Kobayashi Colors

Visual And Multimedia Recognition Course Project

Federico Nocentini

Supervisor: Prof. Alberto Del Bimbo, Ing. Federico Becattini, Ing. Claudio Baecchi

Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Firenze

Index

Introduction

Ours is an age of increasing sensitivity of color. But what meanings are covered by the colors that surround us in everyday life?

The first step toward a more effective use of color is to systematize and classify colors through key-words that express the difference between them.

For single colors this can be done by arranging them sequentially according to their place in the spectrum.

But this is not possible for color combinations, and though it is obvious at a glance that different color combinations convey different images, the problem is how to classify them in a systematic way.

Color Image Scale

In attempting to solve the problem of systematically classifying color combinations Shigenobu Kobayashi wrote a book:
Color Image Scale, a database developed at the Nippon Color and Design Research Institute.

The database was developed by using 130 basic colors to make different three-color combination (triplets). After creating the various triplets Kobayashi associated each one with a key-word (180 in total, e.g. calm, bold, lively) that relate to ways in which people perceive colors.

The 180 key-words are divided into 14 classes (e.g. MODERN, ROMANTIC, CHIC).

How Color Image Scale Works

After a short introduction part we have 130 pages each of which is associated with one of the 130 base colors. Each base color is used to build 9 triplets, each triplet is associated with one of the 180 keywords.

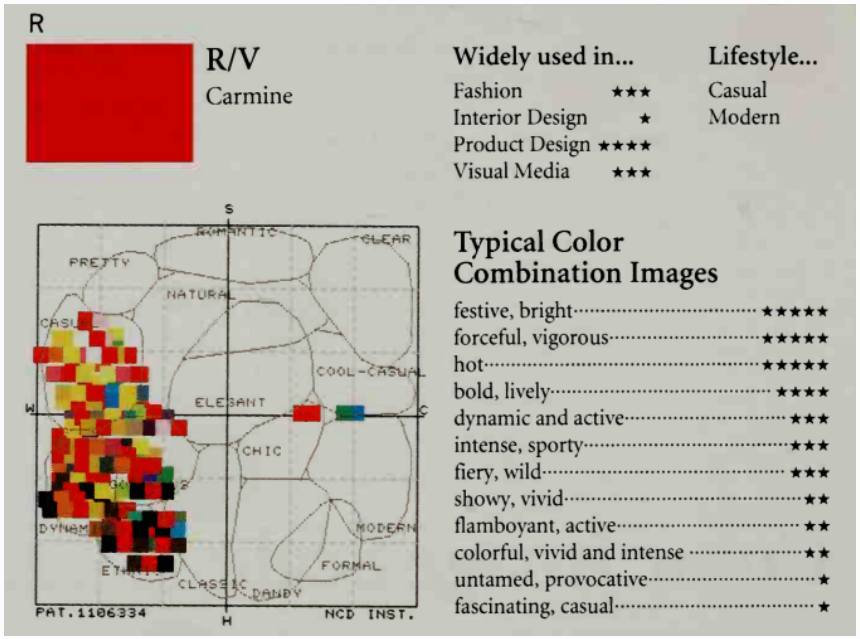

Figure: Top of a page

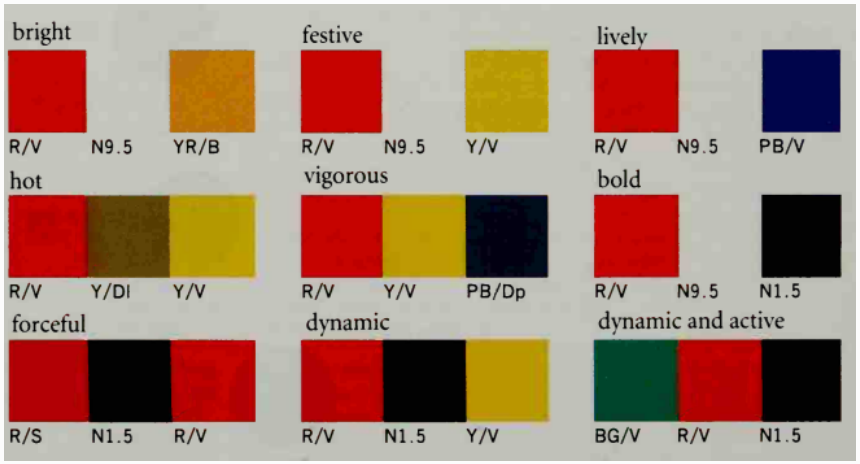

Figure: Bottom of a page

Goal of the Project

The objective of this work is to extract the triplet of principal colors from a top-bottom pair of clothes, an outfit, and compare it to all 1170 Kobayashi triplets.

