國立臺灣大學電機資訊學院資訊工程學系

碩士論文

Department of Computer Science and Information Engineering

College of Electrical Engineering & Computer Science

National Taiwan University

master thesis

以美學為基礎的照片構圖量化分析

Esthetics-based Quantitative Analysis of Photo Composition

高弘政

Hong-Cheng Kao

指導教授：歐陽明 博士

Advisor: Ming Ouhyoung, Ph.D.

中華民國 97 年 6 月

June, 2008

國立臺灣大學
資訊工程學系

碩士論文

以美學為基礎的照片構圖量化分析

高弘政 撰

97
6

致謝

實在無法置信，對於寫作駑鈍的我，也完成了這樣的一本碩士論文，而且是用我破破的英文撰寫出來，它能讀嗎？真是不可思議。

首先感謝我的父母，感謝他們給我的一切以及教導，提供我良好以及自由的學習環境。感謝我的堂兄汶賢，給我藝術方面的啟發，讓我能接觸攝影這塊天地。感謝我的家人。

謝謝我的指導教授歐陽明老師給我在研究上的指導，並且在我的英文寫作不厭其煩地教導。也謝謝其他多媒體實驗室的老師們，特別謝謝徐宏明老師。謝謝多媒體實驗室的學長姐，特別謝謝馬萬鈞學長讓我的論文更加完善。

感謝恒溫、宗佑以及棠為，跟你們一起做研究、有你們的加油打氣及鼓勵，讓我研究的過程更加順利。同時也感謝多媒體實驗室其他的同學們以及實驗室助理們。

最後感謝最親愛的雯惠，有妳的陪伴，陪我度過每個酸甜苦辣的日子，讓我的每一天變得更加美好。

感謝這美好的早晨、美好的一切。

弘政　謹誌于 台灣大學資訊系館

摘要

數位相機的普及率越來越高，因此越來越多人拍照，使得數位照片成指數性成長。面對龐大數量的照片，照片的整理與挑選成了一個很大的問題。

照片構圖的意思是相機位置的擺設以及視野的選擇，決定什麼物體該放進來或該排除在外，一般來說，照片構圖幾乎決定了一張照片的好壞，所以我們針對照片構圖的部分做研究。照片構圖這個問題往往是主觀的，也跟人類的視覺感受有關，雖然是主觀的，但幸運的是，根據攝影學家多年來的經驗，可以萃取出一些比較通用性的原則，這些規則在各種談攝影的書中被談到。我們從書中挑選出幾個定義較明確、且較適用於電腦量化分析的規則，將他們實作成可以自動化執行的程序，用這些規則挑選出比較不好的照片，以及對照片做一些建議與評分。

一張不是很好的照片如果經過很好的裁切，可以讓一張照片起死回生，現在數位相機的能力也越來越強，照片的解析度越來越高，在沒有輸出太大尺寸照片需求的原則下，很有本錢可以對照片做一些裁切，我們以構圖分析的結果來做裁切的原則，有機會能讓照片變的更好。

我們初步的結果是基於 132 張照片，這些照片是針對每個規則（水平線、照片平衡、主體位置、線條及形狀、避免融合）挑選出比較簡化的情況，每個規則的準確率分別在71%、96.8%、73.1%、78.4%以及71.4%。

關鍵字： 照片構圖、美學規則、照片平衡、三等分原則、感興趣區域、照片裁切。

Abstract

Digital camera is very popular and the number of digital photos grows exponentially. Faced to huge number of photos, the collection and selection of photos becomes a big problem.

Photo composition means the placement of camera and the selection of the field of view. It determines whether objects should be placed inside the photo or be excluded outside. Most people agree that photo composition almost determines whether a photo is good or not. Therefore, our research is focus on photo composition. This problem is subjective and relates to human visual perception. Although this problem is subjective, it is fortunate that from the experience accumulated by photographers in recent years, certain common rules were extracted. We select the rules that have clear definition and suitable for automatic quantitative analysis to pick out bad photos or to make recommendations and scores.

It is said that a not very good photo after doing a good cropping will let a photo turn from death into life. The performance of digital camera nowadays is more and more powerful and image resolution is higher and higher, and so images can be cropped without serious loss of resolution. Our results in automatic cropping support this observation.

Our preliminary results for 132 photos are based on five rules (the horizon, photo balance, location of main object, revealing line patterns and shapes, and avoiding mergers), and the precision of each rule is 71%, 96.8%, 73.1%, 78.4%, and 71.4%, respectively.

Keywords: Photo composition, esthetics rules, photo balance, the rule of thirds, region of interest (ROI), photo cropping.

Contents

List of Figures

Chapter 1

Introduction

Photography is the process of recording pictures by means of capture light on a light-sensitive medium, such as electronic sensor. The word "photography" comes from the French *photographie* which is based on the Greek words (*phos*) "light" + (*graphis*) "stylus", "paintbrush" or (*graphê*) "representation by means of lines" or "drawing", together meaning "drawing with light." (Photography definitions at Wikipedia. `http://en.wikipedia.org/wiki/Photography`)

There are several important issues when taking photos: lighting, exposure, photo composition, and photo quality. Lighting and exposure represents the environment's lighting and how the camera catches the light (combination of shutter speed, aperture of the lens, ISO speed, etc). Photo composition means the selection and placement of elements into a photo, and determining the view direction and how wide the view. Generally speaking, the exposure issue can be considered as part of photo composition. Photo quality is affected by the quality of camera equipments (lens quality, sensor quality, etc.), and user's operation issues (out of focus, hand shake, etc.).

In this work, we will focus on the issues of photo composition. Why we want to focus on this topic? Because we are interested in photography and we want to find some interesting and useful problems in photography. [SEL00] also shows that composition is one of the most important factor to photo appearance. Digital camera becomes more and more common. When we finished a journey, a lot of photos were taken especially when we saw the most beautiful scenes. In this case, many almost identical photos were produced. When we want to select the most beautiful one from these similar photos, it is often hard to determine which one should be chosen. If we have a tool that can do some determination based on photo composition, it tells us whether a photo is relatively worse and should be left out or has some good features thus should be reserved, that might avoid many annoying determination. The tool also can give some recommendations and do correction to photos.

This problem is subjective and related to artists' points of view and related to the human visual perception. Although this problem is subjective, it is fortunate that from the experience accumulated by photographers in recent years, some common rules were extracted. These rules are commonly mentioned in many photography books. We surveyed several photography books [Tsa98b, Tsa98a, Yam06, Pet03, BC99] and generalized several rules. The rules have clear definitions and are suitable for computing and quantitative analysis. What we want to do is using some basic rules to pick out bad photos or to make recommendations and scores. It seems that there are few formal researches focusing on esthetics-based quantitative analysis of photo composition.

Camera users may be able to learn some methods to appreciate photos and some tricks to take good photos. Although some professional photographers doing artistic creations do not based on general rules, since they want to express some special meanings and break the limitation of rules, these types of photos are not considered in our work. We only focus the photos taken by armature photographers; their photos often can become better when guided by basic rules.

Many people say that a not very good photo after doing a good cropping will let a photo turn from death into life. We are also motivated by book [Yam06] that we can get multiple composition results by zooming or cropping in a single view (Figure 1.1). The performance of digital camera nowadays is more and more powerful and image resolution is higher and higher. If you don't need to print out your photos with too large size, images can be cropped with slight affects. If a photo after proper cropping can get better composition, this kind of tool may be very useful.

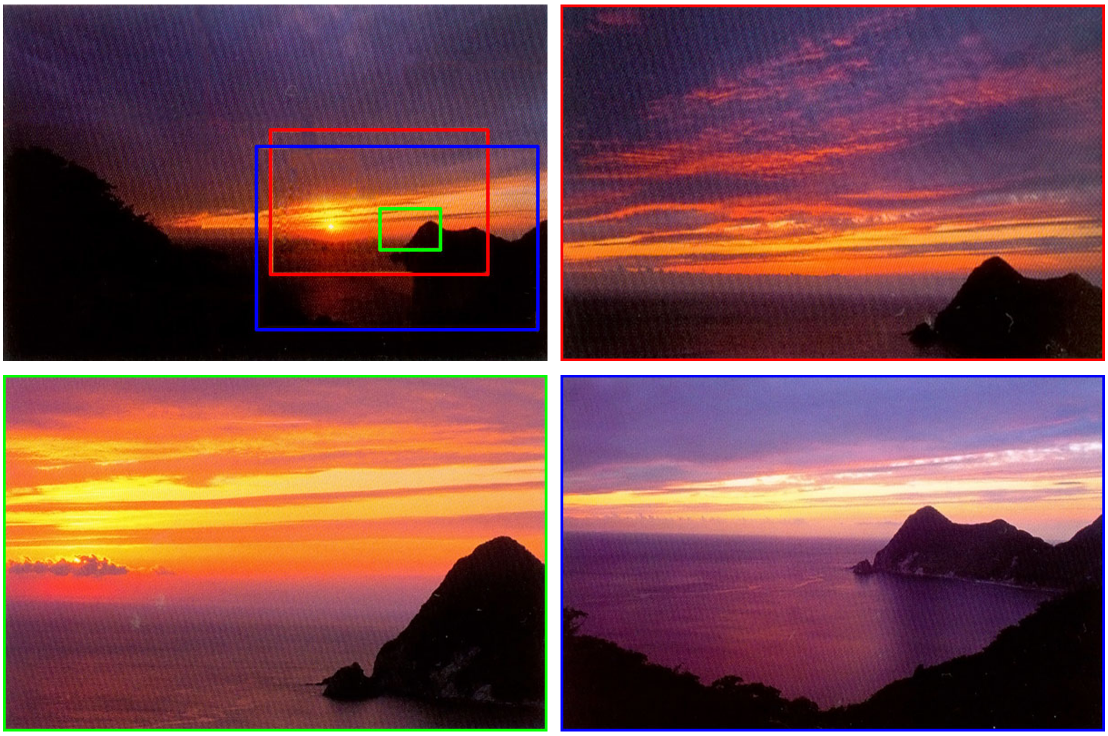


Figure 1.1: Multiple composition results in a single view. (Motivated by book [Yam06])

Chapter 2

Related work

2.1 Recomposition, information preserving, and automatic cropping

[BE07] is the most related work with our work. Their goal is improving the photos taken by amateur photographers according to repositioning the main object, or by blurring the background to distinguish foreground and background, or by blurring the background objects that will merge with the main object. They focus on the ROI that has strong edge and if the background is blurred.

