Modeling Perspective Effects in Photographic Composition

Zihan Zhou[†] Siqiong He[†] Jia Li [‡] James Z. Wang[†]

[†]College of Information Sciences and Technology
[‡]Department of Statistics
The Pennsylvania State University, University Park, PA, USA
zzhou@ist.psu.edu, hesiqiong@gmail.com, {jiali, jwang}@psu.edu

ABSTRACT

Automatic understanding of photo composition is a valuable technology in multiple areas including digital photography, multimedia advertising, entertainment, and image retrieval. In this paper, we propose a method to model geometrically the compositional effects of linear perspective. Comparing with existing methods which have focused on basic rules of design such as simplicity, visual balance, golden ratio, and the rule of thirds, our new quantitative model is more comprehensive whenever perspective is relevant. We first develop a new hierarchical segmentation algorithm that integrates classic photometric cues with a new geometric cue inspired by perspective geometry. We then show how these cues can be used directly to detect the dominant vanishing point in an image without extracting any line segments, a technique with implications for multimedia applications beyond this work. Finally, we demonstrate an interesting application of the proposed method for providing on-site composition feedback through an image retrieval system.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis

General Terms

Algorithms; Experimentation; Human Factors

Keywords

Perspective Effect; Composition Modeling; Image Segmentation; Vanishing Point Detection; Photo Retrieval

1. INTRODUCTION

With the rapid advancement of digital camera and mobile imaging technologies, we have witnessed a phenomenal increase of both professional and amateur photographs in the past decade. Large-scale social media companies, *e.g.*, Flickr, Instagram, and Pinterest, enable their users to share photos with people all around the world. As millions of new photos are added daily on the Internet, the demand increases for creating automatic systems to manage, assess, and edit such content. Consequently, *photo composition understanding* has emerged as a new research area, attracting attention of the multimedia community lately.

In photography, composition is the art of positioning or organization of objects and visual elements (*e.g.*, color, texture, shape, tone, motion, depth) within an photo. Principles of organization include balance, contrast, Gestalt perception and unity, geometry, rhythm, perspective, illumination, and viewing path. Automated understanding of photo composition has been shown to benefit a number of applications such as summarization of photo collections [25] and assessment of image aesthetics [26]. It can also be used to render feedback to the photographer on the photo aesthetics [36, 35], and suggest improvements to the image composition through image re-targeting [21, 4, 7]. Moreover, the capability to understand compositions has broad implications beyond photography. For example, in multimedia advertising, designers often need to come up with interesting compositions to attract the attention or to influence the emotions of the viewers. Also, composition analysis can potentially be used to assist movie directors by providing feedback on the scene composition.

In the literature, most work on image composition understanding has focused on design rules such as the simplicity, visual balance, golden ratio, the rule of thirds, and the use of diagonal lines. These rules are mainly concerned with the 2D rendering of objects or simple ways of dividing the image frame. They are by no means exhaustive for capturing the wide variations in photographic composition.

The Use of Perspective Effect in Photography. Perspective effect is one of the most commonly used techniques in photo composition. As illustrated by the examples in Figure 1, during photographic creation, experienced photographers often make use of the linear perspective effect to emphasize the sense of 3D space in a 2D photo. It not only reveals the appearance of objects in space and their relationships to each other, but also tells us our angle of perception and our position as an observer. In this regard, perspective is credited with recognizing the viewer as a specific unique individual in a distinct place with a point of view [18].

According to the perspective camera geometry, all parallel lines in 3D converge to a single point, the vanishing point, in the image. However, only the vanishing point lying within or near the image frame and associated with the

MM'15, October 26–30, 2015, Brisbane, Australia.

DOI: http://dx.doi.org/10.1145/2733373.2806248.

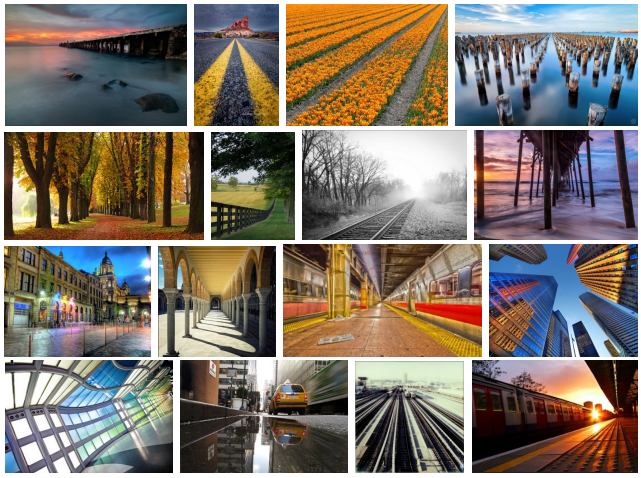
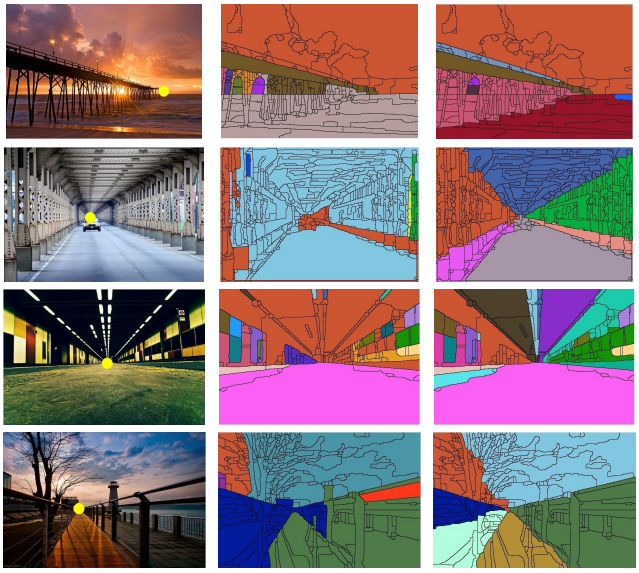


Figure 1: The use of perspective effects in photography. Top two rows: natural landscape. Bottom two rows: urban scenes.

dominant structures of the scene (*e.g.*, grounds, large walls, bridges) convey a strong impression of 3D space or depth to the viewers. Figure 2(a) shows some examples. We regard such a vanishing point as the *dominant vanishing point* of the particular image. By placing the dominant vanishing point at different image locations and choosing how each composition region in the image relates to this point, an experienced photographer could produce various image compositions that convey different messages or impressions to the viewers.