Once the Kobayashi triplet most similar to the outfit triplet is determined associate the corresponding key-word and class with the outfit image.

Figure: Example of a top-bottom pair

Goal of the Project

Figure: Outfits mapped in Kobayashi triplets space

Kobayashi triplets to RGB triplets

To convert the 1170 triplets of kobayashi to 1170 triplets of RGB we used the RGB Table that associates each Kobayashi base color to an RGB.

Color Image Scale
with RGB values

R/Lgr 213,182,166
hue/tone color sample R,G,B

tone \ hue	R	YR	Y	GY	G	BG	B	PB	P	RP	Neutral
vivid V	231,47,39	238,113,25	255,200,8	170,198,27	19,166,50	4,148,87	1,134,141	3,86,155	46,20,141	204,63,92	N9.5 244,244,244
strong S	207,46,49	226,132,45	227,189,28	162,179,36	18,154,47	6,134,84	3,130,122	6,113,148	92,104,163	175,92,87	N9 236,236,236
bright B	231,108,86	241,176,102	255,228,15	169,199,35	88,171,45	43,151,89	0,147,159	59,130,157	178,137,166	209,100,109	N8 206,206,206
pale P	233,163,144	242,178,103	255,236,79	219,220,93	155,196,113	146,198,131	126,188,209	147,184,213	197,188,213	218,176,176	N7 180,180,180
very pale Vp	236,217,202	245,223,181	249,239,189	228,235,191	221,232,207	209,234,211	194,222,242	203,215,232	224,218,230	235,219,224	N6 152,152,152
light grayish Lgr	213,182,166	218,196,148	233,227,143	209,116,73	179,202,157	166,201,163	127,175,166	165,184,199	184,190,189	206,185,179	N5 126,126,126
light L	211,142,110	215,145,96	255,203,88	195,202,101	141,188,90	140,195,110	117,173,169	138,166,187	170,165,199	205,154,149	N4 86,86,86
grayish Gr	171,131,115	158,128,110	148,133,105	144,135,96	143,162,121	122,165,123	130,154,145	133,154,153	151,150,139	160,147,131	N3 60,60,60
dull Dl	162,88,61	167,100,67	139,117,65	109,116,73	88,126,61	39,122,62	24,89,63	53,109,98	44,77,143	115,71,79	N2 38,38,38
deep Dp	172,36,48	169,87,49	156,137,37	91,132,47	20,114,48	23,106,43	20,88,60	8,87,107	58,55,119	111,61,56	N1.5 10,10,10
dark Dk	116,47,50	115,63,44	103,91,44	54,88,48	30,98,50	27,86,49	18,83,65	16,76,84	40,57,103	88,60,50	
dark grayish Dgr	79,46,43	85,55,43	75,63,45	44,60,49	34,62,51	31,56,45	29,60,47	25,62,63	34,54,68	53,52,48	

Figure: RGB Table

K-means to extract the triplets

Before extracting the triplet from the outfit image, it was necessary to remove the white background.

After that we used K-means to extract the main color triplet, in the form of RGB triplet, from the image of the outfit without white background.

