Photo Composition Feedback and Enhancement
— Exploiting Spatial Design Categories and the Notan Dark-Light Principle

Jia Li, Lei Yao and James Z. Wang

1 Introduction

Cameras on mobile phones are becoming the primary means of photo creation for common people. Because of the convenience of mobile phones, it is effortless to take snapshots and share with others. As a result, pictures are being created at a much faster pace. It is estimated that as many as one trillion photos will be taken in the year of 2015. Software tools that make it easier for average photographers to improve photo taking will likely have broad acceptance. Understanding visual aesthetics (Datta et al., 2006) can aid various applications including summarization of photo collections (Obrador et al., 2010), selection of high quality images for display (Fogarty et al., 2001), and extraction of aesthetically pleasing images for retrieval (Obrador et al., 2009). It can also be used to render feedback to the photographer on the aesthetics of his/her photographs.

In order to make image aesthetic quality assessment more dynamic and to reach out to the general public with a practical perspective, we conducted research to develop new technologies that can provide on-site feedback to the photographers (Yao et al., 2012). We focused on feedback from a high-level composition perspective. *Composition* is the art of putting components together with conscious thoughts. In photography, it concerns the arrangement of various visual elements, such as line, color, texture, tone, and space. Composition is closely related

Jia Li
Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania, USA. e-mail: jiali@psu.edu

Lei Yao
Houzz Inc., Palo Alto, California, USA. The work was done when she was with College of Information Sciences and Technology, The Pennsylvania State University, University Park, Pennsylvania, USA. e-mail: simplely@gmail.com

James Z. Wang
College of Information Sciences and Technology, The Pennsylvania State University, University Park, Pennsylvania, USA. e-mail: jwang@psu.edu

to the aesthetic qualities of photographs. Partly because the problem is not well defined, insufficient research efforts have been placed on photographic composition within technical fields such as image processing and computer vision. We studied photographic composition from the perspective of spatial design, *i.e.*, how visual elements are geometrically arranged in a picture.

Providing instant feedback on the composition style can help photographers reframe the subject leading to an aesthetically composed image. We recognized that the abstraction of composition can be done by analyzing the arrangement of the objects in the image. This led us to identify five different forms of compositions, namely, textured images, and diagonally, vertically, horizontally, and center composed images. In our work, these composition types are recognized by three classifiers, *i.e.* the "textured" vs. "non-textured" classifier, the diagonal element detector, and the k-NN classifier for "horizontal", "vertical", and "centered" composition categories. Understanding the composition layout of the query image facilitates the retrieval of images that are similar in composition and content.

Many other applications have been built around suggesting improvisations to the image composition (Bhattacharya et al., 2010; Liu et al., 2001) through image retargeting, and color harmony (Cohen-Or et al., 2006) to enhance aesthetics. These applications are more off-line in nature. Although they are able to provide useful feedback, it is not on the spot, and requires considerable input from the user. On-site professional feedback that we propose can accomplish image improvements that are impossible once the photographer moves away from the photo-taking location.

Building upon our feedback framework, we developed a new method to provide tonal adjustment function based on exemplar pictures chosen by the user. The retrieved images provided by the composition feedback serve as candidates for the exemplar. With a simple click, even on a mobile device a user can pick an exemplar from a short list of images. Particularly in the current work, we make use of an important composition or design concept of dark and light arrangement of masses, sometimes referred to as *"Notan"* by artists. The Notan is fundamental to a composition that artists are advised to examine the Notan of a painting before heading out to paint (Raybould, 2014).

In the tonal adjustment, we try to reach a chosen Notan design by transforming the tonal values. This is in some measure like the dodging and burning operations performed in the darkroom by analog photographers. In dodging and burning, the photographer chooses an area to darken or brighten so that details in such areas can be brought out to enhance the overall composition. In our work, for the consideration of both the limitation of the mobile device and the fact that general users are not necessarily knowledgeable in photography, the computer system automatically determines the areas that should be brightened or darkened, as well as the level of adjustment. The decision is guided by a Notan design, which can be either automatically suggested by the computer or selected by the user from a number of candidates. The involvement of the user is minimal. While tonal adjustment has been a common image processing technique, our approach offers a new perspective because it is based on high-level composition concept of Notan rather than low-level features such as contrast and dynamic range.

Future generations of digital cameras are expected to have access to the high-speed mobile network and possess substantial internal computational power, the same way as today's smart phones. Camera phones can already send photos to a remote server on the Internet and receive feedback from the server (Sorrel, 2010). As a photographer composes, the photos in a lower resolution are streamed via the network to a cloud server. Our software system on the server analyzes the photos and sends on-site feedback to the photographer so that immediate recomposition can be possible. We propose a system comprising of the modules described below.

Given an input image, the **composition analyzer** evaluates its composition properties from different perspectives. For example, visual elements with great compositional potential, such as diagonals and curves, are detected. Photographs are categorized by high-level composition properties. Composition-related qualities, *e.g.*, visual balance and simplicity of background, are also evaluated. Images similar in composition as well as content can be retrieved from a database of photos with high aesthetic ratings so that the photographer can learn through examples.

In the **retrieval module**, a ranking scheme is designed to integrate the composition properties into a content-based retrieval system. In our experiments, we used SIMPLIcity, an image retrieval system based on color, texture and shape features (Wang et al., 2001). Images with high aesthetic ratings, as well as similar composition properties and visual features, are retrieved. An effective way to learn photography as a beginner is often through observing master works and imitating. Practicing good composition in the field helps develop creative sensibility and even unique styling. Especially for amateur photographers, well-composed photographs can be valuable learning resources. By retrieving high-quality similarly composed photographs, our approach can provide users with practical assistance in improving photography composition.

In the **enhancement module**, tonal adjustment can be made to achieve better composition. We explore the concept of Notan, a crucial factor in composition regarding the arrangement of dark and light masses in an image. A new tonal transformation method is developed to achieve the desired Notan design with minimal required user interactions.

The rest of the chapter is organized as follows. The categorization of spatial design is presented in Section 2, with corresponding evaluation results in Section 3. We describe our Notan-guided tonal transform in Section 4. Experiments on the tonal transform method are provided in Section 5. We summarize in Section 6.

2 Spatial Design Categorization

After studying many guiding principles in photography, we find that there are several typical spatial designs. Our goal is to automatically classify major types of spatial designs. In our work, we consider the following typical composition categories: horizontal, vertical, centered, diagonal, and textured.

According to long-existing photography principles, lines formed by linear elements are important because they lead the eye through the image and contribute to the mood of the photograph. Horizontal, vertical, and diagonal lines are associated with serenity, strength, and dynamism respectively (Krages, 2005). We thus include horizontal, vertical, and diagonal in the composition categories. Photographs with a centered main subject and a clear background fall into the category called centered. The photos in the textured category appear like a patch of texture or a relatively homogeneous pattern, for example, a brick wall.

The five categories of composition are not mutually exclusive. We apply several classifiers sequentially to an image: textured versus non-textured, diagonal versus non-diagonal, and finally a possibly overlapping classification of horizontal, vertical, and centered compositions. For example, an image can be classified as non-textured, diagonal, and horizontal. We use a method in (Wang et al., 2001) to classify textured images. It has been demonstrated that retrieval performance can be improved for both textured and non-textured images by first classifying them (Wang et al., 2001). The last two classifiers are developed in the current work, with details to be presented later.

A conventional image retrieval system returns images according to visual similarity. However, photographers often need to search for pictures based on composition rather than visual details. To accomodate this, we integrate composition classification with the SIMPLIcity image retrieval system (Wang et al., 2001). Furthermore, we provide the option to rank retrieved images by their aesthetic ratings so that the user can focus on highly-rated photos.

2.1 The Dataset

The spatial composition classification method is tested on a dataset crawled from `photo.net`, a photography community where peers can share, rate, and critique photos. These photographs are mostly general-purpose pictures and have a wide range of aesthetic quality. Among the crawled photos, a large proportion have frames which can distort the visual content in image processing and impact analysis results. We remove frames from the original images in a semi-automatic fashion. The images containing frames are picked manually and a program is used to remove simple frames with flat tones. Frames embedded with pattern or text usually cannot be correctly removed. These photos are simply removed from the dataset when we re-check the cropped images in order to make sure the program has correctly removed the frames from images. We construct a dataset with $13,302$ unframed pictures. Those pictures are then rescaled so that the long side of the image has at most 256 pixels. We manually labeled 222 photos, among which 50 are horizontally composed, 51 are vertically composed, 50 are centered, and 71 are diagonally composed. Our classification algorithms are developed and evaluated based on this manually-labeled dataset. The entire dataset are used in system performance evaluation.

2.2 Textured vs. Non-textured Classifier

We use the textured vs. non-textured classifier in SIMPLIcity to separate textured images from the rest. The algorithm is motivated by the observation that if pixels in a textured area are clustered using local features, each cluster of pixels yielded are scattered across the area due to the homogeneity appearance of texture. For non-textured images, on the other hand, the clusters tend to be clumped. An image is divided evenly into $4 \times 4 = 16$ large blocks. The algorithm thus calculates the proportion of pixels in each cluster that belong to any of the 16 blocks. If the cluster of pixels is scattered over the whole image, the proportions over the 16 blocks are expected to be roughly uniform. For each cluster, the χ^2 statistic is computed to measure the disparity between the proportions and the uniform distribution over the 16 blocks. The average value of the χ^2 statistics for all the clusters is then thresholded to determine whether an image is textured or not.

2.3 Diagonal Design Element Detection

Diagonal elements are strong compositional constituents. The diagonal rule in photography states that a picture appears more dynamic if the objects fall or follow a diagonal line. Photographers often use diagonal elements as the visual path to draw viewers' eyes through the image.[1] The visual path is the path of eye movement when viewing a photograph (Warren, 2002). When such a visual path stands out in the picture, it also has the effect of uniting individual parts in a picture. The power of the diagonal lines in composition was exploited very early on by artists. For instance, Speed (1972) discussed in great details how Velazquez used the diagonal lines to unite a picture in his painting "The Surrender of Breda".

