International Journal of Multimedia and Ubiquitous Engineering
Vol. 10, No. 3 (2015), pp. 127-134
http://dx.doi.org/10.14257/ijmue.2015.10.3.13

Measuring Image Similarity Based on Shape Context

Canlin Li and Shenyi Qian

School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, China
lcl_zju@aliyun.com

Abstract

Measuring image similarity is important for a number of image processing applications. The goal of research in objective image similarity assessment is to develop quantitative measures that can automatically predict perceived image similarity. In this paper, we propose a new objective approach of measuring image similarity based on shape context. We take the geometric structures of objects into account during measuring the image similarity by virtue of shape context which is a robust and compact, yet highly discriminative descriptor. Firstly we find visual salient regions of images by virtue of a regional contrast based saliency extraction algorithm and employ shape context to describe the shape of visual salient region. Then we detect shape deformations of visual salient regions between two images through estimating shape context distances, and accordingly compute the image similarity values. Real data have been used to test the proposed approach and very good results have been achieved, validating it.

Keywords: *Image Similarity; Shape Context; Visual Salient Region; Shape Distance*

1. Introduction

Measuring the distance or similarity between images is a fundamental and open problem in a large number of image processing and computer vision applications, which include image coding, restoration, denoising, halftoning, segmentation, communication, target detection, image registration, and object recognition. For applications in which images are ultimately to be viewed by human beings, an obvious and accurate method of measuring image similarity is the subjective evaluation based on the human perception. In practice, however, subjective evaluation is usually too inconvenient, time-consuming and expensive. It is necessary to use an objective measure to evaluate image similarity.

Objective image similarity metrics can be roughly classified into intensity-based metrics and geometry-based metrics [1]. Intensity-based similarity metrics assume that the images being compared are at the same scale and are perfectly registered, and their similarity is determined from a comparison of the corresponding pixel intensities. As for this category of metrics, the most commonly used metric is Euclidean distance, which converts images into vectors according to gray levels of each pixel, and then compares intensity differences pixel by pixel. Since Euclidean distance discards image structures, it cannot properly represent the real similarity between images. If a small variation occurs in similar images, a large Euclidean distance between the images could arise. Another commonly used intensity-based metrics are mean squared error (MSE) and peak signal-to-noise ratio (PSNR). It has been shown that MSE and PSNR lack a critical feature: the ability to assess image similarity across distortion types. Generally, a common drawback of existing intensity-based metrics is their high sensitivity to geometric and scale distortions. This becomes a big problem when there are small translations, rotations, or scale differences between the images being compared.

All the intensity-based metrics described above are point operations. In other words, the similarity evaluation at one pixel is independent of all other pixels in the image. However, neighboring image pixels are highly correlated with each other. To take advantage of such correlations, geometry-based metrics establish pixel correspondences between the images based on intensity, and then determine similarity by comparing the geometric transformations between corresponding pixels. Recently, the structural similarity (SSIM) [2] was introduced by Wang, Bovik, et al that also accounts for spatial correlations. In SSIM, the structural information of an image is defined as those attributes that represent the structures of the objects in the visual scene, apart from the mean intensity and contrast. Thus, the SSIM metric compares local patterns of pixel intensities that have been normalized for mean intensity and contrast, and measures similarity across distortion types. In addition, some geometry-based similarity metrics compare edge images. These involve Pixel Correspondence Metric (PCM) [3], Closest Distance Metric (CDM) [3, 4], Figure of Merit (FOM) [5] and Partial Hausdorff Distance Metric (PHDM) [6] *etc*. All of these metrics allow for small localization errors between the structures being compared. Most of these metrics operate in the spatial domain. For these geometry-based methods, correspondences between pairs of pixels in the two images is not assumed, but is established before the metric is computed. This process can be computationally complex.

Intensity-based similarity metrics are common, but they have limited performance within a distortion type. Geometry-based metrics measure similarity with greater accuracy and across distortion types, but incur greater computational cost.

