Vis Comput (2009) 25: 431–439
DOI 10.1007/s00371-009-0346-0

ORIGINAL ARTICLE

An improved belief propagation method for dynamic collage

**Yingzhen Yang · Yichen Wei · Chunxiao Liu ·
Qunsheng Peng · Yasuyuki Matsushita**

Published online: 11 March 2009

Abstract This paper presents a new photo browsing technique, *dynamic collage*. Although previous photo collage techniques have innate advantages for viewing several photos collectively, they only focus on a static two-dimensional arrangement of photos so that the scalability is limited. In dynamic collage, new photos are incrementally inserted into the collage one by one while the old photos are removed accordingly, the positions of photos in the canvas are updated with a local and incremental manner to form a new layout so as to maximize visibility of all the important information embedded in the current collage meanwhile maintaining the visual continuity of two successive collages. To achieve this goal, a carefully designed optimization method based on belief propagation is employed. Unlike most traditional applications of belief propagation on pairwise MRF, we apply belief propagation on factor graph to optimize terms which cannot be represented by pairwise restricted belief propagation. We propose a novel approximate method to reduce the computation complexity, and this approximate method suggests a direction for using belief propagation on factor graph to optimize high order potential functions similar to ours.

Keywords Dynamic collage · Photo collage · Belief propagation · Factor graph · Optimization

Y. Yang · Y. Wei · Y. Matsushita
Visual Computing Group, Microsoft Research Asia, Beijing, China

Y. Yang
e-mail: yzyang@microsoft.com

Y. Wei
e-mail: yichenw@microsoft.com

Y. Matsushita
e-mail: yasumat@microsoft.com

Y. Yang · C. Liu · Q. Peng (✉)
State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China
e-mail: peng@cad.zju.edu.cn

Y. Yang
e-mail: yangyingzhen@cad.zju.edu.cn

C. Liu
e-mail: liuchunxiao@cad.zju.edu.cn

1 Introduction

Following the tremendous development of digital technology and the popularity of multimedia devices, there are a huge number of digital photos in both personal computer and Internet servers today. It is a challenging work to provide a desirable way for users to view a large amount of photos. Recently, a new photo browsing technique, namely, collage, has come into use and it provides a compact and palatable representation for several photos displayed on the same canvas. The main problem of the current collage approaches is the restriction on scalability. All these works can only represent a small number of photos in a canvas. For users who relish the compact representation of collage and also wish to browse massive photo collections, this paper proposes a new method, called dynamic collage. Dynamic collage maintains the benefits of traditional collage while providing a framework to dramatically extend the scalability, which is demonstrated in its application for browsing massive photo collections.

It should be emphasized that the formulation of dynamic collage is also restricted to a two-dimensional canvas. But unlike existing collage techniques, we adopt a different optimization method to adjust the locations of all photos on the canvas. During each iteration, the change of central position

Fig. 1 The results using our photo browsing system as described in Sect. 5 to skim the photo collection 'butterfly.' Photo marked with a *white cycle* is the newly added one, and photo marked with a *red cycle* will be removed in the next collage. After each photo insertion or removal, the photos on the canvas are adjusted in a local and incremental manner to form a new photo collage in real time with an efficient optimization method based on belief propagation on factor graph

of each photo is confined to a predefined neighborhood. This feature of optimization has two main advantages. First, because it is essentially a local optimization problem for each iteration, we can use an efficient local optimization method such as belief propagation algorithm in which every node is mainly affected by its neighbors. Second, this feature ensures the update of the canvas in a visually continuous way. Figure 1 illustrates some frames of dynamic collage generated by our new approach.

Belief propagation has proven to be successful in many early computer vision problems. To avoid exponential computation complexity, many previous applications of belief propagation are restricted on pairwise Markov Random Fields. However, pairwise model is insufficient to represent the information loss due to photo occlusion and blank area which is indispensable in our case. Therefore, we adopt the belief propagation on the factor graph to process high order functions and develop a novel method to approximately compute the messages, reducing the computation complexity significantly.

2 Prior work

Our work of dynamic collage is closely related to traditional photo collage and belief propagation.

