prior $p(x)$ in Equation (1) as:

$$p(x) = p(x_s)p(x_o)p(x_l), \quad (6)$$

where $x_s = \{s_i\}_{i=1}^N$, $x_o = \{o_i\}_{i=1}^N$, and $x_l = \{l_i\}_{i=1}^N$. To avoid introducing any bias, it is natural to model the position prior $p(x_s)$ as a uniform distribution over the canvas.

To globally measure the orientation diversity of all images, we compute the average of the absolute orientation differences between any two images I_i and I_j :

$$o_g = \sqrt{\frac{\sum_{i=1, j=1}^N |o_i - o_j|^2}{N(N-1)}}. \quad (7)$$

To encourage orientation diversity, we model $p(x_o)$ as

$$p(x_o) \propto N(o_g; m_g, \sigma_g^2) \prod_i N(o_i; 0, \sigma_o^2), \quad (8)$$

where $\{m_g, \sigma_g\}$ controls the global diversity of orientations. The second term $N(o_i; 0, \sigma_o)$ encourages each image to be placed upright, where the variance σ_o controls the individual diversity.

In order to assign a unique layer index to each image, we model $p(x_l)$ as

$$p(x_l = l_1, \dots, l_N) = \frac{\prod_i \prod_{j \neq i} \delta(l_i, l_j)}{n!}, \quad (9)$$

where $\delta(l_i, l_j) = 1$ if $l_i \neq l_j$, 0 otherwise.

3 Optimization

Our objective function is the combination of Equation (2) and (6). It is a high dimensional, non-convex combinatorial optimization problem which is difficult to be effectively solved. Therefore, we adopt the MCMC algorithm for the optimization since the posterior (1) can be evaluated at a very cheap computational cost, using our approximated saliency representation.

3.1 Markov chain Monte Carlo

Given a distribution $\pi(x)$ of variables x , in our case $\pi(x) = p(x|z)$, MCMC is a strategy for generating samples $\{x^k\}_{k=1}^K$ of $\pi(x)$ by exploring the state space of x using a Markov chain mechanism. This mechanism constructs a chain which spends more time in the regions with higher probability density. The stationary distribution of the chain will be the target distribution $\pi(x)$. In this paper, we pick the sample

$$x^{k*} = \arg \max_{x^k} p(x^k|z)$$

as the MAP solution.

Most MCMC methods are based on the *Metropolis-Hastings* (MH) algorithm. In MH sampling, the proposal function $q(x^*|x^k)$ (also called transition kernel) can be an arbitrary distribution which is used to sample a candidate sample x^* given the current state x . As pointed in [8], it is the key factor that affects sampling efficiency. In other words, whether or not an MCMC approach can effectively sample the target distribution $\pi(x)$ completely depends on how well the proposal function $q(x^*|x^k)$ is designed.

3.2 Proposal Design

There must be many local optimums in our high dimensional, non-convex combinatorial optimization problem. To avoid sticking at local minima, we design a mixture of proposals to deal with this difficulty: 1) a local proposal q_l that discovers finer details of the target distribution, 2) a global proposal q_g that can explore vast regions of the state space of x , 3) a pairwise proposal q_p that has the inbetween behavior. The mixture of proposal $q(x^*|x^k)$ is defined as:

$$q(x^*|x^k) = v_l q_l(x^*|x^k) + v_g q_g(x^*|x^k) + v_p q_p(x^*|x^k), \quad (10)$$

where v_l, v_g and v_p are three weights with $v_l + v_g + v_p = 1$, and they will be dynamically adjusted. Both global and pairwise proposals are critical to make the algorithm jump out from a local minimum.

For clarity, let $\pi(x_i, *) \equiv \pi(x_i, x \setminus x_i)$ when only the state x_i is involved for update. Similarly, $\pi(s_i, *) \equiv \pi(s_i, x \setminus x_i)$ when only the position of image I_i is involved.

3.2.1 Local Proposal

Local proposal only changes the state of one image once. The proposal $q_l(x^*|x^k)$ should determine which image to be selected for update and how to propose a good state for the selected image, in a probabilistic manner.

For the first issue, we compute a weight w_i for each image. This weight is inversely proportional to the visible saliency ratio $r_i = a_i^{vis}/a_i$:

$$w_i = \frac{(r_i + \epsilon)^{-1}}{\sum_i (r_i + \epsilon)^{-1}}, \quad (11)$$

where $\epsilon = 0.2$ is a constant to dilute the influence of this weighting. We select the image I_i with the probability w_i .

