MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
13-32

Adaptive Image Translation for Painterly Rendering

Kenji Hara Kohei Inoue Kiichi Urahama
Department of Visual Communication Design, Kyushu University
4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan.
{hara, k-inoue, urahama}@design.kyushu-u.ac.jp

Abstract

In the paper, we present a new method of converting a photo image to a synthesized painting image following the painting style of an example painting image. The proposed method uses a hierarchical and adaptive patch-based approach to both the synthesis of painting styles and preservation of scene details. This approach can be summarized as follows. The input photo image is represented as a set of patches divided adaptively using a distance transform technique. Then the mapping between the input photo and example painting images is efficiently inferred using Bayesian belief propagation recursively.

1 Introduction

An important task of computer vision and/or pattern recognition is to render an image in a different style. For example, one may want to modify image appearance into a more artistic one, while preserving the content. For this problem, given an input image (photograph) and a source image (painting), we want to estimate an underlying image which has the content of the input image and the painting style of the source image (Figure 1). This estimate is important for various vision problems; for example photograph-painting discrimination (differentiating paintings from photographs), image restoration, and image super-resolution (estimating high frequency details from a low-resolution image).

Several researchers have applied statistical learning approaches to the image translation problem. Typically, the input and inferred images are devided into spartially overlapping patches (local images), and each inferred image patch is connected to its corresponding input image patch and to its spatial neighbors based on a Markov assumption [1]. Then, each patch of the inferred image is estimated by learning the network parameters using Bayesian belief propagation [4], but this approach requires an aligned image pair consisting of the original and translated version of an image for training, despite that an aligned image pair is not often available. Recently, an extension of the previous methods was developed by Rosales et al [3]. They used a finite set of patch transformations to remove this limitation. Their method requires only one image with the desired style (i.e., source image) instead of a pair of perfectly aligned original/translated images. One common disadvantage of the above methods is that they are limited to an image representation based on a set of uniformly sized patches, since each patch is assigned to one node of a single Markov network based on a four-connected neighborhood. In this case, patch size uniformity may lead to a significant decrease in image quality; for example relatively larger patches (low-resolution patches) are mapped to high frequency details (*e.g.*, edges) of the input image (Figure 2: left) or, for too small sized patches (high-resolution patches), the painting style of the source image may not appear in the output image (Figure 2: right). Our approach, applying belief propagation based on the Markov assumption, is somewhat similar to those of previous learning-based methods. However, our algorithm builds multiple Markov networks to solve the above problem.

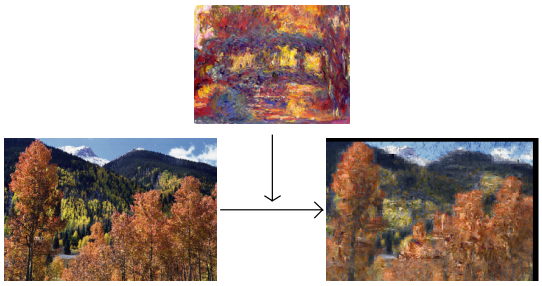


Figure 1. Input, source and output images.

We propose a hierarchical and adaptive technique that converts an input image to a synthesized image following the style of a source image. The input image is represented as a non-uniform adaptive patch resolution using a multi-level hierarchy of uniform patches based on an edge-based distance transform [2]. Each of the new images is generated by solving the corresponding Markov network from the coarser levels to the finer levels. We apply belief propagation in each Markov network.

Our method has similarities to the hierarchical and patch-based method of Wang et al. [4]. However, our method is fully automatic; their method must manually segment the input image into regions, as well as select stroke textures from a source image. The goal of our method is to render a new image in the style of the source image while preserving the edges and details of the input image. We present several synthesized images that are compared to the input and source images.

Figure 2. Images rendered using uniformly sized patches (left: large sized (low-resolution) patches, right: small sized (high-resolution) patches).

2 Image Translation by Belief Propagation

2.1 Formulation

We divide the input and output images into a set of overlapping $N \times N$ block images and then we decompose each block image into the length (called the *block intensity*) and the normalized unit length N^2 dimensional vector (called the *block pattern vector*). Also, we extract a set of $N \times N$ normalized block images from the source image. The similarity between two block images is defined as the Euclidian norm between their block pattern vectors. Then, we assign each block of the output image to one network node. At each node, we select as candidates a set of (10 or 15) block pattern vectors from source image which are the most similar to the corresponding block of the input image. We find the best candidates X= $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ from the finite set of candidates based on the framework of discrete optimizaation.