Collage creates a compact and visually delectable summary of a photo collection by occluding the unimportant regions of each photo with other photos or letting them out of the canvas, so only the salient and informative part of each photo can be visible on the canvas. In this way, the finite canvas space is sufficiently utilized and a compact representation can be attained [4, 7, 16, 17, 21]. The work of Geigel and Loui [7] can automatically generate a collage from a set of photographs, and this can be viewed as the preliminary work for photo collage. But the collage generated by their method is not compact enough and there exists a lot of blank area. Rother et al. [16, 17] proposed a digital tapestry method that automatically produces a dense and seamless collage from a set of photos which can also be seen as a composite image. Wang et al. [21] retained picture frames just as in Google Picasa [13] and use a MCMC method to optimize a posterior. Diakopoulos and Essa [4] provided a method to create collages with user-defined scenarios.

Apart from photo collage, some researchers extended available collage techniques to video. Christel et al. proposed a method to automatically produce a video collage from news videos [3]. Wang et al. succeeded to include home videos using different arrangement strategies [22].

It is worth mentioning the work for visualizing photo collections. Slideshow is used in many applications [1, 13, 14]. Chen et al. proposed a tiling slideshow that uses both spatial and temporal arrangement of photos for the presentation [2].

Belief Propagation provides a useful tool in dealing with many early computer vision problems such as image restoration, image denoising and shape from shading [5, 6] and it has also proved successful in many other applications [18, 19]. Nevertheless they are mostly restricted to pairwise MRF to avoid exponential computation complexity. Potetz [15] provided a method to use belief propagation on factor graph to process high order potential functions but it requires that these functions should have some special forms. In our dynamic collage problem, it is difficult to express our function in the required form. Instead we propose an approximate method to compute the message effectively, which enables us to use belief propagation on factor graph to optimize our object functions.

3 Problem description

Since our method is primarily based on the photo collage technique, it is necessary to outline the photo collage problem.

3.1 Photo collage problem

The goal of photo collage is to find an optimal arrangement of several photos on a two-dimensional canvas. An optimal arrangement should satisfy the following requirements [16, 21]:

(1) *Visual Information Maximization.* The total information revealed by the visible area of all photos in canvas should be maximized. Several method like [8, 10–12] have been developed for extracting the regions of interest (ROIs) of a photo. In addition, important objects in photos can be detected by methods such as [20]. Photo collage should compute a layout of the candidate photos by which all ROIs and important objects are displayed as much as possible.
(2) *Blank Space Minimization.* The blank space is the region on the canvas not covered by any photo.
(3) *Single Photo Visibility.* For any photo displayed on the canvas, its visible area should be no less than a predefined percentage. The purpose of this constraint is to make sure no photo is seriously occluded by others.

Although previous photo collage can effectively create a compact summary of a photo collection, its scalability is limited due to the finite area of the canvas. To overcome this difficulty, we propose dynamic collage. It provides an efficient way to browse a large number of photos in a collage manner without visual abruptness. New photos are incrementally inserted into the collage one by one while the old photos are removed accordingly, the positions of photos in the canvas are updated with a local and incremental manner to form a new layout. Most importantly, we adopt belief propagation on factor to get rid of the pairwise constraint of belief propagation on MRF so that we could design object functions to meet the above requirements as shown in Sect. 4.1.

3.2 Belief propagation on factor graph

To approach the goal we choose the optimization method based on belief propagation on factor graph. Belief propagation is a method for maximizing the posterior of multivariate probability distribution of the form

$$p(\mathbf{X}) = \prod \phi_i(\mathbf{x_i}), \quad \mathbf{x_i} \subset \mathbf{X}. \tag{1}$$