Because of the importance of diagonal visual paths for composition, we create a spatial composition category for diagonally composed pictures. More specifically, there are two subcategories, diagonal from upper left to bottom right ($\backslash$) and from upper right to bottom left ($/$). We declare the composition of a photo as diagonal if diagonal visual paths can be detected.

Detecting the exact diagonal visual paths is challenging. Typically, segmented regions or edges provided by image processing techniques can only be viewed as *ingredients*, aka local patterns, either because of the nature of the picture or the limitation of the processing algorithms. In contrast, an *element* refers to a global pattern, *e.g.*, a broken curve (multiple detectable edges) that is present in a large area of the image plane.

There has been literature on the general principles regarding visual elements, to be briefly described below. We designed our algorithm for detecting diagonal visual paths according to these principles. While we present these principles using the

[1] http://www.digital-photography-school.com/using-diagonal-lines-in-photography

diagonal category as an example, they apply in a similar way to other directional visual paths.

Fig. 1 Photographs of diagonal composition.

1. *Principle of multiple visual types*: Lines are effective design elements in creating compositions, but perfectly straight lines rarely exist in the natural world. Lines we perceive in photographs usually belong to one of these types: outlines of forms, narrow forms, lines of arrangement, and lines of motion or force (Feininger, 1973). We do not restrict diagonal elements to actual diagonal lines of an image plane. They can be the boundary of a region, a linear object, or even an imaginary line along which different objects align. Linear objects, such as pathways, waterways, and the contour of a building, can all create visual paths in photographs. When placed diagonally, they are generally perceived as more dynamic and interesting than other compositions. Figure 1 shows examples of using diagonal compositions in photography.
2. *Principle of wholes, or Gestalt Law*: Gestalt psychologists studied early on the phenomenon of human eyes perceiving visual components as organized patterns or wholes, known as the Gestalt law of organization. According to the Gestalt Law, the factors that aid in human visual perception of forms include proximity, similarity, continuity, closure, and symmetry (Sternberg et al., 2008).
3. *Principle of tolerance*: Putting details along diagonals creates more interesting compositions. Visual elements such as lines and regions slightly off the ideal diagonal direction can still be perceived as diagonal and are usually more natural and interesting.[2]
4. *Principle of prominence*: A photograph can contain many lines, but dominant lines are the most important in regard to the effect of the picture (Folts, 2005).[3] Visual elements need sufficient span along the diagonal direction in order to strike a clear impression.

Following the above principles, we first find diagonal ingredients from low-level visual cues using both regions obtained by segmentation and connected lines obtained by edge detection. Then, we apply the Gestalt Law to merge the ingredients into elements, *i.e.*, more global patterns. The prominence of each merged entity is

[2] http://www.picture-thoughts.com/photography/compos-ition/angle/

[3] http://www.great-landscape-photography.com/photography-composition.html

then assessed. We now describe the algorithms for detecting diagonal visual paths using segmented regions and edges, respectively.

Diagonal Segment Detection: Image segmentation is often used to simplify the image representation. It can generate semantically meaningful regions that are easier for analysis. We describe below our approach to detecting diagonal visual paths based on segmented regions. We use a recent image segmentation algorithm (Li, 2011) because it achieves state-of-the-art accuracy at a speed sufficiently fast for real-time applications. The algorithm also ensures that the segmented regions are spatially connected, a desirable trait many algorithms do not possess.

After image segmentation, we find the orientation of each segment, defined as the orientation of the moment axis of the segment. The moment axis is the direction along which the spatial locations of the pixels in the segment have maximum variation. It is the first principal component direction for the set of pixel coordinates. For instance, if the segment is an ellipse (possibly tilted), the moment axis is simply the long axis of the ellipse. The orientation of the moment axis of a segmented region measured in degrees is computed according to Russ (2006).

Next, we apply the Gestalt Law to merge certain segmented regions in order to form visual elements. Currently, we only deal with a simple case of disconnected visual path, where the orientations of all the disconnected segments are diagonal.

Let us introduce a few notations before describing the rules for merging. We denote the normalized column vector of the diagonal direction by $\mathbf{v}_d$ and that of its orthogonal direction by $\mathbf{v}_d^c$. We denote a segmented region by S, which is a set of pixel coordinates $\mathbf{x} = (x_h, x_v)^t$. The projection of a pixel with coordinate $\mathbf{x}$ onto any direction characterized by its normalized vector $\mathbf{v}$ is the inner product $\mathbf{x} \cdot \mathbf{v}$. The projection of S onto $\mathbf{v}$, denoted by $\mathscr{P}(S, \mathbf{v})$, is a set containing the projected coordinates of all the pixels in S. That is, $\mathscr{P}(S, \mathbf{v}) = \{\mathbf{x} \cdot \mathbf{v} : \mathbf{x} \in S\}$. The length (also called spread) of the projection $|\mathscr{P}(S, \mathbf{v})| = \max_{\mathbf{x}_i, \mathbf{x}_j \in S} |\mathbf{x}_i \cdot \mathbf{v} - \mathbf{x}_j \cdot \mathbf{v}|$ is the range of values in the projected set.

The rules for merging, *i.e.*, similarity, proximity, and continuity, are listed below. Two segments satisfying all of the rules are merged.

- *Similarity*: Two segments S_i, $i = 1,2$, with orientations e_i, $i = 1,2$, are similar if the following criteria are satisfied:
  1. Let $[\check{\varphi}, \hat{\varphi}]$ be the range for nearly diagonal orientations. $\check{\varphi} \le e_i \le \hat{\varphi}$, $i = 1,2$. That is, both S_1 and S_2 are nearly diagonal.
  2. The orientations of S_i, $i = 1,2$, are close:

$$|e_1 - e_2| \le \beta \text{ ,where } \beta \text{ is a predefined threshold.}$$

  3. The lengths of $\mathscr{P}(S_i, \mathbf{v}_d)$, $i = 1,2$, are close:

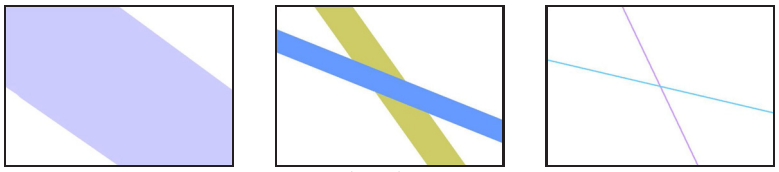
$$r = \frac{|\mathscr{P}(S_1, \mathbf{v}_d)|}{|\mathscr{P}(S_2, \mathbf{v}_d)|} \text{ , } r_1 \le r \le r_2 \text{ ,}$$

     where $r_1 < 1$ and $r_2 > 1$ are predefined thresholds.

- *Proximity*: Segments S_i, $i = 1,2$, are proximate if their projections on the diagonal direction, $\mathscr{P}(S_i, \mathbf{v}_d)$, $i = 1,2$, are separated by less than p, and the overlap of their projections is less than q.
- *Continuity*: Segments S_i, $i = 1,2$, are continuous if their projections on the direction orthogonal to the diagonal, $\mathscr{P}(S_i, \mathbf{v}_d^c)$, $i = 1,2$, are overlapped.

We select the thresholds according to the following:

1. $\beta = 10^\circ$.
2. $r_1 = 0.8$, $r_2 = 1.25$.
3. The values of p and q are decided adaptively according to the sizes of S_i, $i = 1,2$. Let the spread of S_i along the diagonal line be $\lambda_i = |\mathscr{P}(S_i, \mathbf{v}_d)|$. Then $p = k_p \min(\lambda_1, \lambda_2)$ and $q = k_q \min(\lambda_1, \lambda_2)$, where $k_p = 0.5$ and $k_q = 0.8$.
   The value of p determines the maximum gap allowed between two disconnected segments to continue a visual path. The wider the segments spread over the diagonal line, the more continuity they present to the viewer. Therefore, heuristically, a larger gap is allowed, which is why p increases with the spreads of the segments. On the other hand, q determines the extent of overlap allowed for the two projections. By a similar rationale, q also increases with the spreads. If the projections of the two segments overlap too much, the segments are not merged because the combined spread of the two differs little from the individual spreads.
4. The angular range $[\check{\varphi}, \hat{\varphi}]$ for nearly diagonal orientations is determined adaptively according to the geometry of the rectangle bounding the image.



(a) Single stripe (b) $\frac{1}{6} \rightarrow \frac{1}{3}$ stripes (c) Angular range

Fig. 2 Diagonal orientation bounding conditions.

As stated in (Lamb et al., 2010), one practical extension of the diagonal rule is to have the objects fall within two boundary lines parallel to the diagonal. These boundary lines are one-third of the perpendicular distance from the diagonal to the opposite vertex of the rectangular photograph. This diagonal stripe area is shown in Figure 2(a). A similar suggestion is made in an online article,[2] where boundary lines are drawn using the so-called sixth points on the borders of the image plane. A sixth point along the horizontal border from the upper left corner locates on the upper border and is away from the corner by one-sixth of the image width. Similarly, we can find other sixth (or third) points from any corner and either horizontally or vertically.

Suppose we look for an approximate range for the diagonal direction going from the upper left corner to the bottom right. The sixth and third points with respect to the two corners are found. As shown in Figure 2(b), these special points are used to create two stripes marked by lime and blue colors respectively. Let the orientations of the lime stripe and the blue stripe in Figure 2(b) be φ_1 and φ_2. Then we set $\check{\varphi} = \min(\varphi_1, \varphi_2)$, and $\hat{\varphi} = \max(\varphi_1, \varphi_2)$. A direction $\mathbf{v} \in [\check{\varphi}, \hat{\varphi}]$ is claimed nearly diagonal. Similarly, we can obtain the angular range for the diagonal direction from the upper right corner to the bottom left. The orientations of the stripes is used, instead of nearly diagonal bounding lines, because when the width and the height of an image are not equal, the orientation of a stripe twists toward the elongated side to some extent.

