Chapter 30
Identifying Colors of Products and Associated Personalized Recommendation Engine in e-Fashion Business

Keiichi Zempo and Ushio Sumita

Abstract One of the important factors ignored in the literature in e-marketing is "the color" of a product. While one may be able to identify the dominating color of a product based on the overall impression, it is not easy to mechanize the process to determine the dominating color. Accordingly, in many applications, the color of a product is defined subjectively by those who enter the data. Consequently, the color of a product has been a missing link in e-marketing. The purpose of this research is to fill this gap by developing an algorithmic procedure for identifying the dominating color of a product by analyzing a digital image of the product. The algorithmic procedure enables one to reveal color preferences of consumers by analyzing the digital images of the products obtained from the purchasing records. A recommendation engine is also developed based on color class preference vectors of individual consumers.

30.1 Introduction

In recent years, as pointed in [1], there has been an increasing interest in Web usage mining as a means to capture Web user behavioral patterns and to derive e-business intelligence. In [2–4], for example, automatic personalization was proposed based on clustering of user transaction and page-views. A prevalent alternative approach for building personalized recommendation engines would be collaborative filtering. Given a record of activity of a target user, the collaborative filtering approach compares that record with the historical records of other users so as to find the top users who have similar taste or interest. However, it is known that the collaborative filtering approach has some deficiency, see e.g. [5–8], and some optimization strategies have been proposed in [9–11] to overcome such shortcomings. More

K. Zempo (✉) • U. Sumita
Faculty of Engineering, Systems and Information, University of Tsukuba, Tsukuba, Ibaraki 305-8573 Japan
e-mail: zempo@iit.tsukuba.ac.jp

recently, a personal browsing assistant system is developed in [12], where the pre-fetched resources from the hyper-linked Web pages are compared so as to recommend which Web page should be requested next. As an application, there is recommendation engine specialized to fashions, e.g. [13, 14]. To the best knowledge of the authors, however, the color information has not been incorporated in the literature for developing better personalized recommendation engines.

The reason why colors of products has been ignored in e-marketing can be found in that a product typically involves many different colors. While one dominating color of a product may be identified in the eye of human based on the overall impression, it is difficult to mechanize the process for identifying the dominating color. Accordingly, in many applications, the color of a product is defined subjectively by those who enter the data. Furthermore, terms for describing a color are often quite vague and too many. Consequently, the color of a product has been a missing link in e-marketing. The purpose of this paper is to fill this gap by developing an algorithmic procedure for identifying the dominating color of a product by analyzing a digital image of the product. The algorithmic procedure enables one to reveal the color preference of a consumer by analyzing the digital images of the products purchased by the consumer. A recommendation engine is also developed based on color class preference vectors of individual consumers as shown in Fig 30.1.

Throughout the paper, vectors and matrices are indicated by underbar and doubleunderbar respectively, e.g. $\underline{\xi}$, $\underline{\underline{P}}(t)$, etc.
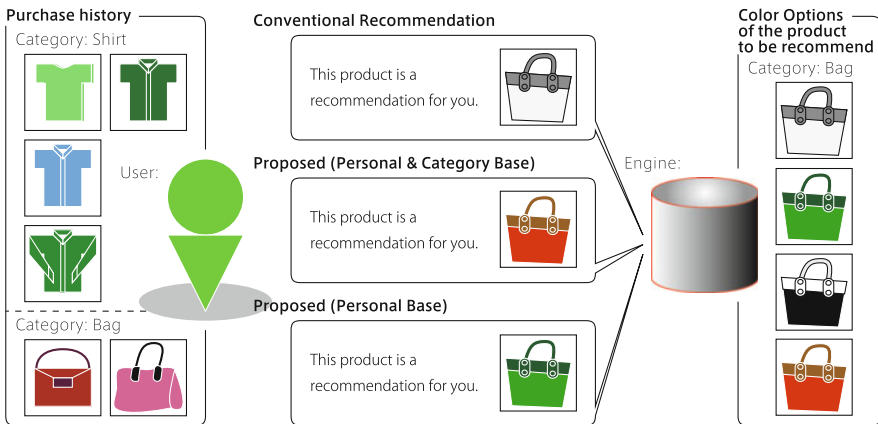