Such probability distributions are often represented in the form of a factor graph. A factor graph is a bipartite graph in which each potential function $\phi_i(\mathbf{x_i})$ is represented as a factor node f and each element of $\mathbf{X}$ is represented as a variable node v. The max-product belief propagation maximizes the posterior by iteratively computing the messages along each edge in the graph according to the following equations:

$$m^t_{i\to f}(x_i) = \prod_{g\in N(i)\backslash f} m^{t-1}_{g\to i}(x_i), \tag{2}$$

$$m^t_{f\to i}(x_i) = \max_{\mathbf{x}_{N(f)\backslash i}} \left(\phi_f(\mathbf{x}_{N(f)}) \prod_{j\in N(f)\backslash i} m^t_{j\to f}(x_j) \right), \tag{3}$$

$$b^t_i(x_i) \propto \prod_{g\in N(i)} m^t_{g\to i}(x_i), \tag{4}$$

where f and g are factor nodes, i and j are variable nodes, and $N(i)$ is the set of neighbors of node i. It is evident that the class of functions that can be processed by belief propagation on factor graph is dramatically extended if we account for only pairwise MRF, but the computation complexity is always prohibitively high. In Sect. 4.2, we propose a approximation method to reduce the computation complexity of (3) from exponential to quadratic.

4 Dynamic collage

The problem of dynamic collage can be described as follows. Given N input photos $\{I_i\}_{i=1}^N$, their importance map $\{A_i\}_{i=1}^N$ (importance map will be discussed in Sect. 4.1) and their initial states $\mathbf{X} = \{x_i\}_{i=1}^N$, for each photo I_i, it has state $x_i(t_i, \theta_i, l_i)$, where $t_i = (t_i^x, t_i^y)$ is 2D spatial coordinates of center of I_i, θ_i is its orientation angle, and $l_i \in 1, 2, \ldots, N$ is its layer which determines its display order. In order to reduce the size of label set, we choose to split label space,defining the object function for central coordinates and orientation angle respectively. Then we adopt belief propagation to optimize the object functions. To reduce the combination complexity, we divide the optimization process into three stages which optimize the central coordinates, orientation angle and layer consecutively. At each stage we consider that other components of the state as constant. The goal of each stage of optimization is demonstrated below.

4.1 Saliency computation

Before performing dynamic collage, it is necessary to extract the salient part from the input photos. Unlike the ROI-based visual attention model employed by previous photo collage methods [4, 16, 21, 22], which assigns a uniform importance value inside the rectangular ROI, we adopt a more general attention model which can assign pixel-wise importance values using the method described in [9]. Then we generate an importance map for each input photo, each pixel on the map is assigned to an importance value of the corresponding pixel in the original photo. It is therefore possible to capture more than one ROI in a single photo. We apply a state-of-art face detector [20] and set a high importance value to the area of faces. In addition, we calculate an integral map (similar to integral image) for each importance map so that we can easily get the sum of importance values in a rectangle area, namely the information that rectangle contains, in one photo in $O(1)$ time. The information loss due to occlusion or out of the canvas boundary can be easily calculated by inquiring the integral map of the photo that has a lower layer or is not entirely contained in the canvas.

4.2 Central coordinates optimization

At the stage of central coordinates optimization, we restrict the moveable range of each photo to a neighborhood of its initial position for the sake of maintaining a relatively small label set. We define a vector $(\Delta x_i, \Delta y_i)$ for photo I_i, so that the possible values for the central coordinates of photo I_i is restricted within a rectangle defined by $[t_i^x - \Delta x_i, t_i^x + \Delta x_i] \times [t_i^y - \Delta y_i, t_i^y + \Delta y_i]$. Actually we can define a step S so that the number of labels for each variable node is $[(2\Delta x_i/S+1)(2\Delta y_i/S+1)]$. We define the object function as

$$E(\mathbf{X}) = M(\mathbf{X}) + \sum_i O(x_i; \mathbf{x}_{N(i)}) + \sum_{i=1}^{N} B_i(x_i), \tag{5}$$

where $M(\mathbf{X})$ means the area of blank space in canvas which is not covered by any photo. For each photo I_i, $N(i)$ represents the indices of the set of photos that can probably occlude part of photo I_i in their moveable range with a higher layer than that of I_i, and $O(x_i; \mathbf{x}_{N(i)})$ is the information loss of photo I_i due to occlusion. If $N(i)$ is empty, then the corresponding term vanishes. $B_i(x_i)$ means the information loss due to out of boundary. The goal of central optimization is to find the optimal states $\mathbf{X}^*$ which can minimize (5).