[SAD*06] shows a semi-automatic, interactive photo cropping approach, using rough image segmentation and eye tracking information to get ROI's location and size, and then automatic photo cropping. Cropping principles are: (1) keep the complete set of ROI and all ROIs, (2) background cropping on segment boundary, (3) bigger ROI is better, and (4) the ROI location shall follow the rule of thirds. Every principle has its objective function and weighting, and they need to minimize these objective functions to get the best cropping.

[CXF*03] and [SLBJ03] want to show most information in limited display resolution. They don't promise all the ROIs in an image will be displayed completely. They not only doing resize but also doing crop with resizing, considering many attention models (saliency, face, and text) to get attention regions and attention values, finally obtain the optimized cropping result. [SLBJ03] also defines the recognizability for objects. They prefer to maximize the recognizability to get the cropping result.

[AS07] presents a technique called seam carving that supports content-aware image resizing functionality for both image reduction and expansion. The method that changes the layout of a photo to change image resolution is different from the traditional resizing and cropping.

2.2 Feature extraction tools for photo composition

We discuss several important image feature detection tools which will be used in this work. Experienced readers in image processing can skip this section.

2.2.1 Canny edge detection

In the analysis phase, we often need image's edge information. We can use Canny edge detection algorithm [Can86] to extract edges. The Canny edge detection algorithm is known to many as one of the optimal edge detectors. It has only one response to a single edge. The distance between the edge pixels as found by the detector and the actual edge is set to be at a minimum. Our implementation also followed "Canny Edge Detection Tutorial" (`http://www.pages.drexel.edu/~weg22/can_tut.html`) authored by Bill Green. Next, we will introduce the details of Canny edge detection algorithm.

Smoothing the image

The first step is smoothing the image in order to eliminate image noises. This step can simply apply Gaussian filter or any other low pass filter. Gaussian filter can be performed using standard convolution methods. The detector's sensibility to noise is lower when the width of the Gaussian mask (σ) is larger. We need to extract larger features and skip smaller, low contrast noise, so the width of the Gaussian mask we choose is relatively larger. We conducted a simple experiment in order to determine the value of σ. When applying Gaussian filter with different σ to a photo that contains the horizon, according to the situation of edge extraction and the accuracy of horizontal line extraction, we found that a too small σ generates many small lines and so line detection fails. Too large σ lets many lines disappeared and the existing lines become discontinued. Figure 2.1 shows the experiment results. The ideal σ we choose is $image\ size/100$.

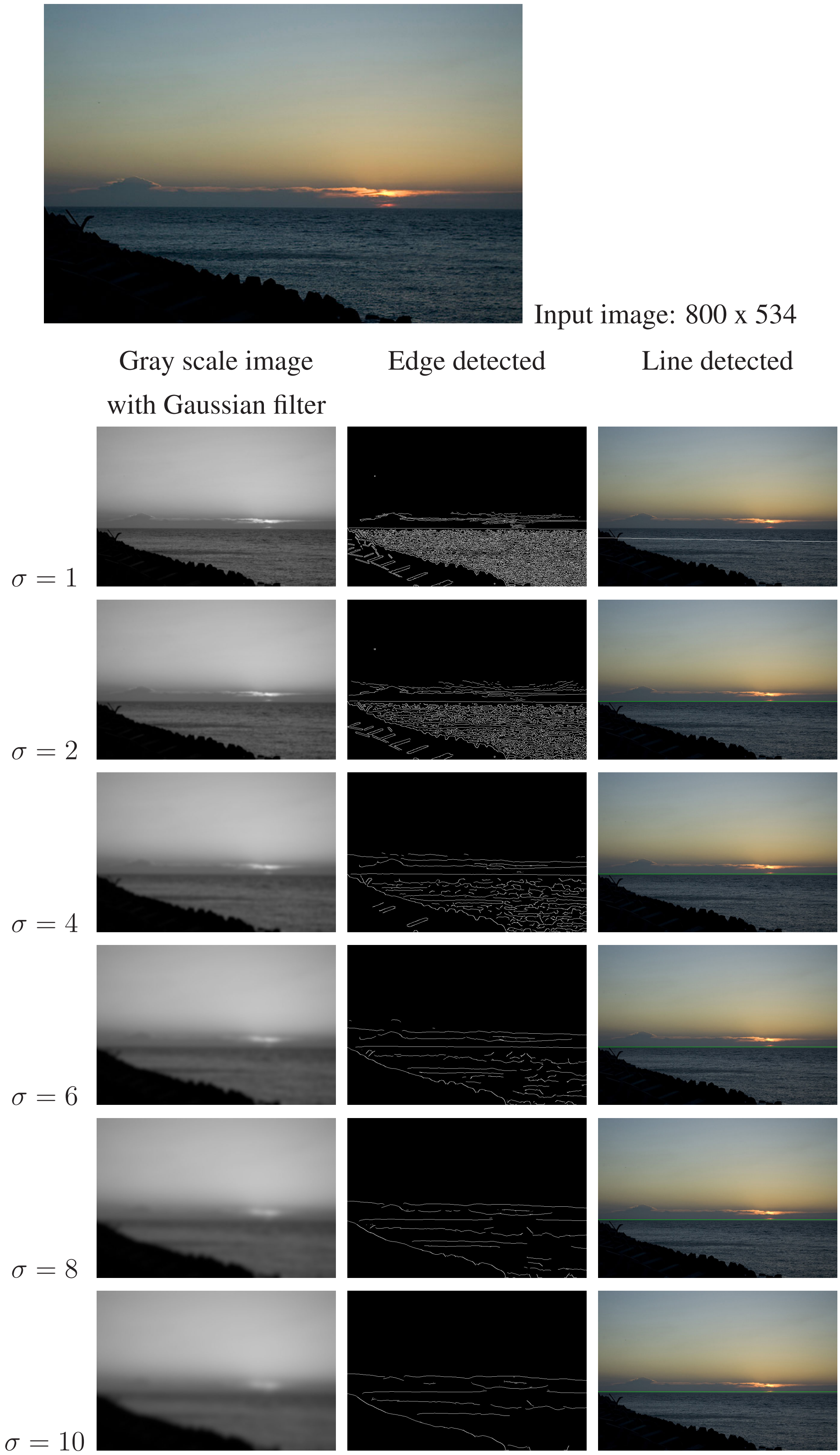


Figure 2.1: The experiment to get the sigma value for smoothing the image.

Getting edge strength

The next step is to find the edge strength by taking the gradient of the image. The Sobel operator performs a 2D spatial gradient measurement on an image. Then, the approximate absolute gradient magnitude (edge strength) at each point can be found. The Sobel operator uses a pair of 3x3 convolution masks, one estimating the gradient in the x-direction (columns) and another estimating the gradient in the y-direction (rows). They are shown below:

$$M_x = \begin{pmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{pmatrix}, M_y = \begin{pmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}.$$

The convoluted images are G_x and G_y. The magnitude, or edge strength, of the gradient is then approximated using the formula:

$$|G| = |Gx| + |Gy|.$$

The result of getting edge strength is shown in Figure 2.2.

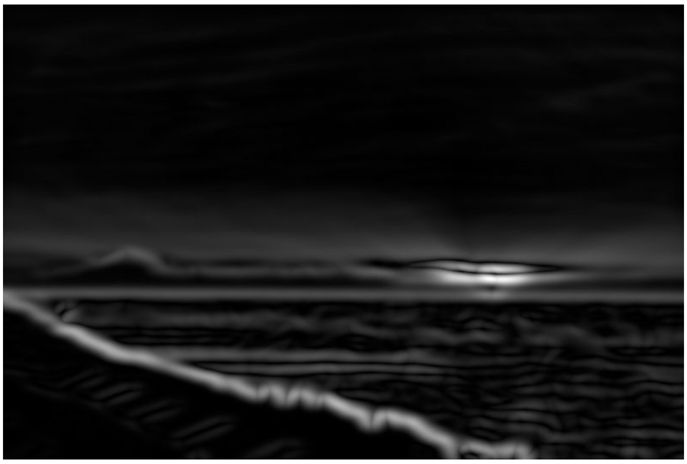


Figure 2.2: The result of getting edge strength.

Getting edge direction

Finding the edge direction is trivial once the gradient in the x and y directions are known. However, you will generate an error whenever G_x is equal to zero. We set the degree to

90^0 when G_x is equal to zero. The formula for finding the edge direction is:

$$\theta = \begin{cases} \tan^{-1}(\frac{G_y}{G_x}), & \text{if } G_x \neq 0 \\ 90^0, & \text{otherwise.} \end{cases}$$

To the convenience of calculation at discrete images, we can quantize the four directions to the degrees of 0^0, 45^0, 90^0, and 135^0. In our implementation, we set $[0^0, 22.5^0]$ and $(157.5^0, 180^0]$ to 0^0, $(22.5^0, 67.5^0]$ to 45^0, $(67.5^0, 112.5^0]$ to 90^0, and $(112.5^0, 157.5^0]$ to 135^0.

Nonmaximum suppression

Nonmaximum suppression is used to trace along the edge in the edge direction and suppress the pixel value that is not the maximum to zero. This will give a thin line in the output image. The result of nonmaximum suppression is shown in Figure 2.3.

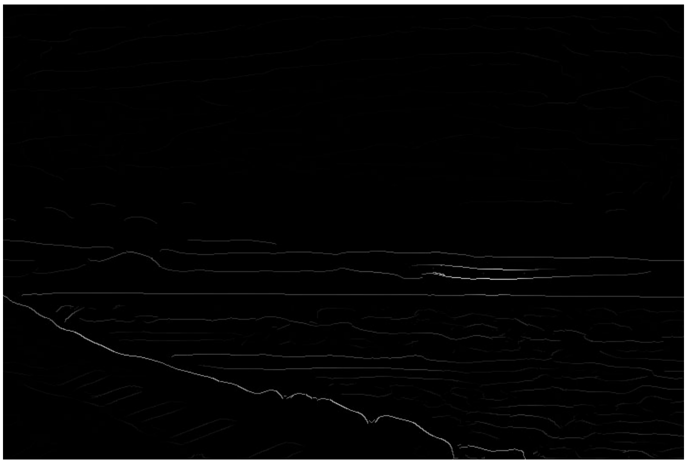
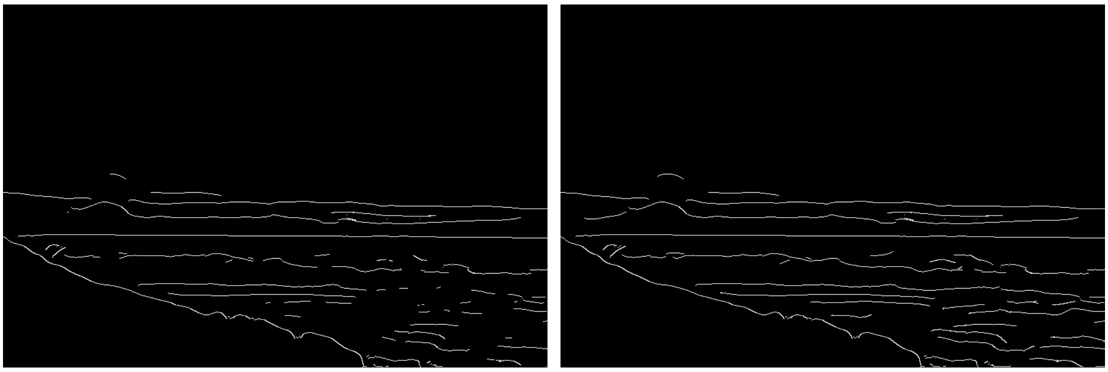


Figure 2.3: The result of nonmaximum suppression.