Fig. 30.1 Overview of the proposing personalized recommendation engine

30.2   Personalized Recommendation Engine Based on the Color of the Product

30.2.1   Development of Algorithm for Identifying the Dominating Color of a Product

A typical digital image of a product used in e-fashion business consists of a number of pixels, which would be too many to define the single dominating color of the product. In order to overcome this difficulty, we introduce $\underline{\Phi}(v_p) \in \mathscr{R}^6$, which we call a CCPV (Color-Class Profile Vector) of a digital image v containing a product p. In the eye of human, however, the Euclidean distance in RGB does not necessarily reflect the way humans differentiate different colors sensuously. Because of this reason, CIE (Commission International de l'Éclairage), the international commission on illumination, proposed the space denoted by CIE-L*a*b* in 1978. In CIE-L*a*b* space, RED, GREEN, YELLOW, BLUE, WHITE and BLACK are extremums of the axes as representative colors[15–17]. By defining the closeness of their representative six colors, we converted each pixel to facilitate by clustering. Based on this idea, we transform a set of pixels constituting a digital image of a product, denoted by v_p in CIE-L*a*b*, into the set of six dimensional vectors. By measuring the Euclidean distances between each of the transformed vectors and six fixed points in CIE-L*a*b* representing RED, GREEN, YELLOW, BLUE, WHITE and BLACK and then taking the average over the pixels in v_p, the sensuous color of the product in the eye of human is represented by a vector $\underline{\Phi}(v_p)$ in CIE-L*a*b*. $\underline{\Phi}(v_p)$ are calculated through the following steps.

Step 1:   **Extraction of the pixels of the product image from the background**
Every digital image obtained from the data has the background constructed by the unique pixel for representing "NON-COLOR". This pixel is different from the pixel corresponding to "WHITE" and never appears in digital images of products. Accordingly, the set of pixels exactly constituting the digital image of the product p can be extracted. The resulting set of pixels is denoted by v_p, and the number of pixels in v_p is written as N_{v_p}.

Step 2:   **Transformation of RGB vectors into CIE-L*a*b* vectors**
In Step 2, this transformation is conducted. Transformation $\mathscr{T}_1$ of the pixel $\underline{\gamma} = {}^t(\gamma_R, \gamma_G, \gamma_B) \in$ RGB into $\underline{\eta} = {}^t(\eta_L, \eta_a, \eta_b) \in$ CIE-L*a*b* is constructed in three stages. In the first stage, $\underline{\gamma}$ is mapped into an intermediate vector $\underline{X} = {}^t(X_1, X_2, X_3)$ via the liner transformation defined by,

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7151 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{pmatrix} \times \begin{pmatrix} \gamma_R \\ \gamma_G \\ \gamma_B \end{pmatrix}. \tag{30.1}$$

The second stage constructs $\underline{f} = {}^t(f_1, f_2, f_3)$ from $\underline{X} = {}^t(X_1, X_2, X_3)$ through the following definition. For $i = 1, 2, 3$, let f_i be defined by,

$$f_i = \begin{cases} X_i^{\frac{1}{3}} & \text{if } X_i > 0.008856 \\ \dfrac{903.3X_i + 16}{116} & \text{else} \end{cases}. \tag{30.2}$$

Finally, $\underline{f}$ is mapped into $\underline{\eta}$ by ,

$$\begin{pmatrix} \eta_L \\ \eta_a \\ \eta_b \end{pmatrix} = \begin{pmatrix} 0 & 116 & 0 \\ 500 & -500 & 0 \\ 0 & 200 & -200 \end{pmatrix} \times \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} + \begin{pmatrix} -16 \\ 0 \\ 0 \end{pmatrix}. \tag{30.3}$$