Modeling Perspective Effects Automatically. Based on the above discussion, we argue that, in order to obtain a comprehensive understanding of composition, it is necessary to develop automatic systems to examine the perspective effects in the photos. To this end, we propose to *partition an image into photometrically and geometrically consistent regions according to the dominant vanishing point.* In our work, we assume that each geometric region can be roughly modeled by a flat surface, or a plane. As shown in Figure 2(c), such a partition naturally provides us with a novel *holistic* yet *compact* representation of the 3D scene geometry that respects the perspective effects of the scene the image captured in, and allows us to derive a notion of relative depth and scale for the objects. Nevertheless, obtaining such a representation is a challenging problem for the following reasons.

First, given any two adjacent geometric regions in an image, there may not be a distinguishable boundary in terms of photometric cues (*e.g.*, color, texture) so that they can be separated. For example, the walls and the ceiling in the second photo of Figure 2 share the same building material. Because existing segmentation algorithms primarily depend on the photometric cues to determine the distance between regions, they are often unable to separate these regions from each other (see Figure 2(b) for examples). To resolve this issue, we propose a novel hierarchical image segmentation algorithm that leverages significant geometric information about the dominant vanishing point in the image. Specifically, we compute a geometric distance between any two adjacent regions based on the similarity of the angles of the two regions in a polar coordinate system, with the dominant



(a) (b) (c)

Figure 2: Geometric image segmentation. (a) The original image with the dominant vanishing point detected by our method (shown as a yellow round dot). (b) Region segmentation map produced using a state-of-the-art method. (c) Geometric image segmentation map produced by our method.

vanishing point being the pole. By combining the geometric cues with conventional photometric cues, our method is able to preserve essential geometric regions in the image.

Second, detecting the dominant vanishing point from an image itself is a nontrivial task. Typical vanishing point detection methods assume the presence of a large number of strong edges in the image. However, for many photos of natural outdoor scenes, such as the image of an arbitrary road, there may not be adequate clearly-delineated edges that converge to the vanishing point. In such cases, the detected vanishing points are often unreliable and sensitive to image noise. To overcome this difficulty, we observe that while it may be hard to detect the local edges in these images, it is possible to directly infer the location of the vanishing point by aggregating the aforementioned photometric and geometric cues over the entire image (Figures 2 and 7). Based on this observation, we develop a novel vanishing point detection method which does not rely on the existence of strong edges, hence works better for natural images.

Application to On-Site Composition Feedback. Since our region-based model captures rich information about the photo composition, it may benefit many composition-based applications. As an illustrative example, we apply it to an image retrieval application which aims to provide amateur users with on-site feedback about the composition of their photos, in the same spirit as Yao *et al.* [35]. In particular, given a query image taken by the user, the system retrieves exemplar photos with similar compositions from a collection of photos taken by experienced photographers. These exemplar photos can serve as an informative guide for the users to achieve good compositions in their photos.

Our Contributions. In summary, we have made the following contributions:

- *Composition modeling:* We model the composition of an image by examining the perspective effects and partitioning the image into photometrically and geometrically consistent regions using our novel hierarchical image segmentation algorithm.
- *Dominant vanishing point detection:* By aggregating the photometric and geometric cues used in our segmentation algorithm, we develop an effective method to detect the dominant vanishing point in an arbitrary image.
- *On-site composition feedback:* We demonstrate that our region-based model can be used to provide amateur users with on-site composition feedback by retrieving images with similar composition as the query image from a collection of photos taken by experienced photographers.

Nevertheless, our framework does not attempt to model *all* potential compositions in the photos, especially when there is a lack of strong perspectivity. While in this paper we focus on the use of perspective geometry in the photography, we point out that there are many works which study other important aspects of composition, including the semantic features (*e.g.*, buildings, trees, roads) [14, 15, 9, 10]. It would be ideal to integrate all these features in order to gain a deeper understanding of the image composition, but such a comprehensive work is beyond the scope of this paper.

2. RELATED WORK

Standard composition rules such as the rule of thirds, golden ratio and low depth of field have played important roles in early works on image aesthetics assessment [5, 22]. Obrador *et al.* [26] later showed that by using only the composition features, one can achieve image aesthetic classification results that are comparable to the state-of-the-art. Recently, these rules have also been used to predict high-level attributes for image interestingness classification [6], recommend suitable positions and poses in the scene for portrait photography [36], and develop both automatic and interactive cropping and retargeting tools for image enhancement [21, 4, 7]. In addition, Yao *et al.* [35] proposed a composition-sensitive image retrieval method which classifies images into horizontal, vertical, diagonal, textured, and centered categories, and uses the classification result to retrieve exemplar images that have similar composition and visual characteristics as the query image. However, as we mentioned before, these features or categories are all about 2D rendering, with 3D impression not taken into account.

Meanwhile, various methods have been proposed to extract 3D scene structures from a single image. The GIST descriptor [27] is among the first attempts to characterize the global arrangement of geometric structures using simple image features such as color, texture and gradients. Following this seminal work, a large number of supervised machine learning methods have been developed to infer approximate 3D structures or depth maps from the image using carefully designed models [14, 15, 9, 31, 24] or grammars [10, 11]. In addition, models tailored for specific scenarios have been studied, such as indoor scenes [19, 12, 13] and urban scenes [3]. However, these works all make strong assumptions on the structure of the scene, hence the types of scene they can handle in practice are limited. Despite the above efforts, obtaining a good estimation of perspective in an arbitrary image remains an open problem.

Typical vanishing point detection algorithms are based on clustering edges in the image according to their orientations. Kosecka and Zhang proposed an Expectation Maximization (EM) approach to iteratively estimate the vanishing points and update the membership of all edges [17]. Recently, a non-iterative method is developed to simultaneously detect multiple vanishing points in an image [33]. These methods assume that a large number of line segments are available for each cluster. To reduce the uncertainty in the detection results, a unified framework has been proposed to jointly optimize the detected line segments and vanishing points [34]. For images of scenes that lack clear line segments or boundaries, specifically the unstructured roads, texture orientation cues of all the pixels are aggregated to detect the vanishing points [29, 16]. But it is unclear how these methods can be extended to general images.

Image segmentation algorithms commonly operate on low-level image features such as color, edge, texture and the position of patches [32, 8, 20, 2, 23]. But it was shown in [30] that given an image, images sharing the same spatial composites can help with the unsupervised segmentation task.