Hysteresis uses two thresholds

If we only use a single threshold T is applied to an image, and there is an edge has an average strength equal to T. The image may has much noise; the edge is not continuous and may become a dashed line. To avoid this, hysteresis uses two thresholds, a high one and a low one. Any pixel's value in the image greater than T_h is marked as an edge pixel immediately. Then, any pixels that are connected to this edge pixel and that have a value greater than T_l are also selected as edge pixels.

In this part, we have to determine the two parameters' value T_h and T_l. In order to let parameters to suit for any kinds of images, we let the parameters have relation with average intensity of whole image. According to some simple experiments, we set T_h to half of average intensity of the image, and set T_l to half of T_h. Figure 2.4 shows the result of hysteresis uses two thresholds.



(a) (b)

Figure 2.4: The result of hysteresis uses two thresholds. (a) Pixel value in the image greater than T_h, and (b) is the final edge detected image.

2.2.2 Hough transform for line detection

We need a tool to get straight lines information in images. We choose Hough transform [DH72], which uses location of points in an image to find the parameters of particular shapes (lines or circles etc.). Every point via one-to-many mapping (image space to parameter space) gets every possible value of parameter. Accumulating every parameter value mapped form all points in an image. Finally, it gets the most effective shape parameters in parameter space.

Straight lines detection and Hough transform

For every point (x, y) in 2D image, lines through it can represent to the equation $f(x, y) = y - ax - b = 0$. a and b are the line's slope and intercept respectively. The above equation can be seen as mutual constraint mapping relationship, mapping from image space (x, y) to parameter space (a, b) or parameter space (a, b) to image space (x, y). In other words, a point (x, y) in image space can be defined as multiple points on a line in parameter space;

the same situation, a point (a, b) in parameter space can be defined as multiple points on a line in image space.

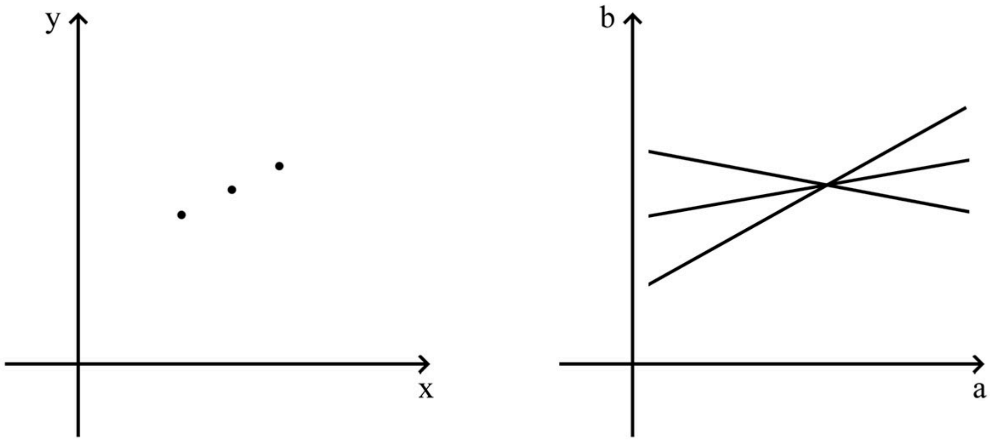


Figure 2.5: Image space (x, y) to parameter space (a, b).

Using accumulator

Because Hough transform maps every image points (x, y) to multiple parameter points (a, b), we can use an accumulator to count the number of times every parameter point (a, b) appears. The highest frequency appearance parameter point (a, b) is the most representative line in image space.

Hough transform algorithm

1. Find every feature point in image space; they are often edge points or skeleton points.

2. For every feature point (x, y) in image space

   (a) For every parameter a, calculate the parameter (a, b) of the line contains (x, y).

   (b) Increment the (a, b) in the accumulator.

3. Find all local maximums in the accumulator.

4. Map every maximum point back to image space will represent a line.

Using polar system

At implementation's point of view, using $f(x, y) = y - ax - b = 0$ does not work. When an vertical line in image space, its slope $a = \infty$ can't count into limited accumulator. Hough transform uses polar coordinate (ρ, σ) to replace (a, b). The mapping equation is $x \cos \theta + y \sin \theta = \rho$. Every line in image space is mapped to a point in parameter space, and lines through a point in image space are mapped to a curve in parameter space.

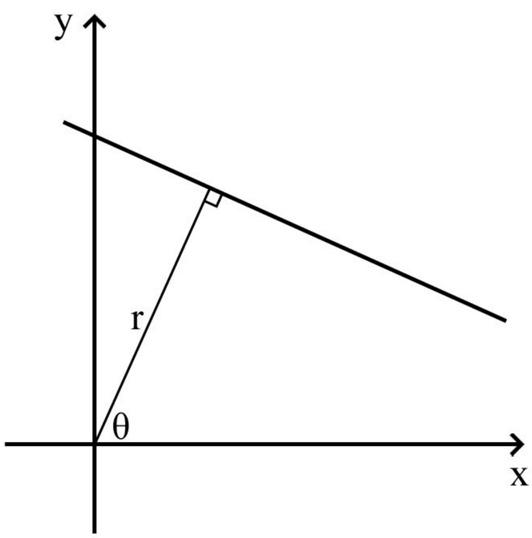


Figure 2.6: Using polar system to represent a line.

2.2.3 Region of interest (ROI)

In this work, we need to know the location of main object (ROI) in an image. The main object is often noticeable, and locates at where it has strong visual stimulation. We use Itti's ROI algorithm [IKN98] to get the salience map (a priority map of visual attention). The above includes the human visual model from psychology. It is a bottom-up approach. It uses intensity contrast, color contrast, and local impulse in every orientation combining every "center-surround" differences in different scales to get the final salience map. Following goes through Itti's ROI algorithm in detail.

Extraction of early visual features

With r, g, and b being the red, green, and blue channels of the input image, an intensity image I is obtained as $I = (r + g + b)/3$. I is used to create a Gaussian pyramid

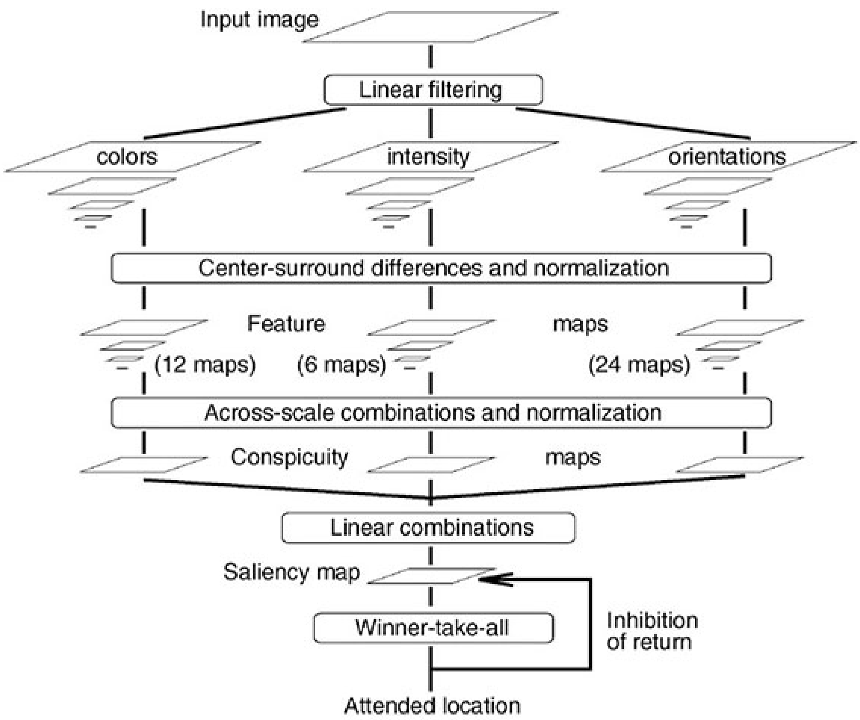


Figure 2.7: Visual attention architectural model. [IKN98]

$I(\sigma)$, where $\sigma \in \{0, 1, ..., 8\}$ is the scale. Four broadly-tuned color channels are created: $R = r - (g + b)/2$ for red, $G = g - (r + b)/2$ for green, $B = b - (r + g)/2$ for blue, and $Y = (r + g)/2 - |r - g|/2 - b$ for yellow (negative values are set to zero). Four Gaussian pyramids $R(\sigma)$, $G(\sigma)$, $B(\sigma)$, and $Y(\sigma)$ are created from these color channels.

Local orientation information is obtained from I using oriented Gabor pyramids $O(\sigma, \theta)$, where $\sigma \in \{0, 1, ..., 8\}$ represents the scale and $\theta \in \{0^0, 45^0, 90^0, 135^0\}$ is the preferred orientation. The Gabor pyramid is created from the image applied Gabor filter then creates a Gaussian pyramid. Gabor filters, which are the product of a cosine grating and a 2D Gaussian envelope, approximate the receptive field sensitivity profile (impulse response) of orientation-selective neurons in primary visual cortex.

The Gabor filter is defined as $g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \psi\right)$ where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$. In this equation, λ represents the wavelength of the cosine factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, ψ is the phase offset, and γ is the spatial aspect ratio, and specifies the ellipticity of the support of the Gabor function. (Gabor filter definition from

Wikipedia. `http://en.wikipedia.org/wiki/Gabor_filter`)

Center-surround differences

Center-surround is implemented in the model as the difference between fine and coarse scales. The center is a pixel at scale $c \in \{2, 3, 4\}$, and the surround is the corresponding pixel at scale $s = c + \sigma$, with $\sigma \in \{3, 4\}$. The across-scale difference between two maps, denoted "$\ominus$", is obtained by interpolation to the finer scale and point-by-point subtraction. Using several scales not only for c but also for $\sigma = s - c$ yields truly multi-scale feature extraction, by including different size ratios between the center and surround regions.