To start the optimization, we should establish the factor graph according to (5). There are two types of nodes: factor nodes which indicate potential functions, namely $M(\mathbf{X})$, $O(x_i; \mathbf{x}_{N(i)})$ and $B_i(x_i)$; variable nodes which indicate elements of $\mathbf{X}$. The factor graph model is shown in Fig. 2. Note

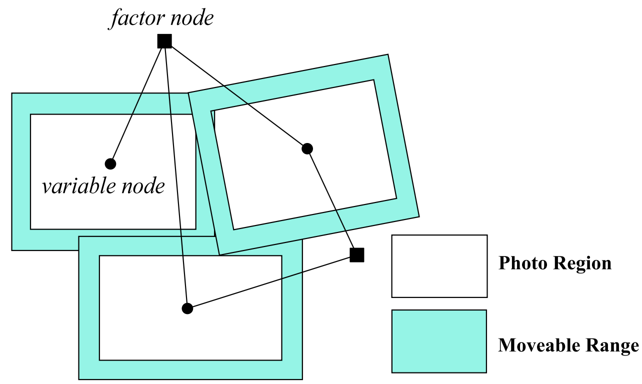

Fig. 2 Illustration of the factor graph model. The state of each photo is represented as a variable node and each potential function is represented as a factor node in the factor graph model. Each factor node is linked to a set of variable nodes which represent the variables of the potential function that factor node represents. The potential functions indicates the information loss and blank area of the canvas which is defined in Sect. 4.2. The movable ranges of all photos form the solution space of the central coordinates optimization problem

that the Belief Propagation on a factor graph can process potential functions with more than two variables, which is impossible in the case of belief propagation on pairwise MRF. The computation of message from a factor node f to a variable has a complexity of $O(M^N)$, where M is the number of labels for each variable, and N is the numbers of the neighbors of f. However, if smartly processed, the complexity of message computation could be reduced to $O(NM^2)$ for some potential functions with special forms described in [15]. Though it is hard for our object functions to be transferred to the form required by [15], we can still show that computation complexity can be significantly reduced. For notational simplicity, we illustrate our method for $N = 3$. We define $M_i \equiv m_{f\to i}$ and $m_i \equiv m_{i\to f}$ (refer to the symbols used in Sect. 3.2). Suppose f represents the potential function $O(x_1; \mathbf{x}_{N(1)})$, so that $N(1) = 2, 3$. We use R_i to denote the region of photo I_i. Suppose that photo R_2 and R_3 do not intersect with each other, then we can split $O(x_1; \mathbf{x}_{N(1)})$ into two parts:

$$O(x_1; \mathbf{x}_{N(1)}) = O(x_1, x_2) + O(x_1, x_3), \tag{6}$$

where $O(x_i, x_j)$ means the information loss of photo I_i due to occlusion by photo I_j. Then we have

$$\begin{aligned} M_1(x_1) &= \inf_{x_2,x_3} \big(O(x_1; \mathbf{x}_{N(1)}) + m_2(x_2) + m_3(x_3)\big) && (7)\\ &= \inf_{x_2,x_3} \big(O(x_1, x_2) + m_2(x_2) \\ &\quad + O(x_1, x_3) + m_3(x_3)\big) && (8)\\ &= \inf_{x_2}\big(O(x_1, x_2) + m_2(x_2)\big) \\ &\quad + \inf_{x_3}\big(O(x_1, x_3) + m_3(x_3)\big). && (9)\end{aligned}$$

It is evident from (9) that the computation complexity of such message is reduced from $O(M^3)$ to $O(M^2)$. When R_2 and R_3 have an intersection, we can approximately calculate the message by fixing x_2 and x_3 in their order.