For the second issue, the most frequently seen method is Random walk sampling. i.e. adding a random disturbance to the current state configuration. However, in Random walk sampling, it is often the case that a small step-size in the proposal will result in exceedingly slow movement of the corresponding Markov chain, whereas a large step-size will result in a very low acceptance rate. To avoid such "blind" sampling, we propose the following sampling algorithm to make large step-sizes without lowering the acceptance ratio based on Multipoint Metropolis method [11].

To update the state $x_i = \{s_i, o_i, l_i\}$, we randomly select one of following proposals: position proposal, orientation proposal, and layer proposal.

Position Proposal Our position proposal is based on the Random-Grid Sampling [8] (RGS):

- Randomly generate a direction e and a grid size r .
- Construct the candidate set as

$$y_m = s_i^k + m \cdot r \cdot e, m = 1, \dots, M$$

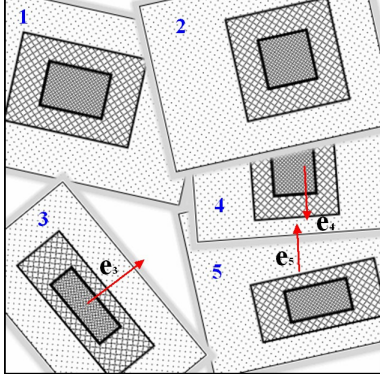


Figure 4. Position Proposal. The direction e_3 is computed for image 3 in blank space driven RGS. The directions e_4 and e_5 are computed for image 4 and image 5 in moveable space driven RGS.

- Draw y from $\{y_m\}_{m=1}^M$ with probability $\pi(y_m)$
- Construct the reference set $\{y_m^r = y - m \cdot r \cdot e\}_{m=1}^M$
- Let $s_i^{k+1} = y$ with probability

$$\min\left\{1, \frac{\sum_{m=1}^M \pi(y_m)}{\sum_{m=1}^M \pi(y_m^r)}\right\}$$

or reject otherwise.

Conceptually, RGS performs a 1D probabilistic search on a random direction such that it can make a large step-size jump from the current state. However, the random sampling of the direction in RGS is still blind. Therefore, we should propose more effective direction e using either *blank space driven RGS* or *moveable direction driven RGS*.

Blank space driven RGS Given a current state configuration, there may be a number of blank regions. In the case there is at least one adjacent blank region B_i for the image I_i (we randomly select one if there are more than one adjacent blank region), we obtain a direction e_i^B from the center of bounding rectangle R_i to the center of the union region $R_i \cup B_i$. One such direction e_3 of image 3 is shown in Figure 4. Then we sample e and r for RGS from two gaussian distributions $N(e; e_i^B, \sigma_e^2)$ and $N(r; m_r, \sigma_r^2)$ respectively. This proposal is in particular useful in the early phase of the sampling when there are many blank regions. We set $M = 10$ in RGS.

Moveable direction driven RGS In the case there is no adjacent blank region for the image I_i , we consider directions $\{e_j\}_{j=1}^n$ from its center to the centers of its n adjacent images $\{I_j\}_{j=1}^n$. First, we denote R_i^a as the bounding rectangle of the saliency region in image I_i . Second, we define a “moveable” distance d_{ij} between images I_i to its neighbor I_j . If the image I_i is above image I_j , the “moveable” distance d_{ij} is the minimal distance between the bounding rectangle R_i^a of the image I_i and the saliency bounding rectangle R_j^a of the image I_j (e.g., from image 4 to image 5 in

Figure 4); Otherwise, d_{ij} is the minimal distance between R_i^a and R_j (e.g., from image 5 to image 4 in Figure 4). Last, we sample a direction e_i^M from the direction set $\{e_j\}_{j=1}^n$ with the probability that is proportional to $\{d_{ij}\}_{j=1}^n$. The final direction e for RGS is again sampled from a gaussian direction $N(e; e_i^M, \sigma_e^2)$. In the case all the distances d_{ij} are 0, a random direction is sampled. This proposal is quite useful in the whole phase of the sampling.

Orientation Proposal The RGS method can be directly applied on orientation proposal because the orientation o_i is a 1D variable. Direction sampling is not necessary. We sample a grid size r from a gaussian distribution $N(r; m_o, \sigma_o^2)$. M is also set as 10.