The first set of feature maps is concerned with intensity contrast, which, in mammals, is detected by neurons sensitive either to dark centers on bright surrounds or to bright centers on dark surrounds. Here, both types of sensitivities are simultaneously computed (using a rectification) in a set of six maps

$$\mathbf{I}(c, s) = |I(c) \ominus I(s)|.$$

A second set of maps is similarly constructed for the color channels, which, in cortex, are represented using a so-called "color double-opponent" system: In the center of their receptive fields, neurons are excited by one color (e.g., red) and inhibited by another (e.g., green), while the converse is true in the surround. Such spatial and chromatic opponency exists for the red/green, green/red, blue/yellow, and yellow/blue color pairs in human primary visual cortex. Accordingly, maps $\mathbf{RG}(c, s)$ are created in the model to simultaneously account for red/green and green/red double opponency and $\mathbf{BY}(c, s)$ for blue/yellow and yellow/blue double opponency:

$$\mathbf{RG}(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))|,$$

$$\mathbf{BY}(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))|.$$

Orientation feature maps, $O(c, s, \theta)$, encode, as a group, local orientation contrast between the center and surround scales:

$$\mathbf{O}(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)|.$$

In total, 42 feature maps are computed: six for intensity, 12 for color, and 24 for orientation.

Normalization

One difficulty in combining different feature maps is that they represent a priori not comparable modalities, with different dynamic ranges and extraction mechanisms. Also, because all 42 feature maps are combined, salient objects appearing strongly in only a few maps may be masked by noise or by less-salient objects present in a larger number of maps.

In the absence of top-down supervision, they propose a map normalization operator, $N(.)$, which globally promotes maps in which a small number of strong peaks of activity (conspicuous locations) is present, while globally suppressing maps which contain numerous comparable peak responses. $N(.)$ consists of (Figure 2.8):

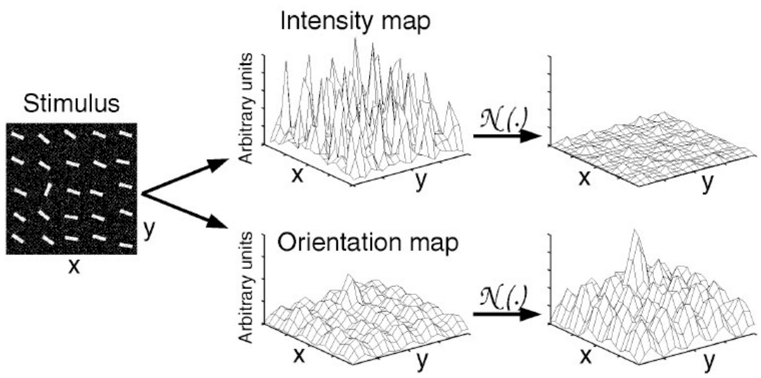


Figure 2.8: The normalization operator $N(.)$. [IKN98]

1. Normalizing the values in the map to a fixed range $[0...M]$, in order to eliminate modality-dependent amplitude differences;

2. Finding the location of the map's global maximum M and computing the average m of all its other local maxima; and

3. Globally multiplying the map by $(M-m)^2$.

Across-scale combinations

Feature maps are combined into three "conspicuity maps," $\bar{I}$ for intensity, $\bar{C}$ for color, and $\bar{O}$ for orientation, at the scale $(s=4)$ of the saliency map. They are obtained through across-scale addition "$\oplus$" which consists of reduction of each map to scale four and point-by-point addition:

$$\bar{I} = \oplus_{c=2}^{4} \oplus_{s=c+3}^{c=4} N(\mathbf{I}(c,s)),$$

$$\bar{C} = \oplus_{c=2}^{4} \oplus_{s=c+3}^{c=4} [N(\mathbf{RG}(c,s)) + N(\mathbf{BY}(c,s))],$$

$$\bar{O} = \sum_{\theta \in \{0^0, 45^0, 90^0, 135^0\}} N(\oplus_{c=2}^{4} \oplus_{s=c+3}^{c=4} N(\mathbf{O}(c,s,\theta))).$$

The motivation for the creation of three separate channels, $\bar{I}$, $\bar{C}$, and $\bar{O}$, and their individual normalization is the hypothesis that similar features compete strongly for saliency, while different modalities contribute independently to the saliency map. The three conspicuity maps are normalized and summed into the final input S to the saliency map:

$$S = \frac{1}{3}(N(\bar{I}) + N(\bar{C}) + N(\bar{O})).$$

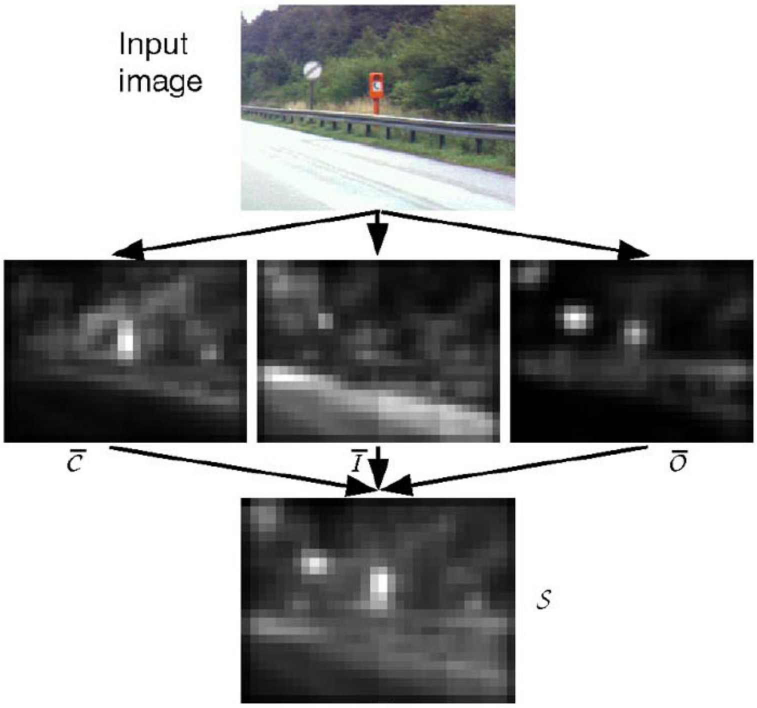


Figure 2.9: Salience map generated from Itti's ROI algorithm. [IKN98]

2.2.4 Face detection

The location of a human face in an image is often noticeable. When we analyze the location of important object, the location of human face is also very important. To do human face detection, we directly use the research result of Chung-Jung Hu [Hu07] . His face detection method is using an ordinary boosting method aided by skin color features. There the performance is better than ordinary boosting ones. After applying the face detector can get number of human face and human face location, and width.

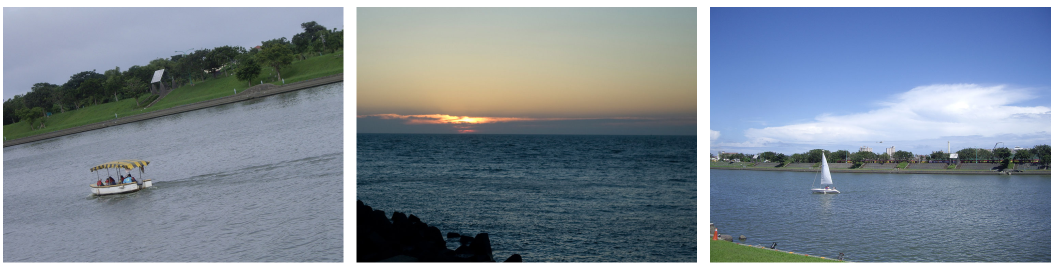
Chapter 3

Rules of esthetics in photo composition

It seems that there are few formal researches focusing on esthetics-based quantitative analysis of photo composition. Here we introduce rules of esthetics in photo composition generalized from books.

3.1 The horizon

Slant horizon will make a photo appear unbalanced and let people feel floating or unstable. A horizontal line placed at the center region so that the photo is divided into top and bottom two equal parts will let a photo dull. A better choice is to place the horizon at the trisection of the photo, or close to 1:1.6 of the golden ratio region. Whether the horizon should be placed on whether the upper trisection or lower trisection part is determined by want part you want to emphasize.



(a) (b) (c)

Figure 3.1: The horizon is slant (a), close to center region (b), and a better one (c).

3.2 Photo balance

If an image cannot be balanced, it must let people feel unstable. According to the feeling of human visual system, generally speaking, saturated color is more heavy than light color, dark is more heavy than bright, big area is more heavy than small area, clear one is more heavy than blurred one, ... , etc. Combining all above, human can get over all feelings to photo weight. In general, skewed center of gravity lets photo unbalance. Expect to left-right unbalanced situation, if all weights aggregate at the top part of a photo, the photo also let people have bad feeling.

Figure 3.2: The balance of left photo is bad and the right one is good.

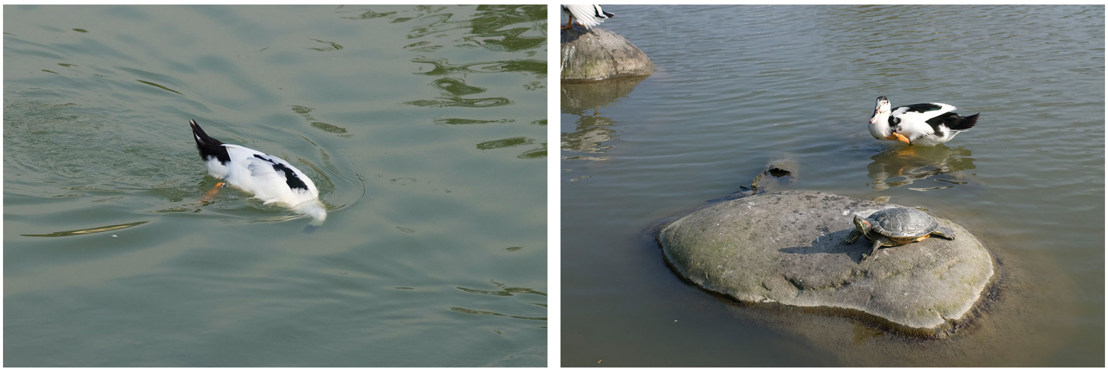
3.3 Location of main object: the rule of thirds

The rule of thirds is by far the best known composition rule. Imagine that your picture area is divided into thirds both horizontally and vertically. The intersections of these imaginary lines suggest four options for placing the center of interest for good composition.