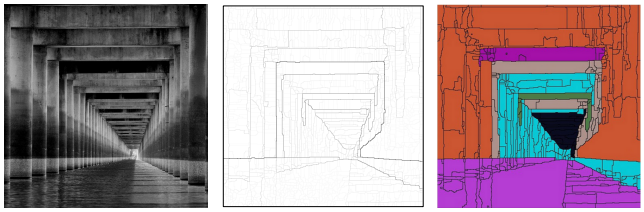
3. GEOMETRIC IMAGE SEGMENTATION

Since our segmentation method follows the classic hierarchical segmentation framework, we give an overview of the framework and some of the state-of-the-art results in Section 3.1. In Section 3.2, we introduce our geometric distance measure for hierarchical image segmentation, assuming the location of the dominant vanishing point is known. The proposed geometric cue is combined with traditional photometric cues in Section 3.3 to obtain a holistic representation for composition modeling. In Section 3.4, we further show how the proposed distance measure, when aggregated over the entire image, can be used to detect the dominant vanishing point in an image.

3.1 Overview of Hierarchical Segmentation

Generally speaking, the segmentation method can be considered as a greedy graph-based region merging algorithm. Given an over-segmentation of the image, we define a graph $\mathcal{G} = (\mathcal{R}, \mathcal{E}, W(\mathcal{E}))$, where each node corresponds to one region, and $\mathcal{R} = \{R_1, R_2, \ldots\}$ is the set of all nodes. Further, $\mathcal{E} = \{e_{ij}\}$ is the set of all edges connecting adjacent regions, and the weights $W(\mathcal{E})$ are a measure of dissimilarity between regions. The algorithm proceeds by sorting the edges by their weights and iteratively merging the most similar regions until certain stopping criterion is met. Each iteration consists of three steps:

1. Select the edge with minimum weight:
$$e^* = \arg\min_{e_{ij} \in \mathcal{E}} W(e_{ij}).$$
2. Let $R_1, R_2 \in \mathcal{R}$ be the regions linked by e^*. Set $\mathcal{R} \leftarrow \mathcal{R} \setminus \{R_1, R_2\} \bigcup \{R_1 \bigcup R_2\}$ and update the edge set $\mathcal{E}$ accordingly.
3. Stop if the desired number of regions K is reached, or the minimum edge weight is above a threshold δ. Otherwise, update weights $W(\mathcal{E})$ and repeat.



(a) (b) (c)

Figure 3: Hierarchical image segmentation using photometric cues only. (a) The original image. (b) The ultrametric contour map (UCM) generated by [2]. (c) The segmentation result obtained by thresholding the UCM at a fixed scale.

Various measures have been proposed to determine the distance between two regions, such as the difference between the intensity variance across the boundary and the variance within each region [8], and the difference in coding lengths [23]. Recently, Arbelaez *et al.* proposed a novel scheme for contour detection which integrates *global photometric information* into the grouping process via spectral clustering [2]. They have shown that this globalization scheme can help identify contours which are too weak to be detected using local cues. The detected contours are then converted into a set of initial regions (*i.e.*, an over-segmentation) for hierarchical image segmentation. We show an example of the segmentation result obtained by [2] in Figure 3. In particular, in Figure 3(b), we visualize the entire hierarchy of regions on an real-valued image called the ultrametric contour map (UCM) [1], where each boundary is weighted by the dissimilarity level at which it disappears. In Figure 3(c), we further show the regions obtained by thresholding the UCM at a fixed scale. It is clear that because the weights of the boundaries are computed only based on the photometric cues in [2], different geometric regions could be merged at early stages in the hierarchical segmentation process if they have similar appearances.

Motivated by this problem, we take the over-segmentation result generated by [2] (*i.e.*, by thresholding the UCM at a small scale 0.05) as the input to our algorithm, and develop a new distance measure between regions which takes both photometric and geometric information into consideration.

3.2 Geometric Distance Measure

We assume that a major portion of the scene can be approximated by a collection of 3D planes parallel to a dominant direction in the scene. The background, *e.g.*, the sky, can be treated as a plane at infinity. The dominant direction is characterized by a set of parallel lines in the 3D space which, when projected to the image, converge to the dominant vanishing point. Consequently, given the location of the dominant vanishing point, our goal is to segment an image so that each region can be roughly modeled by one plane in the scene. To achieve this goal, we need to formulate a dissimilarity measure which yields small values if the pair of adjacent regions belong to the same plane, and large values otherwise.

We note that any two planes that are parallel to the dominant direction must intersect at a line which passes through the dominant vanishing point in the image. Intuitively, this observation provides us with a natural way to identify adjacent regions that could potentially lie on different planes: If the boundary between two regions is parallel to the dominant direction (hence passes through the dominant vanishing point), these two regions are likely to lie on different planes. However, in the real world, many objects are not completely planar, hence there may not be a clear straight line that passes through the dominant vanishing point between them. As an example, if we focus our attention on the three adjacent regions R_1, R_2 and R_3 in Figure 4, we notice that R_1 and R_3 belong to the vertical wall and R_2 belongs to the ceiling. However, the boundaries between the pair (R_1, R_2) and the pair (R_1, R_3) both lie on the same (vertical) line. As a result, it is impossible to differentiate these two pairs based on only the orientation of these boundaries.

To tackle this problem, we propose to look at the angle of each region from the dominant vanishing point in a polar coordinate system, instead of the orientation of each boundary pixel. Here, the angle of a region is represented by the distribution of angles of all the pixels in this region. Mathematically, let the dominant vanishing point P be the pole of the polar coordinate system, for each region R_i, we compute the histogram of the angle value $\theta(X)$ for all the pixels $X \in R_i$, as illustrated in Figure 4.

Let $c_i(\theta)$ be the number of the pixels in R_i that fall into the θ-th bin. We use 360 bins in our experiments. We say that one region R_i *dominates* another region R_j at angle θ if $c_i(\theta) \geq c_j(\theta)$. Our observation is that if one region R_i always dominates another region R_j at almost all angles, these two regions likely belong to the same plane. Meanwhile, if one region has larger number of pixels at some angles whereas the other region has larger number of pixels at some other angles, these two regions likely lie on different planes. This observation reflects the fact a plane converging to the vanishing point often divides along the direction perpendicular to the dominant direction because of architectural or natural structures, e.g., columns and trees. Because perpendicular separation of regions has little effect on the polar angles, the histograms of angles tend to overlap substantially.