Step 3: Construction of a CCPV

Given $\underline{\eta} \in$ CIE-L*a*b*, we consider another transformation $\mathscr{T}_{II}$: CIE-L*a*b* $\to$ CC $\subset \mathscr{R}_+^6$, where $\mathscr{R}_+^6$ is the set of nonnegative vectors in $\mathscr{R}^6$. The space CC, standing for "Color Class", is introduced so as to develop several different color classes as we will see. For constructing CC, the transformation $\mathscr{T}_{II}$ is defined by measuring the inverse of the squared Euclidean distances between $\underline{\eta}$ and six fixed points in CIE-L*a*b* representing RED, GREEN, YELLOW, BLUE, WHITE and BLACK. More formally, we consider the following six fixed points in CIE-L*a*b*.
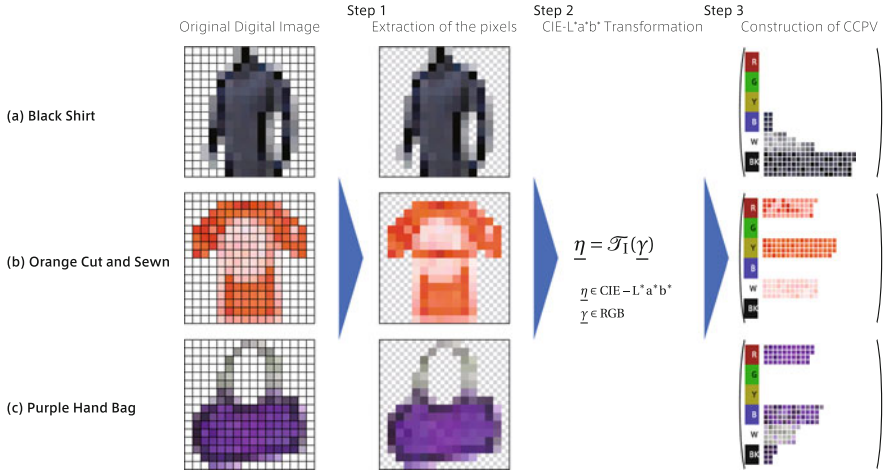
$$\begin{array}{ll} \underline{\eta}_R = {}^t(\ 50, 50,\ 0), & \underline{\eta}_G = {}^t(\ 50, -50,\ \ \ \ 0), \\ \underline{\eta}_Y = {}^t(\ 50,\ 0, 50), & \underline{\eta}_B = {}^t(\ 50,\ \ \ 0, -50), \\ \underline{\eta}_W = {}^t(100,\ 0,\ 0), & \underline{\eta}_{BK} = {}^t(\ \ 0,\ \ \ 0,\ \ \ \ 0), \end{array} \tag{30.4}$$

where each color in RGB are represented as,

$$\begin{array}{ll} \mathscr{T}_I^{-1}(\underline{\eta}_R) = {}^t(\ 0.59, 0.06, 0.18), & \mathscr{T}_I^{-1}(\underline{\eta}_G) = {}^t(-0.04, 0.25, 0.16), \\ \mathscr{T}_I^{-1}(\underline{\eta}_Y) = {}^t(\ 0.29, 0.17, 0.01), & \mathscr{T}_I^{-1}(\underline{\eta}_B) = {}^t(\ \ 0.04, 0.19, 0.55), \\ \mathscr{T}_I^{-1}(\underline{\eta}_W) = {}^t(\ 1.25, 0.95, 0.91), & \mathscr{T}_I^{-1}(\underline{\eta}_{BK}) = {}^t(\ \ \ \ \ 0,\ \ \ 0,\ \ \ 0). \end{array} \tag{30.5}$$

Given $\underline{\gamma} \in v_p$, let $\underline{\eta} = \mathscr{T}_I(\underline{\gamma}) \in$ CIE-L*a*b* and define $\underline{\phi}(\underline{\gamma}) = \mathscr{T}_{II} \circ \mathscr{T}_I(\underline{\gamma}) = \mathscr{T}_{II}(\underline{\eta})$ by

$$\underline{\phi}(\underline{\gamma}) \overset{\text{def}}{=} c \begin{pmatrix} ||\underline{\eta}_R - \underline{\eta}||^{-2} \\ ||\underline{\eta}_G - \underline{\eta}||^{-2} \\ ||\underline{\eta}_Y - \underline{\eta}||^{-2} \\ ||\underline{\eta}_B - \underline{\eta}||^{-2} \\ ||\underline{\eta}_W - \underline{\eta}||^{-2} \\ ||\underline{\eta}_{BK} - \underline{\eta}||^{-2} \end{pmatrix}, \tag{30.6}$$

Fig. 30.2 Algorithm of CCPV construction for typical digital image

where $||\underline{x}||$ denotes the Euclidean norm of $\underline{x}$, and c is the normalization constant. It should be noted that $\underline{\phi}(\gamma)$ is a probability vector, where each component describes the how a typical person would sense the pixel represented by γ to the corresponding color in RED, GREEN, YELLOW, BLUE, WHITE and BLACK.