The main object located in center of a photo often lets a photo felt dull.

3.4 Revealing line patterns and shapes

Line patterns and shapes in photos can let people have special feeling. For example, horizontal lines in a photo let people feel stable and expanded view. So photos that have certain interesting line patterns or shapes are often considered good in photo composition.



(a) (b)

Figure 3.3: Location of ROI, where (a) is centered and (b) is OK, according to the rule of thirds.

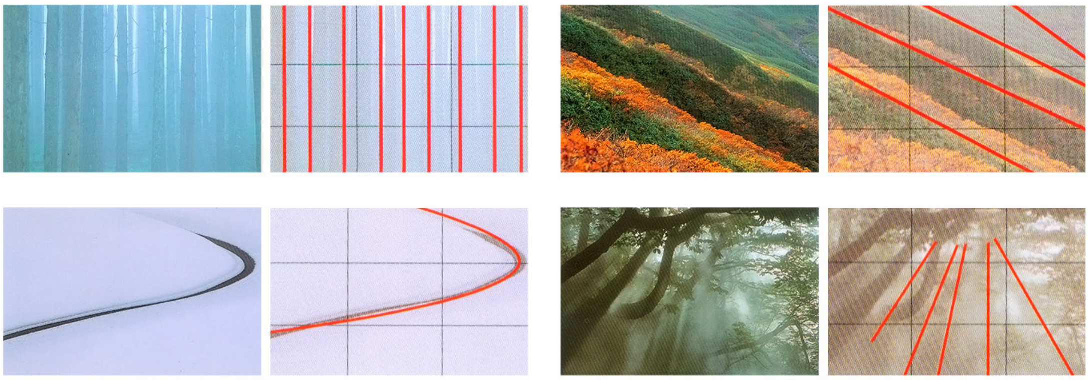


Figure 3.4: Photos contain shapes and line patterns information. (Photos are from [Yam06])

3.5 Avoiding mergers

Avoid mergers means that foreground object should not merge together with background objects. They will also merge with each other when their color is similar. Horizontal line cutting through the main object and tree appears on top of human head results in bad human visual feelings and bad photo composition.

(a) (b)

Figure 3.5: Avoiding mergers, where (a) a horizontal line cuts through the main object and (b) is OK.

3.6 Others: Reserving more space for the viewing direction, etc.

There still many other rules that can be used to analyze. But some rules' descriptions are not concrete, not easy to formulate into automatic computer processes. Other rules such as reserving more space for the viewing direction can be one part of our system if we have good idea to know main object's location and its viewing direction.

Figure 3.6: The space of the bird’s viewing direction should be reserved.

Chapter 4

Implementation

First, all photos were loaded into our program. We use our own CMLAB libgil2 (`https://www.cmlab.csie.ntu.edu.tw/trac/libgil2`) image library to do image I/O, and the library libgil2 also defines some image related data structures and has some basic image processing functions.

We resize all input photos to long edge at 800 pixels first in order to the efficiency of handling photos and the consistency of all photos. Currently we simply use bilinear filter to get resized images.

Rules are implemented to analysis modules for input photos, separately giving some comments and scores. We can also do cropping automatically depend on the information generated from each rule module. Our framework consists of two major components: photograph composition scoring and automatic correction. Next, we will describe the implementations in detail.

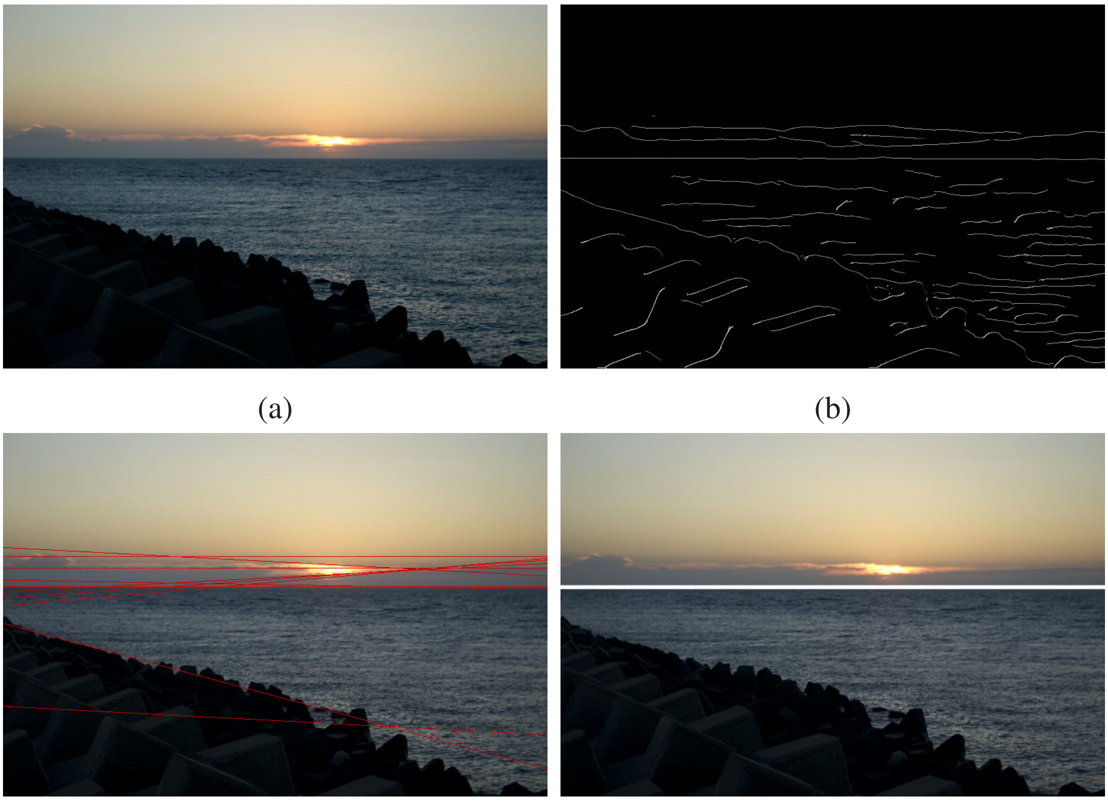
4.1 Scoring system

Based on the five rules described in Chapter 3, we create corresponding quantitative equations for all of them.

4.1.1 The horizon

We use Hough transform [DH72] on edge map to get the most effective line. The edge map is generated by Canny edge detector [Can86]. Figure 4.1 shows the process of horizon detection. Lines are represented by polar coordinate system (r, θ). We can use θ to calculate the angle difference of this line and horizontal 0^0. We denote this tilt angle as θ_t. The horizon tilting score s_{H_t} is defined as:

$$s_{H_t} = \exp(\frac{-{\theta_t}^2}{2{\sigma_{H_t}}^2}) - 1.$$



(a) (b)

(c) (d)

Figure 4.1: Locating horizon. (a) an original image, (b) the edge map created by Canny edge detector, (c) the result from Hough transform, and (d) located horizon.

When the horizon is not slant, then we consider its location. The location of horizontal line can be determined by its top down two parts area ratio. Before calculating the ratio, we transform the polar coordinate system representation of the line to the polynomial equation $ax + by + c$. (a, b) is the tangent vector of the line, the value is $(\cos\theta, \sin\theta)$. c is the distance of the line and the coordinate origin, the value is r. We define the top down two parts ratio R which is the division of two parts' area (the larger one divide to the smaller one). Top down two parts ratio around 1:1 means horizontal line close to center. A better choice is place the horizon at the trisection of the photo (1:2), or close to 1:1.6 the golden ratio region. $DH_{1.0} = [Image\ height]/(1 + 1.0)$, $DH_{1.6} =$

$[Image\ height]/(1+1.6)$, and $DH_{2.0} = [Image\ height]/(1+2.0)$ denote the distances of three horizons and the image boundary. The three horizons' top and down two parts ratio are 1.0, 1.6, and 2.0.

$$s_{H_{r=1.0}} = -\exp(\frac{-(D-DH_{1.0})^2}{2{\sigma_{H_r}}^2}),$$

$$s_{H_{r=1.6}} = \exp(\frac{-(D-DH_{1.6})^2}{2{\sigma_{H_r}}^2}),$$

$$s_{H_{r=2.0}} = \exp(\frac{-(D-DH_{2.0})^2}{2{\sigma_{H_r}}^2}),$$

where

$$\sigma_{H_r} = (DH_{1.6} - DH_{2.0})/2,$$

$$D = \min(r, [Image\ height) - r].$$

The horizon ratio score is denoted as s_{H_b}. If $DH_{1.6} > r > DH_{2.0}$, $s_{H_b} = 1.0$, else s_{H_b} equals to one of the s_{H_r} which has the largest magnitude.

4.1.2 Photo balance

We see every pixel on a photo as a particle. By setting photo center as a pivot point, use moment of force to get result of photo balance. The particle's weight now only considers its intensity. According to some research in visual psychology, the pixel intensity and human visual feeling have some kinds of exponential relation, and the darker pixel is heavier. The simplified equation for pixel weight is defined as $\log(256 - intensity)$, assume the maximum intensity of the image is 255. The gravitation is set toward downward direction. The torque of every pixel is its weight times the X direction offset with pivot point. Sum up all the force to get the total torque, and dividing image pixel number for normalization. If the total torque is large for one direction, the photo is unbalanced. We find the torque τ_I of a image I:

$$\tau_I = \frac{1}{wh} \sum_{x=1}^{w} \sum_{y=1}^{h} \log(256 - I(x,y))(x - \frac{w}{2}),$$

w and h are the width and height of image I. The intensity balance score s_B is defined as:

$$s_B = \exp(\frac{-{\tau_I}^2}{2{\sigma_B}^2}) - 1.$$

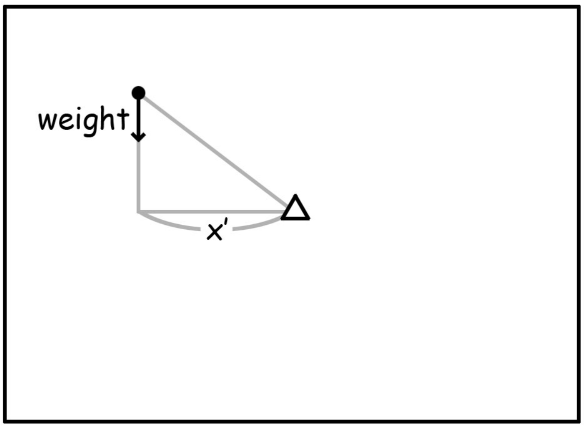


Figure 4.2: Photo balance is seen as physics system.