According to the human perception, the human visual system is very sensitive to shape variations of some visual salient regions when two images are compared. Thus, for measuring the similarity between two images, we should pay more attention to shape preservation of visual salient regions. Therefore, we may take the geometric structures of objects into account during measuring the image similarity as the same way in geometry-based similarity metrics, but we have no need to establish the correspondences between pairs of pixels, since we mainly focus on visual salient regions. As a favorable result, the lower computational cost as well as the better accuracy is expected to achieve.

From the above idea, this paper proposes a novel approach of measuring image similarity based on shape context. Shape context is a more robust and compact, yet highly discriminative descriptor [7]. We employ shape context to describe the shape of visual salient region. We will detect shape deformations of visual salient regions between two images through estimating shape context distances, and accordingly compute the image similarity values. The remainder of this paper is organized as follows. Section 2 illustrates the flowchart of the proposed image similarity measure based on shape context. Section 3 explores how to detect visual salient regions of image. Section 4 introduces shape context and measuring shape context distance. Section 5 elaborates how to determine the proposed image similarity measure based on shape context. Section 6 provides the experimental results. Finally, the paper is concluded in Section 7.

2. Flowchart of the Proposed Approach

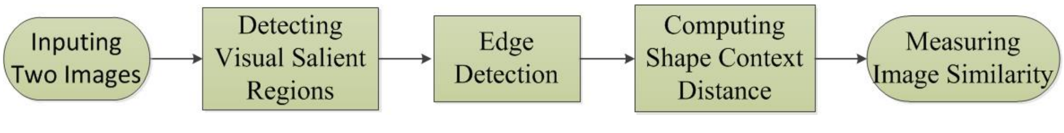
Inputing Two Images
Detecting Visual Salient Regions
Edge Detection
Computing Shape Context Distance
Measuring Image Similarity

Figure 1. Flowchart of the Proposed Approach on Measuring Image Similarity

We illustrate the flowchart of image similarity measure based on shape context proposed by us in Figure 1.

As we can see from Figure 1, the proposed similarity measure based on shape context mainly involves the following steps. After two images to be compared are inputted, their visual salient regions are detected. Then edge detection for two images is conducted, so as to build shape context of visual salient regions. Next, we compute shape context distance for the visual salient regions between two images. At last, the image similarity is measured according to the shape context distance of visual salient regions. The detailed procedure of the approach is as follows.

3. Detecting Visual Salient Regions

In the proposed approach, we detect visual salient regions by virtue of a regional contrast based saliency extraction algorithm, which is simple, efficient, and yields full resolution saliency maps and outperforms existing saliency detection methods [8]. In the first step of this algorithm, the input image is segmented into regions using a graph-based image segmentation method [9], then the color histogram for each region is built. For a region r_k , its saliency value is calculated by measuring its color contrast to all other regions in the image as follows,

$$S(r_k)=\sum_{r_i \neq r_k} w(r_i)D_r(r_i,r_k) \tag{1}$$

where $w(r_i)$ is the weight of region r_i and $D(r_i,r_k)$ is the color distance metric between the two regions. Here the number of pixels in r_i is used as $w(r_i)$ to emphasize color contrast to bigger regions. The color distance between two regions r_1 and r_2 is defined as,

$$D_r(r_1,r_2)=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} f(c_{1,i})f(c_{2,j})D(c_{1,i},c_{2,j}) \tag{2}$$

where $f(c_{k,i})$ is the probability of the i-th color $c_{k,i}$ among all n_k colors in the k-th region r_k , k=1, 2. The probability of a color in the probability density function (i.e. normalized color histogram) of the region is used as the weight for this color to emphasize more the color differences between dominant colors.