Figure: Top-bottom pair

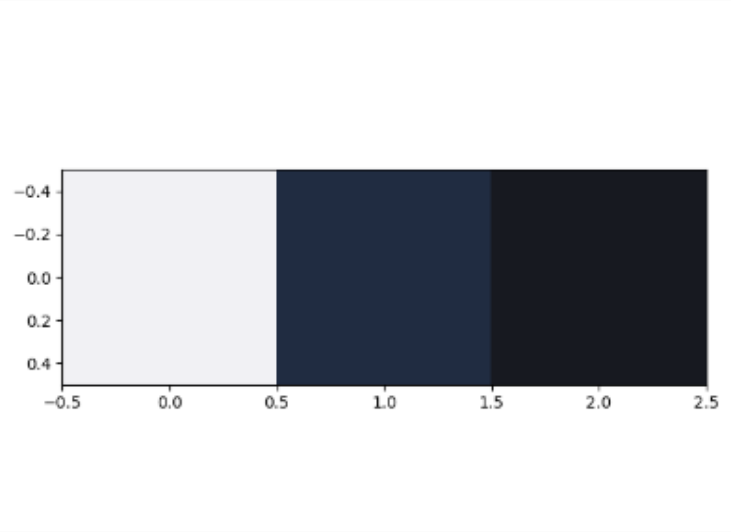

Figure: Triplet extract with K-means

Map to Kobayashi base colors space

We thought to convert the image of the outfit into the Kobayashi base colors space defined by the RGB Table.

Once this was done we simply calculated the histogram of the converted image and extracted the triplet going to see which were the 3 most present Kobayashi base colors.

Figure: Conversion

Dataset build

Once we had the triplet extracted from the image of the outfit we went to compare it with all the triplets of Kobayashi, taking into account the order of the colors.

We then assigned to the outfit image a key-word and a class.

Of course, we discarded the images of the outfit whose triplet was not similar enough to any of Kobayashi's triplets.

In this way we built a dataset with which we trained a CNN classifier.

As a label to build the dataset we used the kobayashi classes since the key-words did not give a homogeneous dataset.

To build our dataset, we used the labels obtained by mapping the images of the outfit into Kobayashi space, as they provided a larger dataset than the labels extracted using K-means.

CNN BUILD


```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1         [-1, 32, 148, 298]             896
       BatchNorm2d-2         [-1, 32, 148, 298]              64
            Conv2d-3          [-1, 64, 72, 147]          18,496
           Dropout-4          [-1, 64, 72, 147]               0
       BatchNorm2d-5          [-1, 64, 72, 147]             128
            Conv2d-6          [-1, 128, 34, 71]          73,856
       BatchNorm2d-7          [-1, 128, 34, 71]             256
            Conv2d-8          [-1, 128, 15, 33]         147,584
           Dropout-9          [-1, 128, 15, 33]               0
      BatchNorm2d-10          [-1, 128, 15, 33]             256
           Conv2d-11           [-1, 256, 5, 14]         295,168
          Dropout-12           [-1, 256, 5, 14]               0
      BatchNorm2d-13           [-1, 256, 5, 14]             512
           Conv2d-14            [-1, 256, 1, 6]         262,400
           Linear-15                 [-1, 1000]       1,537,000
           Linear-16                  [-1, 500]         500,500
           Linear-17                  [-1, 128]          64,128
           Linear-18                   [-1, 14]           1,806
================================================================
Total params: 2,903,050
Trainable params: 2,903,050
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.51
Forward/backward pass size (MB): 43.64
Params size (MB): 11.07
Estimated Total Size (MB): 55.23
----------------------------------------------------------------
```

Figure: CNN

CNN Training and Results

Our dataset consists of 87906 images with associated class labels. We split our dataset into training set (77390) and validation set (10516).

To train our net we have used minibatch of 64 images, learning rate 10^{-4} and we have made the comparison between 2 types of loss: Cross Entropy Loss and Focal Loss.

We have trained our network for 200 epochs and we have obtained an accuracy of 78.34% for the Focal Loss and an accuracy of 77.9% for the Cross Entropy Loss.

CNN Predictions

CLASSIC

CHIC

COOL-CASUAL

ETHNIC

MODERN

FORMAL

DANDY

DYNAMIC

CLEAR

CNN Evaluation

To evaluate the quality of the results obtained by our CNN, we produced results with other techniques:

-Nearest Neighbor baseline made using features extracted from a standard ResNet (trained only on Imagenet). Which produces an accuracy of 49,11%

-Use ResNet18 as a CNN and train it it as we did for our CNN. Which produces an accuracy of 78,56%