4.1.3 Location of main object

We first use Itti's ROI algorithm [IKN98] to get the salience map. The algorithm depends on human visual perception. Combining intensity contrast, color contrast and impulse reaction of every orientation will generate a salience map. The map shows the human visual attention scale. After having this map, set two thresholds to get the location of the most attentive regions. Firstly get the regions that the pixel intensity above high threshold, then expend from these regions until the pixel intensity is lower than low threshold. Analyze the location of main object by every region center. The coordinate of region center is denoted as (c_x, c_y). w and h are the image's width and height. Figure 4.3 shows the ROI detection process.

To get the score, first we calculate the x-direction ratio and the y-direction ratio of ROI center coordinate:

$$r_x = \frac{\max(w - c_x, c_x)}{\min(w - c_x, c_x)},$$

$$r_y = \frac{\max(h - c_y, c_y)}{\min(h - c_y, c_y)}.$$

If $r_x < \gamma$ and $r_y < \gamma$ ($\gamma = \frac{4}{3}$ in this case). The location score s_L is defined as:

$$s_x = \exp(\frac{-(D_x - DWx_{1.0})^2}{2{\sigma_{Lx}}^2}),$$

$$s_y = \exp(\frac{-(D_y - DHy_{1.0})^2}{2{\sigma_{Ly}}^2}),$$

$$s_L = -s_x s_y,$$

otherwise

$$s_x = \max(\exp(\frac{-(D_x - DWx_{2.0})^2}{2{\sigma_{Lx}}^2}), \exp(\frac{-(c_x - DWx_{1.6})^2}{2{\sigma_{Lx}}^2})),$$

$$s_y = \max(\exp(\frac{-(D_y - DWy_{2.0})^2}{2{\sigma_{Ly}}^2}), \exp(\frac{-(c_y - DWy_{1.6})^2}{2{\sigma_{Ly}}^2})),$$

$$s_L = s_x s_y,$$

where

$$DWx_R = \frac{w}{1+R},$$

$$DWy_R = \frac{h}{1+R},$$

$$\sigma_{Lx} = (DWx_{1.6} - DWx_{2.0})/2,$$

$$\sigma_{Ly} = (DWy_{1.6} - DWy_{2.0})/2,$$

$$D_x = \min(c_x, w - c_x),$$

$$D_y = \min(c_y, h - c_y).$$

If the image exhibits multiple ROIs, s_L becomes the weighted sum of all the individual ROI score. The weight is determined by the size of the ROI.

Human face in a photo also is often interested by people. We can see them as main object and do the same analysis. How to detect the location of human face will be mentioned in avoid merger section. After having human face location, we can analyze it and get it locates in center of a photo or other results.

4.1.4 Revealing line patterns and shapes

This work firstly only focuses on straight lines. They are horizontal lines, vertical lines, slope lines, and radial lines. Horizontal lines, vertical lines, slope lines are form by parallel lines. We can see them as a group.

Similar to horizontal line extraction, we can use Canny edge detector [Can86] and Hough transform [DH72] to get the most effective ten lines. Analyze if there are parallel lines among them. Among these ten lines, we calculate their all pair angle differences. The angle difference θ_d is

$$\min(|\theta_i - \theta_j|, |180 - |\theta_i - \theta_j||).$$

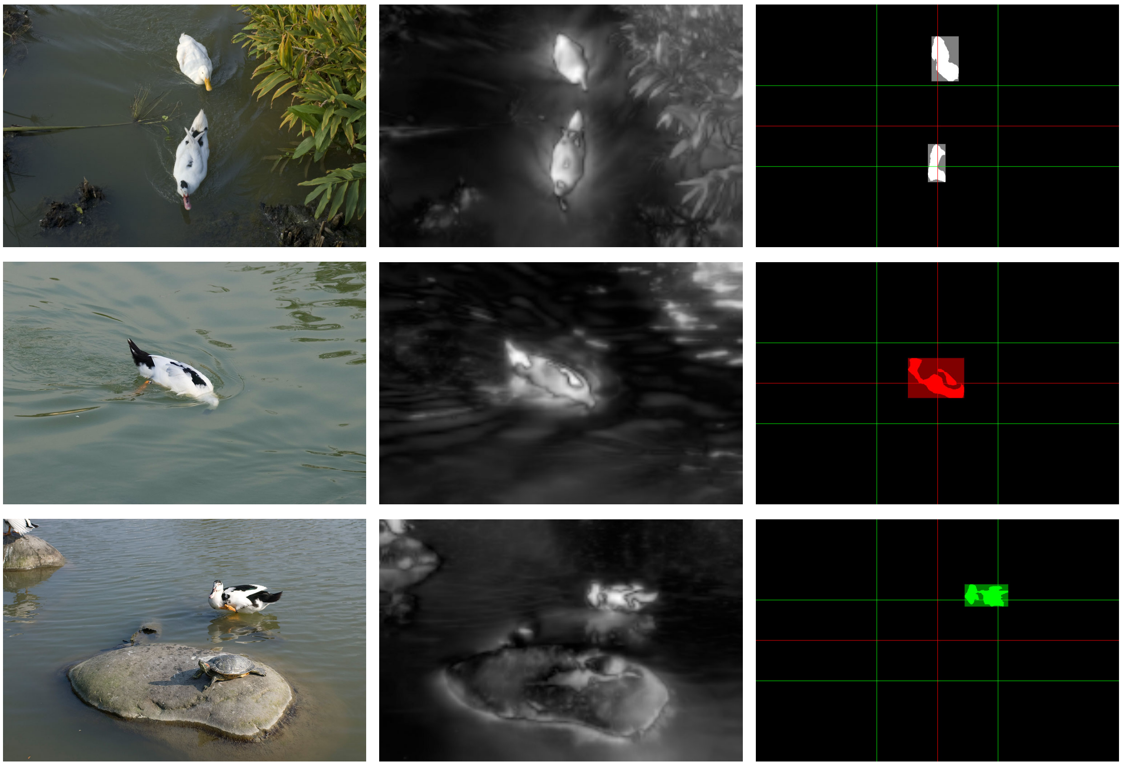


Figure 4.3: ROI detection. Left column, original images. Center column, salience map generated by Itti's ROI algorithm. Right column, found ROI regions.

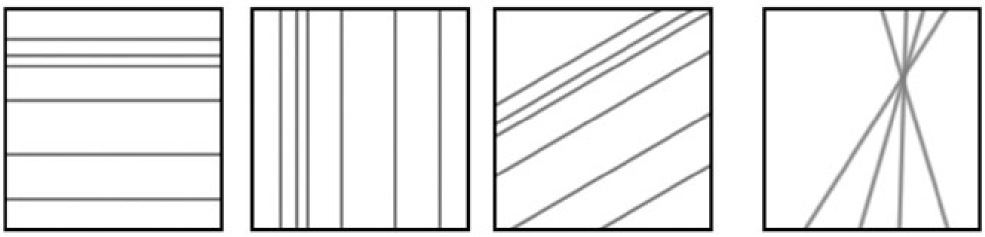


Figure 4.4: Horizontal lines, vertical lines, slope lines, and radial lines.

Count the number of small difference pair (we define the maximum difference is 4^0), and if there are more than three parallel line pairs, existing one group of parallel lines. If parallel lines exist, whether a photo has horizontal lines, vertical lines, or slope lines depends on the orientation of lines.

$$\text{Parallel lines type is} \begin{cases} horizontal, & \text{if } \theta_d < 3 \text{ or } \theta_d > 177 \\ vertical, & \text{if } \theta_d > 87 \text{ and } \theta_d < 93 \\ slope, & \text{otherwise.} \end{cases}$$

In the part of radial lines, calculate the all-pair intersection points of these ten lines. Points out the two intersection points pair that their distance is small than 30 pixels. If there are more than three these kinds of pairs, the image contains a set of radial lines that more than three lines among it. Their intersections are within a radius of 30 pixels.

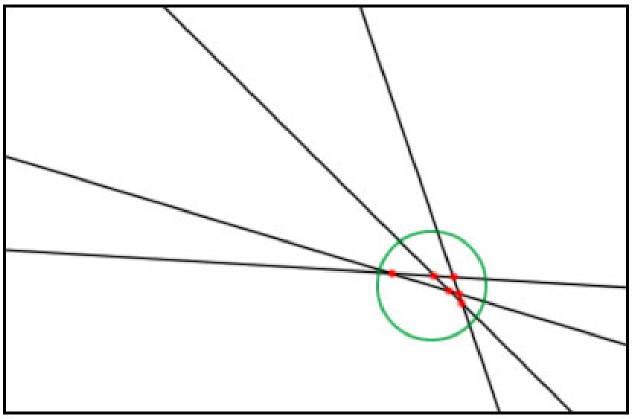


Figure 4.5: Radial lines and their intersection points, intersection points are within a radius of 30 pixels.

The score of revealing line patterns and shapes is the summation of all lines' strength in found line patterns. The line strength is the detected edge's length on the line divides to the possible whole line length in the image. Assume the strength of detected line is ls_i, and there are found patterns with n lines. The revealing score s_R is defined as:

$$s_R = \sum_{i=1}^{n} ls_i.$$

4.1.5 Avoiding mergers

This work firstly only focuses on the main object is human, detects if straight lines cut through human. The method to get face location is applying the face detector by Chung-Jung Hu [Hu07]. His face detection method is using ordinary boosting method and aid by skin color features. After applying the face detector can get number of human face and human face location, and width r. Now we can check if there are lines cut through them. See the human face as a radius r circle and projects the center of circle O to the line that passes the image origin and orthogonal to the cutting line. The distance of projection point O' and the image origin point P is $|\overline{PO'}|$. The coordinate of the center of circle O is notated as (O_x, O_y), and $|\overline{PO'}| = \cos\theta * O_x + \sin\theta * O_y$. The distance of the cutting line and the image origin is R. If $|\overline{PO'}| - r < R < |\overline{PO'}| + r$, this cutting line cuts through human face. Above method is used because we denote the line as (r, θ). If we have the line equation $ax + by + c$, we can just perform the simple line and circle intersection test. Figure 4.7 shows the merger detection result. If we want to detect lines through human body, suppose there is a virtual circle below the face, use the same method to calculate if lines cut through it. Assume the distance between O and detected line l_i is d_i, and there are n found lines. The merger score s_M is defined as:

$$s_M = -\sum_{i=1}^{n} \sqrt{1 - (\frac{2d_i}{w_f})^2}.$$

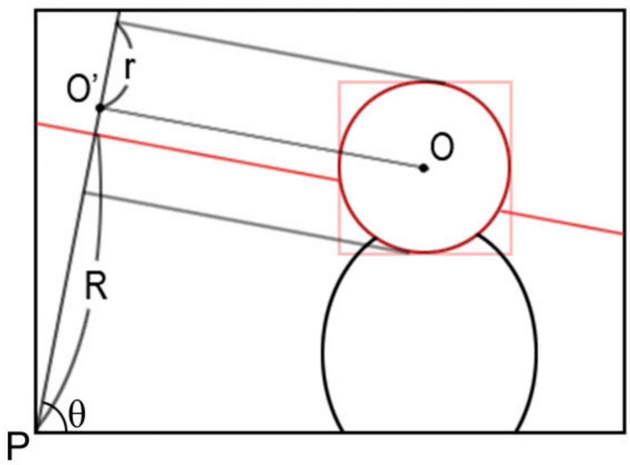


Figure 4.6: The diagram shows face and line intersection.