From now on, a "segment" can be a merged entity of several segments originally provided by the segmentation algorithm. For brevity, we still call the merged entity a segment. Applying the principle of tolerance, we filter out a segment from diagonal if its orientation is outside the range $[\check{\varphi}, \hat{\varphi}]$, the same rule that was applied to the smaller segments before merging.

After removing non-diagonal segments, at last, we apply the principle of prominence to retain only segments with a significant spread along the diagonal direction. For segment S, if $|\mathscr{P}(S, \mathbf{v}_d)| \geq k_l \times l$, where l is the length of the diagonal line and $k_l = \frac{2}{3}$ is a threshold, the segment is declared a diagonal visual path. It is observed that a diagonal visual path is often a merged entity of several small and not prominent individual segments originally produced by the segmentation algorithm.

Diagonal Edge Detection: According to the principle of multiple visual types, besides segmented regions, lines and edges can also form visual paths. Moreover, segmentation can be unreliable sometimes because over-segmentation and under-segmentation often cause diagonal elements to be missed. We observe that among photographs showing diagonal composition, many contain linear diagonal elements. Those linear diagonal elements usually have salient boundary lines along the diagonal direction, which can be found through edge detection. Therefore, we use edges as another visual cue, and combine the results obtained based on both edges and segments to increase the sensitivity of detecting diagonal visual paths.

We use the Edison algorithm for edge detection (Meer and Georgescu, 2001). It has been experimentally demonstrated that the edge detection can generate cleaner edge maps than many other methods. We examine all the edges to find those oriented diagonally and significant enough to be a visual path.

Based on the same set of principles, the whole process of finding diagonal visual paths based on edges is similar to the detection of diagonal segments. The major steps are described below. We denote an edge by E, which is a set of coordinates of pixels located on the edge. As with segments, we use the notation $\mathscr{P}(E, \mathbf{v})$ for the projection of E on a direction $\mathbf{v}$.

1. *Remove non-diagonal edges*: First, edges outside the diagonal stripe area, as shown in Figure 2(a), are excluded. Second, for every edge E, compute the spread of the projections $s_d = |\mathscr{P}(E, \mathbf{v}_d)|$ and $s_o = |\mathscr{P}(E, \mathbf{v}_d^c)|$. Recall that $\mathbf{v}_d$ is the diagonal direction and $\mathbf{v}_d^c$ is its orthogonal direction. Based on the ratio s_d/s_o, we

compute an approximation for the orientation of edge E. Edges well aligned with the diagonal line yield a large value of s_d/s_o, while edges well off the diagonal line have a small value. We filter out non-diagonal edges by requiring $s_d/s_o \geq \zeta$. The choice of ζ will be discussed later.

2. *Merge edges*: After removing non-diagonal edges, short edges along the diagonal direction are merged into longer edges. The merging criterion is similar to the proximity rule used for diagonal segments. Two edges are merged if their projections onto the diagonal line are close to each other but not excessively overlapped.
3. *Examine prominence*: For edges formed after the merging step, we check their spread along the diagonal direction. An edge E is taken as a diagonal visual element if $|\mathscr{P}(E, \mathbf{v}_d)| \geq \xi$, where ξ is a threshold to be described next.

The values of thresholds ζ and ξ are determined by the size of a given image. ζ is used to filter out edges whose orientations are not quite diagonal, and ξ is used to select edges that spread widely along the diagonal line. We use the third points on the borders of the image plane to set bounding conditions. Figure 2(c) shows two lines marking the angular range allowed for a nearly diagonal direction from the upper left corner to the lower right corner. Both lines in the figure are off the ideal diagonal direction to some extent. Let ζ_1 and ζ_2 be their ratios of s_d to s_o, and ξ_1 and ξ_2 be their spreads over the diagonal line. The width and height of the image are denoted by w and h. By basic geometry, we can calculate ζ_i and ξ_i, $i = 1, 2$, using the formulas:

$$\zeta_1 = \frac{h^2 + 3w^2}{2hw}, \;\; \zeta_2 = \frac{3h^2 + w^2}{2hw}, \;\; \xi_1 = \frac{h^2 + 3w^2}{3\sqrt{h^2 + w^2}}, \;\; \xi_2 = \frac{3h^2 + w^2}{3\sqrt{h^2 + w^2}}\,.$$

The thresholds are then set by $\zeta = \min(\zeta_1, \zeta_2)$ and $\xi = \min(\xi_1, \xi_2)$.

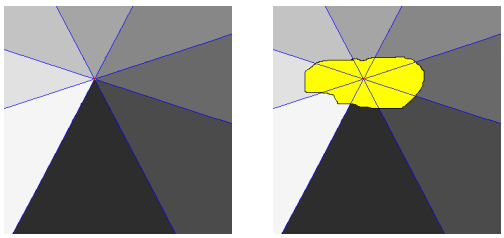
2.4 Horizontal, Vertical and Centered Compositions

Now we present our method for differentiating the remaining three composition categories: horizontal, vertical, and centered. Photographs belonging to each of these categories have distinctive spatial layouts. For instance, a landscape with blue sky at the top and a grass field at the bottom conveys a strong impression of horizontal layout. Images from a particular category usually have some segments that are characteristic of that category, *e.g.*, a segment lying laterally right to left for horizontal photographs, and a homogeneous background for centered photographs.

In order to quantitatively characterize spatial layout, we define the *spatial relational vector* (SRV) of a region to specify the geometric relationship between the region and the rest of the image. The spatial layout of the entire image is then represented by the set of SRVs of all the segmented regions. The dissimilarity between spatial layouts of images is computed by the IRM distance (Li et al., 2000). Ideally, we want to describe the spatial relationship between each semantically

meaningful object and its surrounding space. However, object extraction is inefficient and extremely difficult for photographs in general domain, regions obtained by image segmentation algorithms are used instead as a reasonable approximation.

The SRV is proposed to characterize the geometric position and the peripheral information about a pixel or a region in the image plane. It is defined at both the pixel level and the region level. When computing the pixel-level SRV, the pixel is regarded as the reference point, and all the other pixels are divided into eight zones by their relative positions to the reference point. If the region that contains the pixel is taken into consideration, SRV is further differentiated into two modified forms, inner SRV and outer SRV. The region-level inner (outer) SRV is obtained by averaging pixel-level inner (outer) SRVs over the region. Details about SRV implementation are given below. SRV is scale-invariant, and depends on the spatial position and the shape of the segment.



(a) (b)

Fig. 3 Division of the image into eight angular areas with respect to a reference pixel.

At a pixel with coordinates (x,y), four lines passing through it are drawn. As shown in Figure 3(a), the angles between adjacent lines are equal and stride symmetrically over the vertical, horizontal, 45° and 135° lines. We call the eight angular areas of the plane upper, upper-left, left, bottom-left, bottom, bottom-right, right, and upper-left zones. The SRV of the pixel (x,y) summarizes the angular positions of all the other pixels with respect to (x,y). Specifically, we calculate the area percentage v_i of each zone, $i = 0,...,7$, with respect to the whole image and construct the pixel-level SRV $V_{x,y}$ by $V_{x,y} = (v_0, v_1, ..., v_7)^t$.

The region-level SRV is defined in two forms, called inner SRV, denoted by V', and outer SRV, denoted by V'', respectively. At any pixel in a region, we can divide the image plane into eight zones by the above scheme. As shown in Figure 3(b), for each of the eight zones, some pixels are inside the region and some are outside. Depending on whether a pixel belongs to the region, the eight zones are further divided into 16 zones. We call those zones within the region as inner pieces and those outside as outer pieces. Area percentages of the inner (or outer) pieces with respect to the area inside (or outside) the region form the inner SRV $V'_{x,y}$ (or outer SRV $V''_{x,y}$) for pixel (x,y).

The region-level SRV is defined as the average of pixel-level SRVs for pixels in that region. The outer SRV V_R'' of a region R is $V_R'' = \sum_{(x,y)\in R} V_{x,y}''/m$, where m is the number of pixels in region R. In practice, to speed up the calculation, we may subsample the pixels (x,y) in R and compute V_R'' by averaging over only the sampled pixels. If a region is too small to occupy at least one sampled pixel according to a fixed sampling rate, we compute V_R'' using the pixel at the center of the region.

We use the outer SRV to characterize the spatial relationship of a region with respect to the rest of the image. Then an image with N segments R_i, $i = 1, ..., N$, can be described by N region-level outer SRVs, V_{R_i}'', $i = 1, ..., N$, together with the area percentages of R_i, denoted by w_i. In summary, an image-level SRV descriptor is a set of weighted SRVs: $\{(V_{R_i}'', w_i), i = 1, ..., N\}$. We call this descriptor the *spatial layout signature*.

We use k-NN to classify the three composition categories: horizontal, vertical, and centered. Inputs to the k-NN algorithm are the spatial layout signatures of images. The training dataset includes equal number of manually-labeled examples in each category. In our experiment, the sample size for each category is 30. The distance between the spatial layout signatures of two images is computed using the IRM distance. The IRM distance is a weighted average of the distances between any pair of SRVs, one in each signature. The weights are assigned in a greedy fashion so that the final weighted average is minimal. Details about IRM are referred to (Li et al., 2000; Wang et al., 2001).

We conducted our experiments on a single compute node with two quadcore Intel processors running at 2.66 GHz and 24 GB of RAM. For the composition analysis process, the average time to process a 256×256 image is three seconds, including image segmentation (Li, 2011), edge detection (Meer and Georgescu, 2001), and the composition classification as described.