Based on this observation, we define the geometric distance between any two regions R_i and R_j as follows:

$$W_g(e_{ij}) = 1 - \max\left(\frac{\sum_\theta \min(c_i(\theta), c_j(\theta))}{|R_i|}, \frac{\sum_\theta \min(c_i(\theta), c_j(\theta))}{|R_j|}\right),$$

where $|R_i|$ and $|R_j|$ are the total numbers of pixels in regions R_i and R_j, respectively. For example, as illustrated in Figure 4(c), R_1 dominates R_3 at all angles and hence we have $W_g(e_{1,3}) = 0$. Meanwhile, R_1 and R_2 dominate each other at different angles and their distributions have very small overlap. As a result, their geometric distance is large: $W_g(e_{1,2}) = 0.95$. In Figure 4(d), we show all the boundaries weighted by our geometric distance. As expected, the boundaries between two regions which lie on different planes tend to have higher weights than other ones. This suggests that, by comparing the angle distributions of two adjacent regions, we can obtain a more robust estimate of the boundary orientations than directly examining the orientations of boundary pixels.

Here, a reader may wonder why we don't simply normalize the histograms and use popular metrics like KL divergence or the earth mover's distance to compare two regions. While our intuition is indeed to compare the distributions of angles of two regions, we have found in practice that computing the normalized histograms could be highly unstable for

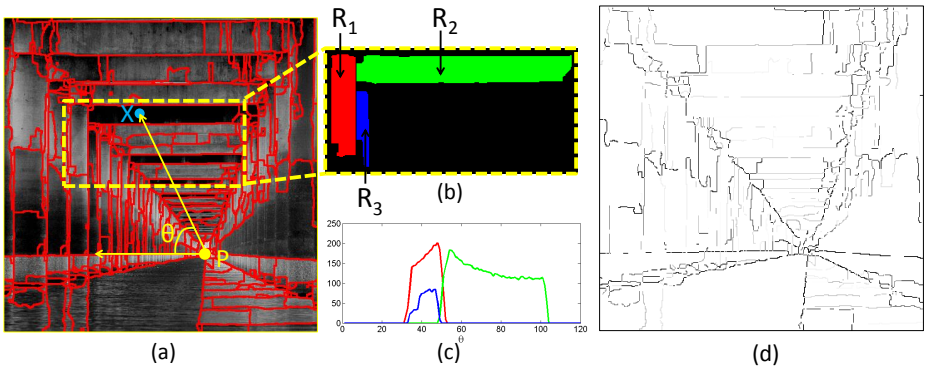


Figure 4: Illustration of the the computation of the geometric distance. (a) The over-segmentation map with the polar coordinate system. (b) Three adjacent regions from the image. (c) The histograms of angle values for the three regions. (d) The boundary map weighted by the geometric distance between adjacent regions.

small regions, especially at the early stages of the iterative merging process. Thus, in this paper we propose an alternative geometric distance measure which avoids normalizing the histograms, and favors large regions during the process.

3.3 Combining Photometric and Geometric Cues

While our geometric distance measure is designed to separate different geometric structures, *i.e.*, planes, in the scene, the traditional photometric cues often provide additional information about the composition of images. Because different geometric structures in the scene often have different colors or texture, the photometric boundaries often coincide with the geometric boundaries. On the other hand, in practice it may not always be possible to model all the structures in the scene by a set of planes parallel to the dominant direction. Recognizing the importance of such structures to the composition of the image due to their visual saliency, it is highly desirable to integrate the photometric and geometric cues in our segmentation framework to better model composition. In our work, we combine the two cues by a linear combination:

$$W(e_{ij}) = \lambda W_g(e_{ij}) + (1-\lambda) W_p(e_{ij}), \qquad (1)$$

where $W_p(e_{ij})$ is the photometric distance between adjacent regions, and can be obtained from any conventional hierarchical image segmentation method. Here we adopt the contour map generated by [2].

In Figure 5, we show the segmentation results of an image using our method with different choices of λ and a fixed number of regions K. Note that when $\lambda = 1$, only the geometric cues are used for segmentation; when $\lambda = 0$, the result is identical to that obtained by the conventional method [2]. It can be seen that using the geometric cues alone ($\lambda = 1$), we are able to identify most of the structures in the scene. Some of the boundaries between them may not be accurate enough (*e.g.*, the boundary between the bridge on the left and the sky area). However, when $\lambda = 0$, the algorithm tends to merge regions from different structures early in the process if they have similar colors. By combining the two cues (*e.g.*, λ is 0.4 or 0.6), we are able to eliminate the aforementioned problems and obtain satisfactory result. Additional results are provided in Figures 6. Our method typically achieves the best performance when λ is in the range of $[0.4, 0.6]$. We fix λ to 0.6 for the remaining experiments.

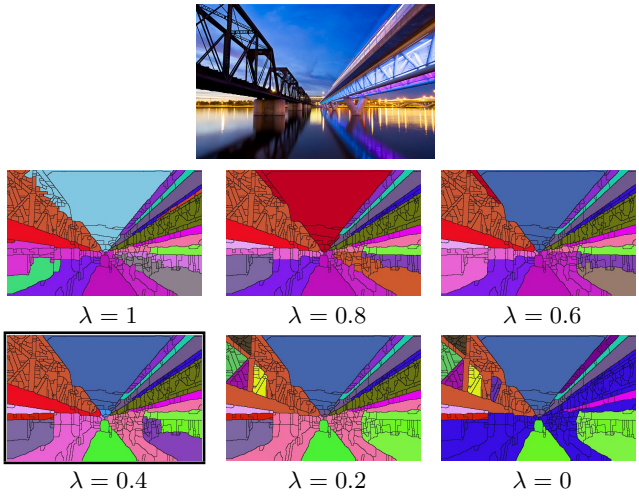


Figure 5: Image segmentation results by integrating both photometric and geometric cues. Different values of weighting parameter λ have been used.

3.4 Enhancing Vanishing Point Detection

In the previous subsection we demonstrated how the knowledge about the dominant vanishing point in the scene can considerably improve the segmentation results. However, detecting the vanishing point in an arbitrary image itself is a challenging problem. Most existing methods assume that (1) region boundaries in the image provide important photometric cues about the location of the dominant vanishing point, and (2) these cues can be well captured by a large number of line segments in the image. In practice, we notice that while the first assumption is generally true, the second one often fails to hold, especially for images of natural outdoor scenes. This is illustrated in Figure 7: although human can easily infer the location of the dominant vanishing point from the orientations of the aggregated region boundaries, existing line segment detection algorithms may fail to identify these boundaries. For this reason, any vanishing point detection method relying on the detected line segments would also fail.