(a) (b)

Figure 4.7: Merger detection. (a) the face detection result and (b) lines retrieved by Hough transform overlaid with the face ROI.

4.1.6 Parameters

Now we have six dimensional score vector $(s_{H_t}, s_{H_b}, s_B, s_L, s_R, s_M)$. We manually choose several parameters which are used in the scoring equations. In our implementation, $\sigma_{H_t} = 1.0$ and $\sigma_B = 5.0$.

4.2 Automatic correction

In the previous section, we have described the quantitative metrics based on photo composition. A six dimensional vector will be reported as a first order description of the input image. As one of our goals is to correct badly shoot photographs, we would like to use the composition descriptor to make existed photographs look better. We perform rational correction, cropping correction for photo balance and object centered step by step, but we have no solution for the problem that after fixing a problem from one rule, the other problems from the other rules may occur.

4.2.1 Rotational correction

Rotational correction is simply done by rotating the image $-\theta$ degrees so that the image is horizontally balanced. Once the image is rotated, we then find the rectangle which has the maximum area within the rotated region and the original aspect ratio is maintained. Figure 4.8 shows the steps of rotational correction.

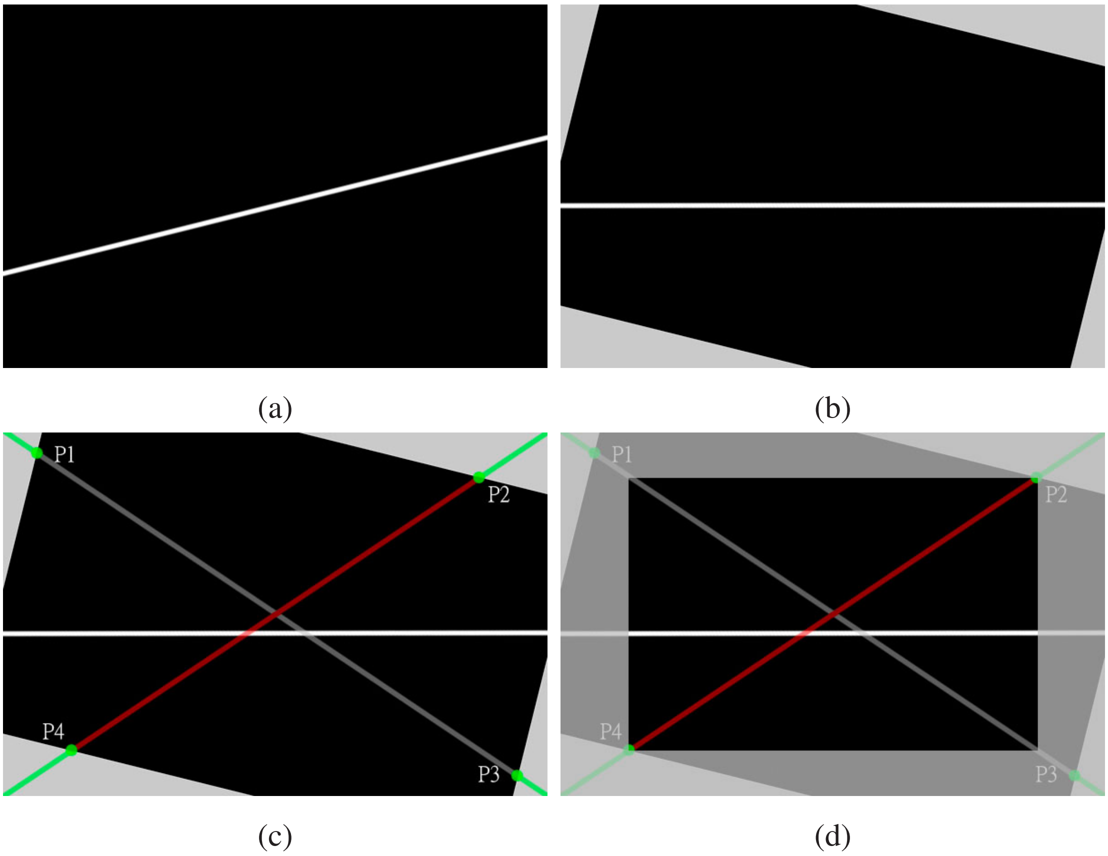


Figure 4.8: Steps of rotational correction.

(a) Input image.

(b) Rotate image by $-\theta$.

(c) P_1 to P_4 are the intersections of two diagonal lines and original image boundaries.

(d) Choosing P_2 and P_4 to be result bounding box, however $|\overline{P_2P_4}| < |\overline{P_1P_3}|$.

4.2.2 Cropping correction

If the analysis of photo composition results in bad comment, it may have chance to refine by doing crop. For example, the major horizontal line is centered, the photo is unbalance, and the main object is centered, etc. Try to do crop an image to break the bad results. There are many possible cropping results. We apply them to other rules step by step to get the final one.

Firstly focusing on the part of photo balance, we use a greedy algorithm to get the cropping result. We continuously cut off one slice at the left side or the right side of the image determined by which result reserves more ROI, and then calculates the torque of whole image again. The slice width we choose is 5 pixels. In order to maintain the width and height ratio of original image, we need to cut off a slice of the upper or the lower side of the image. We select the side that has more refinement deflection after cutting off it. Iterative cutting slice and calculating new torque until the photo balance is fitted in certain threshold.

Figure 4.9: Cropping correction for unbalanced photo. Cutting off a slice of right side of the image and the upper or the lower side of the image.

In the part of "centered," no matter the major horizontal line or the main object is using the same solution. We already know the ratio of the two sides is 1:1. We can just crop one side to let two sides ratio of image become 1:2 or 2:1. There are upper left, upper right, lower left, and lower right four cases. Apply to the photo balance rule and choose the best balance one.

Chapter 5

Results

In this part, we test some photos that show good or bad in each rule and have scores show the difference between rules. s_{H_t}, s_{H_b}, s_B, and s_L are normalized to the range of $[0, 1]$, where 1 represents the best score. If the scoring system cannot find any response from a particular rule, we assign the score of that rule as a null number. In order to compare the precision of our system, we do some thresholdings from experiments to determine which type the photo is. We select the 132 testing photos because they have features related to our five rules and focus on simpler cases. The precision of the following five categories are 71%, 96.8%, 73.1%, 78.4%, 71.4%, respectively. Next, we show some results for every part in detail.

5.1 The horizon

In the part of the horizon, we select 31 photos that contain at least one obvious horizon or the horizon is occluded most half of it. We say that the horizon is slant if the tilt angle of the horizon is more than 3^0, and the horizon is centered if the ratio of top part and down part is less than 1.1. We say that the horizon is placed at good place if the ratio is between $\frac{9}{6}$ and $\frac{14}{6}$. Among these 31 photos, 22 photos are detected that contain the horizon and the locations are correct. Figure 5.1 shows the horizon detection results and their skew angle and top down two parts ratio and the score vector.

Category 1: 22 out of 31 test cases are correct. (71.0%)

$[0.000, 0.003, 0.989, -, -, -]$
10^0, ratio: 1.246
The horizon is slant.

$[1.000, 0.003, 0.997, 0.594, -, -]$
0^0, ratio: 1.020
The horizon is very close to center

$[0.882, 1.000, 0.972, 0.520, -, -]$
1^0, ratio: 1.911
The horizon is placed at good place

Figure 5.1: Results of the horizon.

5.2 Photo balance

In the part of photo balance, we detect photos that are visually unbalanced. If the total torque of the photo is more than 10, the photo is unbalanced and right heavy. If the total torque of the photo is less than -10, the photo is unbalanced and left heavy.

$[0.002, 0.500, 0.011, 0.544, -, -]$
Torque: 15.040761 (right heavy)

$[-, -, 0.020, -, -, -]$
Torque: -13.949572 (left heavy)

$[-, -, 0.955, -, -, -]$
Torque: -1.520265 (no comment)

Figure 5.2: Results of photo balance.

Category 2: 29 out of 31 test cases are correct. (93.5%)

5.3 Location of main object

In the part of location of region of interest (ROI), we select 26 photos that contain obvious ROI. Among these 26 photos, 19 photos are detected that contain major ROI and its location is correct.

$[-, -, 0.996, 0.050, -, -]$
ROI is centered.

$[-, -, 0.858, 0.830, -, -]$
ROI is placed at good place.

$[-, -, 0.999, 1.000, -, -]$
ROI is placed at good place.

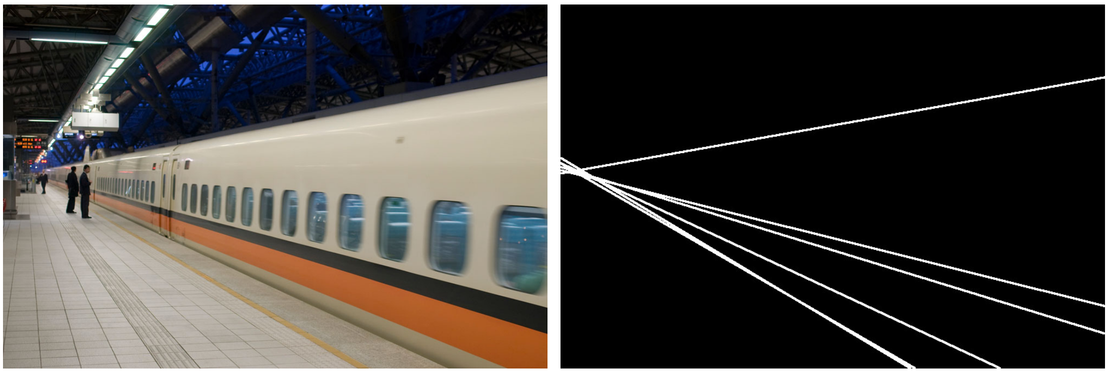
Figure 5.3: Results of the location of ROI.

