

國立交通大學

資訊科學與工程研究所

碩士論文

網站舒適度的評價與估量

An Evaluation and Measure for Websites Visual Comfort

研究生：張家駿

指導教授：謝筱齡/黃廷祿 教授

中華民國九十八年八月

網站舒適度的評價與估量  
An Evaluation and Measure for Websites Visual Comfort

研究生：張家駿

Student : Chia-Chun Chang

指導教授：謝筱齡/黃廷祿

Advisor : Sheau-Ling Hsieh/ Ting-Lu Huang

國立交通大學  
資訊科學與工程研究所  
碩士論文

A Thesis

Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

August 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年八月

學生：張家駿

指導教授：謝筱齡

黃廷祿

國立交通大學資訊科學與工程所 碩士班

## 摘要

人們從傳播媒體取得訊息的來源已逐漸從電視、報章雜誌轉變為網際網路，網路漸漸成為現今人主要的獲取資訊來源。在長時間的使用下，容易讓使用者產生疲倦等感覺，其中一個影響因素來自於視覺。

本研究探討在網站瀏覽過程中的明暗度變化所帶給使用者的舒適度觀感，藉由錄製網站瀏覽的影像片段，使用影像處理分析明暗度變化程度，並透過 SVM 與使用者問卷的評分做比較分析，試圖建立影像處理過後的預測模型來模擬使用者對於視覺舒適程度的心理觀感。

結果顯示，影像分析的標準與使用者的問卷結果呈現正相關。並以此影像分析方法完成一視覺化工具進而可用來分析畫面變動程度與給予視覺舒適程度評鑑。

關鍵字：視覺舒適度，易用性，支援向量機

# An Evaluation and Measure for Websites Visual Comfort

Student : Chia-Chun Chang

Advisors : Dr. Sheau-Ling Hsieh

Dr. Ting-Lu Huang

## Abstract

The way People retrieve information has been changed from traditional media, e.g., TV or newspaper, to surfing Internet as the mainstream method. Normally, after surfing for a long period, people feel tired. One of the causes is the visual effect received by our eyes.

The study focuses on the visual comfort caused by luminance changing frequency during the navigating behavior. We capture surfing video clips and use image processing to analyze luminance modification signals. At the same time, utilizing Support Vector Machine to construct an evaluation tool including a predict model depending on people rating results.

The results illustrate that signal sequences generated by Matlab analyzer have positive correlation with people rating results. Moreover, we design and develop a prototype of video analyzer tool with graphic user interfaces.

Keywords: visual comfort, accessibility, Support Vector Machine, Matlab

# Contents

摘要 .....	i
Abstract .....	ii
Contents .....	ii
List of Table .....	iii
List of Figure .....	iv
Chapter 1 Introduction .....	1
Chapter 2 Backgrounds .....	2
2.1 Flicker, Epilepsy and Visual Comfort .....	2
2.2 Previous research .....	2
2.3 Web Accessibility .....	5
2.4 Evaluation Tools .....	9
Chapter 3 Methodology .....	10
3.1 Experiment Data Preparation .....	11



<b>3.2</b>	<b>Video Processing Model .....</b>	<b>12</b>
<b>3.3</b>	<b>K-means clustering and Support vector machine learning .....</b>	<b>14</b>
3.3.1	K-means clustering.....	14
3.3.2	Support vector machine .....	15
<b>3.4</b>	<b>Evaluation threshold .....</b>	<b>16</b>
<b>Chapter 4</b>	<b>Experimental Results .....</b>	<b>18</b>
<b>4.1</b>	<b>Questionnaire results.....</b>	<b>18</b>
<b>4.2</b>	<b>SVM arguments decision .....</b>	<b>21</b>
<b>4.3</b>	<b>Evaluation tool prototype .....</b>	<b>23</b>
<b>Chapter 5</b>	<b>Conclusions .....</b>	<b>29</b>
<b>5.1</b>	<b>Conclusions .....</b>	<b>29</b>
<b>5.2</b>	<b>Future Work .....</b>	<b>29</b>
<b>References.....</b>	<b>.....</b>	<b>30</b>

## List of Table

<b>Table 3-1</b>	<b>Hardware and software environment we record a web surfing video clip.....</b>	<b>12</b>
------------------	--	-----------

<b>Table 4-1 Detailed questionnaire rating scores table.....</b>	<b>19</b>
<b>Table 4-2 result of partition label of rating scores clustered by k-means clustering.....</b>	<b>21</b>
<b>Table 4-3 Positive and negative classification using questionnaire rating and statistical value from luminance variable pixel sequence .....</b>	<b>23</b>

## List of Figure

<b>Figure 2-1 The TMTF plots the limits of the ability to perceive flicker for various frequencies (x-axis) and modulations (y-axis).....</b>	<b>3</b>
<b>Figure 2-2 Broca-Sulzer effect.....</b>	<b>3</b>
<b>Figure 2-3 Components of Web Accessibility.....</b>	<b>5</b>
<b>Figure 2-4 Cycle of Accessibility Implementation .....</b>	<b>6</b>
<b>Figure 2-5 Components and Guidelines for Web Accessibility .....</b>	<b>6</b>
<b>Figure 2-6 Screenshot of PEAT .....</b>	<b>9</b>
<b>Figure 3-1 Architecture of questionnaire, screen capturing, image processing, classification and SVM training model.....</b>	<b>10</b>
<b>Figure 3-2 Luminance translation from sRGB color space into relative luminance domain .....</b>	<b>12</b>
<b>Figure 3-3 Extraction process to calculus pixel variation in relative luminance domain</b>	<b>13</b>

**Figure 3-4 Architecture of support vector machine training and predict model..... 15**

**Figure 4-1 Pie chart for all questionnaire results ..... 18**

**Figure 4-2 Difference of mode of rating between CS students and non-CS student .....20**

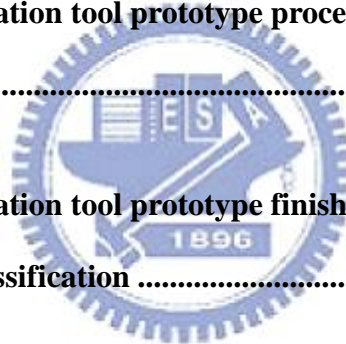
**Figure 4-3 Sample stairs chart of luminance variation pixel sequence .....24**

**Figure 4-4 Screenshot of evaluation tool prototype initiation .....25**

**Figure 4-5 Screenshot of evaluation tool prototype after loading a input video.....26**

**Figure 4-6 Screenshot of evaluation tool prototype processing and playing input video  
file.....27**

**Figure 4-7 Screenshot of evaluation tool prototype finishing a detect session and the  
corresponding predict classification .....28**





# Chapter 1 Introduction

Current World Wide Web becomes more and more rich information including picture, sound and video clips. Even in the normal surfing behavior, our eyes looking at the monitor is receiving signals from screen. Some people may have the experience of getting tired when using computer for a long time. That's because the visual images may affect brains without notice.

In Chapter 2, we introduce backgrounds of flicker, visual comfort and related issues. The W3C web accessibility standards are referred in order to introducing current issue on the Internet. In the end of this chapter, we introduce tools developed for visual effect evaluations.

In Chapter 3, the whole architecture of our study is explained in detail, including experiment data preparation, video processing model and support vector machine. Moreover, we discuss the threshold of visual flicker in web accessibility.

In Chapter 4, we show the results of our experiment. Detailed parameter decisions are explained there. The evaluation tool designed and developed in this study is illustrated in the Chapter as well.

Conclusions and future work are described in the last chapter.

## Chapter 2 Backgrounds

In this chapter, we introduce visual comfort, accessibility and other related background knowledge. First, we introduce why visual comfort is important in section 2.1. Second, web accessibility is described in section 2.2. In section 2.3, evaluation tools will be mentioned.

### 2.1 Flicker, Epilepsy and Visual Comfort

Eyes as to be observer to real world is the most important information receptor of our body. And the light through our pupil relay on our retina translated to the signal passed along the nerve system. When the brain processes these signal, it form the image we have seen. One of the major visual perceptions is luminance and flash contrast. Luminance is related to the volume of lightness that our retina received.

Once the situation switches high to low brightness rapidly, it causes flicker, which may trigger photosensitive epilepsy is the problem we face now.

What are these triggers for people with photosensitive epilepsy? One is the screen that flickers slowly and main factors are frequency, brightness, the amount of field of vision exposed to the light and the background light levels. Photosensitive epilepsy affects approximately one in four thousand people [1].

In 1997, the famous animation pokemon episode, "Dennou Senshi Porygon", consisting of many flashing images, was broadcast in Japan, causing many children to report seizure-like syndromes. In this event, the public starts to pay more attention on photosensitive epilepsy. Some signals receiving from eyes do make people illness [2].