Now, we define the cost function to be minimized is defined as follows:

$$\hat{\mathcal{X}} = \underset{\mathcal{X}}{\operatorname{argmin}} \left(cost_1(\mathcal{X}) + \eta \, cost_2(\mathcal{X}) \right)$$

where η is a weighting factor, and the first term $cost_1(\cdot)$ prevents the same patterns from being assigned to neighboring nodes as follows:

$$cost_1(\mathcal{X}) = \sum_{(i,j)} \delta(\mathbf{x}_i, \mathbf{x}_j)$$

where (i, j) indicates neighboring nodes i and j. $\delta(\cdot;\,)$ is the delta function.

The second term $cost_2(\cdot)$ enforces the constraint that the corresponding pixel values in the overlap region between neighbors agree as follows:.

$$cost_2(\mathcal{X}) = \sum_{(i,j)} \psi(\mathbf{x}_i, \mathbf{x}_j)$$

where $\psi(\cdot;\,)$ is defined as:

$$\psi(\mathbf{x}_i, \mathbf{x}_j) = |\mathbf{x}_{ij} - \mathbf{x}_{ji}|^2$$

where $\mathbf{X}_{ij}$ is the vector representing a set of the pixels belonging to node i, overlapping with node j. $|\cdot|$ is the Euclid norm.

2.2 Belief Propagation

In our work, we find an approximate solution to the optimization problem with a Bayesian belief propagation [6] [10]. Solving the above problem is equivalent to maximizing the joint probability based on a Markov random field as:

$$\begin{aligned} P(\mathcal{X}) &\propto \exp\left(-\frac{cost_1(\mathcal{X}) + \eta \, cost_2(\mathcal{X})}{2\sigma^2}\right) \\ &= \exp\left(-\frac{\sum_{(i,j)} (\delta(\mathbf{x}_i, \mathbf{x}_j) + \eta \, \psi(\mathbf{x}_i,}{2\sigma^2}\right. \\ &= \prod_{(i,j)} \Psi(\mathbf{x}_i, \mathbf{x}_j) \end{aligned}$$

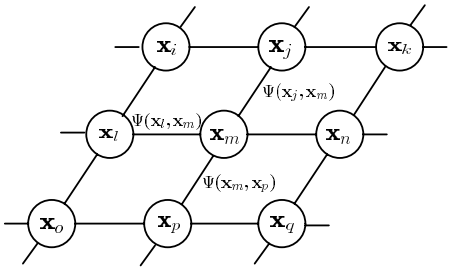


Figure 3. Markov network used for belief propagation.

where $\psi(x_i, x_j)$ is defined as:

$$\Psi(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\delta(\mathbf{x}_i, \mathbf{x}_j) + \eta \, \psi(\mathbf{x}_i, \mathbf{x}_j)}{2\sigma^2}\right)$$

We connect each block image to its spatial neighbors (Figure 3) and then find the following solution

$$\hat{\mathbf{x}}_j = \underset{\mathbf{x}_j}{\operatorname{argmax}} \prod_k M_j^k$$

where k runs over all node neighbors of node j, and M_j^k is the message from node k to node j. We calculate M_j^k from:

$$M_j^k = \max_{\mathbf{x}_k} \Psi(\mathbf{x}_j, \mathbf{x}_k) \prod_{l \neq j} \tilde{M}_k^l$$

where $\tilde{M}_k^l$ is M_k^l from the previous iteration. The initial M_j^k's are set column vectors of 1's, of the dimensionality of the variable x_j. While the expression for the joint probability does not generalize to a network with loops, we nonetheless found good results using these update rules.

3 Extending to Adaptive Image Translation

The methods in the previous section is limited to an image representation based on a set of uniformly sized patches. this may lead to a significant decrease in image quality. We propose a image translation method of generating a non-uniform adaptive patch resolution using a multi-level hierarchy of uniform patches based on an edge-based distance transform.