Category 3: 19 out of 26 test cases are correct. (73.1%)

5.4 Revealing line patterns and shapes

In the part of revealing line patterns and shapes, we select 37 photos that contain obvious line patterns. Among these 37 photos, 29 photos are detected that contain line patterns and its type is correct.

$[0.000, 0.915, 1.000, 0.500, 3.497, -]$ Slope line pattern is detected.



$[-, -, 0.999, 0.500, 1.451, -]$ Slope radial pattern is detected.

Figure 5.4: Results of revealing line patterns and shapes.

Category 4: 29 out of 37 test cases are correct. (78.4%)

5.5 Avoiding mergers

In this part, the number of related photo we collected is few at present. We select 7 photos that contains human faces with crossing straight lines. Figure 5.5 shows some results.

Category 5: 5 out of 7 test cases are correct. (71.4%)

5.6 Better photo composition by cropping

Figure 5.6 shows an unbalanced photo and its cropped result.

$[0.607, 0.578, 0.919, 0.500, 1.085, -3.434]$
A line cuts through the human face.

$[1.000, 0.001, 0.890, 0.500, -, -1.552]$
A line cuts through the human face.

$[0.607, 0.5, 0.924, 0.500, -, 0]$

Figure 5.5: Results of avoiding mergers.

Original photo   After cropping

Figure 5.6: Cropped results for photo unbalance.

Figure 5.7 shows the results of tilted horizon, because the horizon is tilted.

Original photo After cropping

Figure 5.7: Cropped results - slant horizon.

Figure 5.8 shows the results of centered object. The main object locates around the center of a photo. Reposition the main object and consider photo balance.

Original photo After cropping

Figure 5.8: Cropped results - centered.

Chapter 6

Conclusion and future work

6.1 Conclusion

We implement esthetics rules in photography to automatically analyze five major rules, and use these rules to determine whether the photo composition is good or not, since we can get some scores and recommendations. We hope that a user can gain his/her knowledge during this process and so can take better photos in the future. Photo composition is just one reason, perhaps the most important one, that determines whether a photo is good or not. According to these analysis results, we can get better photos by cropping.

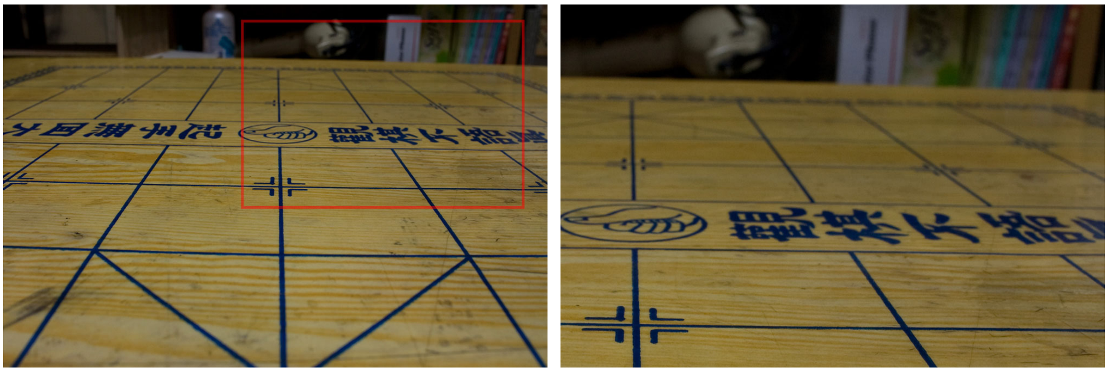
6.2 Discussion

In out implementation, we have six dimensional score vectors after analyzing from five rules. For a user's point of view, one may hope to obtain a single result for a photo. It is hard to determine the final result from multiple scores, it may not be a good solution by just use linear weight for every scores. Sometimes one bad score from a single rule may dominate the final score.

In the part of photo correction, we focus on one single rule that reports bad score for photo correction, and step by step correction to get the final result. If we can consider all the rules at once, maybe we can get better result. But there may exist conflicts among corrections from rules.

We can also optimize a photo based on one rule. For example, a photographer takes a photo according to the rule of thirds and places the main object to the right place, but the placement of the main object may have offset to the best place initially, and therefore we can correct this kind of small error to the optimal one.

Professor Rung-Huei Liang of NTUST found a problem at our cropping correction section. He said that the cropping will affect the photo's perspective and field-of-view. If the cropping region is not at the center of photo, the cropped image cannot be produced by normal camera shot, and the image looks a little strange. This situation is more visible when the photo is taken using wide angle lens. Figure 6.1 shows an example for this problem. We can see that image (b) has a right angle at the image's lower left. It is impossible to have this image when using a normal camera.



(a) (b)

Figure 6.1: Perspective problem from photo cropping. (a) the original image, (b) is cropping from (a).

6.3 Future work

In the part of photo balance analysis, there are still more features that can be used to evaluate the weight (color saturation, sharpness, etc.) for unbalance calculation. Extract these information of an image is not too difficult but how to combine these information to evaluate the weight is a big problem. Maybe we can find more discussions in psychology researches. Furthermore, it may be evaluated block by block, not only pixel by pixel. Simultaneous contrast effect and assimilation effect [Chi02] in psychology shows that pixels do affect each other. It still needs more investigation to implement and apply these theories.

In the part of line patterns and shapes, lines in photo are not always formed by simple edges, but can be formed by texture of lines. There may be certain kinds of asymptotic lines. If we can extract that kind of lines or textures, more types of line patterns can be detected. Furthermore, we only focus on straight lines now, curved lines and shapes still need more investigation.

Figure 6.2: Lines are not formed by simple edges. (This photo is from Flickr)

In the part of photo correction, we can make more improvements which are mentioned in discussion section.

We can also implement other esthetics rules such as the space of the viewing direction, simplicity, framing, etc, to add more rules to let the analysis of the photo composition problem becomes more complete. Our implementations of five rules also allow other implementation methods to get better results.

We also need more user studies to evaluate the performance of our work. We need to evaluate the rules inside a photo and to know which one is better between the original photo and the photo corrected by our system.

Currently, our program interface contains simply photo input, and will output a report that contains all parts of detection results and recommendations. We think that these information is not clear enough for users. Providing better user interface can also bring in more interesting applications.

Figure 6.3: A report shows complex information.

Bibliography

[AS07] AVIDAN S., SHAMIR A.: Seam carving for content-aware image resizing. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers* (New York, NY, USA, 2007), ACM, p. 10.

[BC99] BURIAN P., CAPUTO R.: *National Geographic Photography Field Guide (中譯：國家地理攝影精技)*, 2nd ed. National Geographic, 1999, pp. 24–33.

[BE07] BANERJEE S., EVANS B.: In-camera automation of photographic composition rules. *IEEE Trans. Image Processing 16*, 7 (July 2007), 1807–1820.

[Can86] CANNY J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell. 8*, 6 (1986), 679–698.

[Chi02] CHIANG T. J.: *Perception and Photography (觀念攝影 - 影像的視覺心理剖析)*. Xue Ling culture Co. (雪嶺文化), 2002, pp. 82–87.

[CXF*03] CHEN L.-Q., XIE X., FAN X., MA W.-Y., ZHANG H., ZHOU H.-Q.: A visual attention model for adapting images on small displays. *Multimedia Syst. 9*, 4 (2003), 353–364.

[DH72] DUDA R. O., HART P. E.: Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM 15*, 1 (1972), 11–15.

[Hu07] HU C.-J.: A real-time skin-color-enhanced face detection algorithm. *Master thesis, Dept. of Computer Science and Information Engineering, National Taiwan University* (2007).

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 20*, 11 (1998), 1254–1259.

[Pet03] PETERSON B.: *Learning to See Creatively (中譯：拍照前先學會看)*, revised ed. Amphoto Books, 2003, pp. 88–129.

[SAD*06] SANTELLA A., AGRAWALA M., DECARLO D., SALESIN D., COHEN M.: Gaze-based interaction for semi-automatic photo cropping. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems* (New York, NY, USA, 2006), ACM, pp. 771–780.

[SEL00] SAVAKIS A. E., ETZ S. P., LOUI A. C.: Evaluation of image appeal in consumer photography. In *Human Vision and Electronic Imaging V* (June 2000), Rogowitz B. E., Pappas T. N., (Eds.), vol. 3959 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pp. 111–120.

[SLBJ03] SUH B., LING H., BEDERSON B. B., JACOBS D. W.: Automatic thumbnail cropping and its effectiveness. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2003), ACM, pp. 95–104.

[Tsa98a] TSAI M.-F.: *Better Photography (攝影常犯的錯誤)*. Recreation Press (渡假出版社), 1998, pp. 9–15, 17–21, 23–24, 32.

[Tsa98b] TSAI M.-F.: *How to Compose a Picture (構圖的訣竅)*. Recreation Press (渡假出版社), 1998, pp. 24–32, 76–88, 108–117, 131.

[Yam06] YAMAGUCHI T.: *Photo Composition Determines All! Methods to Take Great Photos (中譯：構圖決定一切！高水準攝影完全攻略)*. Gakken Co., 2006, pp. 5–18.

Hong-Cheng Kao (高弘政)

photo@cmlab.csie.ntu.edu.tw
TEL: 886-958232992

Education:

MS: CSIE, National Taiwan University (2006/9 – 2008/6)
Thesis: Esthetics-based Quantitative Analysis of Photo Composition
Advisor: Prof. Ming Ouhyoung
BS: CSIE, National Chong Cheng University (2002/9 – 2006/6)
HS: The Affiliated Senior High School of National Taiwan Normal University (1999/9–2002/6)

Working experience:

中正大學資訊工程學系網頁設計 (2004/11 – 2005/6)
中正大學圖書館電腦軟硬體維護

Skill:

C/C++, Java, HTML, PHP, Computer graphics, Digital image processing, Photoshop

Contest award:

2004 趨勢百萬程式競賽第五名隊伍成員
初賽題目：Central Computer Activity Monitor System
決賽題目：Intelligent Computer Problem Analysis System

Paper publication:

Hong-Cheng Kao, Wan-Chun Ma and Ming Ouhyoun, "Esthetics-based Quantitative Analysis of Photo Composition," in Pacific Graphics 2008 Conference Posters, 2008, Tokyo, Japan

Hong-Cheng Kao and Ming Ouhyoung, "Esthetics-based Quantitative Analysis of Photo Composition - how do you know if a photo is beautiful," The Cooperative Workshop of CyberLink and CML on Communications and Multimedia, 2008, Taipei, Taiwan. (Second price)