### 2.2 Previous research

The luminance change frequency [3] acts as a function of time and is characterized by the temporal modulation transfer function. Two main parameters are described and measured the person's temporal MTF. First, amplitude of modulation indicates the difference between maximum and minimum luminance from the mean. We can compute the percentage depth

modulation referred as temporal contrast. The other is temporal frequency in Hertz that means the number of cycles per second.

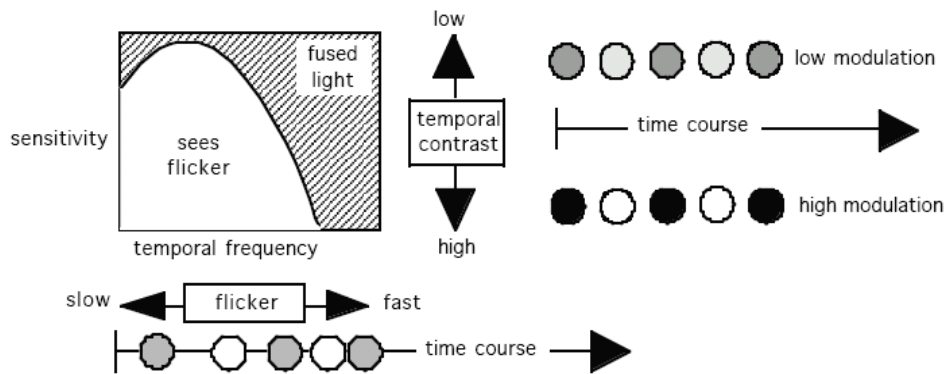


Figure 2-1 The TMTF plots the limits of the ability to perceive flicker for various frequencies (x-axis) and modulations (y-axis).

The Broca-Sulzer effect (phenomenon) states that a light with fixed luminance and duration of 50-100 msec will appear brighter than if it were left on for a shorter or longer time

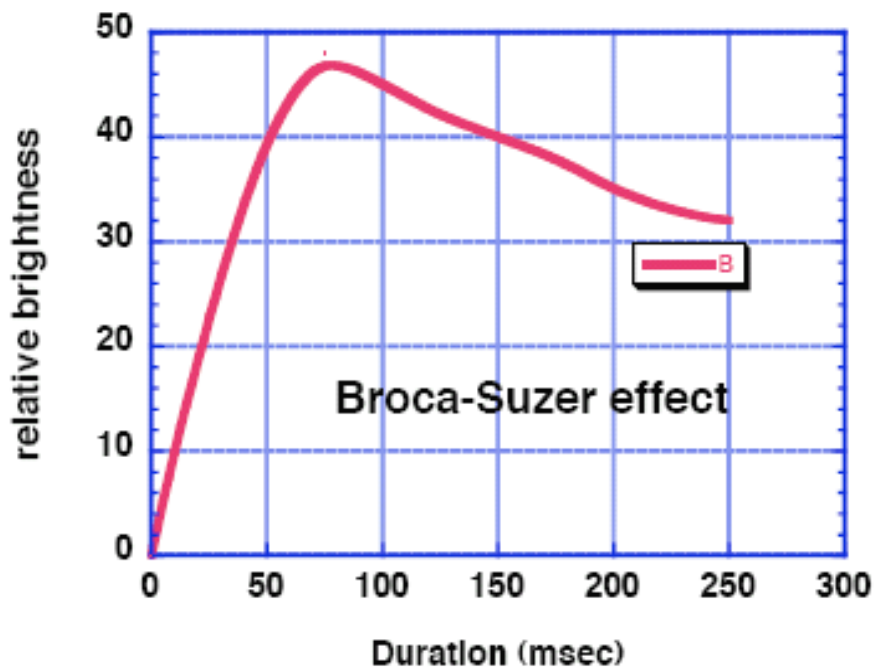


Figure 2-2 Broca-Sulzer effect.

Since the volume of light we receive stand important role in this topic, we have to explain how luminance is measured. Standard unit of Luminance is candela per square meter ( $\text{cd/m}^2$ ) [4].

$$Lv = \frac{d^2F}{dAd\Omega\cos\theta}$$

Where  $L_v$  is the luminance ( $\text{cd/m}^2$ ),  $F$  is the luminous flux or luminous power (lm),  $\theta$  is the angle between the surface normal and the specified direction,  $A$  is the area of the surface ( $\text{m}^2$ ), and  $\Omega$  is the solid angle (sr). Luminance is the amount of visible *light* leaving a point on a surface in a given direction.

In this study, we focus on computer monitors. Thus, we utilize relative luminance. Relative luminance follows the definition of luminance, but the value is normalized to 1 for a reference white [14].

The relative luminance could be calculated from RGB color spaces. For RGB color spaces that use the ITU-R BT.709 primaries (or sRGB, which defines the same primaries), relative luminance can be calculated from linear RGB components [13]:

$$Y = 0.2126 R + 0.7152 G + 0.0722 B$$

The formula reflects the luminosity function: green light contributes the most to the intensity perceived by humans, and blue light the least.

In addition, gamma correction [11] [12] is a nonlinear operation used to code/decode luminance in image system.

$$\text{Output luminance} = 255 * (\text{RGB}/255) ^{\text{gamma}}$$

Gamma value is used to quantify contrast. In the case of gamma equal 1, the cure is totally linear. When gamma value is greater then 1, it's usually called gamma expansion and is the application of the expansive power-law nonlinearity. The gamma value less than 1 is called gamma compression and vise verse.

Furthermore, [15] shows that information processing time takes huge role in visual fatigue effects. It makes the reaction time and subjective rating of visual fatigues vary as well as

working period.

### 2.3 Web Accessibility

Since the rapid growing web become popular as time goes by, the main information communication shifted from the traditional television to the Internet. The longer usage time, the more exposed time we spend and rich media act as role as television.

The World Wide Web become rich-providing information source since people sharing and interaction activities is getting more and more. However, the usability isn't equally increasing for people with disability to access the web effectively. There are three main categories to indicate: authoring tools, browser and context [5].

1. Context is designed to be accessible.
2. Browser and multimedia player provide usable and accessible experience.
3. Authoring tool include content management system (CMS) such as blogging system and other tool to generate web content



Figure 2-3 Components of Web Accessibility

Figure 2-3 illustrates the component of web accessibility. Developer is referred to be the content producer using authoring tools such as CMS. User is the information consumer as the end point of data flow using browser media players. It's clear that content is the middle product

and is required whole componenets are accesiable.

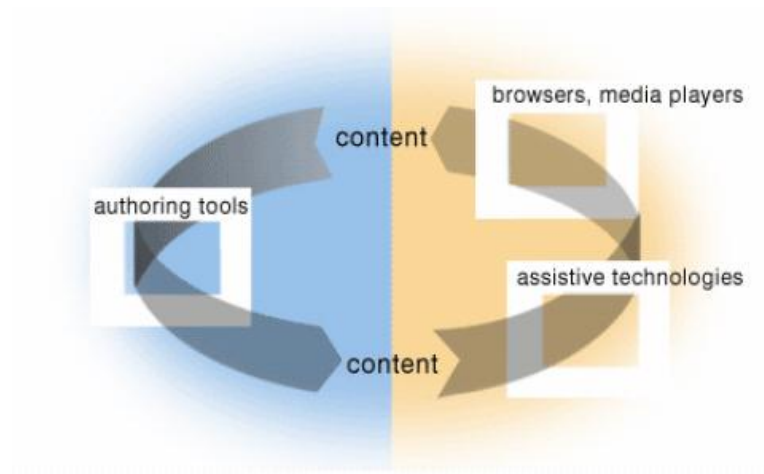


Figure 2-4 Cycle of Accessibility Implementation

Figure 2-4 shows us how to implement accessibility. It's a cycle of chicken and egg. New feature implementation by authoring tools needs support by browser and media player. In the other hand, browser and media player wait for a critical mass of authoring tools to use a new feature to get accessibility better.