The schematic diagram of the above steps are shown in Fig. 30.2. The color-class profile vector of v_p can now be defined by,

$$\underline{\Phi}(v_p) \stackrel{\text{def}}{=} \frac{1}{N_{v_p}} \sum_{\underline{\gamma} \in v_p} \underline{\phi}(\gamma) \ . \tag{30.7}$$

We may say that $\underline{\Phi}(v_p)$ describes how a typical person would sense the six different colors RED, GREEN, YELLOW, BLUE, WHITE and BLACK from the overall impression of the digital image v_p of product p.

30.2.2 Development of Color-Classes via Clustering of CCPVs

The algorithmic procedure described in Sect. 26.2.1 enables one to represent each digital image v_p of product p by the corresponding CCPV, $\underline{\Phi}(v_p)$. The data obtained from X Corporation contain 5665 such digital images, to each of which one of 425 colors was assigned by X Corporation. The purpose of this section is to develop a reasonable number of color classes by clustering these 425 colors, so that the effects of color in marketing can be analyzed efficiently. For this purpose, we represent each color defined by X Corporation by a CCPV in CIE-L*a*b*. More specifically, let x

be a color given by X Corporation and define,

$$V(x) = \{v_p : \text{the color } x \text{ is assigned to product } p\} . \tag{30.8}$$

The number of elements in $V(x)$ is denoted by $N(x) = |V(x)|$. The color x is then represented by $\underline{\Phi}_x \in$ CIE-L*a*b* where,

$$\underline{\Phi}_x = \frac{1}{N(x)} \sum_{v_p \in V(x)} \underline{\Phi}(v_p) . \tag{30.9}$$

30.2.3   Color Class Preference Vectors of Customer

In order to define a color class preference vector of a consumer, we introduce the following sets.

$CUST = \{i : 1 \le i \le N_c\}$: the set of customers
$S = \{j : 1 \le j \le N_s\}$: the set of product categories
$S(j)$: the number of products in the product category $j \in S$
$q_r(j)$: the set of products which are identical having the same product ID but belong to different color classes in the rth product in the product category $j \in S$
$Q(j) = \{q_1(j), \cdots , q_{S(j)}(j)\}$: the set of product groups in $S(j)$, where each group consists of identical products having different color classes
N_{CC} : the number of color classed to be combined
$CC = \{1, \cdots , N_{CC}\}$: the set of color classes
$n(i,j,x)$: the number of products, purchased by consumer $i \in CUST$, which belong to $Q(j)$ having color class $x \in CC$

For $l \in CC, l = 1, \cdots , m$, let the color class distribution vector, $\underline{\theta}(i,j)$, be defined by

$$\underline{\theta}(i,j) = [\theta(i,j,1), \cdots , \theta(i,j,N_{CC})]; \quad \theta(i,j,l) = \frac{n(i,j,l)}{\sum_{k=1}^{N_{CC}} n(i,j,k)} . \tag{30.10}$$

The corresponding mean and variance vectors, $\underline{\mu}(j), \underline{\sigma}(j)$, can be obtained as

$$\underline{\mu}(j) = \frac{1}{N_c} \sum_{i \in CUST} \underline{\theta}(i,j) , \tag{30.11}$$

$$\underline{\sigma}(j) = [\sigma(j,1), \cdots , \sigma(j,N_{CC})]; \quad \sigma(j,l) = \sqrt{\frac{1}{N_c - 1} \sum_{i \in CUST} \{\theta(i,j,l) - \mu(j,l)\}^2} . \tag{30.12}$$

Then the color class preference vector of consumer $i \in CUST$ for the product category $j \in S$ can be defined in the following manner.

$$\underline{z}(i,j) = [z(i,j,1), \cdots, z(i,j,N_{CC})]; \quad z(i,j,l) = \frac{\theta(i,j,l) - \mu(j,l)}{\sigma(j,l)} . \qquad (30.13)$$