Layer Proposal To sample layer index l^{k+1} , we do not consider the previous layer index l^k because the layer change will often cause large change of the likelihood. Therefore, we generate the layer index using Multiple-Try Metropolised Independence Sampling (MTMIS) [8]. The basic process is:

- Uniformly draw a trail set of layer index samples $\{y_m\}_{m=1}^M$ for the set $[1, 2, \dots, N]$. Compute $W = \sum_{m=1}^M \pi(y_m, *)$
- Draw a layer index y from the trail set $\{y_m\}_{m=1}^M$ with probability proportional to $\pi(y_m, *)$
- Let $l^{k+1} = y$ with probability

$$\min\left\{1, \frac{W}{W - \pi(y, *) + \pi(l^k, *)}\right\}$$

and let $l^{k+1} = l^k$ otherwise.

We set the number $M = 2N$ so that we have a good chance to search a better layer index in a probabilistic manner. Another big advantage of using multiple-try sampling is that we can incrementally compute $\{\pi(l_i = 1, *), \dots, \pi(l_i = N, *)\}$ in a top-down way such that the computation cost of multiple-try sampling is just twice the cost of a Random walk sampling.

3.2.2 Global Proposal

In global proposal, we also have three proposals for the position, orientation, and layer index x_s , x_o , and x_l .

Position Proposal To make the new sample x_s^{k+1} jump far way from the local minimum, we sample the positions for all images independent of current state x_s^k .

Roughly speaking, all images in a good picture collage should be well separated on the canvas, as shown in Figure 1(c). Therefore, we firstly divide the canvas C into a number of $N^* > N$ squares and randomly select a number of N centers s_i^c of squares without drawback. Then, we sample x_s from the distribution $\prod_i N(s_i; s_i^c, \sigma_{s_i^c}^2)$, where $\sigma_{s_i^c}$ is 1/6 width of the square.

Oriental Proposal The orientation x_o is sampled based on the prior of orientation:

$$q(x_o) \propto \prod_i N(o_i; 0, \sigma_o^2), \quad (12)$$

Layer Proposal The layer index is sampled from its prior distribution as shown in Equation (9). Based on the conditional probability property $p(l_1, \dots, l_N) = p(l_1)p(l_2, \dots, l_N|l_1)$, the whole layer index can be sampled by sequentially sample l_1 and l_2, \dots, l_N . $p(l_1)$ is easily proved to be a uniform distribution and hence can be directly sampled. In a similar way l_2, \dots, l_N with the condition of l_1 are sequentially sampled.

3.2.3 Pairwise Proposal

The acceptance rate of global proposal usually is low compared with local proposal. In order to make the Markov chain have the ability to partially jump away from the local minimum, a pairwise proposal is designed for this goal. It can be viewed as a compromise of local proposal and global proposal. In each iteration, it swaps the positions, orientations, or layer indexes of uniformly selected two different images. This proposal is in particular useful in the early and intermediate phases of the sampling.

3.2.4 Dynamic Weighting

The three weights v_l , v_g and v_p in Equation (10) represent our expectation on the frequencies of the local, global, and pairwise proposals being utilized. On the one hand, when the local proposal can not improve the result in a longer time, the global and pairwise proposals should have larger probabilities to be utilized. So we set $v_l = \exp(-\frac{t^2}{2\sigma_t^2})$, where t is the iteration number that the local proposal does not improve the result continuously, σ_t controls the probability that the local proposal is utilized, and is set as $6N$ so that three types of local proposals have a good chance to be utilized. On the other hand, the pairwise proposal has the potential to partially jump away from local minimum and find local details. So it is desirable to encourage the pairwise proposal to have a higher probability to be utilized than the global proposal. Therefore, we set $v_p = 3v_g$.

3.3 Parallel Sampling

If a large number of images ($N > 32$) are to be placed, the canvas is firstly partitioned into several sub-canvases. Then we perform collage inference in parallel on each sub-canvas. Afterwards, those sub-canvases are packed into the original large canvas. Then, two steps are sequentially performed to refine the collage: 1) view each sub-canvas as a single image and run the local sampling, 2) run the local

and pairwise samplings for the all images on the canvas. An example is shown in Figure 7, in which Figure 7(a) is the packed collage and Figure 7(b) is the refined collage.

4 Experimental Result

We test our approach on visual image summarization for photo collections and image search results. The following experimental results are all obtained with a fixed set of parameters as mentioned before. The canvas is square and its size is set so that its area is about half of the total area of all input images.

To show the efficiency, we present an example to show the changing curve of the energy (minus log of posterior) and the collage evolution process. An example as shown in Figure 5 is performed on 10 images. Parallel sampling is not used for this example. In the curve of Figure 5(d), the x-axis represents the iteration number, and the y-axis is the lowest energy up to the current iteration. From this curve, we can see that the lowest energy decreases quite fast around iteration 10, 600, and 1200. This phenomenon demonstrates that our sampling algorithm can jump away from the local minimum and hence can effectively explore the vast state space. The whole sampling process only takes 8 seconds for this collage on a 2.8GHz desktop PC.