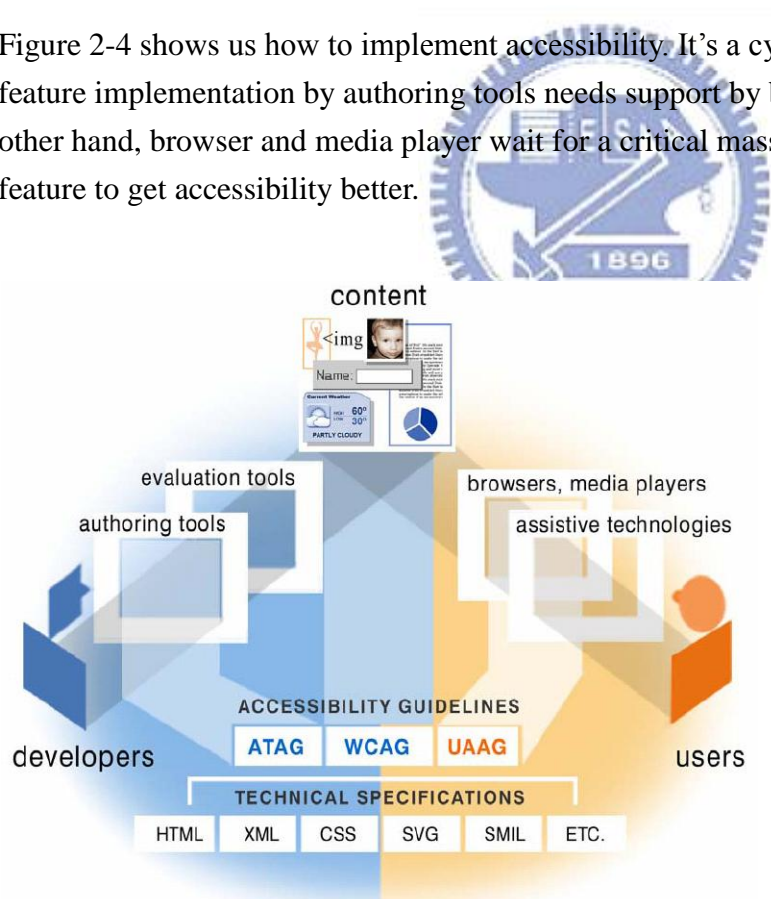


Figure 2-5 Components and Guidelines for Web Accessibility

Figure 2-5 shows the entire components of web accessibility. These components are bringing together by the web accessibility initiative (WAI) of World Wide Web consortium (W3C). W3C is helping coordinate international web accessibility effort. WAI work with three guidelines for implementation of W3C specification.

1. Authoring Tool Accessibility Guidelines (ATAG),
2. User Agent Accessibility Guidelines (UAAG);
3. Web Content Accessibility Guidelines (WCAG);

ATAG 1.0 was published in February 2000 and ATAG 2.0 Working Draft was released in 21 May 2009 [16]. ATAG is primitively written for authoring tools to produce accessible content via an accessible authoring interface. Intending to meet the need of many different audiences, ATAG and supporting documents is very important for people who want to choose more accessible authoring tools and people who want their authoring tools to improve accessibility in the future version.

UAAG 1.0 was published in February 2000 and UAAG working Draft was released in 11 March 2009 [17]. UAAG explain how to make user agent such as browser and media player more accessible to people with disability. UAAG is primitively for developer of web browser, media player and other user agent used by user. UAAG and supporting documents intend to meet the need of people who want to choose user agent more accessible and people who want their use agent improve accessibility.

WCAG 1.0 was published in May 1999 [18]. WCAG explain how to make web content more accessible to people with disability. It's not only targeted primitively for content producer but also used by authoring tools developer, user agent developer and evaluation tools developer. Authoring tools developer use it to create tools that generate accessible content, user agent developer use it to create tool s utilize accessible content and evaluation tools developer use it to create tools discovering accessibility issue of the content.

Since WCAG 1.0 was published, many organizations adopted WCAG 1.0 in their policy, and these feedbacks from, developer or policy maker form the basis of WCAG 2.0 that was released in 11 December 2008. WCAG 2.0 Is different from WCAG 1.0 by introducing a new concept called priority scheme. Priority scheme define testable success criteria, remove technology-specific information and provide testable information in checklist. In additional to

the testable information, it also provides an overview of materials for user to orient the suite of documentation.

Here list flicker related guideline referred in WCAG 1.0 and WCAG 2.0.

<p>WCAG 1.0</p>	<p>Checkpoints: 7.1 Until user agents allow users to control flickering, avoid causing the screen to flicker. [Priority 1]</p> <p>Note. People with photosensitive epilepsy can have seizures triggered by flickering or flashing in the 4 to 59 flashes per second (Hertz) range with a peak sensitivity at 20 flashes per second as well as quick changes from dark to light (like strobe lights).</p>
<p>WCAG 2.0</p>	<p>Guideline 2.3 Seizures: Do not design content in a way that is known to cause seizures.</p> <p>2.3.1 Three Flashes or Below Threshold: Web pages do not contain anything that flashes more than three times in any one second period, or the flash is below the general flash and red flash thresholds. (Level A)</p> <p>Note: Since any content that does not meet this success criterion can interfere with a user's ability to use the whole page, all content on the Web page (whether it is used to meet other success criteria or not) must meet this success criterion. See Conformance Requirement 5: Non-Interference.</p> <p>2.3.2 Three Flashes: Web pages do not contain anything that flashes more than three times in any one second period. (Level AAA)</p>

The detail descriptions in guideline 2.3.1 of WCAG2.0 have defined some keyword such like relative luminance, blinking, flash and red flash. Relative luminance is the relative brightness of any point in a color space and usually normalized to 0 for darkest black and 1 for lightest white. Blinking is a visual effort switch back and forth between two visual states in a way that is meant to draw attention. Flash is a pair of opposing changes in relative luminance that can cause seizures in some people. What make blinking differ from flash is the volume and



frequency. Flash is more dangerous to cause seizures if it is large enough and in the right frequency range.

## 2.4 Evaluation Tools

Tools for detect flicker are focusing on the video signal processing. One of the old measuring tool is Harding Flash and Pattern analyzer, it convert each pixel into luminance value and compute the difference between adjacent frames and use Spectrum analyzers to detect the contrast changing frequency.

Another is Photosensitive Epilepsy Analysis Tool (PEAT) offered by the TRACE center [7]. PEAT is free, downloadable and designed to reduce risk of Seizures. Figure 2-6 is the screenshot of PEAT tool.

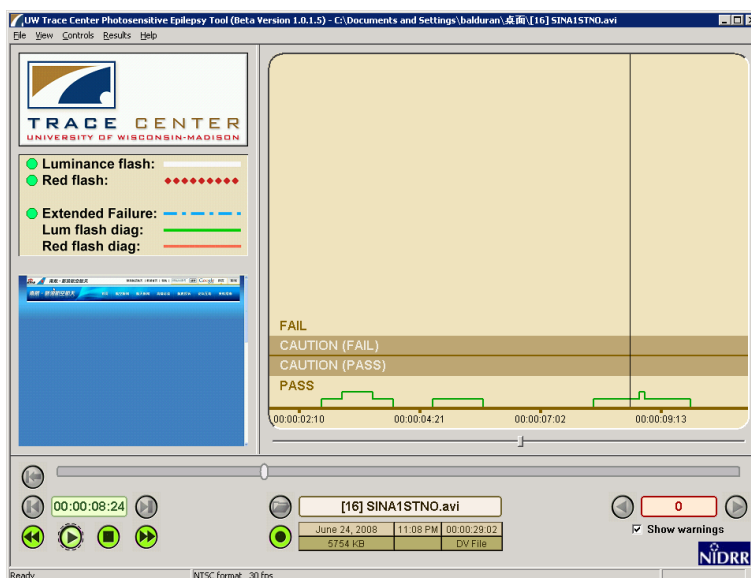


Figure 2-6 Screenshot of PEAT

The other way is WWW text parsing. When the browser is loading HTML, it parses animated GIFs, embedded object and third party plug-in. analyzing animated gifs is easier and combined with advertisement block. The flashing gifs is bothersome and divert user. Lots' of plug-in such as Adblock plus have been developed.

However, embedded object, i.e., Flash, is more difficult to detect visual effect of context than texts and animated Gifs. Therefore, in the study, we adapt video capture approach.

## Chapter 3 Methodology

A two-phase study was designed to explore the identification and classification of visual comfort and flicker factor. Figure 3-1 is the architecture we design to verify the relation between visual comfort degree and the video-processing module. Our method designed to evaluate and measure visual comfort is focused on flicker making viewer feel uncomfortable. There are two major portions: Matlab video processing code for signal processing and questionnaire result analyzing for people feeling.