Confusion Matrix

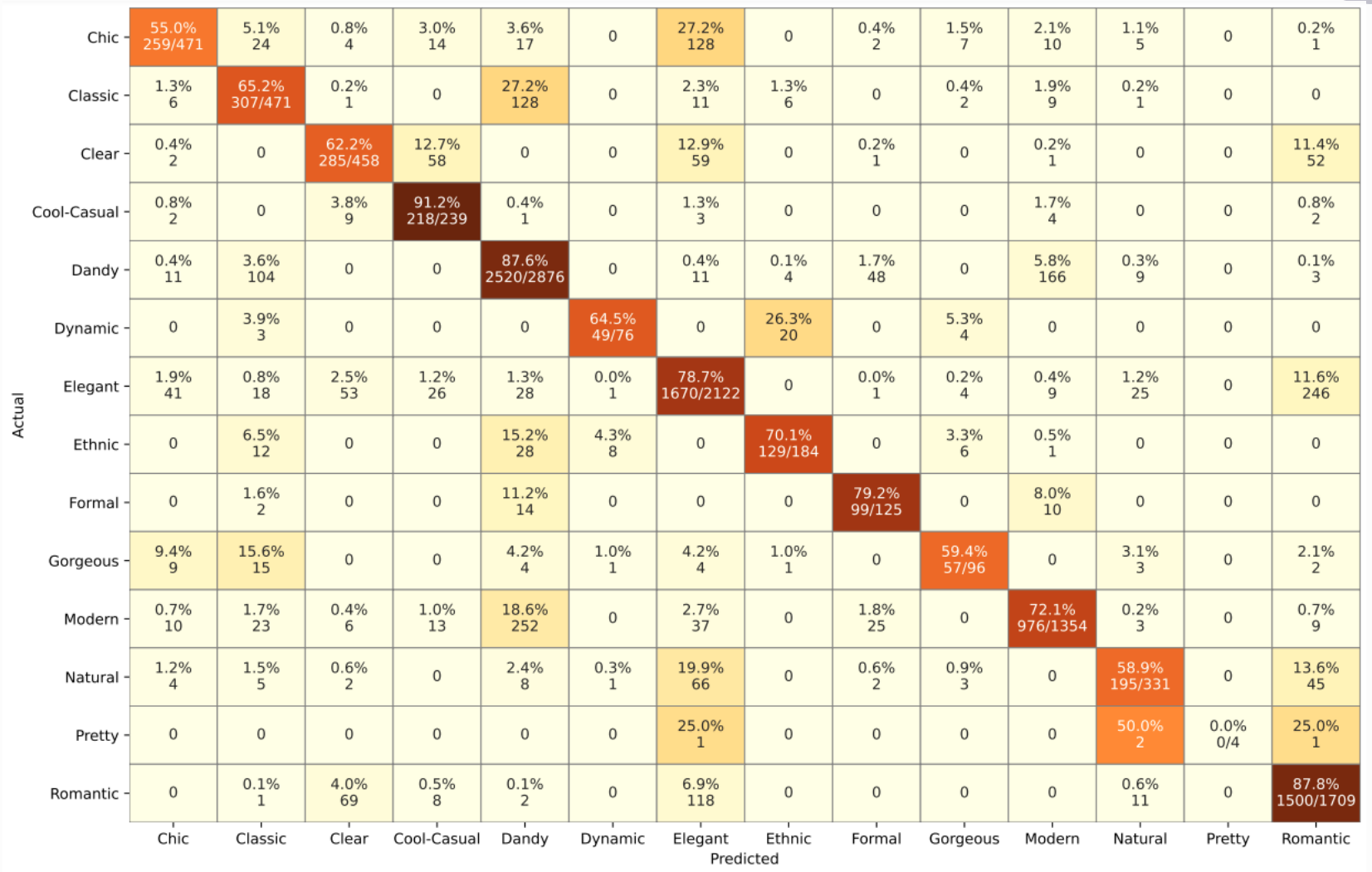

Figure: ResNet18 Confusion Matrix

Conclusions

In this work we looked for various ways to associate the various classes introduced by Kobayashi in Color Image Scale with outfits.

To do this we have used various techniques to extract the main colors from the various outfits, compared the triplets extracted with the triplets of kobayashi and finally determined the class of each outfit.

We believe that this classification of colors based on images is the key to understanding the way in which color combination are perceived, and that this systems open a path to the future, when sensitivity to color will continue to grow.