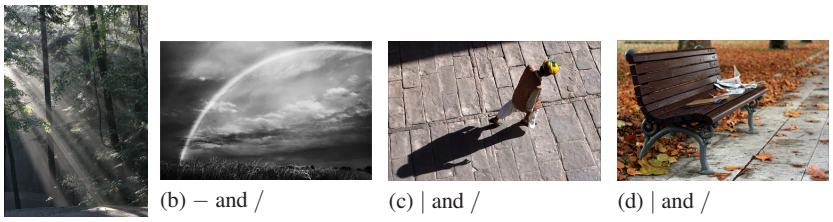
2.5 Composition-sensitive Photo Retrieval

The classic approach taken by many image retrieval systems (Datta et al., 2008) is to measure the visual similarity based on low-level features. A large family of visual descriptors have been proposed in the past to characterize images from the perspectives of color, texture, shape, interesting points, etc. However, due to the fact that many visual descriptors are generated by local feature extraction processes, the overall spatial composition of the image is usually lost. In semantic content oriented applications, spatial layout information of an image may not be critical. But for photography applications, the overall spatial composition can be a critical factor affecting how an image is perceived. For photographers, it is often more interesting to search for photos with similar composition or design style rather than visual details. As described above, our algorithms capture strong compositional elements in photos and classify them into six composition categories, with five main categories named textured, horizontal, vertical, centered,

and diagonal, and the diagonal category is further divided into two categories diagonal$_{ulbr}$ (upper left to bottom right) and diagonal$_{urbl}$ (upper right to bottom left). The composition classification is used in the retrieval system to return images with similar composition.

We use the SIMPLIcity system to retrieve images with similar visual content, and then re-rank the top K images by considering their spatial composition and aesthetic ratings. SIMPLIcity is a semantic-sensitive region-based image retrieval system. IRM is used to measure visual similarity between images. For a thorough description of algorithms used in SIMPLIcity, readers are referred to the original publication (Wang et al., 2001). In our system, the rank of an image is determined by three factors: its visual similarity to the query, the spatial composition categorization, and the aesthetic rating. Since these factors are of different modality, we use a ranking scheme rather than a complicated scoring equation.

Given a query, we first retrieve K images through SIMPLIcity, which gives us an initial ranking. When composition is taken into consideration, images with the same composition categorization as the query are moved to the top of the ranking list.



(a) | and \ (b) − and / (c) | and / (d) | and /

Fig. 4 Photographs classified into multiple categories. Categories are shown with symbols.

The composition classification is non-exclusive in the context of image retrieval. For instance, a textured image can be classified concurrently into horizontal, vertical, or centered categories. We code the classification results obtained from the classifiers by a six-dimensional vector c, corresponding to six categories (recall that the diagonal category has two subcategories diagonal$_{ulrb}$ and diagonal$_{urbl}$). Each dimension records whether the image belongs to a particular category, with 1 being yes and 0 no. Note that an image can belong to multiple classes generated by different classifiers. The image can also be assigned to one or more categories among horizontal, vertical, and centered, if neighbors belonging to the category found by k-NN reach a substantial number (in our experiments $k/3$ is used). Non-exclusive classification is more robust than exclusive classification in practice because a photograph may be reasonably assigned to more than one composition category. Non-exclusive classification can also reduce the negative effect of misclassification into one class. Figure 4 shows example pictures that are classified as more than one category.

The compositional similarity between the query image and another image can be defined as

$$s_i = \sum_{k=0}^{3} I(c_{qk} = c_{ik} \text{ and } c_{qk} = 1) + 2\sum_{k=4}^{5} I(c_{qk} = c_{ik} \text{ and } c_{qk} = 1),$$

where c_q and c_i are categorization vectors for the query image and the other image, and I is the indicator function returning 1 when the input condition is true, 0 otherwise. The last two dimensions of the categorization vector correspond to the two diagonal categories. We multiply the matching function by 2 to encourage matching of diagonal categories in practice. Note that the value of s_i is between 0 and 6, because one image can at most be classified into five categories, which are textured, diagonal$_{ulbr}$, diagonal$_{urbl}$ and two of the other three. Therefore by adding composition classification results, we divide the K images into 8 groups corresponding to compositional similarity from 0 to 7. The original ranking based on visual similarity remains within each group. Although the composition analysis is performed on the results returned by SIMPLIcity, we can modify the influence of this component in the retrieval process by adjusting the number of images K returned by SIMPLIcity. The larger K is, the stronger factor composition is to overall retrieval.

3 Evaluation Results on Composition Feedback

The spatial design categorization process was incorporated as a component into our OSCAR (On-Site Composition and Aesthetics feedback through exemplars) system (Yao et al., 2012). User evaluation was conducted on composition layout classification, similarity and aesthetics quality of retrieved images, and the helpfulness of the feedback for improving photography. We only present results for the study on composition classification here. Interested readers are referred to that paper for comprehensive evaluation results. Professional photographers or enthusiasts would have been ideal subjects for such studies. However, due to time constraints, we were unable to recruit professionals. Instead, we recruited around 30 students, most of whom were graduate students at Penn State with practical knowledge of digital images and photography. All photos used in these studies are from `photo.net`.

A collection of around 1,000 images were randomly picked to form the dataset for the study on composition. Each participant is provided with a set of 160 randomly-chosen images and is asked to describe the composition layout of each image. At an online site, the participants can view pages of test images, next to each of which are seven selection buttons: “Horizontal”, “Vertical”, “Centered”, “Diagonal (upper left, bottom right)”, “Diagonal (upper right, bottom left)”, “Patterned”, and “None of the above”. Multiple choices are allowed. We used “Patterned” for the class of photos with homogeneous texture (or the textured class in our earlier description). We added the “none of the above” choice to allow more flexibility for the user’s perception. A total of 924 images were voted each by three or more users.

In order to understand compositional clarity, we examine the variation in users' votes on composition layout. We quantify the ambiguity in the choices of composition layout using entropy. The larger the entropy in the votes, the higher the ambiguity is in the composition layout of the image. The entropy is calculated by the formula $\sum p_i \log 1/p_i$, where $p_i, i = 0,...,6$, is the percentage of votes for each category. The entropy was calculated for all 924 photos and its value was found to range between 0 and 2.5 . We divided the range of entropy into five bins. The photos are divided into seven groups according to the composition category receiving the most votes. In each category, we compute the proportion of photos yielding a value of entropy belonging to any of the five bins. These proportions are reported in Table 1. We observe that among the seven categories, horizontal and centered categories have the strongest consensus among users, while "none of the above" is the most ambiguous category.

Table 1 Distribution of the entropy for the votes of users. For each composition category, the percentage of photos yielding a value of entropy in any bin is shown. h: horizontal, v: vertical, c: centered, ulbr: diagonal (upper left, bottom right), urbl: diagonal (upper right, bottom left), t: textured, none: none of the above.

	[0,0.5]	(0.5,1.0]	(1.0,1.5]	(1.5,2.0]	(2.0, 2.5]
h	36.12	29.96	17.18	15.42	1.32
v	12.98	45.67	19.71	20.19	1.44
c	25.36	45.48	13.12	14.87	1.17
ulbr	12.99	44.16	19.48	19.48	3.90
urbl	16.87	43.37	18.07	20.48	1.20
t	10.77	36.92	10.77	36.92	4.62
none	6.59	39.56	17.58	34.07	2.20

We evaluate our composition classification method in the case of both exclusive classification and non-exclusive classification. The users' votes on composition are used to form the ground truth, with specifics to be explained shortly. We consider only six categories, *i.e.* horizontal, vertical, centered, diagonal$_{ulbr}$, diagonal$_{urbl}$ and textured for this analysis. The "none of the above" category was excluded for the following reasons.

- The "none of the above" category is of great ambiguity among users, as shown by the above analysis.
- Only a very small portion of images is predominantly labeled as "none of the above". Among the 924 photos, 17 have three or more votes for "none of the above".
- We notice that these 17 "none of the above" photos vary greatly in visual appearance; and hence it is not meaningful to treat such a category as a compositionally-coherent group. It is difficult to define such a category. A portion of images in this category shows noisy or complex scenes without clear centers of attention. This can be a separate category for consideration in future work.

We conducted exclusive classification only on photos of little ambiguity according to users' choices of composition. The number of votes a category can receive ranges from zero to five. To be included in this analysis, a photo has to receive three or more votes for one category (that is, the ground-truth category) and no more than one vote for any other category. With this constraint, 494 out of the 924 images were selected. Table 2 is the confusion matrix based on this set of photos.

Table 2 The confusion matrix for exclusive classification of 494 images into six composition categories. Each row corresponds to a ground truth class. h: horizontal, v: vertical, c: centered, ulbr: diagonal (upper left, bottom right), urbl: diagonal (upper right, bottom left), t: textured, none: none of the above.

	h	v	c	ulbr	urbl	t
h	**107**	0	20	3	8	4
v	1	**32**	39	3	2	10
c	10	7	**132**	8	11	12
ulbr	4	0	5	**18**	0	2
urbl	2	1	13	0	**22**	1
t	0	2	6	0	0	**9**

We see that the most confusing category pairs are vertical vs. centered and diagonal$_{urbl}$ vs. centered. Figure 5(a) shows some examples labeled as vertical by users while classified as centered by our algorithm. We observe that the misclassification is mainly caused by the following: (1) vertical images in the training dataset cannot sufficiently represent this category; (2) users are prone to label images with vertically elongated objects as vertical although such images may be classified as centered in the training data; and (3) the vertical elements fail to be captured by image segmentation. Figure 5(b) gives diagonal$_{urbl}$ examples mistakenly classified as centered. The failure to detect diagonal elements results mainly from: (1) diagonal elements located beyond the diagonal tolerance set by our algorithm; and (2) imaginary diagonal visual paths, *e.g.*, the direction of an object's movement.

In non-exclusive classification, the criterion for a photo being assigned to one category is less strict than in the exclusive case. A photo is labeled as a particular category if it gets two or more votes on that category. In total there are 849 out of the 924 photos with at least one category voted twice or more. The results reported below is based on these 849 photos.