Let $CCQ\left(j, \check{j}\right)$ be the set of color classes which products in $q_{\check{j}}(j) \in Q(j)$ possess. If consumer i is to purchase a product $p \in q_{\check{j}}(j) \in Q(j)$, then the color $\tilde{x}(i,j)$ to be recommended is determined by

$$\tilde{x}(i,j) = \arg \max_{x \in CCQ(j,\check{j})} \{z(i,j,x)\} . \qquad (30.14)$$

In the approach discussed above, the color class preference vector of consumer $i \in CUST$ is defined for each product category $j \in S$. As an alternative approach, the single color class preference vector of consumer $i \in CUST$ may be employed for all the products in $Q = \bigcup_{j=1}^{N_S} Q(j)$. In this case, in place of Eq. (30.10), we define

$$\underline{\theta}(i) = [\theta(i,1), \cdots, \theta(i,N_{CC})]; \quad \theta(i,l) = \frac{\sum_{j=1}^{N_S} n(i,j,l)}{\sum_{j=1}^{N_S} \sum_{k=1}^{N_{CC}} n(i,j,k)} . \qquad (30.15)$$

Then the color class preference vector of $i \in CUST$ for all the product in Q can be defined by

$$\underline{z}(i) = [z(i,1), \cdots, z(i,N_{CC})]; \quad z(i,l) = \frac{\theta(i,l) - \mu(l)}{\sigma(l)} , \qquad (30.16)$$

where the mean and the variance vectors are also changed accordingly as $\underline{\mu}$ and $\underline{\sigma}$. Let $CCQ(j,\check{j})$ be defined as before. If consumer $i \in CUST$ is to purchase a product $p \in q_{\check{j}}(j) \in Q(j)$, then the color $\tilde{x}(i)$ to be recommended is determined by,

$$\tilde{x}(i) = \arg \max_{x \in CCQ(j,\check{j})} \{z(i,x)\} . \qquad (30.17)$$

The latter approach may work better because of the larger data volume involved in constructing $\underline{z}(i)$. If the color class was not defined because of the lack of the purchase history for the customer or the lack of the color options, the engine would recommend the default choice of the color.

30.3 Numerical Experiments Based on Real Data

30.3.1 Data Description

Sumita Research Laboratory at the University of Tsukuba has been working with a TV shopping company, hereafter called X Corporation, for developing a CRM (Customer Relationship Management) support engine based on real data. X Corporation has been in retail business worldwide, offering a variety of products ranging from Apparel products, Jewelries, and Home electronics appliances to foods. A typical digital image used in the e-business consists of $400 \times 400 = 160{,}000$ pixels, which would be too much to define the dominating color-class of the image.

The data obtained from X corporation consist of demographic information of those consumers who purchased at least one product during the period between September 1st, 2004 and August 31st, 2007, as well as their purchasing records and channels, product records and TV programs during the period. The amount of consumers, N_c, was 455,415, the amount of product categories, N_s, was 34 and the data consisted of about 2.3 million records. The average number of purchase occasions per customer and purchased quantity per customer were 3.70 and 5.33, respectively. The digital images collected from the data obtained from X Corporation amount to 6762, involving 1782 types of products spread over 34 small categories. The structure and the key components of these records are in Fig. 30.3. The database of X Corporation defines 430 colors appear for the products corresponding to the 6762 digital images. However, five of them are clearly useless (e.g. NON-COLOR, CLEAR) and eliminated. Consequently, the data to be used

Genre	Category			
1 Fashion Wear	10 Inner-Lingerie	13 Skirt	16 Men's	19 Suit-One piece
	11 Jacket-Coat	14 Knit	17 Ensemble	
	12 Shirt-Blouse	15 Pants	18 Cut and Sewn	
2 Bag	20 Hand Bag	22 Tote Bag	24 Racksack/Organizer Bag	
	21 Shoulder Bag	23 Boston Bag	25 Pouch/Pochette 26 Bag etc.	
3 Fashion Accessory	30 Ring	32 Earring/Ear Clip	33 Bracelet/Anklet	
	31 Necklace/Pendant/Brooch		34 Set etc.	
4 Brand Accessory	30 Ring	32 Earring/Ear Clip	33 Bracelet/Anklet	
	31 Necklace/Pendant/Brooch		34 Set etc.	
5 Fashion Gadget	50 Presbyopia glasses	52 Stole/Scarf/Belt	54 Shoes/Umbrella/Hat	
	51 Watch	53 Wallet/Key Purse	55 Wig/Hair Accessory 56 Others	