Image search result summarization As an interesting application, the picture collage can be used to summarize image search results. Google's and Yahoo's image search engines using keywords usually return some undesired images (e.g., figure, cartoon). So in our experiment, we just discard such improper images. Two picture collages from Yahoo's image search are shown in Figure 6(a) and Figure 6(b). Picture collages from Google's image search are shown in Figure 5(c) and Figure 6(c). The former two are obtained using the parallel scheme and the latter are not. The above four results are all obtained in 10 - 20 seconds.

Photograph collections summarization Figure 1(c) shows a collage from 8 images obtained within several seconds. A collage from a larger number of images is a more challenging problem. Figure 7(b), Figure 8(a) and Figure 8(b) are picture collages respectively from 32 images, 32 images, and 35 images. All the results are obtained using parallel sampling with four sub-canvases, in 1-2 minutes.

5 Conclusion

In this paper, a Bayesian picture collage framework is proposed to automatically create a visual image summarization for a set of images. Moreover, a very efficient MCMC optimization method is designed to infer the collage.

The future work will include: 1) a better saliency analysis technique, e.g., incorporating face detection and object recognition; 2) collage creation for hundreds or thousands of input images; 3) allowing scale change of input images.



Figure 5. Optimization Process. (a) shows the initial collage where the shaded area is blank space. (b) is the collage after 1000 iterations. (c) is the final collage after 1800 iterations. (d) shows the changing curve of the minus log of posterior. The fast decreasing around iteration 10, 600 and 1200 demonstrates that our algorithm can jump away from the local minimum.

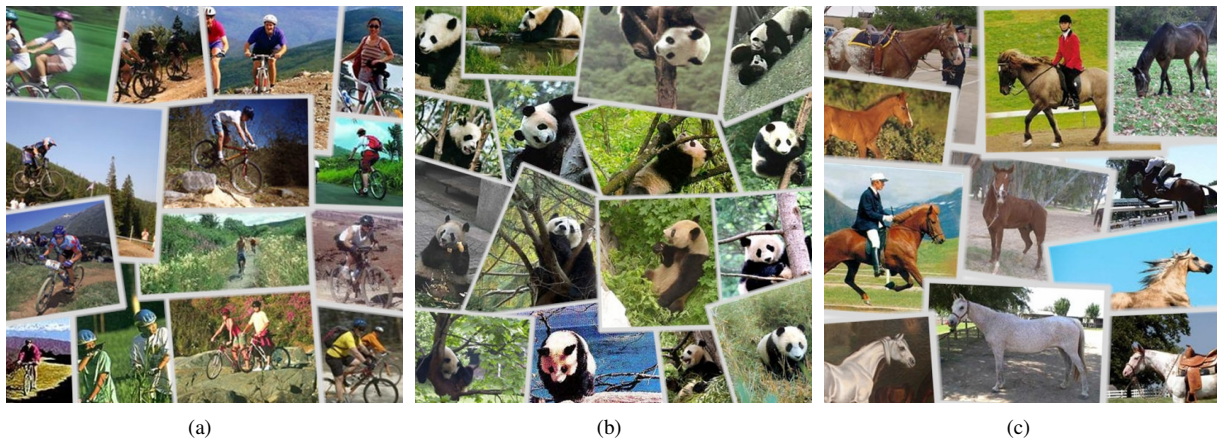


Figure 6. Image Search Results Summarization. (a) the collage of 14 images selected among the first five pages returned by Yahoo’s image searching using the keyword ”cycling mountain”. (b) the collage of 16 ”Panda” images from Yahoo’s image search. (c) the collage formed by 12 ”horse” images from Google’s image search.

Furthermore, we will investigate a quantitative methodology to evaluate collage results by conducting a user study.

Acknowledgments

The work was done when Jingdong Wang was a visiting student at Microsoft Research Asia. Jingdong Wang and Long Quan are supported by the Hong Kong UGC AOE-IT Project AoE/E-01/99 and Hong Kong RGC Grant HKUST6182/04E and HKUST6190/05E.