The composition categorization of a photo is represented by a six-dimensional binary vector, with 1 indicating the presence of a composition type, and 0 the absence. Let $M = (m_0, ..., m_5)$ and $U = (u_0, ..., u_5)$ denote the categorization vector generated by our algorithm and by users respectively. The value m_0 is set to 1 if and only if there are 10 or more nearest neighbors (among 30) labeled as horizontal. The values of m_1 and m_2, corresponding to the vertical and centered categories, are set similarly. For the diagonal categories, m_i, where $i = 3, 4$, is set to 1 if any diagonal

(a)

(b)

Fig. 5 Photo examples mistakenly classified as centered by our algorithm. (a) Photos labeled as vertical by users. (b) Photos labeled diagonal$_{urbl}$ by users.

element is detected by our algorithm. Finally, m_5 is set to 1 if the textured versus non-textured classifier labels the image as textured. Three ratios are computed to assess the accuracy of the non-exclusive classification.

- Ratio of partial detection r_1: the percentage of photos for which at least one of the user labeled categories is declared by the algorithm. Based on the 849 photos, $r_1 = 80.31\%$.
- Detection ratio r_2: the percentage of photos for which all the user labeled categories are captured by the algorithm. Define $M \succ U$ if $m_j \geq u_j$ for any $j \in [0,5]$. So r_2 is the percentage of images for which $M \succ U$. We have $r_2 = 66.00\%$.
- Ratio of perfect match r_3: the percentage of photos for which $M = U$. We have $r_3 = 33.11\%$.

4 Notan-guided Tonal Transform

The tonal value, *i.e.* the luminance, in a picture is a major factor for the visual impression conveyed by the picture. In art, the luminance at a location is simply called the value. Artists have remarked on the prominent role of values even for color paintings. Speed (1972) wrote:

> "By drawing is here meant the expression of form upon a plane surface. Art probably owes more to form for its range of expression than to color. Many of the noblest things it is capable of conveying are expressed by form more directly than by anything else. And it is

> interesting to notice how some of the world's greatest artists have been very restricted in their use of color, preferring to depend on form for their chief appeal."

While recognizing the importance of color, Payne (2005) remarked "Perhaps color might be called a non-essential factor in composition, since unity may be created without it." Regarding values, Payne (2005) wrote:

> "Dark and light usually refers to the range of values in the entire design while light and shade generally denote the lighted and shaded parts of single items. Both light and dark and light and shade are active factors in composition."

The use of light and shade to create the sense of solidity or relief on a plane surface, a technique called *chiaroscuro*, is an invention in the West. The giants in art, Leonardo Da Vinci, Raphael, Michelangelo, and Titian, are masters of this technique. The art of the East has a very different tradition, emphasizing the arrangement of dark and light in the overall design. Speed (1972) called this approach of the East *mass drawing*. Again quoting from (Speed, 1972),

> "The reducing of a complicated appearance to a few simple masses is the first necessity of the painter. . . . The art of China and Japan appears to have been more influenced by this view of natural appearances than that of the West has been, until quite lately. . . . Light and shade, which suggest solidity, are never used, a wide light where there is no shadow pervades everything, their drawing being done with the brush in masses. (referring to the East art)"

Until fairly modern time, Chinese paintings were mostly done in black ink, and even the colored ones have very limited range in chroma. In Chinese ink painting, a graceful juxtaposition of dark and light is a preeminent principle for aesthetics, called *Nong-Dan*. "Nong" literally means high concentration in liquid solution, while "Dan" means thin concentration. For ink, Nong-Dan refers to the concentration of black pigment. Hence, "Nong" leads to dark, and "Dan" leads to light. The same concept is used in Japanese painting and the Japanese imported directly the two Chinese characters in Kanji. The English translation from Kanji is *Notan*.

Relatively recently, Notan has been used in the West as a compact word meaning the overall design in black and white, or a small number of tonal scales. Mass Notan study focuses on the organization of simplified tonal structure rather than details. For example, a scene is reduced to an arrangement of major shapes (mass) with different levels of tonal values. The goal of a mass Notan study is to create a harmonious and balanced design (or "big picture"). Raybould (2014) recommends strongly the practice of mass Notan study as an initial step in painting to secure balanced and pleasing composition.

The essence of Notan is also well recognized in photography. Due to the difficulty in controlling light, especially in outdoor environments, photographers use dodging and burning techniques to achieve desired exposures for regions that cannot be reached by a single shot. Traditionally, dodging and burning are darkroom techniques applied during the film-to-paper printing process to alter the exposure of certain areas without affecting the rest of the photo. Specifically, dodging brightens an area, and burning darkens. Ansel Adams extensively used dodging-and-burning

techniques in developing many of his famous prints. He mentioned in his book *The Print* (Adams, 1995) that most of his prints are not the reproduction of the scenes but instead his visualization of the scenes. As Ansel Adams put it, "dodging and burning are steps to take care of mistakes God made in establishing tonal relationships."

In the digital era, to realize one's personal visualization, a photographer can modify the tonal structure using photo editing software. However, applying dodging and burning digitally can be time-consuming and requires a considerable level of mastery in photography, both technically and artistically.

In our work, we aim at developing a system that performs dodging and burning kind of adjustments on the tonal values of photographs with minimum user involvement. This is motivated by the need to enhance photos on mobile devices and to reach a broader set of users. The restrictive interface of the mobile device prohibits extensive manual photo editing. Moreover, an average user may not have sufficient art understanding and professional patience to improve the composition effectively, as the process can be much more sophisticated than a mere change of dynamic range or contrast. Although most people are clear about whether they find a photo aesthetically pleasing, it is a different matter when it comes to creating an aesthetically pleasing picture. This is the gap between an amateur and an artist.

Our system, targeting an average user, makes photo composition editing nearly automatic. In fact, the only involvement of a user is to input his/her judgment on whether a picture or a design is appealing or desired. It is a small step to turn the system fully automatic, but we feel that it is actually beneficial to inject some personal taste as allowed by the amount of interaction on the mobile device. Specifically, two strategies are exploited. First, to enhance a picture, a collection of Notan structures are created based on the original picture. A user can select a favorite Notan or the system chooses one closest to the Notan structure of an exemplar picture. This helps the user pinpoint easily a favored design. Second, in order to make the altered picture convey such a design, tonal transform is applied. This step is automatic by matching the tonal value distributions with those of the exemplar picture. The differences between our system and some existing tonal transform methods will be discussed at a more technical level in a short moment. In the current work, we assume a given exemplar picture. As an extension to the work, we can invoke a search engine using text and/or images to suggest exemplar pictures. A plethora of highly-aesthetic online photo collections exist.

Prior research most relevant to ours includes style transfer and tone reproduction. As a particular type of style, color transfer studies the problem of applying the color palette of a target image to a source image, essentially reshaping the color distribution of the source image to accord with the target at some cost. The histogram matching algorithm derives a tone-mapping function from the cumulative density functions of the source and the target. Various techniques have been developed (Reinhard et al., 2001; Abadpour and Kasaei, 2006; Xiao and Ma, 2006; Pitie and Kokaram, 2007; Pitie and Dahyot, 2007; Xiao and Ma, 2009; Papadakis et al., 2011; Pouli and Reinhard, 2011). These methods process the color distribution globally and do not consider spatial information. Pixels of the same color are subject to the same transformation regardless of whether they are in dark or light regions.

Artifacts can be easily brought in when the source histogram is very different from the target. Wen et al. (2008) conducted color transfer between corresponding regions chosen by the user in the source image and the target image. Tai et al. (2005) formed correspondence between segmented regions in the source image and the target before color transfer.

4.1 Method Overview

Let us first define a few terminologies. *Source image* is the image to be altered, while the *exemplar image* serves as a good example for the luminance distribution and possibly the Notan as well. The Notan we intended to obtain for the source image is *source Notan*, while the Notan of the exemplar image is called *exemplar Notan*. The tonal value or luminance will also be referred to as intensity in the sequel.

The outline of the Notan-guided tonal transform is as follows.

- Identify the source Notan and exemplar Notan.
- Perform Notan-guided region-wise histogram matching between the source image and the exemplar image.
- Postprocess the transformed image to remove possible artifacts at region boundaries.

The source and exemplar images are subject to segmentation by the algorithm in (Li, 2011). The average luminance of each segment is computed. To obtain the exemplar Notan, we first obtain a binarization threshold for the luminance using Otsu's method (Otsu, 1979) which assumes a bimodal distribution and calculates the optimum threshold such that the two classes separated by the threshold have minimal intra-class variance. This threshold decides whether any segmented region in the exemplar image is either dark (below threshold) or light (above). The source Notan can be obtained by different schemes. When the luminance threshold slides from small to large, more segmented regions in the source image are marked as dark. Because there are only finitely many segments, there are only finitely many possible Notans by thresholding at different values. With n segments, there are at most $n+1$ Notans. We can either let the algorithm choose a Notan automatically for the source image or let the user select his favorite Notan from the candidates. In the fully automatic setting, we have tested two schemes. We can either use Otsu's method to decide the threshold between dark and light (Automatic Scheme 1) or choose the source Notan with the proportion of dark area closest to that of the exemplar Notan (Automatic Scheme 2).

The algorithm for Notan-guided region-wise histogram matching will be presented later. The proposed approach differs from existing work in several ways. Instead of deriving a global tone-mapping function from two intensity distributions, a mapping function is obtained for each region in the source image. The mapping function is parameterized by the generalized logistic function. Although the regions are subject to different transforms, the parameters in the region-wise mapping

functions are optimized simultaneously to minimize an overall matching criterion between the source and the exemplar images. The approach does not require a correspondence established between regions in the two images. Furthermore, as elaborated in the next subsection, the spatial arrangement of dark and light, as embedded in Notan, plays an important role in determining the transform. In another word, the tonal transform is not just for matching two intensity histograms, but also an attempt to reach certain spatial patterns of dark and light.