To alleviate this issue, we propose to use our geometric distance measure $W_g(e_{ij})$ to obtain a more robust estimation of the orientation of each boundary and subsequently develop a simple exhaustive search scheme to detect the

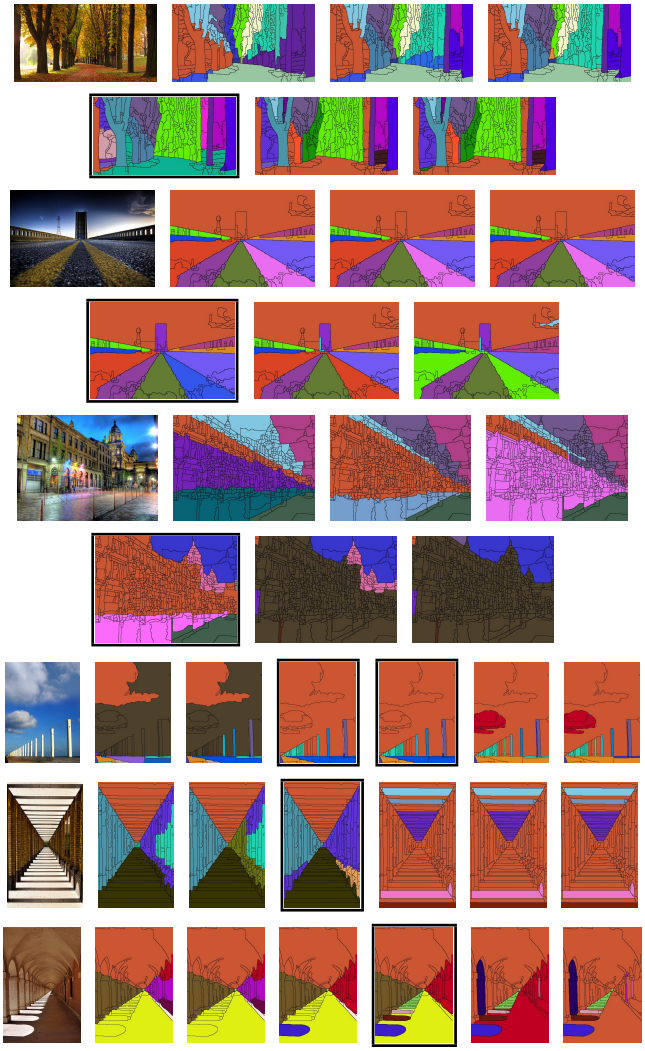


Figure 6: Additional segmentation results. For each original image, we show the results in the order of λ =1, 0.8, 0.6, 0.4, 0.2 and 0.

dominant vanishing point. In particular, given a hypothesis of the dominant vanishing point location, we can obtain a set of boundaries which align well with the converging directions in the image by computing $W_g(e_{ij})$ for each pair of adjacent regions. These boundaries then form a "consensus set". We compute a score for the hypothesis by summing up the strengths of the boundaries in the consensus set (Figure 7(c) and (d)). Finally, we keep the hypothesis with the highest score as the location of the dominant vanishing point (Figure 7(e)). Our algorithm can be summarized as follows:

1. Divide the image by an $m \times n$ uniform grid mesh.
2. For each vertex P_k on the grid, we compute the geometric distance $W_g(e_{ij})$ for all the boundaries in an over-segmentation of the image. Then, the consensus score for P_k is defined as: $f(P_k) = \sum_{e_{ij} \in \mathcal{E}} W_p(e_{ij}) W_g(e_{ij})$.
3. Select the point with the highest score as the detected dominant vanishing point: $P^* = \arg\max f(P_k)$.

Here, the size of the grid may be chosen based on the desired precision for the location of the vanishing point. In practice, our algorithm can find the optimal location in about one minute on a 50×33 grid on a single CPU. We also note that the time may be reduced using a coarse-to-fine procedure.

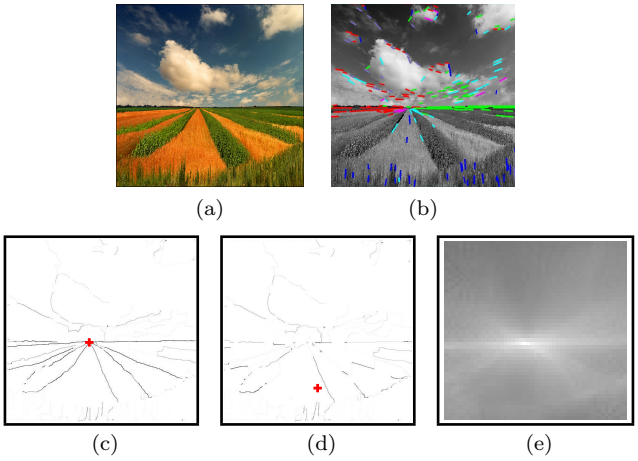


Figure 7: Enhancing vanishing point detection. (a) Original image. (b) Line segment detected. (c) and (d) The weighted boundary map for two different hypotheses of the dominant vanishing point location. (e) The consensus score for all vertices on the grid.

In addition, we assume that the dominant vanishing point lies in the image frame because, as we noted before, only the vanishing point which lies within or near the frame conveys a strong sense of 3D space to the viewer. But our method can be easily extended to detect vanishing points outside the frame using a larger mesh grid.