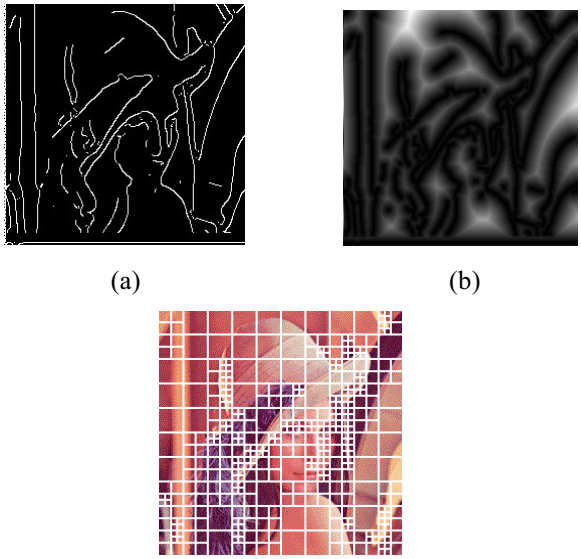
3.1 Adaptive Block Rearrangement on DistanceTransform

We provide a detailed description of different levels of resolution by using image blocks of different sizes. The detailed procedure is described in the following, together with an example illustrated in Figures 4-5.

1. Estract edges from the input image (Figure 4(a)).
2. Compute the distance transform of the binary image (edge points and non-edge points) (Figure 4(b)).
3. Divide initially the distance transformed image into the overlapping block images (Section 2).
4. Subdivide recursively each block image into four overlapping block images (Figure 5) until the

average pixel values within a block image is less than the user-defined threshold.

5. Divide the original input image using the obtained block images (Figure 4(c)).



(a) (b) (c)

Figure 4. Reconstruction of block images.

3.2 Successive Belief Propagation

In this section, applying successively the belief propagation for the multi-level hierarchy of uniform blocks, we will synthesize painting styles while preserving scene details.

First, for a set, $\{\hat{x}_1^{(0)},\cdots,\hat{x}_P^{(0)}\}$, of blocks comprising the lowest resolution, we use the belief propagation scheme described in Section 3.1 (Figure 5(a)). For the resulting region y_0, we transform the next lowest resolution by solving

$$\hat{\mathcal{X}}_1 = \underset{\mathcal{X}_1}{\operatorname{argmin}} (cost_1(\mathcal{X}_1) + \eta\, cost_2(\mathcal{X}_1) + \eta\, cost_3(\mathcal{X}_1, \mathcal{Y}_0))$$

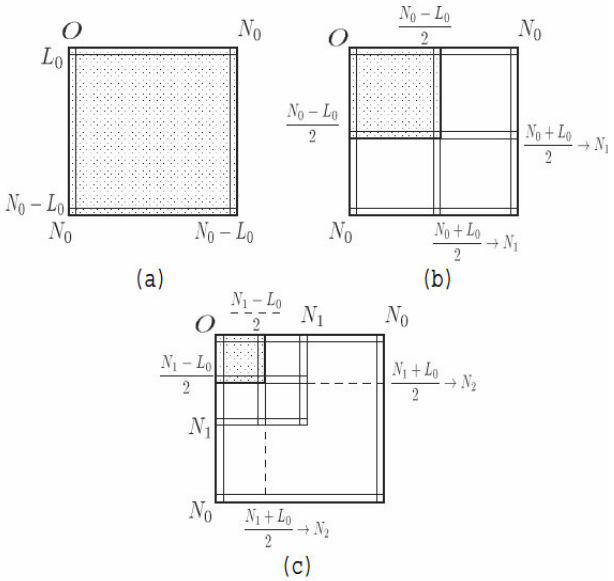
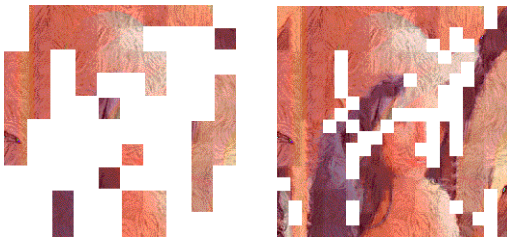


Figure 5. Recursive division for patches.