Compared with traditional histogram-manipulation algorithms, one advantage of applying transformation functions in a region-wise fashion is to avoid noisy artifacts within regions. However, its performance depends on region segmentation to some extent. If the same object is mistakenly segmented into several regions, different transformation functions applied on its parts may cause artifacts. In real dodging and burning practice, a similar situation can be remedied by careful localized motion of the covering material during the darkroom exposure development or applying a subtle dodging/burning brush over a large area in digital photo editing software. We use fuzzy region maps to cope with this problem. Bilateral filter is employed to generate fuzzy maps for regions. Bilateral filter is well known for its edge-preserving property. It considers both spatial adjacency and intensity similarity. We use the fast implementation in (Paris and Durand, 2009).

4.2 Region-wise Histogram Matching

The intensity histogram records the proportion of pixels at a series of tonal scales, but not where the tonal values locate in the image. In this subsection, we describe the method for region-wise histogram matching between the source and exemplar images. A certain level of spatial coherence is obtained by the region-wise approach in comparison to the existing methods of global histogram matching. In the next subsection, we will revise the histogram-matching criterion to take into account Notan, thereby attempting directly to achieve a favored spatial design.

A *sub-histogram* is defined as the intensity histogram of a region. The image segmentation algorithm in (Li, 2011) is used to divide an image into semantically meaningful regions. The image is converted into the CIELab color space and the luminance channel is extracted to build the per region sub-histogram. The range of the intensity values is $[0,1]$. In the discussion below, the histogram is in fact a probability density function. We use the terminology "histogram" loosely here to be consistent with the often-used term "histogram matching."

Let $H_i(x)$, $x \in [0,1]$ be the sub-histogram for the ith region and n be the number of regions. Let $H(x)$ be the histogram for the entire image. We parameterize $H_i(x)$ by a single Gaussian or a two component Gaussian mixture. The main reason to use a Gaussian mixture instead of the usual histogram obtained by discretization is to ensure smoothness of H, a necessity for applying an optimization software package used in the region-wise histogram-matching algorithm. Although $H_i(x)$ should have finite support, we ignore the tail of the Gaussian distribution because the variance of

X is usually small in a single region obtained by similarity-based segmentation. The two-component option is provided to accommodate intensity distributions of clearly textured regions. Suppose the number of components for $H_i(x)$ is $K_i \in \{1,2\}$. We have

$$H(x) = \sum_{i=1}^{n} H_i(x) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} p_{ij} \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp - \frac{(x-u_{ij})^2}{2\sigma_{ij}^2} \ .$$

We use an unsupervised clustering algorithm (Bouman, 1997) to estimate K_i and the mean μ_{ij} and the variance σ_{ij} of each component. Similarly, the intensity distribution of the exemplar image $\tilde{H}$ is also approximated by GMM. Instead of summing over sub-histograms, a single GMM with $\tilde{K}$ components is used to represent the entire image. $\tilde{K}$ is also estimated by the algorithm in (Bouman, 1997).

To measure the distance between two distributions with support on $[0,1]$, we use the integrated difference between their cumulative density functions (Werman and Rosenfeld, 1985):

$$D(H,\tilde{H}) = \int_0^1 \left(\int_0^\lambda H(x)dx - \int_0^\lambda \tilde{H}(x)dx \right)^2 d\lambda \ . \tag{1}$$

We adopt a special case of the generalized logistic function as the tone-mapping function. The generalized logistic function is defined as

$$Y(x) = A + \frac{K-A}{(1+Qe^{-B(x-M)})^{1/v}} \ .$$

The general expression above provides a high degree of flexibility. We retain only two parameters b and m to allow changes in curvature and translation of the inflection point (Verhulst, 1838).

$$Y(x) = \frac{1}{1+e^{-b(x-m)}} \ . \tag{2}$$

The reason for choosing the above function is that it can accomplish different types of tonal adjustment by setting different parameters, allowing a unified expression for the transformation functions. Moreover, the logistic curve tends to preserve contrast. Figure 6 illustrates some tone-mapping curves generated by (2) with different values of b and m.

We constrain the parameter space of b and m such that $Y(x)$ in Equation (2) is monotonically increasing and the intensity range after transformation is not compressed too much. The first condition can be met provided $b > 0$. For the second condition, we set two thresholds t_0 and t_1 such that:

$$Y(0) = \frac{1}{1+e^{bm}} \leq t_0 \ , \ Y(1) = \frac{1}{1+e^{-b(1-m)}} \geq t_1 \ . \tag{3}$$

A right (left) translation of the inflection point, *i.e.* $m >> 0.5$ ($m << 0.5$), will darken (brighten) the region, causing a burning (dodging) effect.

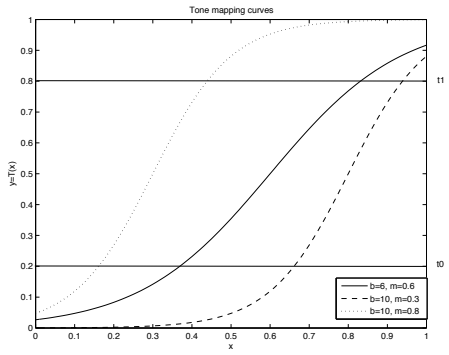


Fig. 6 Tone-mapping curves with various parameters.

Let the parameters of the transform $Y(x)$ for the ith region be b_i and m_i. For the overall image, the tonal transformation is then parameterized by $T = \{m_1, b_1, ..., m_n, b_n\}$. After we apply the transformation functions on individual regions, the intensity distribution of the modified image becomes

$$H(y;T) = \sum_{i=1}^{n} \frac{dX_i(y)}{dy} H_i(X_i(y);T) \ ,$$
$$\text{where } X_i(y) = Y_i^{-1}(y) \ . \tag{4}$$

We cast region-wise histogram matching as an optimization problem. The objective function $F(T)$ measures the distance between the intensity distributions of the transformed source image and the exemplar image. Suppose the source image contains n regions with average intensities μ_i, $i = 1, ..., n$, and the average intensities of the regions after tone mapping become μ_i', $i = 1, ..., n$. The optimization problem for the region-wise histogram matching is:

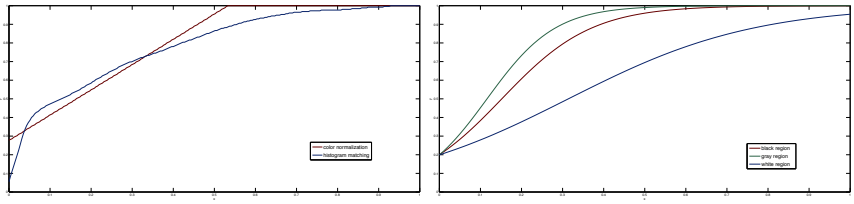
$$F(T) = \min_T D(H(y;T), \tilde{H}(y)) \ ,$$
$$\text{s.t. } (\mu_i - \mu_j)(\mu_i' - \mu_j') \geq 0 \ ,$$
$$\forall 1 \leq i \leq n, \ 1 \leq j \leq n \ . \tag{5}$$

Recall D is the distance defined in (1). The optimization is constrained so that the original order of region intensities is retained (the relative brightness of the regions will not be reversed). We use the package called CFSQP developed at the University of Maryland (Lawrence et al., 1994) to solve the optimization.

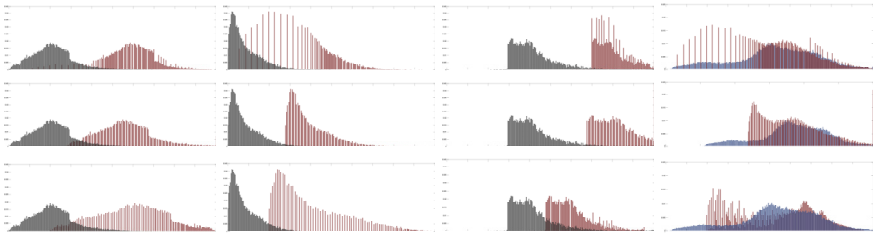
The major problem with the global tone-mapping function is the complete loss of the spatial information. The approach of transferring color between matched regions is intuitive but requires correspondence between regions, which is only meaningful for images very similar in content. For example, Figure 7 (a) shows a pair of images taken as the source image and the exemplar image. Their intensity

(a) Left to right: the source image, the exemplar image, the intensity histograms (gray for the source image and blue for the exemplar).

(b) First three images from left to right: the modified image by histogram matching, by color normalization, and by region-wise adjustment. Last image: the segmented regions.



(c) Tone-mapping curves. Left: histogram matching (blue) and color normalization (red). Right: Transformation functions for different regions (red curve for black region; green for gray region; and blue for white region).



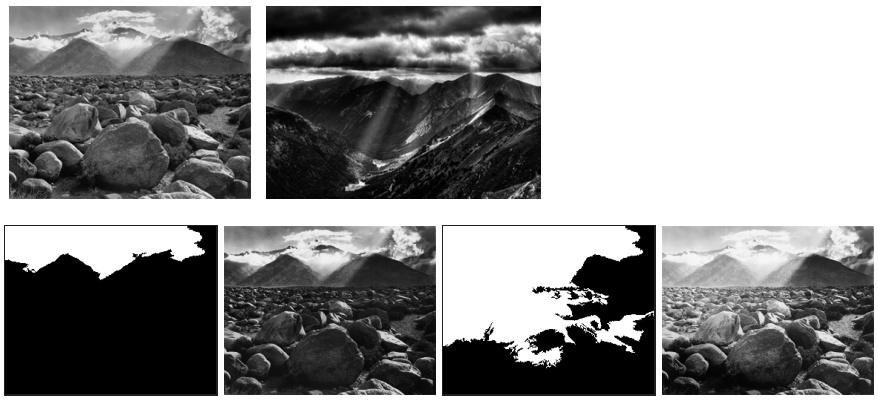
(d) Left to right: histograms for the segmented region shown in black, region in gray, region in white, and the entire image before and after matching. The histograms in gray are for the original image before matching; red for the modified image; and blue for the exemplar image. Row 1: results for global histogram matching. Row 2: color normalization. Row 3: region-wise histogram matching.