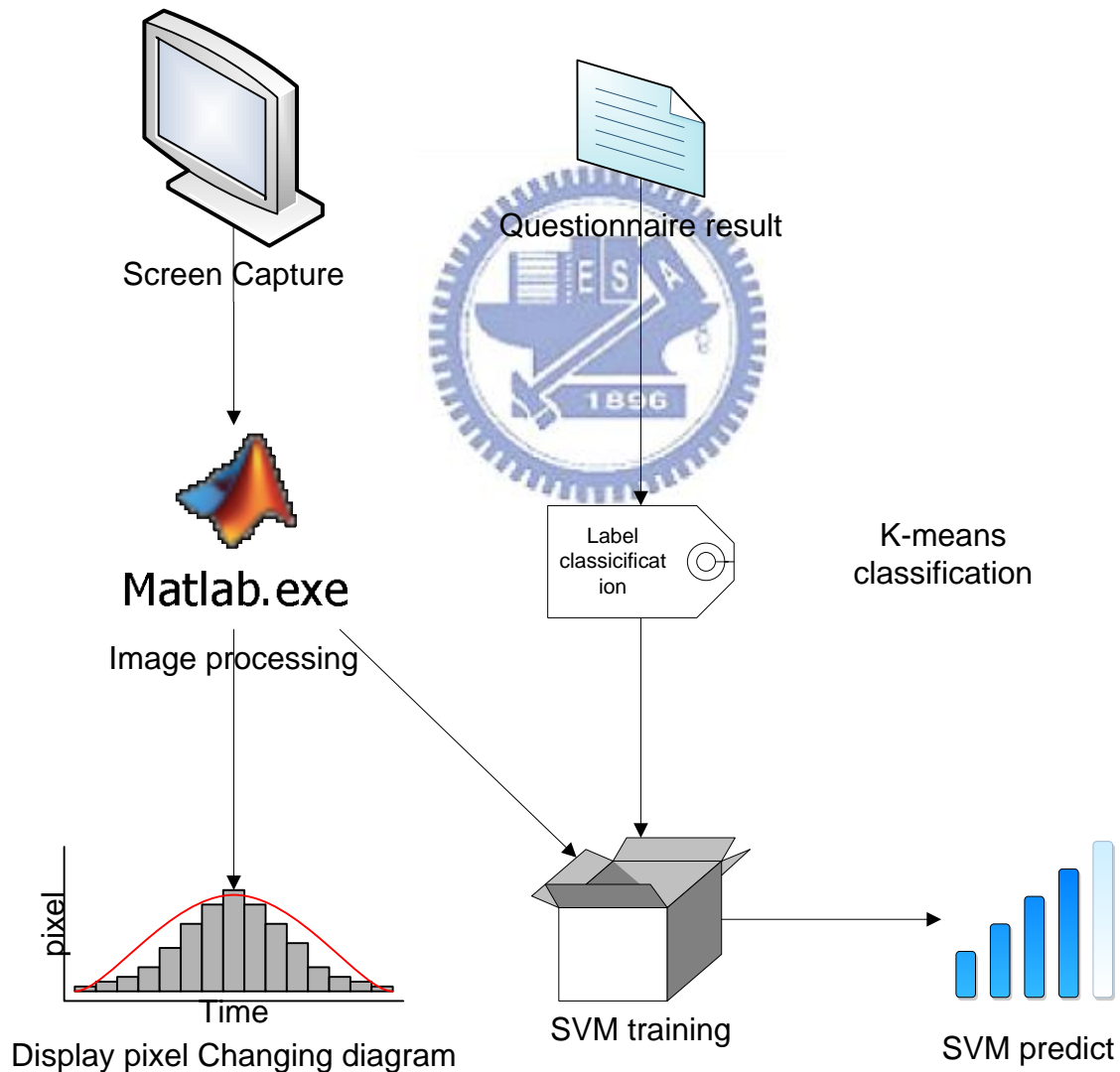


Figure 3-1 Architecture of questionnaire, screen capturing, image processing, classification

and SVM training model

Here is the process we are going to construct the evaluation tools. First we capture a web-surfing video clip to record a website visual performance. The follow steps is human rating steps, collecting what people feel comfortable or not while browsing this website. We have questionnaires for student to answer. In another hand, according to WCAG guideline 2.0, we use image processing approach to processing each frame of this video clip. We will get a visual graph showing the flash occur or not and corresponding analyzed data set. In order to make our evaluation tool to predict human rating, we choose SVM for classification. This corresponding analyzed data set will be the feature and result of questionnaire will be the label to train SVM model. SVM model is used to predict the rating of next incoming video clip and will be embedded in our evaluation tool.

### 3.1 Experiment Data Preparation

In order to measure website visual comfort, we choose top 30 Alexa's traffic ranking website for candidate. Using screen capture software to record 30s' video clip and choose average flicker rating clips for questionnaire. The video clips contain normal Internet surfing progress using Microsoft IE. Each website we record screencast three times in different from cached or not. All these website are grouped into 23 individual websites since some of them belongs to the same company.



The hardware equipment list below:

SW/HW	Specification
OS	Windows XP SP3
CPU	Intel Core 2 6300 @ 1.86GHz
Browser	Firefox 3.0.3

Screen-Casting Tool	AutoScreen Recorder Pro 3.0
Frame rate	30 frames /second

Table 3-1 Hardware and software environment we record a web surfing video clip

We replay the 23 captured video clips to 42 student majored in computer science and 46 student majored in non-computer science. After playing one clips, we ask these viewer to rate from 1 to 5 and repeated until all is done. Outliner of those rating scores is removed and left over 81 copies is available. These rating scores is used to label these web site good or not.

### 3.2 Video Processing Model

Figure 3-2 shows the process we transform color space into luminance space.

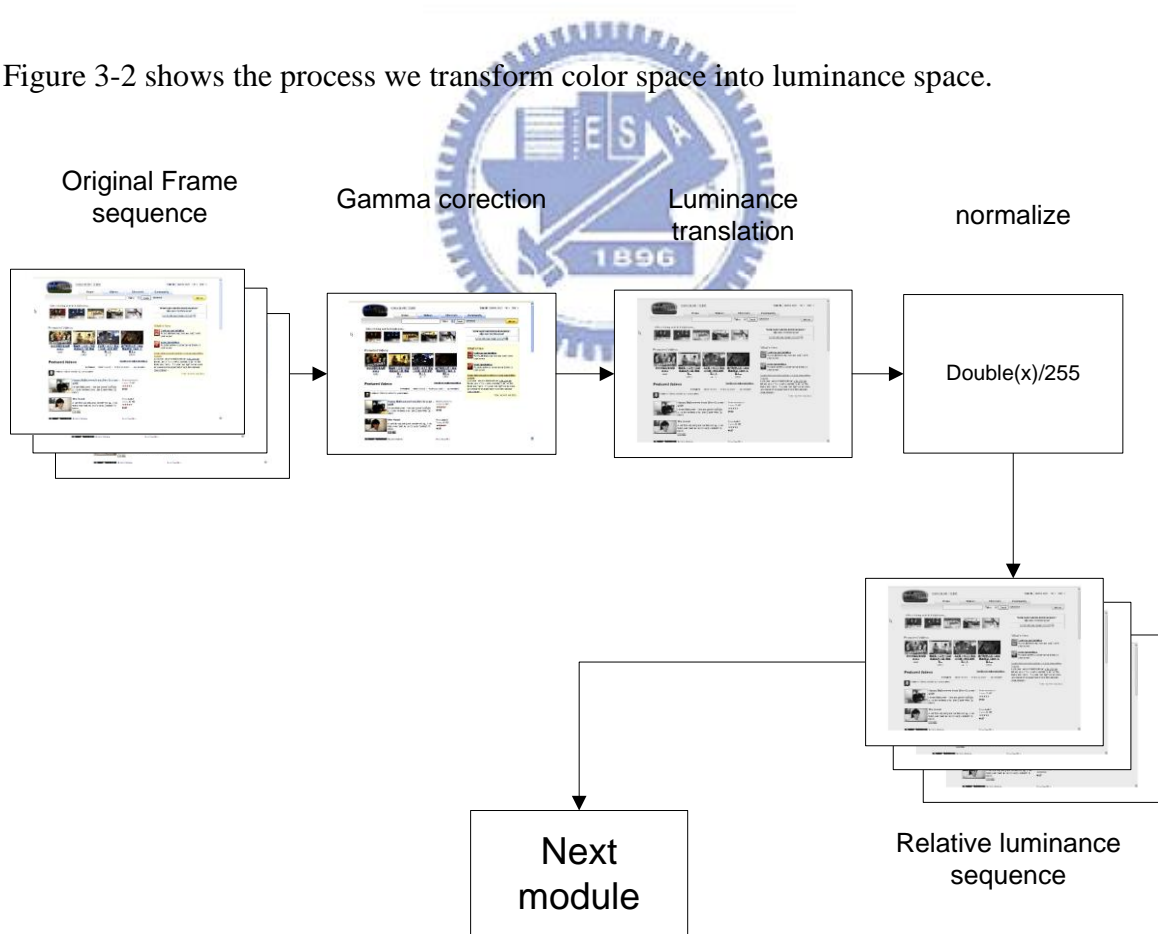


Figure 3-2 Luminance translation from sRGB color space into relative luminance domain

A video clips is a sequence of frames. Flicker is relative luminance changing ratio along time. First, we process each frame using gamma correction. Second, we translate each frame from standard RGB color spaces into YUV color space. Since we focus on luminance information, we drop color data but the luminance value. The last step is normalizing luminance value into 0 to 1. Now whole screen is grayscale image and each pixel is stored in double type.