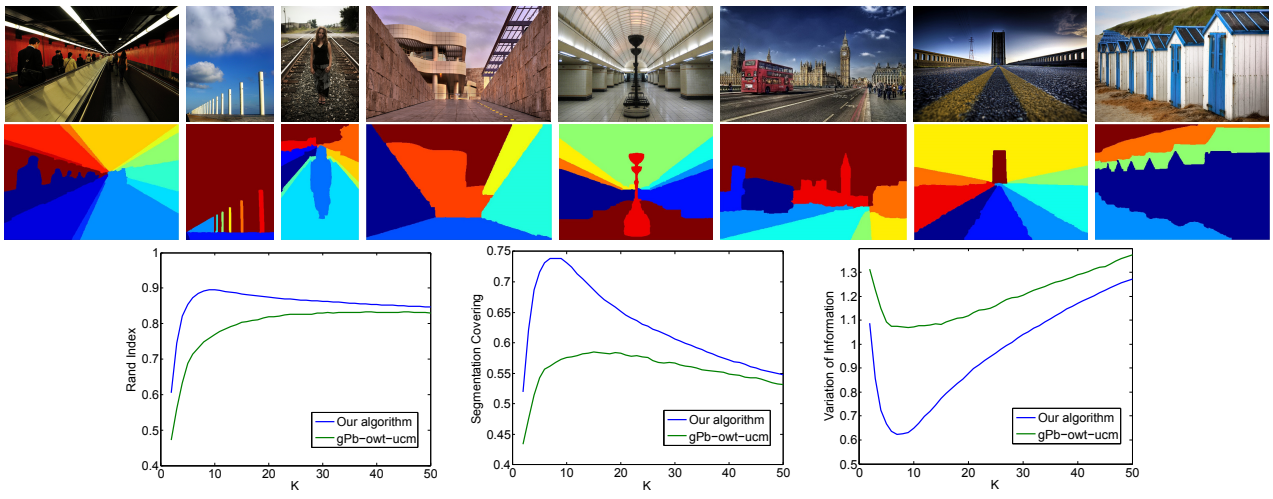
4. QUANTITATIVE EVALUATIONS

We systematically evaluate the proposed geometric image segmentation algorithm and vanishing point detection algorithm in Section 4.1 and 4.2, respectively.

4.1 Evaluation on Image Segmentation

In this section, we compare the performance of our method with the state-of-the-art image segmentation method, *gPb*-owt-ucm [2]. For this experiment, we assume known dominant vanishing point locations. We emphasize that our goal here is not to compete with [2] as a generic image segmentation algorithm, but to demonstrate that information about the vanishing point (*i.e.*, the geometric cue), if properly harnessed, can empower us to get better segmentation results.

To quantitatively evaluate the methods, we use three popular metrics to compare the result obtained by each algorithm with the manually-labeled segmentation: Rand index (RI), variation of information (VOI) and segmentation covering (SC). First, the RI metric measures the probability that an arbitrary pair of samples have the same label in two partitions. The range of RI metric is $[0, 1]$, higher values indicating greater similarity between two partitions. Second, the VOI metric measures the average condition entropy of two clustering results, which essentially measures the extent to which one clustering can explain the other. The VOI metric is non-negative, with lower values indicating greater similarity. Finally, the SC metric measures the overlap between the region pairs in two partitions. The range of SC metric is $[0, 1]$, higher values indicating greater similarity. We refer interested readers to [2] for more details about these metrics.



(a) Rand Index (b) Segmentation Covering (c) Variation of Information

Figure 8: Segmentation benchmarks. K is the number of regions. Top rows show some example test images with the manually labeled segmentation maps.

For this experiment, we manually labeled 200 images downloaded from `flickr.com`. These images cover a variety of indoor and outdoor scenes and each has a dominant vanishing point. During the labeling process, our focus is on identifying all the regions that differ from their neighbors either in their geometric structures or photometric properties. We show some images with the hand-labeled segmentation maps in Figure 8.

Figure 8 also shows the benchmark results of both methods. Our method significantly outperforms gPb-owt-ucm on all metrics. This is consistent with the example results in Figures 5 and 6, suggesting that our method is advantageous in segmenting the geometric structures in the scene.

4.2 Evaluation on Vanishing Point Detection

Next, we compare our vanishing point detection method with two state-of-the-art methods proposed by Tardif [33] and Tretiak *et al.* [34], respectively. As discussed earlier, both methods rely on the line segments to generate vanishing point candidates. Then, a non-iterative scheme similar to the popular RANSAC technique is developed in [33] to group the line segments into several clusters, each corresponding to one vanishing point. Using the vanishing points detected by [33] as an initialization, [34] further propose a non-linear optimization framework to jointly refine the extracted line segments and vanishing points.

In this experiment, we use 400 images downloaded from `flickr.com` whose dominant vanishing points lie within the image frame. All images are scaled to size 500×330 or 330×500. To make the comparison fair, for [33] and [34] we only keep the vanishing point with the largest support set among all hypotheses that also lie within the image frame. We consider a detection successful if the distance between the detected vanishing point and the manually labeled ground truth is smaller than certain threshold t, and plot the success rates of all methods w.r.t. the threshold t in Figure 9. As one can see, our method outperforms existing methods as long as the threshold is not too small ($t \geq 10$ pixels), justifying its effectiveness for detecting the dominant vanishing point in arbitrary images. When t is small, our method does not perform well because its precision in locating the vanishing point is limited by the size of the grid mesh. Nevertheless,

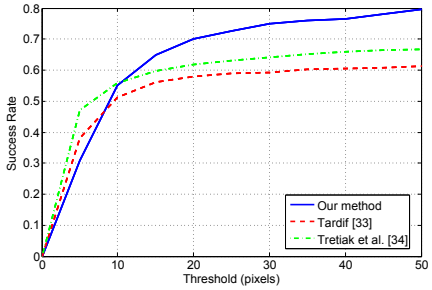


Figure 9: Comparison of vanishing point detection algorithms.

this issue can be alleviated using a denser grid mesh at the cost of more computational time. Also, we note that while the joint optimization scheme proposed in [34] can recover weak line segments and vanishing points for urban scenes, its improvement over [33] is quite small in our case.

5. ON-SITE COMPOSITION FEEDBACK TO PHOTOGRAPHERS

The segmentation results obtained by our method capture rich information about the composition of images, hence can be used to facilitate various composition-driven applications. We now discuss one such application, which aims at providing amateur users with on-site feedback about the composition of their photos.

As we know, good composition highlights the object of interest in a photo and attract the viewer's attention immediately. However, it typically requires years of practice and training for a photographer to master all the necessary composition skills. An effective way for an amateur or an enthusiast to learn photography is through observing masterpieces and building models about photography. Nowadays, thanks to increased popularity of online photo sharing services such as `flickr.com` and `photo.net`, one can easily access millions of photos taken by people all around the world. Such resource naturally provides us with opportunities to develop new and more efficient ways for the beginners to learn the composition skills.

Specifically, given a photo taken by the user, we propose to find photos with similar compositions in a collection of photos taken by experienced or accomplished photographers. These photos are rendered as feedback to the user. The user can then examine these exemplar photos and consider re-composing his/her own photo accordingly. Yao *et al.* pioneered this direction in [35], but the types of composition studied there are limited to a few categories which are predefined based on simple 2D rules.