Fig. 7 Comparison between global and region-wise tone mapping.

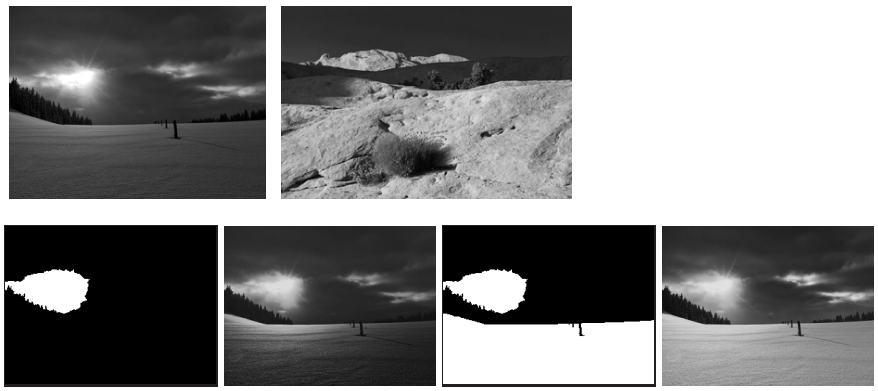
distributions are very different from each other. Figure 7 compares two global approaches, global histogram matching and color normalization (Reinhard et al., 2001), with the proposed region-wise approach. When the source image is low-keyed and the exemplar is high-keyed, a global mapping function tends to remove too many details in the dark areas and overexpose the light areas. With region-wise adjustments, however, each transformation function contributes to the overall

histogram matching while its transformed range is not severely constrained by other regions. For example, the tone-mapping curve of a dark region can have a higher growth rate than light regions (Figure 7 (b)).

4.3 Notan-guided Matching



(a) Example 1



(b) Example 2

Fig. 8 Modification by different Notan patterns for two example images. Top row in each example: the source image (left) and the exemplar image (right). Bottom row in each example: two source Notan patterns and the modified images (on the right of the corresponding Notan).

The objective function for region-wise histogram matching provided in (5) ignores the spatial arrangement of dark and light. We thus introduce a new objective function dependent on the Notan. Consequently, the revised image tends to yield a

Notan appearance close to the specified source Notan. Let H_{dark} and H_{light} be the intensity distributions for the dark and light areas of the source image respectively, where the dark and light areas are decided by the source Notan. Similarly, let $\tilde{H}_{dark}$ and $\tilde{H}_{light}$ be the intensity distributions for the dark and light areas of the exemplar image respectively. The new optimization problem is

$$\begin{aligned} &F_n(T) = \min_T \left(D(H_{dark}(y;T), \tilde{H}_{dark}(y)) + D(H_{light}(y;T), \tilde{H}_{light}(y)) \right) , \\ &\text{s.t. } (\mu_i - \mu_j)(\mu'_i - \mu'_j) \geq 0 \text{ , for any } 1 \leq i \leq n, 1 \leq j \leq n \ . \end{aligned} \tag{6}$$

Comparing optimization (6) with (5), we see that the new objective function is the sum of two separate distribution distances, one involving only the dark areas in the two images and the other only the light areas. However, because of the constraints to retain the intensity ordering of the regions, the optimization problem cannot be decoupled into one for the dark areas and one for the light areas.

Figure 8 illustrates the impact of the chosen source Notan on the modified image under the same exemplar image and exemplar Notan. Two different Notans are shown for each source image in Figure 8. The Notans are accompanied by their corresponding modified images. By imposing different Notans, the modified images generated by optimization (6) present quite different dark-light compositions. On a mobile device, we can potentially show users a few options of source Notans and let them pick what they find most appealing.

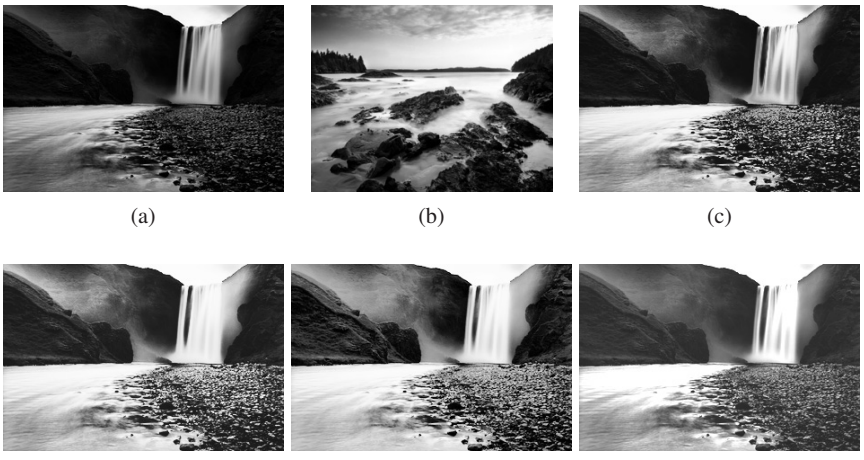


Fig. 9 Contrast comparison. (a) Source image. (b) Exemplar image. (c) Notan-guided region-wise histogram matching (optimization (6)). (d) Modified image generated by region-wise histogram matching (optimization (5)). (e) Global histogram matching. (f) Color normalization.

A side benefit of Notan-guided matching is to better keep contrast. When the proportions of dark and light areas differ substantially between the source image and the exemplar, matching without Notan often results in over reduced contrast

(an overall whitened or blackened look). The effect of large disparity in dark-light proportion is mitigated by the Notan, which enforces matching the dark areas and light areas separately. For example, the exemplar image in Figure 9 (b) has a proportionally small dark area (rocks) which contrasts with a large light area, while the source image has a relatively large dark area. In this example, we used the threshold given by Otsu's method to generate the source Notan. The modified image obtained by region-wise histogram matching without Notan (optimization (5)), shown in Figure 9 (d), seems to be overexposed with much reduced contrast. This issue is more serious with modified images obtained by global histogram matching in (e) and color normalization in (f). The result of Notan guided matching in (c) keeps the best contrast.

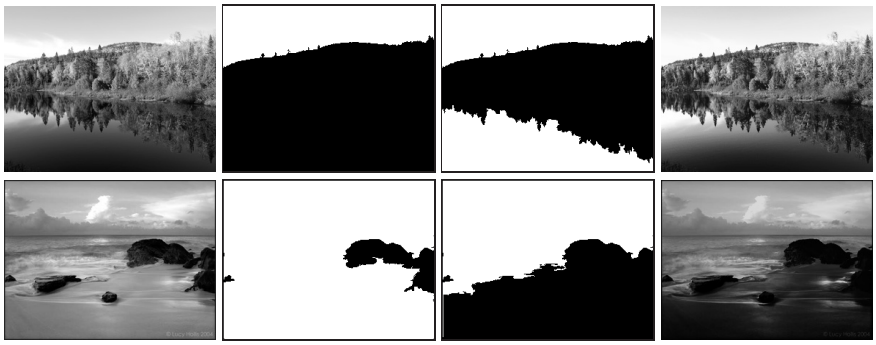


Fig. 10 Modifying images by choosing a favored Notan without using an exemplar image. Left to right: original image (serving as both source and exemplar), exemplar Notan, source Notan (manually selected), modified image.

Considering the stringent interface on a mobile device, we explore a scenario when an exemplar image is not available. Interestingly, we may enhance the composition of an image by just specifying a desired Notan. In Figure 10, the source image itself serves as the exemplar image. The exemplar Notan is obtained using the threshold of Otsu's method. The source Notan is manually chosen, supposedly more appealing than the automatically picked Notan. The results demonstrate that the modified images indeed seem better composed. This self-boosting method may seem surprising at first glance. To better understand this, note that the exemplar Notan will have a more contrasted dark and light because of the way the threshold is chosen. It should also be closer to what the Notan of the source image without modification appears to be. However, the spatial arrangement of the dark and light is not as pleasant as what is specified by the manually chosen Notan. What is essentially done by our algorithm is to make the manually set dark and light areas appear better divided and hence more obvious to the eye. This is achieved by histogram matching with the exemplar dark and light areas, which by set up are well contrasted.

This experiment of self-boosting composition enhancement hints that choosing a source Notan is more important than an exemplar image. Here, we used the source image as the exemplar image. We may also generate artificial intensity distributions for dark and light and plug them into optimization (6), thereby bypassing completely exemplar image and exemplar Notan. This can be interesting to investigate.

As explained in Section 4.1, we allow a fully automatic setting where the Notan of the source image is chosen among a set of possible Notans generated by different thresholds between dark and light. This is motivated by the need of mobile devices where minimal user interaction is desired. In this setting, we exploit the exemplar image not only for histogram matching but also for selecting a source Notan. The underlying assumption is that the exemplar image is well composed in the two tonal scales of dark and light. The source Notan closest to the exemplar Notan in terms of dark and light proportions is used. This is no doubt a rather simple similarity defined for two Notans. In future work, we can employ a more sophisticated similarity measure between two Notans. For the experimental results in Section 5, this automatic setting is employed.

5 Experimental Results in the Automatic Setting

In Figure 11, we show results by our Notan-guided region-wise histogram-matching algorithm and compare with global histogram matching and color normalization. The source Notan is automatically chosen (see description in the previous section). Our new method tends to generate smoother histograms and better-controlled dynamic range. The other methods more often yield burned out areas.

Figure 12 presents more examples. In the experiments, the number of segments is set to 3 for simple scenes and 6 for complex scenes. Note that more segments require more parameters to be estimated and therefore more computation. We observe that the global histogram matching often yields the artifact of abrupt changes in intensity. The color normalization method uses a linear mapping function whose growth rate is determined by the variances of the source and the exemplar distributions. A high (or low) growth rate can burn out (or flatten) the final image. Our new method controls better the extreme cases by regulating the transformation parameters.