To illustrate this, we rewrite (6) as

$$O(x_1;\mathbf{x}_{N(1)}) = O(x_1,x_2) + O(x_1,x_3|x_2), \tag{10}$$

$O(x_1,x_3|x_2)$ means the information loss of photo I_1 due to occlusion by photo I_3 excluding its intersection with photo I_2. Because that all the possible central coordinates of photo I_2 are within a neighborhood of its original value, we can treat x_2 as a constant x_2^* in this neighborhood for approximation which will be determined later. Therefore we can rewrite (7) as follows:

$$\begin{aligned} M_1(x_1) &\approx \inf_{x_2,x_3}\big(O(x_1,x_2)+O(x_1,x_3|x_2^*)+m_2(x_2) \\ &\quad + m_3(x_3)\big) \qquad (11)\\ &= \inf_{x_2}\big(O(x_1,x_2)+m_2(x_2)\big) \\ &\quad + \inf_{x_3}\big(O(x_1,x_3|x_2^*)+m_3(x_3)\big). \qquad (12)\end{aligned}$$

After the evaluation of the first inf operation in (12), we can have an optimal value of x_2 which can be assigned to x_2^*.

We then should remove the intersection area of R_1, R_2 from R_1 before evaluating the second inf operation because x_2 is fixed. In order to evaluate $O(x_1,x_3|x_2^*)$, we divide the residual part of R_1 into small k rectangles, namely, $\{R_i^1\}_{i=1}^k$, which do not intersect with each other, as shown in Fig. 3. Then we can have

$$O(x_1,x_3|x_2^*) = \sum_{i=1}^{k} O_R\big(R_i^1,x_3\big). \tag{13}$$

Here $O_R(R_i^1,x_3)$ denotes the information loss of rectangle R_i^1 due to occlusion by photo I_3, which can be evaluated by inquiring the integral photo of I_1.

With this approximate method we can compute the message as if R_2 and R_3 do not intersect. It is easily verified that this method can be extend to arbitrary N. Similarly we write

$$\begin{aligned} O(x_1;\mathbf{x}_{N(1)}) &= O(x_1,x_2)+O(x_1,x_3|x_2) \\ &\quad + O(x_1,x_4|x_2,x_3)+\cdots \\ &\quad + O(x_1,x_N|x_2,\ldots,x_{N-1}), \qquad (14)\end{aligned}$$

where $N(1)=2,\ldots,N$.

And then by performing the same routine, we can get

$$\begin{aligned} M_1(x_1) &\approx \inf_{x_2}\big(O(x_1,x_2)+m_2(x_2)\big) \\ &\quad + \inf_{x_3}\big(O(x_1,x_3|x_2^*)+m_3(x_3)\big) \\ &\quad + \inf_{x_4}\big(O(x_1,x_4|x_2^*,|x_3^*)+m_4(x_4)\big) \\ &\quad + \cdots + \inf_{x_N} O(x_1,x_N|x_2^*,\ldots,x_{N-1}^*). \qquad (15)\end{aligned}$$

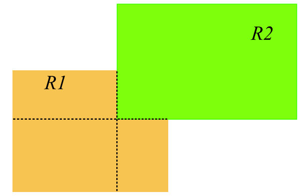

Fig. 3 Illustration of rectangle division. Divide the region of photo 1 excluding its intersection with the photo 2 into small rectangles by *dash line*

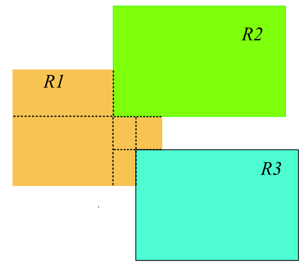

Fig. 4 Illustration of incremental rectangle division. Divide the region of photo 1 excluding its intersection with photo 2 and photo 3 into small rectangles based on former division by *dash line*

We evaluate the inf operation from left to right. For every possible value of x_1, when all the $k-2$ inf operations in (15) are evaluated, we can have a sequence of fixed value $\{x_i^*\}_{i=2}^{k-1}$. Before we evaluate $\inf_{x_k}(O(x_1,x_k|x_2^*,\ldots,x_{k-1}^*)+m_k(x_k))$, we remove the intersection area of R_1, R_{k-1} from R_1 and divide the residual part of R_1 into L_k small rectangles, namely $\{R_i^1\}_{i=1}^{L_k}$. Note that this dividing process is incremental, as shown in Fig. 4. And we have

$$O(x_1,x_k|x_2^*,\ldots,x_{k-1}^*) = \sum_{i=1}^{L_k} O_R\big(R_i^1,x_k\big). \tag{16}$$

This method is computationally efficient. Note that before evaluating the kth inf operation in (15), dividing the residual part of R_1 costs $O(k)$ time and should engender at most $O(k^2)$ small rectangles. We use 2D balanced tree to maintain all these engendered small rectangles, so it takes only $O(k)$ to find the small rectangles which intersect with R_k, and the evaluation of (16) also costs $O(k)$ time. When R_1 is further divided, we add the newly engendered small rectangles into the 2D balanced tree to update it.