In this paper, we take a completely different approach and develop a similarity measure to compare the composition of two images based on their geometric image segmentation maps. Our observation is that, experienced photographers often are able to achieve different compositions by placing the dominant vanishing point at different image locations, and then choosing how the main structures of the scene are related to it in the captured image. In addition, while the difference in the dominant vanishing point locations can be simply computed as the Euclidean distance between them, our geometric segmentation result offers a natural representation of the arrangement of structures with respect to the dominant vanishing point. Specifically, given two images I_i and I_j, let P_i and P_j be the locations of dominant vanishing points and S_i and S_j be the segmentation results generated by our method for these two images, respectively, we define the similarity measure as follows:[1]

$$D(I_i, I_j) = F(S_i, S_j) + \alpha \|P_i - P_j\|, \qquad (2)$$

where $F(S_i, S_j)$ is a metric to compare two segmentation maps. We adopt the Rand index [28] for its effectiveness. In addition, α controls the relative impact of the two terms in Eq. (2). We empirically set $\alpha = 0.5$.

To obtain a dataset of photos which make good use of the perspective effect, we collect 3,728 images from `flickr.com` by querying the keyword "vanishing point". When collecting the photos, we use the sorting criterion of "interestingness" provided by `flickr.com`, so that the retrieved photos are likely to be well-composed and taken by experienced or accomplished photographers. Each photo is then scaled to size 500×330 or 330×500. To evaluate the effectiveness of our similarity measure (Eq. (2)), we manually label the dominant vanishing point and then apply our geometric image segmentation algorithm (with the proposed distance measure $W(e_{ij})$ and the stopping criteria $\delta = 0.55$) to obtain a segmentation for each image.

In Figure 10, we show the retrieved images for various query images. The results clearly show that the proposed measure is not only able to find images with similar dominant vanishing point locations, but also effectively captures how each region in the image is related to the vanishing point. For example, the images in the 4th, 6th, and 7th rows of Figure 10 all have similar vanishing point locations (around the image center), but very different scene structure compositions, hence convey very different impressions to the viewer.

6. DISCUSSION AND FUTURE WORK

We have developed a new method for modeling visual composition by analyzing the perspective effects and segmenting the image based on photometric and geometric cues.

[1]Here, we assume the two images have the same size, after rescaling.

Figure 11: Composition modeling with complex perspective geometry. First row: photos with multiple dominant vanishing points. Second row: photos with foreground objects.

The method has been demonstrated for its effectiveness in detecting the dominant vanishing point from an arbitrary scene. Among a variety of potential applications, we have illustrated how our model can be used to build a composition-sensitive image retrieval system capable of providing on-site feedback to photographers.

Our work opens up several future directions. First, we may extend our geometric image segmentation and vanishing point detection algorithms to images with two or more dominant vanishing points, often found in man-made environments (Figure 11, first row). Here, our goal is to detect all the dominant vanishing points in an image and to group the regions according to the corresponding vanishing points.

Second, one challenge in composition recognition for real-world photos is the presence of large foreground objects (Figure 11, second row). They typically correspond to regions which are not associated with any vanishing point in the image. We will analyze the composition of these images by first separating the foreground objects from the background. We note that, while our analysis of the perspective geometry provides valuable information about the 3D space, many popular composition rules studied in early work, such as the simplicity of the scene, golden ratio, rule of thirds, and visual balance have focused on the arrangement of objects in the 2D image plane. We believe that combining the strength of both approaches will enable us to obtain a deeper understanding of the composition of these images.

Finally, besides providing on-site feedback to photographers, the proposed method has many potential real-world applications. For example, it can be used to retrieve images with similar composition in a large-scale image database. Here, a major challenge is to assess the relevance of the perspective effect in the overall composition of a photo. Possible solutions include using the metadata from photo-sharing websites, and developing new automatic methods (*e.g.*, via the detection of strong vanishing points). As another example, the proposed method can be employed to further summarize the query results. When a query results in a large number of images that have similar levels of visual similarity or aesthetic quality, the query results can be structured as a tree with levels of refinement in terms of composition by grouping the images using a hierarchical clustering scheme.

7. REFERENCES

[1] P. Arbelaez. Boundary extraction in natural images using ultrametric contour maps. In *POCV*, 2006.

[2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, 2011.

[3] O. Barinova, V. Konushin, A. Yakubenko, K. Lee, H. Lim, and A. Konushin. Fast automatic single-view 3-d reconstruction of urban scenes. In *ECCV (2)*, pages 100–113, 2008.

[4] S. Bhattacharya, R. Sukthankar, and M. Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *ACM Multimedia*, pages 271–280, 2010.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *ECCV (3)*, pages 288–301, 2006.

[6] S. Dhar, V. Ordonez, and T. L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *CVPR*, pages 1657–1664, 2011.

[7] C. Fang, Z. Lin, R. Mech, and X. Shen. Automatic image cropping using visual composition, boundary simplicity and content preservation models. In *ACM Multimedia*, pages 1105–1108, 2014.

[8] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[9] S. Gould, R. Fulton, and D. Koller. Decomposing a scene into geometric and semantically consistent regions. In *ICCV*, pages 1–8, 2009.

[10] A. Gupta, A. A. Efros, and M. Hebert. Blocks world revisited: Image understanding using qualitative geometry and mechanics. In *ECCV (4)*, pages 482–496, 2010.

[11] F. Han and S. C. Zhu. Bottom-up/top-down image parsing by attribute graph grammar. In *ICCV*, pages 1778–1785, 2005.

[12] V. Hedau, D. Hoiem, and D. A. Forsyth. Recovering the spatial layout of cluttered rooms. In *ICCV*, pages 1849–1856, 2009.

[13] V. Hedau, D. Hoiem, and D. A. Forsyth. Thinking inside the box: Using appearance models and context based on room geometry. In *ECCV (6)*, pages 224–237, 2010.

[14] D. Hoiem, A. A. Efros, and M. Hebert. Automatic photo pop-up. *ACM Trans. Graph.*, 24(3):577–584, 2005.

[15] D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. *International Journal of Computer Vision*, 75(1):151–172, 2007.

[16] H. Kong, J.-Y. Audibert, and J. Ponce. Vanishing point detection for road detection. In *CVPR*, pages 96–103, 2009.

[17] J. Kosecká and W. Zhang. Video compass. In *ECCV (4)*, pages 476–490, 2002.

[18] D. A. Lauer and S. Pentak. *Design Basics*. Cengage Learning, 2011.