Fig. 30.3 Category structure

for our analysis contain 425 colors (corresponds to 5665 digital images) defined by X Corporation. In what follows, these 425 colors are categorized into several number of newly defined color-classes by analyzing the 6762 digital images. The algorithmic procedure used to establish the color-classes can be applied to a digital image of any product with one of the 425 colors, identifying the dominating color-class of the product automatically. In turn, the algorithmic procedure enables one to canalize the consumers from the perspective of color preferences, thereby filling the missing link in e-marketing.

In order to cluster 425 colors, each represented by $\underline{\Phi}_x$, we employ the group average method in hierarchical clustering [18, 19]. In this approach, a set of vectors would be grouped together one by one based on the nearest Euclidean distance until the predetermined number of clusters would exhaust the original set. In each grouping, the resulting cluster is represented by one vector which can be generated as the weight center of the two clusters to be merged. We terminated the grouping just before the six representative color (RED, GREEN, YELLOW, BLUE, WHITE, BLACK) combined to the other six representative color.

For each cluster generated by the above algorithm, the histogram is constructed by 425 colors over the digital images involved in the cluster. Namely if a cluster consists of $\underline{\Phi}_{x(1)}, \cdots, \underline{\Phi}_{x(T)}$, then the histogram is constructed over the products in $\bigcup_{l=1}^{T} V(x(l))$. The grouping resulted into generate 14 color-classes, (i.e. $N_{CC} = 14$), named as BLACK, BEIGE, WHITE, PINK, BROWN, GRAY, BLUE, NAVY, GREEN, PURPLE, RED, ORANGE, SAXE-BLUE and YELLOW.

30.3.2 Accuracy Test for Color Class Recommendation Engine

In this subsection, we examine the accuracy of the color class recommendation engine developed in Sect. 30.2.3. The data set obtained from X Corporation is decomposed into ten subsets of equal size randomly. Based on the cross validation approach, nine subsets are used to construct $z(i,j)$ in Eq. (30.13) and $z(i)$ in Eq. (30.16), while the remaining subset is used for testing accuracy. In order to provide a basis for comparison, the following random estimation accuracy is considered.

Random Estimation:

If consumer i is to buy a product $p \in q_{\check{j}}(j)$ and a color class is chosen randomly, the probability of its correctness is given by $|CCQ(j, \check{j})|^{-1}$ where $CCQ(j, \check{j})$ is the set of color classes which products in $q_{\check{j}}(j)$ possess.

In Table 30.1, the results for testing accuracy based on $z(i,j)$ in Eq. (30.13) and the results for testing accuracy based on $z(i)$ in Eq. (30.16) are exhibited respectively. One can observe that the color class recommendation engine outperforms the random estimation consistently with only one exception for "51 Watch" in row of table $z(i,j)$. However, even for this product, the color class recommendation

Table 30.1 Accuracy test of the recommendation engine based on $\underline{z}(i,j)$ and $\underline{z}(i)$ for customer i and category j ("ratio" notes acc./rand.)