When the factor node represents $M(\mathbf{X})$, the strategy used above is also applicable, but this time N is the number of photos displayed on the canvas simultaneously. Because the number of photos accommodated by the canvas simultaneously has a upper bound (for example, 30) and for most occasions the N is no greater than 5, the computation complexity of (15) is $O(N^2)$, instead of the original exponential complexity.

Using these higher-order object functions, it can be easily verified that the object function (5) satisfies all the three requirements of photo collage mentioned in Sect. 3.1, which

Fig. 5 The comparison between the results of applying BP on the factor graph (*right*) and on the pairwise MRF (*left*)

cannot be achieved by belief propagation on pairwise MRF, while the computation complexity remains the same. We implement the BP on factor graph with the proposed approximate message computation method as well as BP on pairwise MRF (consider only pairwise occlusion). Figure 5 shows that our method is obviously superior to the latter for the high utilization rate of the canvas space with almost no blank area.

4.3 Orientation angle optimization

The task of orientation angle optimization is to assign each photo an orientation angle assuming that the central positions is fixed. We use predefined discrete angles $\{n\theta\}_{n=-2}^{2}$, where $\theta = 5$ deg. Since the central positions remain unchanged, the blank area will not be changed much at this stage, and the information loss due to occlusion can be reduced to pairwise pattern for approximation, so the objective function for optimization of orientation angle is as follows:

$$E(\mathbf{X}) = \sum_{i,j} O(x_i, x_j) + \sum_{i=1}^{N} B_i(x_i) \tag{17}$$

where $O(x_i, x_j)$ is information loss due to occlusion between photo pair I_i, I_j, and $B_i(x_i)$ is the same as that in (5). Note that it is a simplified version of (5) and the label set for each variable node is extremely small, so this stage is computationally efficient.

If the photos can be rotated, their intersection is no longer rectangular. Since we have predefined the rotation angle set, for each input photo, we rotate its importance map for each angle, then create a new importance map as the tight bounding box of the rotated importance map, with unoccupied area filled with zeros. So the intersection of rotated photos can also be treated as rectangles using their updated importance maps.

4.4 Layer optimization

At this stage, each photo is assigned an new layer number with fixed central coordinates and orientation angle. The goal of the layer optimization is to set the N numbers $1, 2, \ldots, N$ to all photos on the canvas to reduce the information loss. To optimize the order of layers, we develop a variant of topological sort algorithm. First, we construct a layer graph where each photo is represented as a node in the graph and there is an edge between a pair of nodes if their corresponding photos have an intersection. For each edge (i, j) in the layer graph, we compute a direction d of the edge that indicates a favorable layer order between the two photos I_i and I_j, and the amount of information loss e by the layer order. The weight values $w(i)$ and $w(j)$ of photos I_i and I_j are first computed as the sum of importance values in their overlapped area. The direction d is determined as $i \to j$ if $w(i) < w(j)$, and $j \to i$ otherwise. The direction represents the order of layers. The photo with a smaller weight w should have a smaller label, therefore it will be occluded by the other with a larger weight. Finally, the information loss e is computed using the amount of visual information that is occluded in the layer order as $e = |w(i) - w(j)|$.

If the layer graph is acyclic, a topological sorting method can be employed to generate the desirable order of all photos such that all the edge directions are satisfied. When the layer graph is cyclic, we discard the edge with lowest information loss until a node with no in-edge appears so that the topological sorting can proceed. In this way, we can maintain the direction of edges with relatively small information loss so that the final order attained is optimal. Because of the efficiency of topological sorting, this optimization stage is very fast.