[19] D. C. Lee, M. Hebert, and T. Kanade. Geometric reasoning for single image structure recovery. In *CVPR*, pages 2136–2143, 2009.

[20] J. Li. Agglomerative connectivity constrained clustering for image segmentation. *Statistical Analysis and Data Mining*, 4(1):84–99, 2011.

[21] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or. Optimizing photo composition. *Comput. Graph. Forum*, 29(2):469–478, 2010.

[22] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *ECCV (3)*, pages 386–399, 2008.

[23] H. Mobahi, S. Rao, A. Y. Yang, S. S. Sastry, and Y. Ma. Segmentation of natural images by texture and boundary compression. *International Journal of Computer Vision*, 95(1):86–98, 2011.

[24] V. Nedovic, A. W. M. Smeulders, A. Redert, and J.-M. Geusebroek. Stages as models of scene geometry. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9):1673–1687, 2010.

[25] P. Obrador, R. de Oliveira, and N. Oliver. Supporting personal photo storytelling for social albums. In *ACM Multimedia*, pages 561–570, 2010.

[26] P. Obrador, L. Schmidt-Hackenberg, and N. Oliver. The role of image composition in image aesthetics. In *ICIP*, pages 3185–3188, 2010.

[27] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[28] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

[29] C. Rasmussen. Grouping dominant orientations for ill-structured road following. In *CVPR (1)*, pages 470–477, 2004.

[30] B. C. Russell, A. A. Efros, J. Sivic, B. Freeman, and A. Zisserman. Segmenting scenes by matching image composites. In *NIPS*, pages 1580–1588, 2009.

[31] A. Saxena, M. Sun, and A. Y. Ng. Make3d: Learning 3d scene structure from a single still image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(5):824–840, 2009.

[32] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.

[33] J.-P. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *ICCV*, pages 1250–1257, 2009.

[34] E. Tretiak, O. Barinova, P. Kohli, and V. S. Lempitsky. Geometric image parsing in man-made environments. *International Journal of Computer Vision*, 97(3):305–321, 2012.

[35] L. Yao, P. Suryanarayan, M. Qiao, J. Z. Wang, and J. Li. Oscar: On-site composition and aesthetics feedback through exemplars for photographers. *International Journal of Computer Vision*, 96(3):353–383, 2012.

[36] Y. Zhang, X. Sun, H. Yao, L. Qin, and Q. Huang. Aesthetic composition represetation for portrait photographing recommendation. In *ICIP*, pages 2753–2756, 2012.

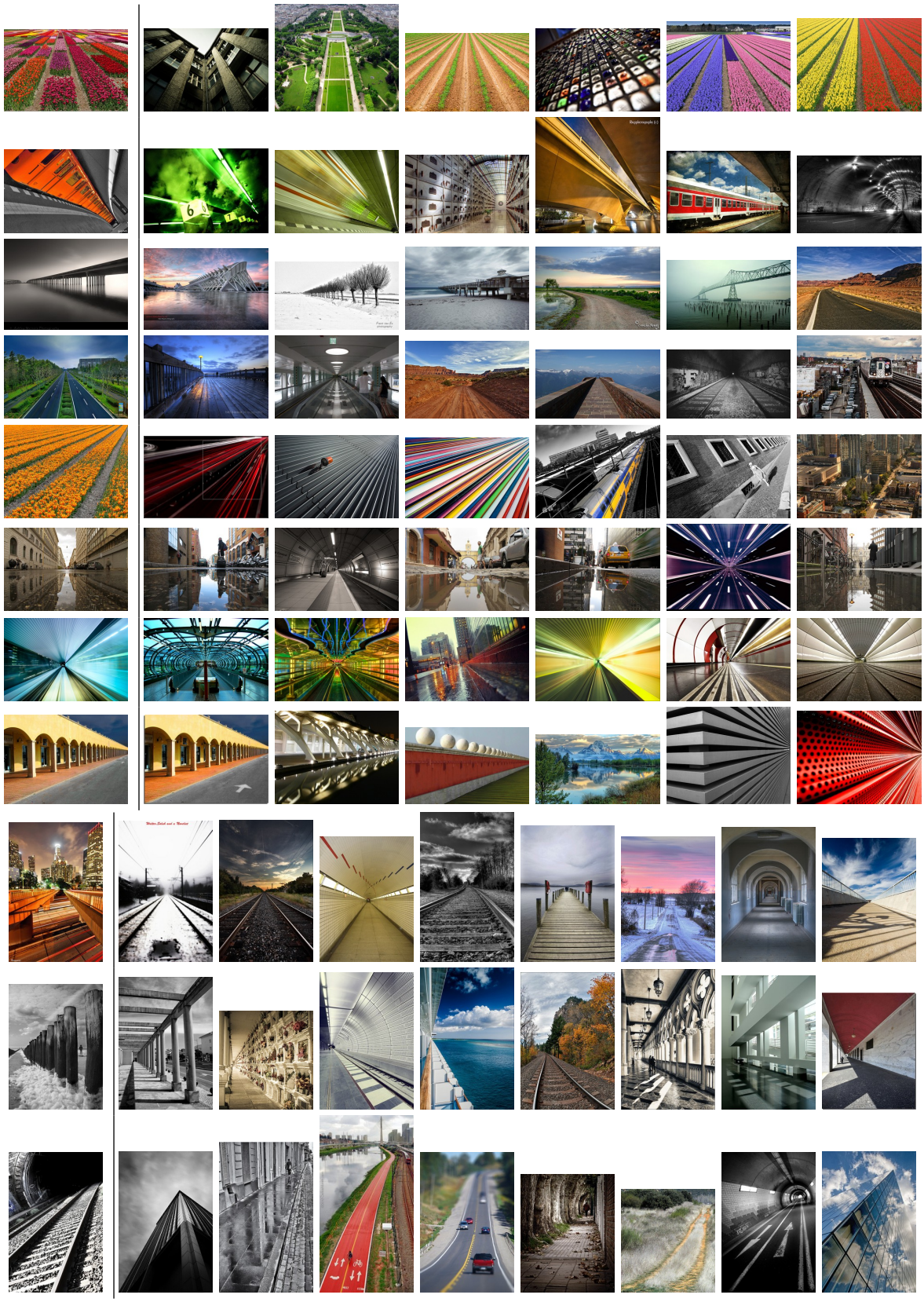


Figure 10: Composition-sensitive image retrieval results. Each row shows a query image (first image from the left) and the top-6 or top-8 ranked images retrieved.