(a) (b)

Figure 5. Hierarchical image translation

where $cost_1(\cdot)$, $cost_2(\cdot)$ and η are the same as those of Section 2.1. $\cos t_3(\cdot)$ is given by

$$cost_3(\mathcal{X}_1, \mathcal{Y}_0) = \sum_{<k>} \phi(\mathbf{x}_k^{(1)}, \mathbf{y}_k^{(1)})$$

where $< k >$ means that k-th image block $x_k^{(1)}$ and y_0 have overlapping, $y_k^{(1)}$ is the overlapping region with $x_k^{(1)}$, $\phi(x_k^{(1)}, y_k^{(1)})$ is the squired error (Figure 6(a)). Then, we estimate $\hat{x}_1$ by maximizing

$$\exp(-\frac{cost_1(\mathcal{X}_1) + \eta\, cost_2(\mathcal{X}_1) + \eta\, cost_3(\mathcal{X}_1, \mathcal{Y}_0)}{2\sigma^2}) = \prod_{(i,j)} \Psi(\mathbf{x}_i, \mathbf{x}_j) \prod_{<k>} \Phi(\mathbf{x}_k, \mathbf{y}_k)$$

where x_i and y_i denote as $x_i^{(1)}$ and $y_i^{(1)}$, respectively, $\Phi(x_k, y_k)$ is given by

$$\Phi(\mathbf{x}_k, \mathbf{y}_k) = \exp(-\frac{\eta\, \phi(\mathbf{x}_k, \mathbf{y}_k)}{2\sigma^2})$$

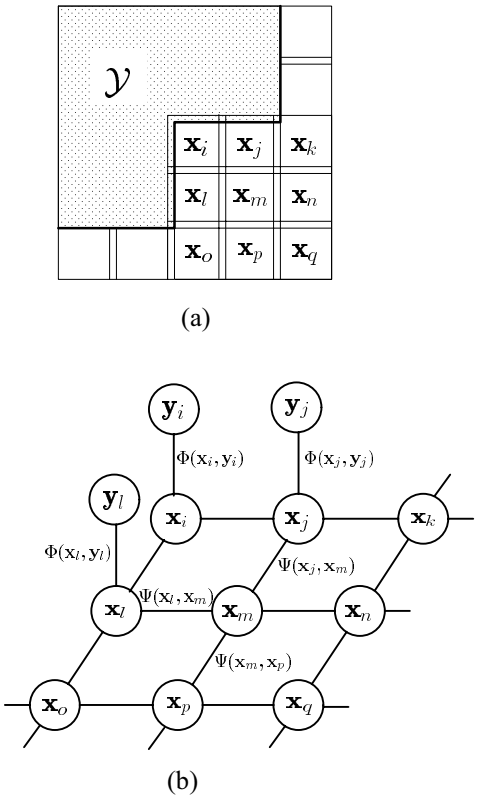


Figure 6. Markov network used for belief propagation in high resolution region.

The estimate is obtained by solving

$$\hat{\mathbf{x}}_j = \arg\max_{\mathbf{x}_j} \Phi(\mathbf{x}_j, \mathbf{y}_j) \prod_k M_j^k$$

where k runs over all node neighbors of node j, and M_j^k is the message from node k to node j. We iteratively calculate M_j^k from:

$$M_j^k = \max_{\mathbf{x}_k} \Psi(\mathbf{x}_j, \mathbf{x}_k) \Phi(\mathbf{x}_k, \mathbf{y}_k) \prod_{l \neq j} \tilde{M}_k^l$$

The resulting region is then merged to y_0 (Figure 6(b)). The above procedure is performed from low resolution to higher resolution until the whole region is transformed.

4 Results

All the following synthetic images have been generated using the overlapping width $l_0 = 2$ and the initial block size $N_0 = 26$. Given the top input image (Lena) and the middle left source image (Gogh: Self-Portrait), the synthetic images at the bottom left and the top right of Figure 7 have been generated using our algorithm and the conventional method [3], respectively. Also, the second synthetic image at the bottom right of Figure 8 has been generated from the top input image and the middle right source image (Cezanne: Mont Sainte-Victoire). Finally, the synthetic image at the bottom of Figure 8 has been generated from the top input image and the middle source image (Monet: The Japanese Bridge).

5 Conclusion

We have proposed a hierarchical and adaptive technique to convert an input image to a synthesized image following the style of a source image. One of disadvantages of the conventional image translation methods is that they are limited to an image representation based on a set of uniformly sized patches, and it may lead to a significant decrease in image quality. To overcome this problem, we have employed a multiresolution analysis approach. As future works we plan to extend our framework to 3D art.