In the following step, we process these luminance images with discrete differentiation as shown in Figure 3-3.

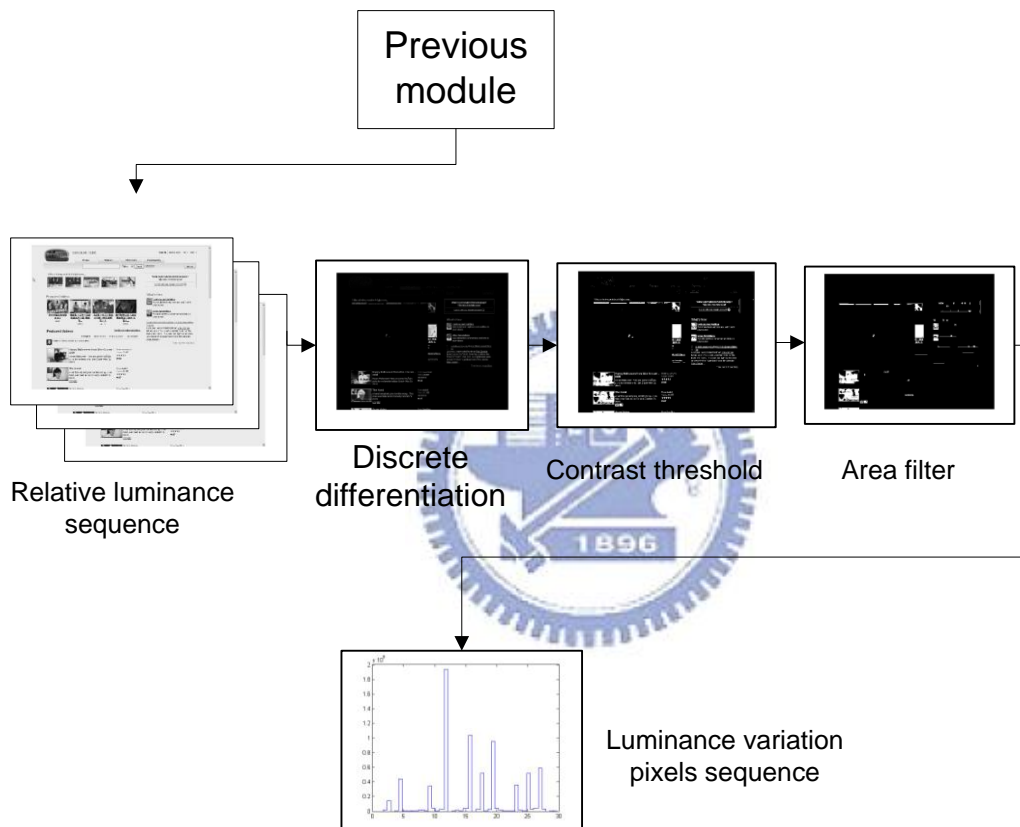


Figure 3-3 Extraction process to calculus pixel variation in relative luminance domain

These sequences of grayscale images are the raw data we are going to calculate. W3C WCAG 2.0 indicated A general flash is defined as a pair of opposing changes in relative luminance of 10% or more of the maximum relative luminance where the relative luminance of the darker image is below 0.80; and where "a pair of opposing changes" is an increase followed by a decrease, or a decrease followed by an increase, therefore we process these image sequence differentiation in adjacent frames first. Filtering out these pixel vary in relative luminance exceed 10 %. That's an mask hiding imperceptibility and the mask will be processed in the

area connectedness. Since the area of visual effect our eyes receives is larger to be noticed. we will produce an mask without pepper noise.

### 3.3 K-means clustering and Support vector machine learning

We will introduce two classification method employed in our experiments. One is k-means clustering and the other is support vector machine.

#### 3.3.1 K-means clustering

K-means clustering is a method of clustering analysis that help to partition n observations into k cluster. Each observations belong to the same cluster is with the nearest mean. K-means group the data by minimizing the sum of square of distance between the data and the corresponding centroids. This algorithm is a standard and popular algorithm for unsupervised learning of classification.

K-means use two-phase iterative algorithm to minimizing the sum of square of point-to-centroid distance. First step is assignment step and the following step is update step. Assignment step labels each observation to the cluster with the closest mean. This step partition these observations into several cluster temporarily. Update step will calculate the new mean to be the new centroid of the observation in each cluster. This algorithm will deem to converge when assignment step make no longer change.

In statistics toolbox of Matlab there is a function and we prepare our raw data fitting the input matrix format [9].

### 3.3.2 Support vector machine

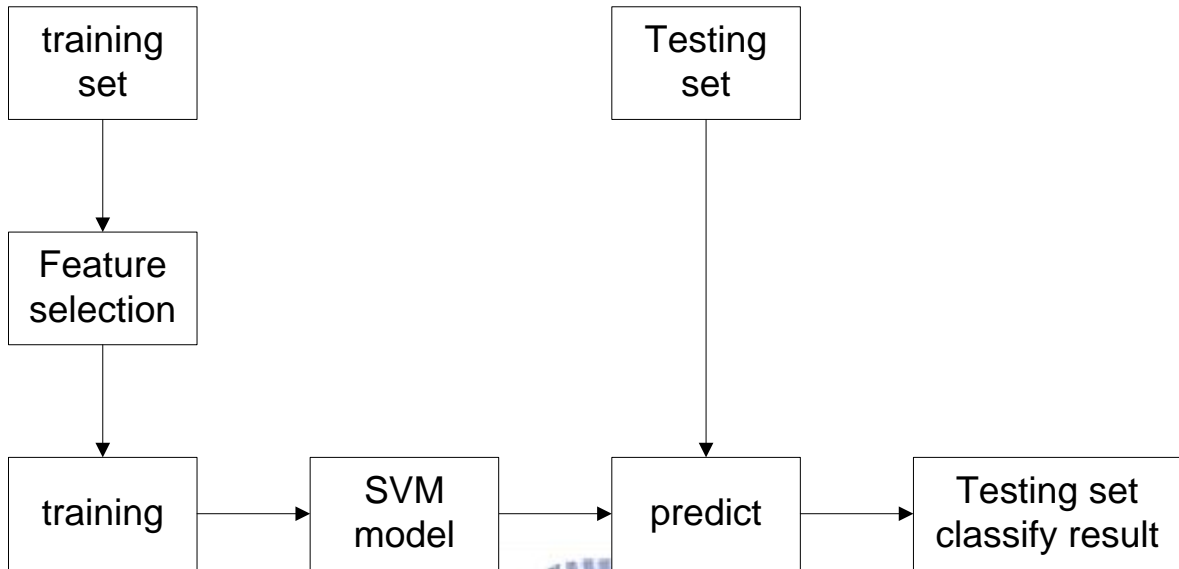


Figure 3-4 Architecture of support vector machine training and predict model

Support Vector Machine is a supervised machine learning method used for classification. Input data of SVM are two sets of vector data in high dimension. SVM will construct a Hyperplane between these two set. Hyperplane, which maximize the margins of these two set data, is used to predict which set that next income data belong to. Without discussing what model is better for our situation, we take SVM as a black box to use. Figure 3-4 show the architecture we use support vector machine in our study.

LIBSVM developed by Chih-Jen Lin in National Taiwan University [10] is a SVM library that is easy to use. It provides simple interface that use can link it with our program. This library includes source code in C++, Java and etc. we develop evaluation tool in Matlab and these is interface for Matlab.

Feature selection is very important portion in classification. In previous research done by Ming-Yu Wei [19], the average of transition count is meaningful. We chose the summation of transition which ideal as exposure. Further the numbers of peak of the signal of luminance difference maybe stand for the changes.

How to classify these website based on questionnaire rating is hard to decide. We can cut up these 23 website into two partitions, one is good and the other is bad. Moreover there are three partition, first is good, second is bad and the rest of them is neutral. In the two partitions, the negative rating is treated as a special index to indicate viewer bad feelings. Once one of these questionnaire are rated as 1 (uncomfortable), we take this website is bad. Regardless of these negative rating in questionnaire, it may be dispersed in the statistical process. A way to divide these websites into two groups is using mean value or median value for partition. The higher rating is good and vice versa, otherwise is bad.

### 3.4 Evaluation threshold

In WCAG 2.0 guideline 2.3.1 it have noted below formula.