4.5 Combining the three optimization stages

We run the above three stage optimization processes in their order as an integral optimization cycle. Each time the cycle is performed, all photos on the canvas change their states, and they move to a new location within the neighborhood of their initial position. We detect the total information loss of all photos in the canvas after each iteration, and terminate the iteration process when the value of total information loss no longer declines. Obviously, the most influential optimization stage is the central coordinates optimization, which directly determines the positions of photos on the canvas. In order to improve the efficiency of the optimization, we define a two-scale method to accelerate the optimization cycle. In a canvas with size of 1024×768, and 12 photos with average size of 300×300 we perform only central coordinates optimization at the first level (coarse level) with a relatively large step ($S = 10$), and set $\Delta x = \Delta y = 40$ for all photos. For the second level (fine level), we set a smaller step ($S = 5$), and set $\Delta x = \Delta y = 10$ for all photos, then conduct the whole optimization cycle. Experiment results show that the first level optimization costs 0.6 seconds and the second level optimization costs 0.2 seconds with Pentium 4, 3 GHz CPU, which evinces the efficiency of our dynamic collage technique. This is compared with Picture Collage [21] which costs 8 seconds to perform sampling process for 10 photos. Because of the efficiency of our optimization, we can guarantee the real-time rendering of the browsing system described in the next section.

5 Application for visualizing massive photo collections

We develop a photo browsing system based on dynamic collage to facilitate users to browse large photo collections. The overview of the system is shown in Fig. 6. There are four modules in the system. Saliency computation module computes the importance maps of the input photos, which can come from either Internet or local disk. The scheduler module is responsible for deciding when to update the canvas and which photo should be inserted into canvas or taken out of canvas. In our implementation, we choose the constant number of photos in a canvas and each time we remove the oldest photo in canvas which has been in it for the longest time and insert a new one. The optimization module runs the optimization process described in Sect. 4 after the new photo is inserted or the old one is removed. Note that we only update the canvas after the dynamic collage optimization module has got the optimal states for each photo, then the rending module renders a smooth transition from the old state to the new one by linear interpolation for each photo on the canvas.

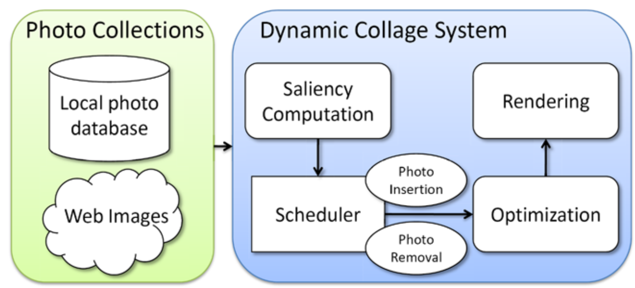

Fig. 6 The overview of our photo browsing system based on dynamic collage

Fig. 7 The results using our photo browsing system as described in Sect. 5 to skim the photo collection 'flower'

6 Experimental results and discussion

We use our photo browsing system mentioned in Sect. 5 to skim two photo sets, 'butterfly' and 'flower.' Figures 1 and 7 show respectively the results. The local and incremental feature of our dynamic collage algorithm ensures the visual continuity when the canvas is dynamically updated, which is demonstrated in our video demos. The experimental results mentioned in Sect. 4.5 show the efficiency of the dynamic collage algorithm, the comparison between belief propagation on pairwise MRF and factor graph as shown in Fig. 5 indicating that belief propagation on factor graph can fully capture the optimal positions of photos on the canvas.

In this paper, we propose a new photo visualization technique, dynamic collage, which employs an optimization method with a local and incremental manner based on belief propagation to form a standard photo collage. Dynamic collage can ensure the visual continuity of dynamic change of the current collage, which is useful for creating modifiable collage and browsing massive photo collections.

Acknowledgements We thank the Microsoft Research Asia for providing us with all the test photo sets and for the stimulating discussions on the algorithm design, especially the idea of incorporating belief propagation with the novel method of layer optimization. Most of the implementation work of the browsing system in Sect. 5 is accomplished when the first author worked in the Microsoft Research Asia as a student.