In the second step of saliency extraction, spatial information is incorporated by introducing a spatial weighting term in Equation 1 to increase the effects of closer regions and decrease the effects of farther regions. Specifically, for any region r_k , the spatially weighted region contrast based saliency is defined as [8]:

$$S(r_k)=\sum_{r_i \neq r_k} \exp(D_s(r_i,r_k) / \delta_s^2)w(r_i)D_r(r_i,r_k) \tag{3}$$

where $D_s(r_i,r_k)$ is the spatial distance between regions r_i and r_k , and δ_s controls the strength of spatial weighting. Larger values of δ_s reduce the effect of spatial weighting so that contrast to farther regions would contribute more to the saliency of the current region. The spatial distance between two regions is defined as the Euclidean distance between their centroids.

4. Computing Shape Context Distance

Shape context is a shape descriptor that is applicable for shape or region matching [7].

4.1. Edge Detection for Shape Context

For shape context, an object is treated as a possibly infinite point set and the shape of an object is essentially assumed to be captured by a finite subset of its points. More practically, a shape is represented by a discrete set of points sampled from the internal or external contours on the object. These can be obtained as locations of edge pixels as found by an edge detector, giving us a set $P=\{p_1,p_2,......,p_n\}$, $p_i \in R^2$, of n points. The shape can be sampled with roughly uniform spacing. Assuming contours are piecewise smooth, a good approximation to the underlying continuous shapes can be obtained by picking n to be sufficiently large.

4.2. Shape Context

Consider the set of vectors originating from a point to all other sample points on a shape. These vectors express the configuration of the entire shape relative to the reference point. Based on this, Belongie [7] defined shape context as the distribution over relative positions, and identified it as a more robust and compact, yet highly discriminative descriptor. For a point p_i on the shape, a coarse histogram h_i of the relative coordinates of the remaining n-1 points is computed as follows.

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in bin(k)\} \tag{4}$$

This histogram is defined to be the shape context of p_i. In log-polar space the uniform bins are used to make the descriptor more sensitive to positions of nearby sample points than to those of points farther away. Consider a point p_i on the first shape and a point q_j on the second shape. Let $C_{ij}=C(p_i,q_j)$ denote the cost of matching these two points. As shape contexts are distributions represented as histograms, it is natural to use the χ^2 test statistic:

$$C_{ij}=C(p_i,q_j) = \frac{1}{2}\sum_{k=1}^{K}\frac{[h_i(k)-h_j(k)]^2}{h_i(k)+h_j(k)} \tag{5}$$

where $h_i(k)$ and $h_j(k)$ denote the K-bin normalized histogram at p_i and q_j, respectively.

Given the set of costs C_{ij} between all pairs of points p_i on the first shape and q_j on the second shape, bipartite graph matching is conducted [7], so as to minimize the total cost of matching,

$$H(\pi)=\sum_i C(p_i,q_{\pi(i)}) \tag{6}$$

subject to the constraint that the matching be one-to-one, i.e π is a permutation. This is an instance of the square assignment (or weighted bipartite matching) problem, which can be solved in $O(N^3)$ time using the Hungarian method [10]. The input to the assignment problem is a square cost matrix with entries C_{ij}. The result is a permutation $\pi(i)$ such that (6) is minimized. When the number of sample points on two shapes is not equal, the cost matrix can be made square by adding dummy nodes to the smaller point set. In this case, a point will be matched to a “dummy” whenever there is no real match available at smaller cost than a constant matching cost of ε_d. As illustrated in [7], shape context matching is proven to be invariant under scaling and translation, and robust under small geometrical distortions, occlusion and presence of outliers.

4.3. Measuring Shape Distance

After shape matching, shape distance is estimated as the weighted sum of three terms: shape context distance, image appearance distance, and bending energy [7]. Shape context distance $D_{sc}(P,Q)$ between shapes P and Q is measured as the symmetric sum of shape context matching costs over best matching points, *i.e.*,

$$D_{sc}(P,Q)=\frac{1}{n}\sum_{p\in P}\underset{q\in Q}{\arg\min}\, C(p,T(q))+\frac{1}{m}\sum_{q\in Q}\underset{p\in P}{\arg\min}\, C(p,T(q)) \tag{7}$$