References

[1] B. Freeman, E. Pasztor, and O. Carmichael: "Learning low-level vision," *International Journal of Computer Vision*, vol.40, no.1, pp.25-47, 2000.

[2] K. Inoue and K. Urahama: "Generation of strokes in oil-painting-like images based on distance transform," *IEICE Trans. on Information and Systems*, vol. J87-A, no.4, pp. 580-582, 2004.

[3] R. Rosales, K. Achan, and B. Frey: "Unsupervised image translation," *Proc. International Conference on Computer Vision, ICCV2003*, vol. I, no.4, pp. 472-478, 2003.

[4] B. Wang, W. Wang, H. Yang, and J. Sun: "Efficient example-based painting and synthesis of 2nd directional texture," *IEEE Trans. Visualization and Computer Graphics*, vol. 10, no.3, pp. 266-277, 2004.

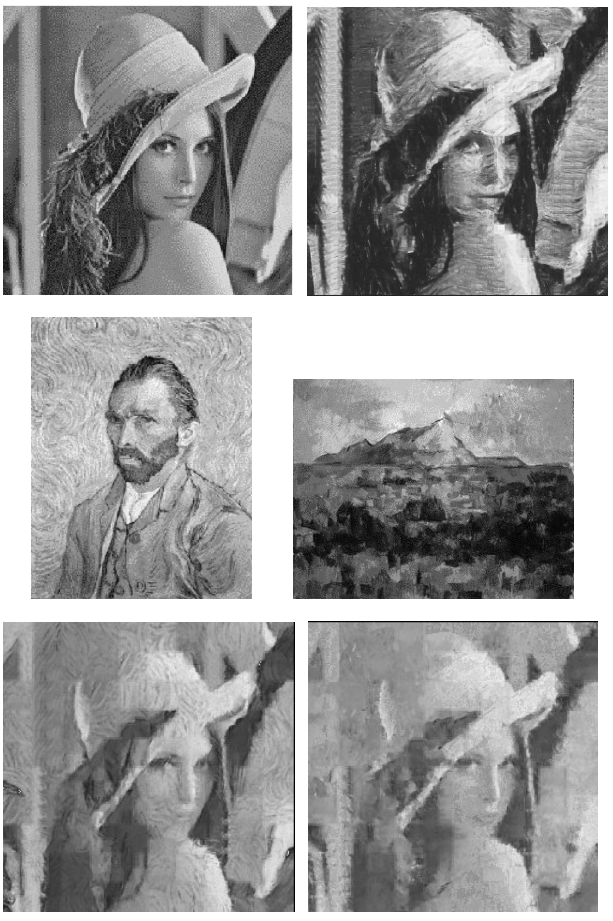


Figure 7. Experimental result.

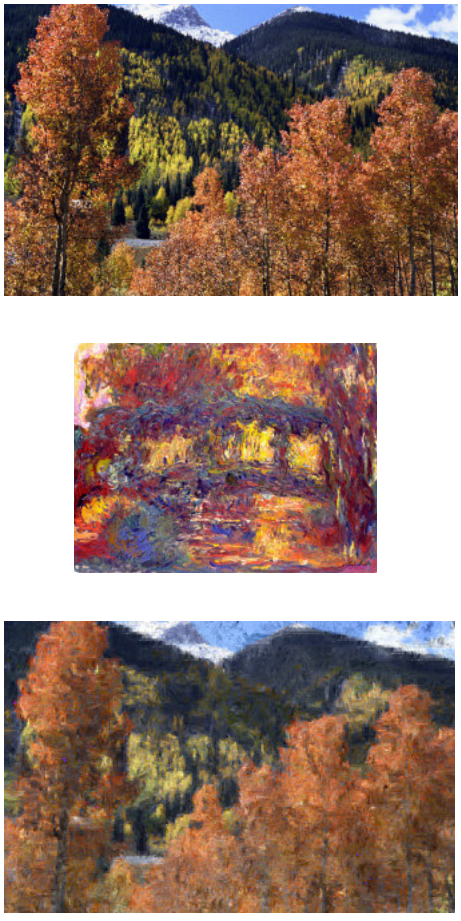


Figure 8. Experimental result.