References

1. ACDSee: http://www.acdsee.com/
2. Chen, J.C., Chu, W.T., Kuo, J.H., Weng, C.Y., Wu, J.L.: Tiling slideshow. In: Proceedings of ACM Multimedia 2006, pp. 25–34 (2006)
3. Christel, M.G., Hauptmann, A.G., Wactlar, H.D., Ng, T.D.: Collages as dynamic summaries for news video. In: Proceedings of ACM Multimedia 2002, pp. 561–569 (2002)
4. Diakopoulos, N., Essa, I.: Mediating photo collage authoring. In: Proceedings of ACM Symposium on User Interface Software and Technology 2005, pp. 183–186 (2005)
5. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. Int. J. Comput. Vis. **70**(1), 41–54 (2006)
6. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning low-level vision. Int. J. Comput. Vis. **40**(1), 25–47 (2000)
7. Geigel, J., Loui, A.: Using genetic algorithms for album page layouts. IEEE Multimed. **10**(4), 16–27 (2003)
8. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)
9. Kang, H., Matsushita, Y., Tang, X., Chen, X.: Space-time video montage. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 2, pp. 1331–1338 (2006)
10. Liu, T., Sun, J., Zheng, N.N., Tang, X., Shum, H.Y.: Learning to detect a salient object. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07), pp. 1–8 (2007)
11. Ma, Y.F., Zhang, H.J.: Contrast-based image attention analysis by using fuzzy growing. In: Proceedings of ACM Multimedia 2003, pp. 374–381 (2003)
12. Mei, T., Zhu, C.Z., Zhou, H.Q., Hua, X.S.: Spatio-temporal quality assessments for home videos. In: Proceedings of ACM Multimedia 2005, pp. 439–442 (2005)
13. Picasa: http://picasa.google.com/
14. PicLens: http://www.piclens.com/
15. Potetz, B.: Efficient belief propagation for vision using linear constraint nodes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07), pp. 1–8 (2007)
16. Rother, C., Bordeaux, L., Hamadi, Y., Blake, A.: Auto-collage. ACM Trans. Graph. **25**(3), 847–852 (2006)
17. Rother, C., Kumar, S., Kolmogorov, V., Blake, A.: Digital tapestry. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 589–596 (2005)
18. Sun, J., Zheng, N.N., Shum, H.Y.: Stereo matching using belief propagation. IEEE Trans. Pattern Anal. Mach. Intell. **25**(7), 787–800 (2003)
19. Tang, K.L., Tang, C.K., Wong, T.T.: Dense photometric stereo using tensorial belief propagation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 132–139 (2005)
20. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01), vol. 1, pp. 511–518 (2001)
21. Wang, J., Sun, J., Quan, L., Tang, X., Shum, H.Y.: Picture collage. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 1, pp. 347–354 (2006)
22. Wang, T., Mei, T., Hua, X.S., Liu, X.L., Zhou, H.Q.: Video collage: a novel presentation of video sequence. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME'07), pp. 1479–1482 (2007)

Yingzhen Yang received the Bachelor degree in the College of Computer Science and Technology in Zhejiang University in 2006, and the Master degree in the same college in 2008. Y. Yang now working in the State Key Lab of CAD&CG of Zhejiang University under the supervision of Professor Qunsheng Peng.

Yichen Wei obtained Ph.D. degree in 2006 from Hong Kong University of Science and Technology and joined Visual Computing Group, Microsoft Research Asia in Sept. 2006. His research interest is in computer vision, including video analysis (tracking, stabilization, annotation, etc.) and structure from motion (stereo matching, image based modeling and camera calibration).

Chunxiao Liu born in 1979, Ph.D. candidate at the State Key Lab of CAD&CG, Zhejiang University, China. His research interests include image and video completion, image and video based rendering and virtual reality.

Yasuyuki Matsushita is the leading researcher in Microsoft Research Asia, his research interest lies in photometric methods in computer vision and computer graphics as well as video analysis for videoquality improvement.

Qunsheng Peng born in 1947. He received a Ph.D. from University of East Anglia, UK in 1983 and is currently a professor at the State Key Lab of CAD&CG, Zhejiang University, China. His research interests include virtual reality, realistic image synthesis, computer animation, and scientific data visualization.