Note 1: For the sRGB colorspace, the relative luminance of a color is defined as  $L = 0.2126 * R + 0.7152 * G + 0.0722 * B$  where R, G and B are defined as:

- if  $R_{sRGB} \leq 0.03928$  then  $R = R_{sRGB}/12.92$  else  $R = ((R_{sRGB}+0.055)/1.055)^{2.4}$
- if  $G_{sRGB} \leq 0.03928$  then  $G = G_{sRGB}/12.92$  else  $G = ((G_{sRGB}+0.055)/1.055)^{2.4}$
- if  $B_{sRGB} \leq 0.03928$  then  $B = B_{sRGB}/12.92$  else  $B = ((B_{sRGB}+0.055)/1.055)^{2.4}$

and  $R_{sRGB}$ ,  $G_{sRGB}$ , and  $B_{sRGB}$  are defined as:

- $R_{sRGB} = R_{8bit}/255$
- $G_{sRGB} = G_{8bit}/255$
- $B_{sRGB} = B_{8bit}/255$

The "^" character is the exponentiation operator [18].

Note 2: Almost all systems used today to view Web content assume sRGB encoding. Unless it is known that another color space will be used to process and display the content, authors



should evaluate using sRGB color space. If using other color spaces, see [8].

Note 3: If dithering occurs after delivery, then the source color value is used. For colors that are dithered at the source, the average values of the colors that are dithered should be used (average R, average G, and average B).



## Chapter 4 Experimental Results

In this chapter, our work will be presented. We will describe detail result of classification of questionnaire rating in section 4.1. The SVM training and predict result is include in sec 4.2 and final we have a evaluation tool prototype providing some mechanic for web developer to evaluation the accessibility.

### 4.1 Questionnaire results

We have total 81 available questionnaire results that don't include improper voting. All these rating is used to classify which website stand for good and which one stand for bad in visual comfort. We draw these websites rating questionnaire result as pie charts. Figure 4-1 is the pie charts show us the percentage of each rating scores.

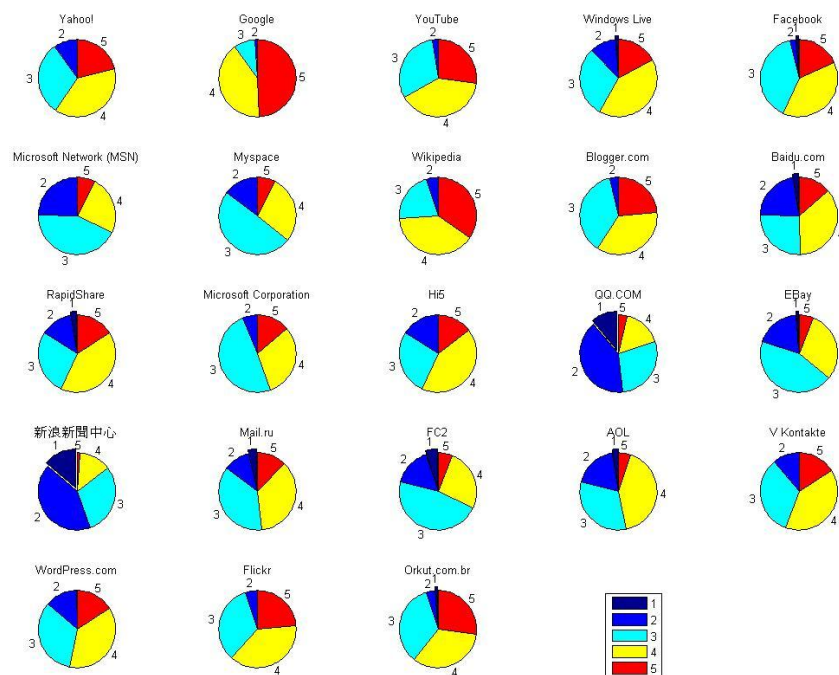


Figure 4-1 Pie chart for all questionnaire results

Figure 4-1 show high rating scores as red and low rating scores as blue. It's clear that QQ.com and sina.com get a lot of rating scores of 1 and a little of rating scores of 5. This result tells us

that both websites of QQ.com and sina.com give terrible performance. In the other hand, google.com and Wikipedia get a lot of rating score of 5 and almost get no rating score of 1. We can say that these two website have good visual comfort performance. As for the other websites, we can't give any conclusion.

Here we list detailed rating scores in Table 4.1.

	rating 1	rating 2	rating 3	rating 4	rating 5	mode
<b>Yahoo!</b>	0	8	25	32	18	4
<b>Google</b>	0	1	8	33	41	5
<b>YouTube</b>	0	2	25	34	22	4
<b>Windows Live</b>	1	9	24	34	15	4
<b>Facebook</b>	1	2	33	31	16	3
<b>Microsoft Network (MSN)</b>	0	20	36	21	6	3
<b>Myspace</b>	0	12	40	24	7	3
<b>Wikipedia</b>	0	4	17	32	30	4
<b>Blogger.com</b>	0	3	30	31	19	4
<b>Baidu.com</b>	2	18	21	30	12	4
<b>RapidShare</b>	2	11	22	34	14	4
<b>Microsoft Corporation</b>	0	5	40	26	12	3
<b>Hi5</b>	0	13	24	34	12	4
<b>QQ.COM</b>	9	34	24	13	3	2
<b>EBay</b>	1	15	38	24	5	3
<b>Sina.com</b>	11	34	26	11	1	2
<b>Mail.ru</b>	3	10	30	30	10	4
<b>FC2</b>	4	13	39	22	5	3
<b>AOL</b>	2	15	26	36	4	4
<b>V Kontakte</b>	0	9	27	33	14	4
<b>WordPress.com</b>	0	11	27	31	14	4
<b>Flickr</b>	0	4	29	31	19	4
<b>Orkut.com.br</b>	1	3	30	27	22	3

Table 4-1 Detailed questionnaire rating scores table

With two different groups of students, we can compare the rating trend between computer science major and non-computer science major. Figure 4-2 show mode of their rating in one

chart. As we can see in this figure, the difference between CS students and non-CS student is no larger than one. People give similar rating for these websites in term of visual comfort and we won't process these two groups of rating individually.

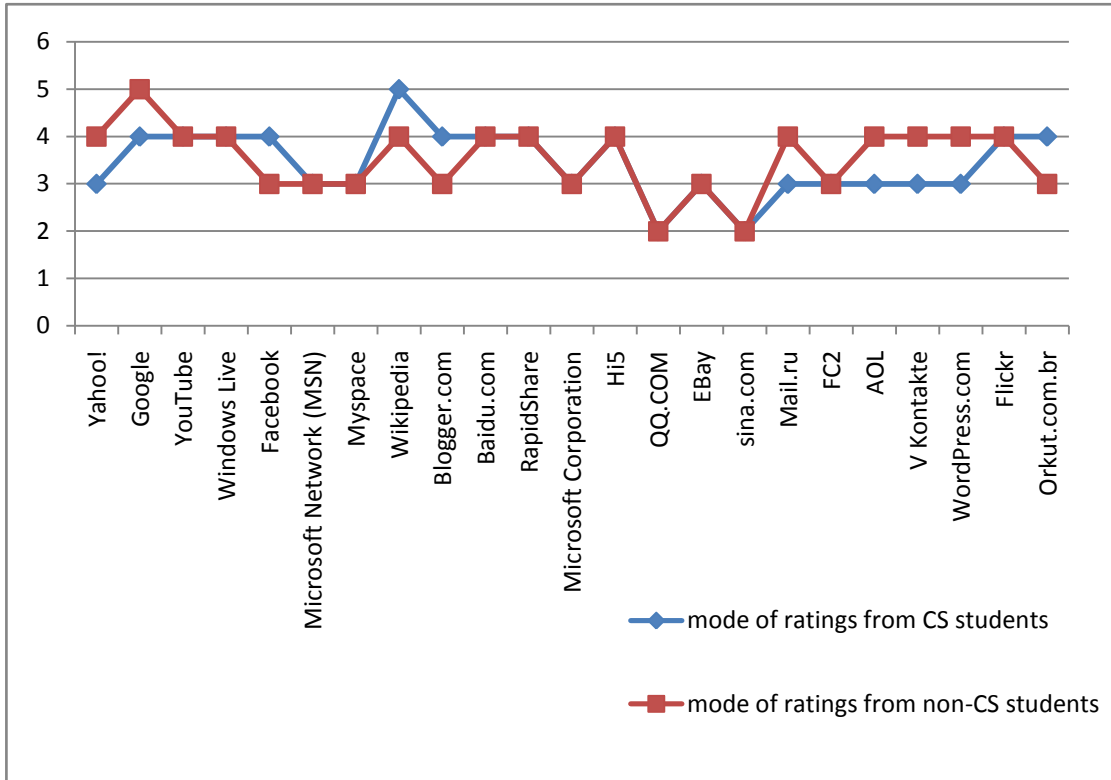


Figure 4-2 Difference of mode of rating between CS students and non-CS student

There are several way to explain those rating result and we try to classify these web site into two group based on human feeling. We select k-means clustering to partition 23 websites into two clusters. This concept is simple and clear, using unsupervised learning to data set and analyzing result come from clustering algorithm. Table 4.2 shows the result of k-mean clustering.