References

- [1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. F. Cohen. Interactive digital photomontage. *ACM Trans. Graph.*, 23(3):294–302, 2004.
- [2] C. B. Atkins. Adaptive photo collection page layout. In *ICIP*, pages 2897–2900, 2004.
- [3] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(8):1026–1038, 2002.
- [4] L. Fei-Fei, R. Fergus, and P. Perona. A bayesian approach to unsupervised one-shot learning of object categories. In *ICCV*, pages 1134–1141, 2003.
- [5] J. Geigel and A. C. Loui. Using genetic algorithms for album page layouts. *IEEE MultiMedia*, 10(4):16–27, 2003.
- [6] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11):1254–1259, 1998.
- [7] N. Jovic, B. J. Frey, and A. Kannan. Epitomic analysis of appearance and shape. In *ICCV*, pages 34–43, 2003.
- [8] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, 2002.
- [9] V. J. Milenkovic. Rotational polygon containment and minimum enclosure using only robust 2d constructions. *Comput. Geom. Theory Appl.*, 13(1):3–19, 1999.
- [10] H. Murata, K. Fujiyoshi, S. Nakatake, and Y. Kajitani. Rectangle-packing-based module placement. In *ICCAD ’95*, pages 472–479, Washington, DC, USA, 1995. IEEE Computer Society.
- [11] Z. Qin and J. Liu. Multi-point metropolis method with application to hybrid monte carlo. Technical report, Department of Statistics, Harvard University, April 2002. appeared in *J. Comp. Phys.*, 172:827-40.
- [12] C. Rother, S. Kumar, V. Kolmogorov, and A. Blake. Digital tapestry. In *CVPR (1)*, pages 589–596, 2005.



Figure 7. Parallel Collage Sampling. (a) shows a large collage packed from 4 parallel collages. The shaded area is blank space. (b) the refined collage by local and pairwise sampling. The blank space is eliminated and the picture collage is much more appealing.

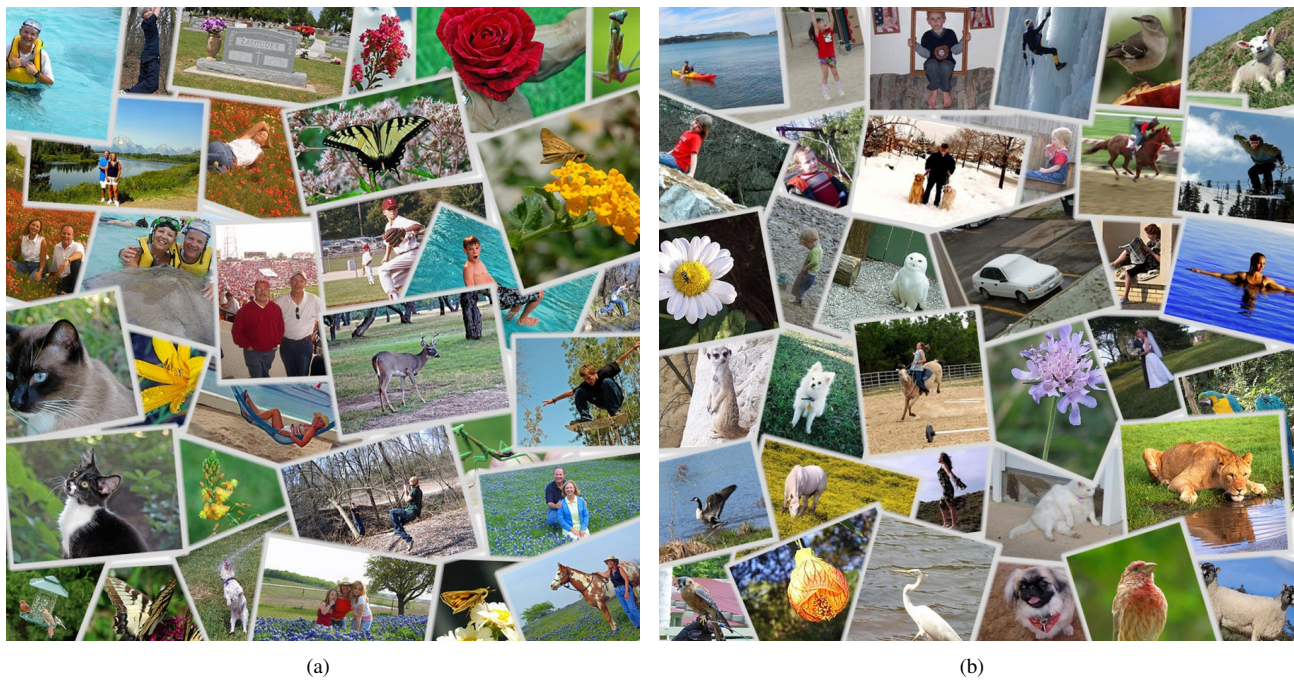


Figure 8. Collage Results for Photograph Collections. The collages in (a) and (b) are constructed respectively from 32 and 35 images selected from two different photograph collections. They are obtained using the parallel sampling within 1-2 minutes.