Numerical Modeling of Image Discriminability for Home Storage and Organization System on a Smart Device

Rui Fukui
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, JAPAN
fukui@ics.t.u-tokyo.ac.jp

Tomosaka Sato
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, JAPAN
tsato@ics.t.u-tokyo.ac.jp

Keisuke Maeda
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, JAPAN
maeda@ics.t.u-tokyo.ac.jp

Masahiko Watanabe
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, JAPAN
masahiko@ics.t.u-tokyo.ac.jp

Masamichi Shimosaka
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, JAPAN
simosaka@ics.t.u-tokyo.ac.jp

Abstract
In Home storage and organization system on a smart device, thumbnail pictures (Tag Image) of daily-use objects are often used. Discriminability of Tag Image is important to realize superior usability. In this paper, we have tried to construct a numerical model of Tag Image’s discriminability. The proposed model is based on simple linear regression from popular image features and their statistics. In addition, web-based data input system has also been developed to collect training data efficiently. Consequently, the input system has acquired substantial number of data and a numerical model has been constructed. The constructed model has substantially good but not perfect performance.

Author Keywords
Home storage and organization, value of image discriminability, tag image

ACM Classification Keywords
H.5.0 [General]: .

General Terms
Algorithms, Design, Experimentation, Human Factors, Measurement
Introduction

Home storage and organization system supported by ubiquitous intelligences has been studied [1, 3, 7]. Recently, several systems for smart devices based on Tag Image have been developed [2, 6]. In these systems, Tag Images of daily-use objects are taken by smart devices. In our target system, the Tag Image and the information of the object are summarized in one screen as shown in Figure 1. The appearance of Tag Images are important to the usability. If the Tag Images are inadequate as shown in Figure 2, a user cannot find out the objective Tag Image quickly.

In this study, image discriminability is defined as an index that indicates how easily a Tag Image can be distinguished and found out. We have tried to construct a numerical model of Tag Image's discriminability. This numerical model enables a user to know the validity of a picture which will be taken. Figure 3 shows an example of a user interface. The green bar and number represent the validity of the image and the interface encourages the user to take picture when the validity is high enough.

Problems and solutions in numerical modeling

Problems and solutions in numerical modeling of discriminability of Tag Images are summarized below.

Problem 1 To make clear the relationship between Image Discriminability and image features
⇒ Solution: Utilization of simple linear regression

Problem 2 To calculate a model with a weak processing resource on a smart device
⇒ Solution: Regression shrinkage and selection of features via “Lasso”

Problem 3 To apply the proposed model to various daily-use objects and home storages
⇒ Solution: Data Collection with web-based data input system

To make clear the relationship between Image Discriminability and image feature

The concept of image feature is well known in computer vision and image processing [3]. Value of Image Discriminability (VID) can be described as the relationship with image feature. Many types of numerical model has been developed, such as linear regression, logistic regression, multiple regression and neural network. In this study, simple linear regression is adopted as the basic model because linear regression has the characteristic to distinguish useful/useless features. Linear regression expresses VID as the linear combination of the image features and the weight parameters as described in equation Equation 1, where $x$ is the image feature vector and $w$ is the weight parameter vector.

$$ y(x, w) = w_0 + w_1 x_1 + \cdots + w_D x_D \quad (1) $$
Those weight parameters are optimized to minimize the sum of squared residuals. The weight parameter represents the dependency of each feature and VID, in other words, it can represent the usefulness of each feature.

**To calculate a model with a weak processing resource**

To deal with the numerical model in an application software on a smart device, reduction of computational complexity is essential. To realize the reduction, dimension reduction is applied to the input data of linear regression. In our frame work, over 800 image features are calculated as input data of linear regression at first, then useless features are reduced with Lasso (least absolute shrinkage and selection operator) \( \|w\|_1 \). Lasso uses the L1-norm of the weight parameter vector \( \|w\|_1 \) as the constraint. In Lasso regression, increasing the penalty will cause more and more weight parameters to be driven to zero. That means features with those weight parameters have no effect on the calculation of VID.

**To apply the model to various objects and storages**

Numerical model must have generality for various daily-use objects and house storages. For the training phrase of the model with Linear Regression, substantial number of dataset of Image features and VID need to be collected as training data. In order to build dataset, various images of daily-use objects and home storage are collected, and VID of those images are evaluated by multiple collaborators. Pairwise comparison is used to convert Image Discriminability into numerical index, VID. In addition, two approaches are taken to collect data efficiently. One is consideration of priority of each comparisons and the other is development of a web-based data input system.

**Experiment**

The parameter, \( \lambda \) in Lasso is optimized to minimize the mean square error (MSE) in leave-one-out cross-validation. In the images of daily-use objects the optimized parameter is \( \lambda = 0.44 \). In the images of house storages the optimized parameter is \( \lambda = 0.32 \). Lasso regression reduces image feature dimension from 827 to 134 in object images, to 99 in storage images respectively. The difference between inner and outer region of the image is found to be effective additional statistics. This may indicate that contrast difference between the center object and the background is a clue to estimate VID precisely.

The Table 1 shows the statistics type of features with non zero weight parameter.

<table>
<thead>
<tr>
<th>Category</th>
<th>Object</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>max</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>min</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>range</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>stdev</td>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td>kurtosis</td>
<td>20</td>
<td>29</td>
</tr>
<tr>
<td>skewness</td>
<td>23</td>
<td>27</td>
</tr>
<tr>
<td>color distance</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>max index</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>min index</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

**Table 1:** Type of statistics

The top three types in each category are described with bold letters. In both category, kurtosis and skewness get high scores. These statistics represent the shape of the distribution. Kurtosis represents whether the data are peaky or flat relative to a normal distribution. Skewness is a measure of symmetry, or more precisely, the lack of symmetry.
Therefore, an image with high discriminability has special shape histograms; e.g. a color histogram has a strong peak or edges are concentrated in the center of an image. The results of double cross-validation are showed in Table 2. RMSE stands for root-mean-square error. Considering the VID distributes with \( \sigma = 25 \), MSE (Mean Square Error) 14.45 and 9.78 means that the numerical models have sufficient accuracy. Figure 4 shows correlation diagram of the estimated value and the answer value.

<table>
<thead>
<tr>
<th></th>
<th>Object Image</th>
<th>Storage Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>14.45</td>
<td>9.78</td>
</tr>
<tr>
<td>RMSE</td>
<td>18.34</td>
<td>16.64</td>
</tr>
</tbody>
</table>

**Table 2: Double cross-validation**

**Conclusion**

In this paper, we build a dataset of VID with pictures of daily-use objects and house storages. Then we tried to construct a numerical model of VID for human. In daily-use object images, the difference between the inner and the outer region of an image have a large effect on the discriminability. In both image categories, the bias of the feature distribution is important. The double cross-validation confirmed that the constructed model have a substantially good performance. As future work, introduction of the image segmentation and principal object detection will be able to improve the precision of the numerical model.

**References**