	Based on $\underline{z}(i,j)$				Based on $\underline{z}(i)$			
	Records	Acc.	Rand.	Ratio	Records	Acc.	Rand.	Ratio
Fashion wear								
10 Inner-lingerie	565,088	0.43	0.32	1.36	551,873	0.48	0.32	1.53
11 Jacket-coat	34,482	0.34	0.30	1.11	56,984	0.45	0.30	1.49
12 Shirt-blouse	50,832	0.36	0.32	1.13	66,679	0.44	0.32	1.36
13 Skirt	25,641	0.40	0.33	1.21	38,999	0.52	0.33	1.54
14 Knit	44,945	0.27	0.25	1.06	63,412	0.33	0.26	1.30
15 Pants	64,242	0.43	0.35	1.22	77,971	0.54	0.36	1.51
16 Men's	168	0.47	0.31	1.51	351	0.50	0.31	1.60
17 Ensemble	29,206	0.34	0.30	1.14	48,895	0.42	0.30	1.42
18 Cut and sewn	136,540	0.58	0.27	2.18	159,947	0.34	0.27	1.25
19 Suit-one piece	24,460	0.39	0.32	1.22	39,179	0.53	0.33	1.61
(Sub total)	975,604	0.43	0.31	1.39	1,104,290	0.45	0.31	1.45
Bag								
20 Hand bag	13,918	0.25	0.20	1.25	29,218	0.48	0.21	2.26
21 Shoulder bag	17,250	0.20	0.17	1.21	35,931	0.39	0.19	2.04
22 Tote bag	27,320	0.25	0.23	1.05	51,454	0.35	0.24	1.47
23 Boston bag	17,318	0.17	0.16	1.06	36,155	0.31	0.17	1.82
24 Rucksack/organizer bag	4153	0.24	0.21	1.18	11,580	0.40	0.22	1.85
25 Pouch/pochette	20,771	0.21	0.19	1.11	34,361	0.33	0.19	1.70
26 Bag all	2865	0.30	0.22	1.37	8183	0.48	0.20	2.42
(Sub total)	103,595	0.22	0.20	1.13	206,882	0.37	0.20	1.83
Fashion accessory								
30 Ring	3288	0.82	0.44	1.86	8152	0.50	0.45	1.12
31 Necklace/pendant/brooch	7058	0.84	0.36	2.31	14,913	0.58	0.38	1.52
32 Earring/ear clip	2066	0.59	0.43	1.36	6246	0.53	0.36	1.47
33 Bracelet/anklet	1241	0.87	0.31	2.86	3187	0.43	0.27	1.59
34 Set etc.	5247	0.58	0.42	1.38	12,740	0.36	0.30	1.18
(Sub total)	18,900	0.74	0.39	1.87	45,238	0.49	0.36	1.35
Brand accessory								
40 Ring	1110	0.72	0.40	1.80	3900	0.61	0.39	1.56
41 Necklace/pendant/brooch	4366	0.79	0.36	2.17	10,699	0.50	0.36	1.38
42 Earring/ear clip	1895	0.56	0.43	1.30	5640	0.59	0.38	1.56
43 Bracelet/anklet	358	0.72	0.24	2.97	1066	0.41	0.25	1.64
44 Set etc.	131	0.60	0.47	1.27	842	0.49	0.36	1.38
(Sub total)	7860	0.72	0.38	1.90	22,147	0.54	0.37	1.47

Table 30.1 (continued)

	Based on $z(i,j)$				Based on $z(i)$			
	Records	Acc.	Rand.	Ratio	Records	Acc.	Rand.	Ratio
Fashion gadget								
50 Presbyopia glasses	2098	0.17	0.17	1.03	7749	0.40	0.20	1.99
51 Watch	2881	0.25	0.36	0.69	6528	0.73	0.36	2.02
52 Stole/scarf/belt	25,785	0.31	0.27	1.18	42,137	0.41	0.26	1.55
53 Wallet/key purse	26,031	0.28	0.18	1.52	45,128	0.38	0.18	2.12
54 Shoes/umbrella/hat	20,126	0.40	0.24	1.68	35,385	0.43	0.23	1.85
55 Wig/hair accessory	876	0.50	0.49	1.00	7,554	0.95	0.49	1.95
56 Others	7046	0.37	0.24	1.55	16,023	0.36	0.25	1.40
(Sub total)	84,843	0.32	0.24	1.37	160,504	0.44	0.24	1.81
Total	1,190,802	0.41	0.29	1.40	1,539,061	0.44	0.29	1.53

engine based on $z(i)$ supersedes the random estimation by a factor of two. It can be seen that, when the volume of test data is high, the color class recommendation engine based on $z(i)$ outperforms the color class recommendation engine based on $z(i,j)$. This implies that color preferences of consumers are reflected beyond product categories for products which are purchased rather often at modest prices, as represented by Fashion Wear (10 through 19) , Bag (20 through 26) and Fashion Gadget (50 through 56). For more expensive products which are likely to be purchased with less frequency, however, color preferences of consumers within the product category prevail over those derived from all products, as can be seen in Fashion Accessory (30 through 34) and Brand Accessory (40 through 44). This result is in agreement that one who have the color to prefer may buy the other color product as an accent color. In any case, one may select whichever the recommendation engine based on $z(i,j)$ or $z(i)$, by considering which is suitable for the genres of product.