6 Summary

This chapter presented two computerized approaches to provide photographers with on-site composition feedback and enhancement suggestions. The first approach is based on spatial design categorization that places a photo into one or more categories including horizontal, vertical, diagonal, textured, and centered. Such categorization enables retrieval of exemplar photos with similar composition. The second approach utilizes the concept of Notan in visual art for tonal adjustment. A user can improve

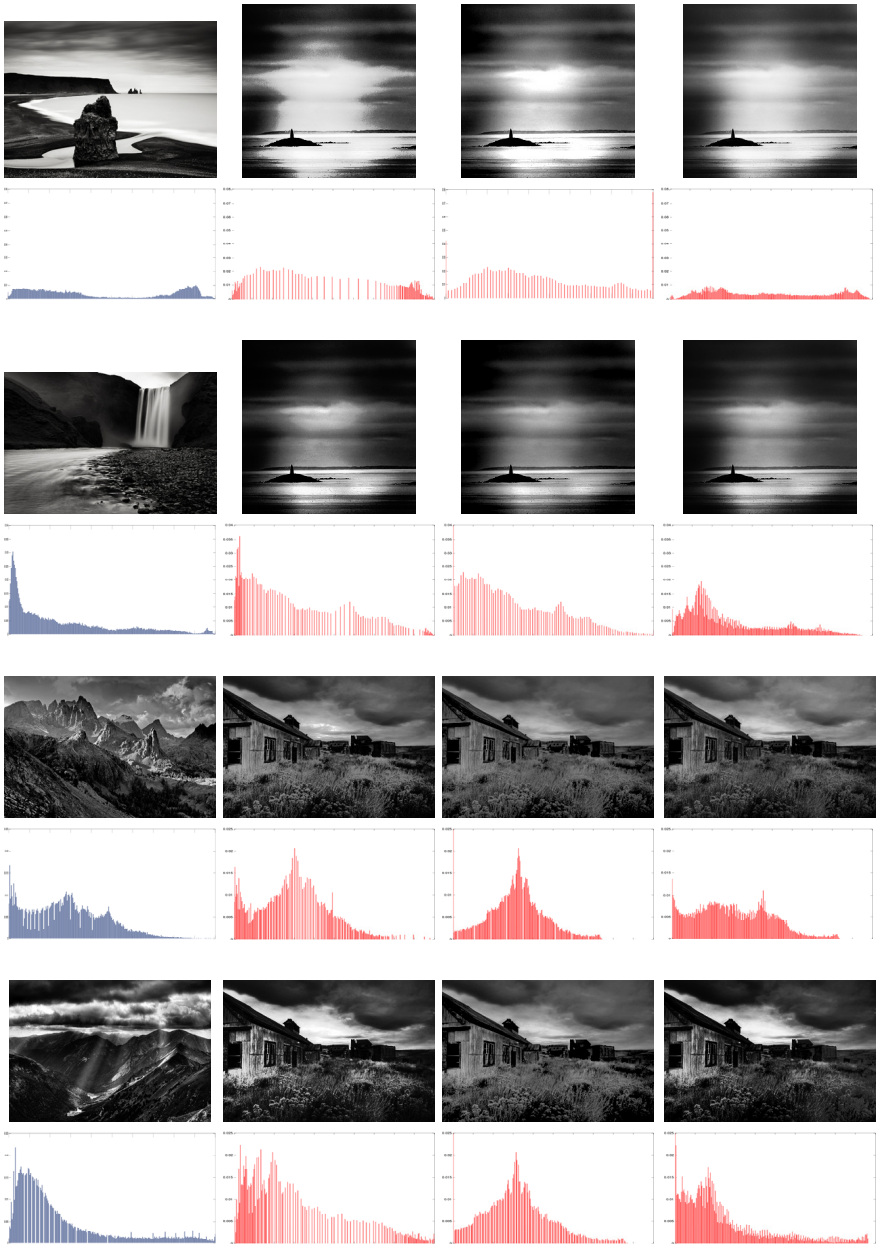


Fig. 11 Comparison of algorithms by modified images and their histograms. (a) Exemplar image. (b) Modified source image by global histogram matching. (c) Color normalization. (d) Notan-guided region-wise histogram matching.

Fig. 12 Additional experimental results. (a) The source image. (b) The exemplar. (c) Global histogram matching. (d) Color normalization. (e) Notan-guided region-wise histogram matching.

the aesthetics of a given photo through transforming the dark-light configuration toward that of a target photo. We view this work as just the beginning of a new direction under which principles of composition in visual art are used to guide the development of computational photography techniques.

Acknowledgements This material is based upon work supported by the National Science Foundation under Grant Nos. 0347148 and 0936948.

References

Abadpour, A. and Kasaei, S. (2006). Color transfer in correlated color space. In *Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and Its Applications*, pp. 305–309, New York, USA.

Adams, A. (1995). *The Print*. Little, Brown.

Bhattacharya, S., Sukthankar, R., and Shah, M. (2010). A coherent framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proceedings of ACM Multimedia*

Conference, pp. 271-280.
Bouman, C. A. (1997). Cluster: An unsupervised algorithm for modeling Gaussian mixtures. http://www.ece.purdue.edu/~bouman .
Cohen-Or, D., Sorkine, O., Gal, R., Leyvand, T. and Xu, Y. (2006). Color harmonization. *ACM Transactions on Graphics*, *25*(3):624-630.
Datta, R., Joshi, D., Li, J. and Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. In *Proceedings of European Conference on Computer Vision*, pp. 288-301.
Datta, R., Joshi, D., Li, J. and Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2), 5, 1-60.
Feininger, A. (1973). *Principles of Composition in Photography*. Thames and Hudson Ltd.
Folts, J. A., Lovell, R. P., and Zwahlen, F. C. (2005). *Handbook of Photography*. Thompson Delmar Learning.
Fogarty, J., Forlizzi, J. and Hudson, S. E. (2001). Aesthetic information collages: Generating decorative displays that contain information. In *Proceedings of ACM Symposium on User Interface Software and Technology*, pp. 141-150.
Krages, B. P. (2005). *Photography: The Art of Composition*. Allworth Press.
Lamb, J., and Stevens, R. (2010). Eye of the photographer. *The Social Studies Texan*, *26*(1):59-63.
Lawrence, C., Zhou, J. L., and Tits, A. L. (1994). User's guide for CFSQP version 2.0: A C code for solving (large scale) constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality constraints. *Technical Report*, *drum.lib.umd.edu/handle/1903/5496*.
Li, J. (2011). Agglomerative connectivity constrained clustering for image segmentation. *Statistical Analysis and Data Mining*, *4*(1):84-99.
Li, J., Wang, J. Z., and Wiederhold, G. (2000). IRM: Integrated region matching for image retrieval. In *Proceedings of ACM Multimedia Conference*, pp. 147-156.
Liu, L., Chen, R., Wolf L. and Cohen-Or, D. (2010). Optimizing photo composition. *Computer Graphic Forum*, *29*(2):469-478.
Meer, P. and Georgescu, B. (2001). Edge detection with embedded confidence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(12):1351-1365.
Obrador, P., Anguera, X., Oliveira, R. and Oliver, N. (2009). The role of tags and image aesthetics in social image search. In *Proceedings of the ACM SIGMM Workshop on Social Media*, 65-72.
Obrador, P., Oliveira, R. and Oliver, N. (2010). Supporting personal photo storytelling for social albums. In *Proceedings of ACM Multimedia Conference*, 561-570.
Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.
Papadakis, N., Provenzi, E., and Caselles, V. (2011). A variational model for histogram transfer of color images. *IEEE Transactions on Image Processing*, 20:1682–1695.
Paris, S. and Durand, F. (2009). A fast approximation of the bilateral filter using a signal processing approach. *International Journal of Computer Vision*, 81(1):24–52.
Payne, E. (2005). *Composition of Outdoor Painting*. 7th Ed. Deru's Fine Arts.
Pitie, A. C. K. F. and Dahyot, R. (2007). Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1-2):123–137.
Pitie, F. and Kokaram, A. (2007). The linear Monge-Kantorovitch colour mapping for example-based colour transfer. In *Proceedings of the IEEE European Conference on Visual Media Production*, pages 1–9.
Pouli, T. and Reinhard, E. (2011). Progressive color transfer for images of arbitrary dynamic range. *Computers and Graphics*, 35(1):67–80.
Raybould, B. J. (2014). Notan painting lessons. Virtual Art Academy. http://www.virtualartacademy.com/notan.html .
Reinhard, E., Ashikhmin, M., Gooch, B., and Shirley, P. (2001). Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34-41.
Russ, J. C. (2006). *The Image Processing Handbook*. CRC Press.
Sorrel, C. (2010). Nadia camera offers opinion of your terrible photos. *WIRED*, online, July 26.

Speed, H. (1972). *The Practice and Science of Drawing*. 3rd Ed. Dover Publications.

Sternberg, R. J. (2008). *Cognitive Psychology*. Wadsworth Publishing.

Tai, Y.-W., Jia, J., and Tang, C.-K. (2005). Local color transfer via probabilistic segmentation by expectation-maximization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 747–754.

Verhulst, P. F. (1838). A note on population growth. *Correspondence Mathematiques et Physiques*, 10:113-121.

Wang, J. Z., Li, J., and Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(9):947-963.

Warren, B. (2002). *Photography: The Concise Guide.* Delmar Cengage Learning.

Wen, C.-L., Hsieh, C.-H., Chen, B.-Y., and Ouhyoung, M. (2008). Example-based multiple local color transfer by strokes. *Computer Graphics Forum*, 27:1765–1772.

Werman, S. P. M. and Rosenfeld, A. (1985). A distance metric for multi-dimensional histograms. *Computer, Vision, Graphics, and Image Processing*, 32:328–336.

Xiao, X. and Ma, L. (2006). Color transfer in correlated color space. In *Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and Its Applications*, pp. 305–309, New York, NY, USA..

Xiao, X. and Ma, L. (2009). Gradient-preserving color transfer. In *Computer Graphics Forum*, vol. 28, pp. 1879–1886.

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