Websites	k-means clustering
Yahoo!	1
Google	1
YouTube	1
Windows Live	1
Facebook	1
Microsoft Network (MSN)	-1

<b>Myspace</b>	1
<b>Wikipedia</b>	1
<b>Blogger.com</b>	1
<b>Baidu.com</b>	1
<b>RapidShare</b>	1
<b>Microsoft Corporation</b>	1
<b>Hi5</b>	1
<b>QQ.COM</b>	-1
<b>EBay</b>	-1
<b>Sina.com</b>	-1
<b>Mail.ru</b>	1
<b>FC2</b>	-1
<b>AOL</b>	-1
<b>V Kontakte</b>	1
<b>WordPress.com</b>	1
<b>Flickr</b>	1
<b>Orkut.com.br</b>	1

Table 4-2 result of partition label of rating scores clustered by k-means clustering

One group of these rating scores get higher centroid rating location about 3.5 and the other one get lower value about 2.5. Therefore, we succeed to partition two groups shown as Table 4-2. The size of positive rating group is seventeen and the size of opposition group is sixth. You could check that this classification result is similar to result come from Ming-Yu Wei [19].

#### 4.2 SVM arguments decision

As we referred SVM above, there are so many undecided parameters such as cost and gamma, the most important arguments for SVM training. These arguments affect the accuracy after predict following corresponding SVM training process. Mostly people use n-fold cross-validation to decide these arguments. N-fold Cross-validation is the process to spilt original data into several groups. Use one group to validate the accuracy and the other groups is used for model training. For example, 5-fold cross-validation partition original dataset into 5 groups, use 4 groups to train correspond model and the final one group is used to predict. After all group have been the validation group, best accuracy will be thought as accuracy with

arguments are corresponding.

In our study we choose cross-validation argument as 5 that is usually set to. We try any combination of cost and gamma and leave the best accuracy as the arguments we will use in our training-predict model. Setting cost and gamma value from  $2^{-10}$  to  $2^{10}$ , it's easy for loop computation to try the best cost and gamma.

The arguments in our model are setting cost as 32 and gamma as 2. The corresponding accuracy after cross-validation is 95% that is very good in performance.

Since we use k-means as classifier to process questionnaire rating, we may thought why we don't partition feature vector into two groups. But the result is terrible to establish an trustable model for prediction. Table 4-3 lists the result that each website is clustered into which partition labeled positive or negative.

Websites	Rating	Statistical value
Yahoo!	1	1
Google	1	1
YouTube	1	1
Windows Live	1	1
Facebook	1	1
Microsoft Network (MSN)	-1	-1
Myspace	1	1
Wikipedia	1	1
Blogger.com	1	-1
Baidu.com	1	1
RapidShare	1	1
Microsoft Corporation	1	-1
Hi5	1	1
QQ.COM	-1	1
EBay	-1	1
sina.com	-1	-1
Mail.ru	1	1
FC2	-1	1

AOL	-1	-1
V Kontakte	1	1
WordPress.com	1	1
Flickr	1	1
Orkut.com.br	1	1

Table 4-3 Positive and negative classification using questionnaire rating and statistical value from luminance variable pixel sequence

The positive or negative label that we use questionnaire rating to decide is similar to previous work done by Ming-Yu Wei [19], but the label decided by statistical value classified by k-means is different. For example, this result shows us QQ.com is positive but actually user rating is negative. Therefore we don't use k-means to classify these website. Using SVM, which is supervised machine learning, is better way in comparison.

#### 4.3 Evaluation tool prototype

Figure 4-3 is a sample chart to present the result of our Matlab image processing code. As shown in the Figure 4-3, we illustrate the screen luminance various pixel sequence as stairs chart. The red line presents the summation of pixel in increasing relative luminance with corresponding timestamp and the blue line presents these in decreasing relative luminance.

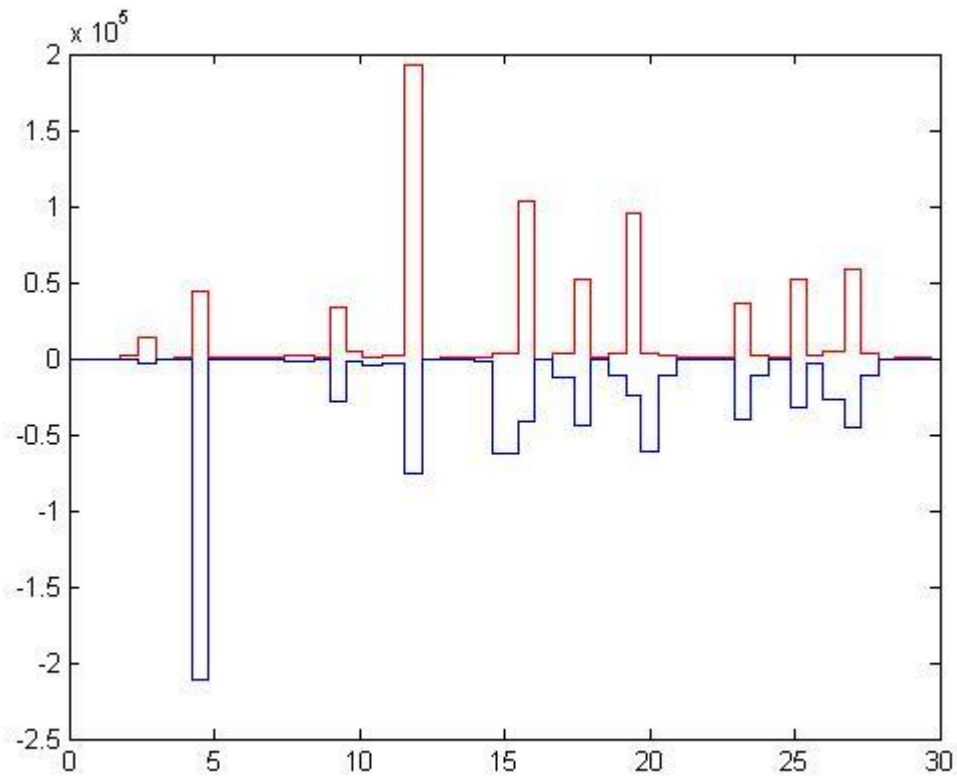


Figure 4-3 Sample stairs chart of luminance variation pixel sequence

Recently we contribute an evaluation tool prototype for website developer focusing in web accessibility. This evaluation tool has two major features: processing input video clip and rating based on SVM predict.

Figure 4-4 is the screenshot of my evaluation tool prototype. This GUI program will load a video clips as input and output the image processing result as Figure 4-3. We will explain the usage as detail as we can in the following pages.



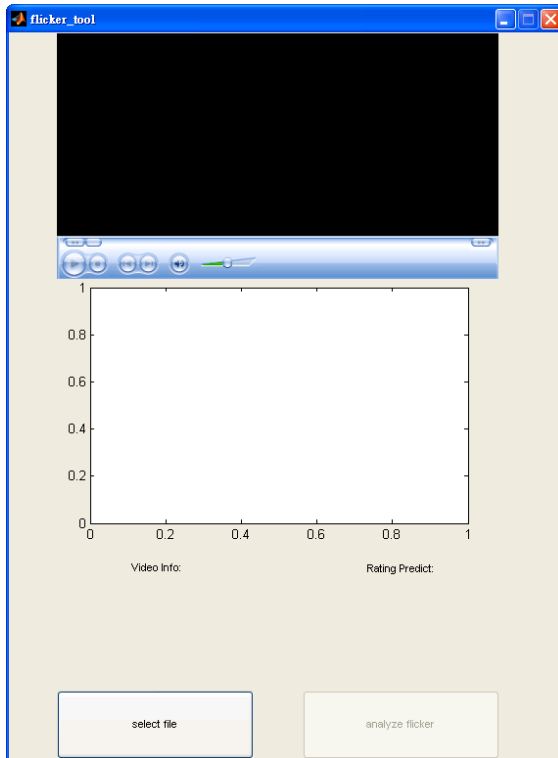


Figure 4-4 Screenshot of evaluation tool prototype initiation

Figure 4-4 is the screenshot of our evaluation tool prototype initiation. The user interface layout is divided into three areas. In the upper area of this user interface there is one media player to play the selected video clip. The middle area of user interface shows the stairs chart of luminance variation pixel sequence and additional information about selected video. There are two main buttons to operate in the bottom of graphic user interface: one is used for loading a video clip and the other is the button to process the selected video clip.