where $T(\bullet)$ denotes the estimated shape transformation of thin plate spline. Image appearance distance $D_{ac}(P,Q)$ is defined as the sum of squared brightness differences in Gaussian windows around corresponding image points,

$$D_{ac}(P,Q)=\frac{1}{n}\sum_{i=1}^{n}\sum_{\Delta\in Z^2}G(\Delta)[I_P(p_i+\Delta)-I_Q(T(q_{\pi(i)})+\Delta)]^2 \tag{8}$$

where I_P and I_Q are the gray-level images corresponding to P and Q , respectively. Δ denotes some differential vector offset and G is a windowing function typically chosen to be a Gaussian, thus putting emphasis to pixels nearby. Thus through summing over squared differences in windows around corresponding points, score the weighted gray-level similarity. This score is computed after the thin plate spline transformation T has been applied to best warp the images into alignment. The third term $D_{be}(P,Q)$ corresponds to the "amount" of transformation necessary to align the shapes. In the case of thin plate spline the bending energy is a natural measure [11].

5. Measuring Image Similarity Based on Shape Context

We assume that the image similarity is measured between the image I_1 and the image I_2, and after the visual salient regions are detected from these two images, S_1 represents the set of visual salient regions of I_1, and S_2 means the set of visual salient regions of I_2. Let $S_1=\{P_1, P_2, \cdots, P_m\}$ and $S_2=\{Q_1, Q_2, \cdots, Q_n\}$. We compute shape distance for the visual salient regions between two images, and accordingly obtain the image similarity values. The details are as follows.

Step 1. For every visual salient region P_i (i=1,…,m) in S_1, we search the nearest neighbor Q_{nn} in the set S_2 of visual salient regions of I_2, which has the shortest shape distance to P_i according to the shape distance in Section 4.3.

Step 2. For the visual salient region Q_{nn} in S_2, we search the nearest neighbor P_{nn} in the set S_1 of visual salient regions of I_1, which has the shortest shape distance to Q_{nn}.

Step 3. Compare P_i and P_{nn}, to see if they are the same visual salient region of I_1. If so, it proves that the shape of P_i is still well maintained. Otherwise, it means that the severe shape deformation of P_i has occurred, so that we cannot find the right corresponding region in S_2 for P_i.

Step 4. We count the shape-maintained regions between these two images on the basis of Step 3, accordingly measure the similarity between these two images.

6. Experimental Results

To validate the proposed similarity measure, we took real images to confront it with the real world. The image dataset in our experiment is picked from the COREL database with 10,000 images, in which the images belong to 100 semantic categories, each of which has 100 images.

As shown in Figures 2-7, we select six typical cases of image similarity measure in the paper due to space limitation, which involve different scenes including “horse”, “cruise”, “flower”, “cascade”, “bus” and “sunset” for illustrating the extensive application of our method. In every case, we measure the image similarity between the first image I_1 and each of the remaining images I_2, I_3, I_4, I_5 and I_6 using our proposed method. Then the remaining images are ranked according to their similarity to the first image I_1 and the top ranked image is considered the most similar to the first one. In detail, we use the algorithm in Section 5 to calculate the above image similarity measures. The calculated results of similarity measure are summarized by Table 1 where “Rank 1” means that the corresponding image has the maximum similarity to the reference image I_1, and “Rank 5” indicates that the corresponding image has the minimum similarity to the reference image

I_1. According to the obtained results in Table 1, the images I_4, I_2, I_4, I_5, I_5 and I_3 are respectively marked as the most similar to the leftmost one in Figures 2-7. We can figure out that the achieved similarity results in Figures 2-7 approximate the subjective human choices and are consistent as well as accurate. Thus our proposed method about similarity measure based on shape context has been validated.