30.4 Conclusion

One of the important factors ignored in the past analyses in e-marketing is "colors" of products. This is so because it is difficult to define a color of a product, which typically consists of many different colors. The purpose of this research is to fill this gap by developing an algorithmic procedure for identifying the dominating color of a product by analyzing a digital image of the product. Since humans tend to clearly distinguish RED from GREEN as well as YELLOW from BLUE, the Euclidean distance in CIE-L*a*b* is more consistent with the sensuous feeling of human for colors than the Euclidean distance in RGB. Accordingly, for analyzing color preferences of consumers in e-marketing, CIE-L*a*b* is more appropriate than RGB. Based on this idea, we proposed the CCPV (Color-Class Profile Vector) which represents the overall impression of a digital image containing a product. Since each product has its color in the data base, these vectors can be utilized to

categorize many different colors, resulting in 14 color classes. This enables one to study color preferences of consumers by segments. Furthermore, it provides a basis for constructing a recommendation engine based on the color classes for enhancing e-commerce. We had also confirmed the effectiveness of personalized recommendation engine with CCPV from the numerical experiments based on real data. This study is still in its infancy. It would be necessary to combine the color analysis proposed in this thesis with other approaches, such as automatic personalization and collaborative filtering, so as to empower the existing recommendation engines. This line of research is underway and will be reported elsewhere in due course.

References

1. Srivastava J, Cooley R, Deshpande M, Tan P (2000) ACM SIGKDD Explorations Newsletter. doi:10.1145/846183.846188
2. Mobasher B, Dai H, Luo T, Nakagawa M (2002) Discovery and evaluation of aggregate usage profiles for web personalization. Data Min Knowl Disc 6(1):61–82. doi:10.1023/A:1013232803866
3. Pinyapong S, Kato T (2005) IEICE Trans Inf Syst. doi:10.1093/ietisy/e88-d.5.938
4. Liu K, Fang B, Zhang W (2011) IEICE Trans Inf Syst. doi:10.1587/transinf.E94.D.542
5. Sarwar BM, Karypis G, Konstan J, Riedl J (2001) Proceedings of the 10th international conference on World Wide Web. doi:10.1145/371920.372071
6. Jung K, Lee J (2002) E-commerce and web technologies. Springer, Heidelberg. doi:10.1007/3-540-45705-4_26
7. Inoue K, Urahama K (2004) IEICE Trans Inf Syst E87-D(7):1987–1988
8. Kang H, Yoo SJ (2007) IEICE Trans Inf Syst. doi:10.1093/ietisy/e90-d.12.2100
9. Aggarwal CC, Wof JL, Yu PS (1999) Proceedings of the 1999 ACM SIGMOD international conference on management of data. doi:10.1145/304182.304188
10. Sarwar B, Karypis G, Konstan J, Riedl J (2000) Proceedings of the 2nd ACM conference on Electronic commerce. doi:10.1145/352871.352887
11. Vartak M, Madden S (2013) Proceedings of the 2013 ACM SIGMOD international conference on management of data. doi:10.1145/2463676.2465270
12. Jung JJ, Lee K, Park S, Jo G (2005) IEICE Trans Inf Syst E88-D(5):843–850
13. Lin Y, Kawakita Y, Suzuki E, Ichikawa H (2012) International symposium on applications and the Internet. doi:10.1109/SAINT.2012.75
14. Huang C, Wei C, Wang Y (2013) IEEE international conference on multimedia and Expo workshops. doi:10.1109/ICMEW.2013.6618318
15. C.I.E. (1971) Recommendations on uniform color spaces, color-difference equations, psychometric color terms. Supplement No. 2 to CIE publication No. 15 (E.-1.3.1)
16. Fairchild MD, Berns RS (1993) Color Res Appl. doi:10.1002/col.5080180308
17. Connolly C (1997) IEEE Trans Image Process. doi:10.1109/83.597279
18. Kaufman L, Rousseeuw PJ (2008) Finding groups in data: an introduction to cluster analysis. Wiley, New York. doi:10.1002/9780470316801
19. Wiggerts TA (1997) Proc Conf Rev Eng. doi:10.1109/WCRE.1997.624574