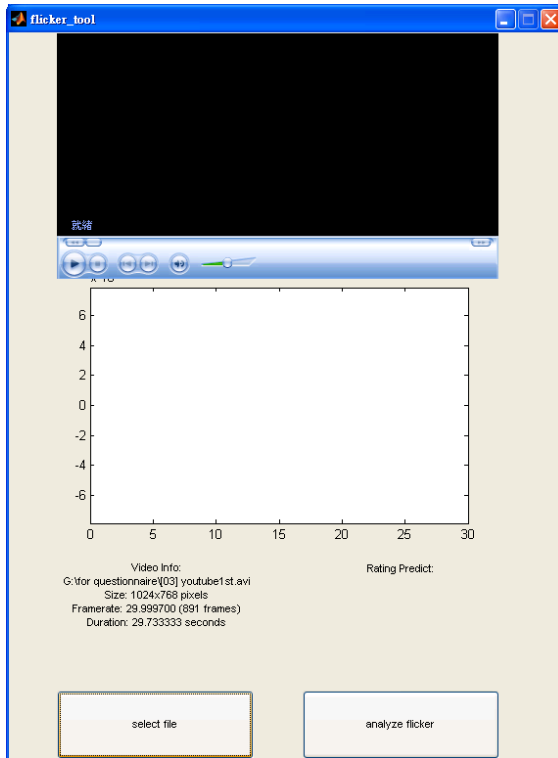


Figure 4-5 Screenshot of evaluation tool prototype after loading a input video

As long as we click the button labeled “select file”, the file select dialog show up. After selecting one video file, the middle area will display detail metadata of selected video file shown in Figure 4-5.

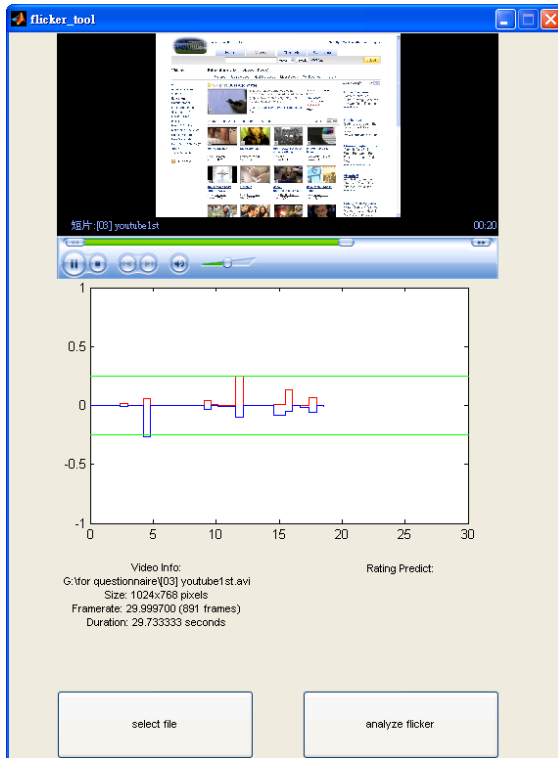


Figure 4-6 Screenshot of evaluation tool prototype processing and playing input video file

Next step is to process the selected video file. Clicking another button labeled “analyze flicker“ will start image processing. In the same time, upper media player will play selected video. Figure 4-6 show the screenshot as demo.

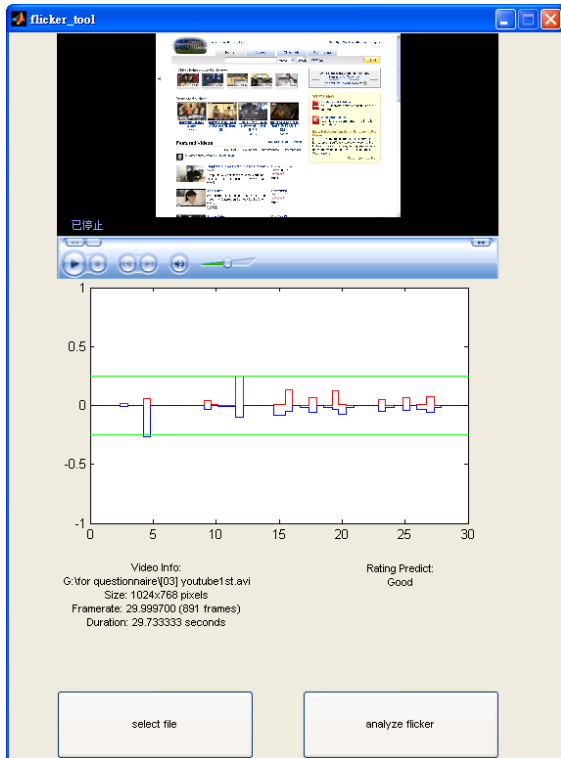
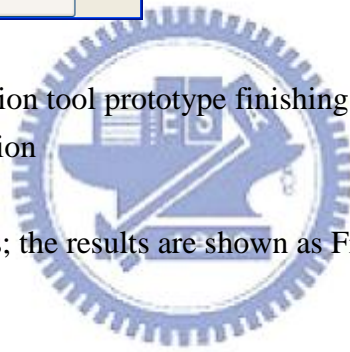


Figure 4-7 Screenshot of evaluation tool prototype finishing a detect session and the corresponding predict classification

The image processing terminates; the results are shown as Figure 4-7.



## Chapter 5 Conclusions

### 5.1 Conclusions

In conclusion, the relation between visual comfort and flicker is highly dependent. We have show that the correlation between statistical values of signal sequences generated by Matlab analyzer code and result of user rating is -0.5 in previous work. This result presents us the visual comfort felt by people get large negative correlation with image processing result. One possible explanation is that the more screen luminance variation pixel, the more uncomfortable feeling and the less user rating.

The present study enhances current visual analyze tool by providing a much more active evaluation model of visual comfort. Our study present a predict user rating and corresponding model which is constructed with real human factor. The result of cross validation lead support to extend that statistical value of result of image processing are related with result of questionnaire processed with k-mean classification.

In addition, it is important to emphasize that the original dataset of measured website in this study limit our sample space of experience. Recording each surfing behavior is impossible for numerous websites. We readily acknowledged that our study is exploratory and there are problems with the questionnaire design. Another problem that often affect user rating is about personal preference or brand, which is hard to measure.

### 5.2 Future Work

We have divided all these capture website into two partitions, one is positive rating and the other is negative rating. Trying to make more detail analyze of human factor and the signal processing, grouping all these capture website into five groups to represent people rating is more precise. Another way to improve is analyze these luminance changing signals using Spectrum analyze. After processing luminance information, the processed signal may be vary in frequency domain.

## References

- [1] J. E. Farrell, Brian L. Benson, and Carl R. Haynie, “Predicting Flicker Thresholds For Video Display Terminals”, 1987 Proceedings of the Society for Information Display, SID, Vo1.28/4, 1987.
- [2] SHERYL WuDUNN “TV Cartoon's Flashes Send 700 Japanese Into Seizures” Thursday, December 18, 1997
- [3] Vision Science II - Monocular Sensory Aspects of Vision Lecture 23 – Temporal Vision Phenomena <http://arapaho.nsuok.edu/~salmonto/vs2.html>
- [4] Luminance *Lighting Design Glossary*:  
<http://www.schorsch.com/kbase/glossary/luminance.html>. Retrieved on Apr. 13, 2009.
- [5] Chisholm, W. and S. Henry, “Interdependent components of Web accessibility”, ACM New York, NY, USA 2005
- [6] Web Content Accessibility Guidelines (WCAG) 2.0 Available at <http://www.w3.org/TR/WCAG20/>
- [7] Photosensitive Epilepsy Analysis Tool (PEAT) Available at <http://trace.wisc.edu/peat/>
- [8]  
<http://www.w3.org/TR/UNDERSTANDING-WCAG20/visual-audio-contrast-contrast.html> as of 07/15/2009
- [9] Seber, G. A. F., *Multivariate Observations*, Wiley, New York, 1984
- [10] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [11] <http://www.digital.idv.tw/DIGITAL/Classroom/MROH-CLASS/oh109/wpyqg3mn.gif>
- [12] <http://graphics.stanford.edu/gamma.html>

[13] <http://www.w3.org/Graphics/Color/sRGB.html>

[14] [http://en.wikipedia.org/wiki/Luminance\\_\(relative\)](http://en.wikipedia.org/wiki/Luminance_(relative)) as of 07/15/2009

[15] 林彥輝、林孟希、謝孟穎，動態資訊處理作業之視覺疲勞衡量，第十三屆中華民國人因工程學會年會暨學術研討會，2006, 4 March

[16] <http://www.w3.org/TR/WAI-AUTOOLS/> as of 07/15/2009

[17] <http://www.w3.org/TR/WAI-USERAGENT/> as of 07/15/2009

[18] <http://www.w3.org/TR/WCAG10/> as of 07/15/2009

[19] Wei, Ming-Yu, College of Computer Science, National Chiao Tung University, MS Thesis: “Visual Comfort Diagnoses on Websites”, Dec. 2008