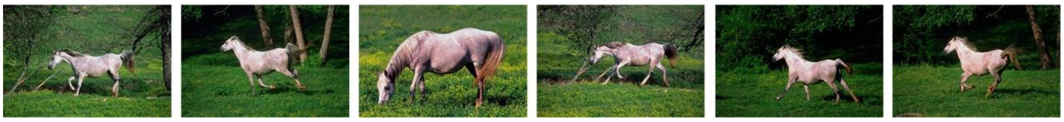

Figure 2. Image Case 1 Involving "Horse" Scene where these Images are Named as I_1, I_2, I_3, I_4, I_5 and I_6 from Left to Right, the First Image I_1 is the Reference

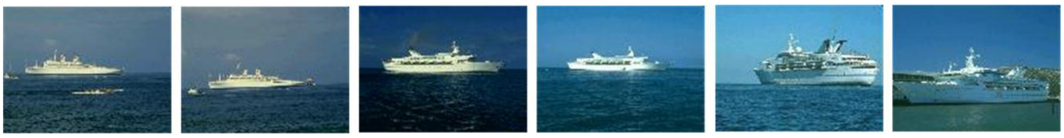

Figure 3. Image Case 2 Involving "Cruise" Scene where these Images are Named as I1, I2, I3, I4, I5 and I6 from Left to Right, the First Image I1 is the Reference

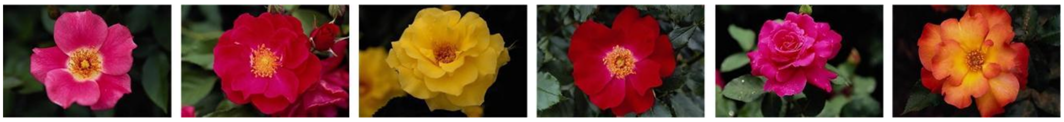

Figure 4. Image Case 3 Involving "Flower" Scene where these Images are Named as I1, I2, I3, I4, I5 and I6 from Left to Right, the First Image I1 is the Reference

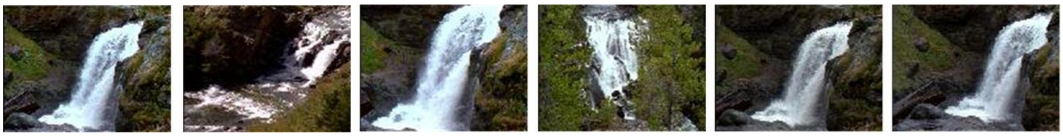

Figure 5. Image Case 4 Involving "Cascade" Scene where these Images are Named as I1, I2, I3, I4, I5 and I6 from Left to Right, the First Image I1 is the Reference

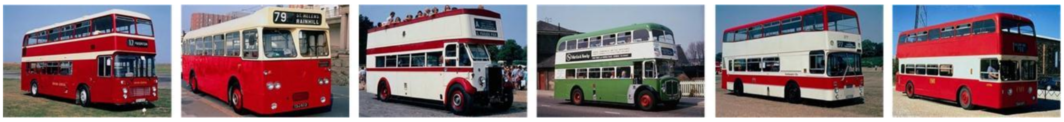

Figure 6. Image Case 5 Involving "Bus" Scene where these Images are Named as I1, I2, I3, I4, I5 and I6 from Left to Right, the First Image I1 is the Reference

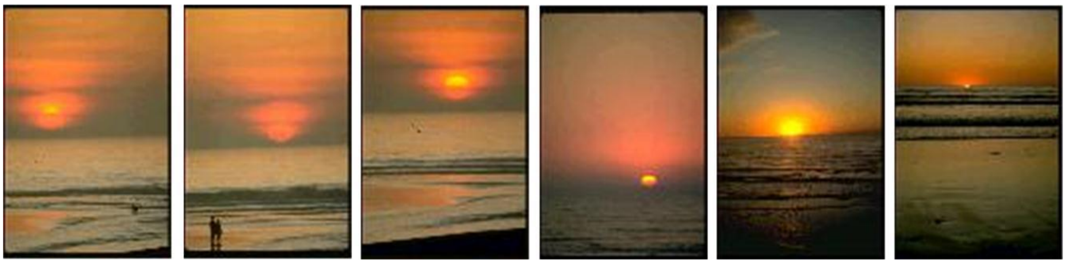

Figure 7. Image Case 6 Involving "Sunset" Scene where these Images are Named as I1, I2, I3, I4, I5 and I6 from Left to Right, the First Image I1 is the Reference

Table 1. The Calculated Rank Results of Similarity Measure for Every Case using the Proposed Approach in Figures 2-7

Similarity Rank	Reference	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
horse	I_1	I_4	I_5	I_6	I_2	I_3
cruise	I_1	I_2	I_4	I_3	I_6	I_5
flower	I_1	I_4	I_2	I_6	I_3	I_5
cascade	I_1	I_5	I_6	I_3	I_2	I_4
bus	I_1	I_5	I_6	I_4	I_3	I_2
sunset	I_1	I_3	I_2	I_6	I_4	I_5

7. Conclusion

In this paper, we proposed an image similarity measure based on shape context. The proposed approach on image similarity measure mainly involves the following steps. In the first, the visual salient regions are detected from two images to be compared through a regional contrast based saliency extraction algorithm. Then we build shape context of visual salient regions. Next, we compute shape context distance for the visual salient regions between two images. Finally, the image similarity is measured according to the shape distance of visual salient regions. Experimental results showed that the achieved similarity ranks for real images closely match the subjective human choices and are consistent as well as accurate. The good results from the experiments illustrated the practicability and effectiveness of our proposed measure approach, validating it.

Acknowledgements

The work is supported by the National Nature Science Foundation of China (No.U1404623), the Natural Science Foundation of Henan Educational Committee (No.12B520068) and the PHD Foundation of Zhengzhou University of Light Industry (No.2011BSJJ002)

References

[1] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index", IEEE Transactions on Image Processing, vol. 18, no. 11, pp. 2385-2401, **(2009)**.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, vol. 13, no. 4, **(2004)**, pp. 600-612.

[3] M. S. Prieto and A. R. Allen. "A similarity metric for edge images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 10, **(2003)**, pp. 1265-1273.

[4] K. Bowyer, C. Kranenburg and S. Dougherty, “Edge detector evaluation using empirical ROC curves”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **(1999)**; Ft. Collins, CO, USA.

[5] W. Pratt. “Digital Image Processing”, Wiley, **(2001)**.

[6] D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge. “Comparing images using the Hausdorff distance”. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 9, **(1993)**, pp. 850-863

[7] S. Belongie, J. Malik and J. Puzicha, “Shape matching and object recognition using shape contexts”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, **(2002)**, pp. 509-522.

[8] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang and S. M. Hu. “Global contrast based salient region detection”. In: IEEE Conference on Computer Vision and Pattern Recognition, **(2011)** June 21-23; Colorado, USA

[9] P. F. Felzenszwalb and D. P. Huttenlocher. “Efficient graph-based image segmentation”. International Journal of Computer Vision, vol. 59, no. 2, **(2004)**, pp. 167-181

[10] C. Papadimitriou and K. Stieglitz, “Combinatorial Optimization: Algorithms and Complexity”, Prentice Hall, **(1982)**.

[11] F.L. Bookstein, “Principal Warps: Thin-Plate Splines and Decomposition of Deformations”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 6, **(1989)**, pp. 567-585.

Authors

Li Canlin received PhD degree in computer science from Shanghai Jiaotong University in 2010. Since 2010 he has been in the School of Computer and Communication Engineering at Zhengzhou University of Light Industry. His research interests include image processing, multimedia, graphics, digital entertainment and software engineering. Dr. Li is a member of IEEE as well as ACM.

Qian Shenyi is an associate professor of School of Computer and Communication Engineering at Zhengzhou University of Light Industry. He received Master's degree in computer science from Huazhong University of Science & Technology in 2001. His research interests include image processing, computer decision support system, computer software and theory and emergency management.