Online Learning of Color Transformation for Interactive Object Recognition under Various Lighting Conditions

Yasushi Makihara, Yoshiaki Shirai, and Nobutaka Shimada
Dept. of Computer-Controlled Mechanical Systems, Graduate School of Engineering,
Osaka University, 2-1 Yamadaoka, Suita, Osaka, 565-0871, Japan
{makihara, shirai, shimada}@cv.mech.eng.osaka-u.ac.jp

Abstract

This paper describes an online learning method of color transformation for interactive object recognition. In order to recognize objects under various lighting conditions, the system estimates a color transformation from the color of an object model by observing the color of a reference object. The system first initializes a general color transformation. Next the system automatically recognizes a target object with the color transformation. When the system fails in recognition, the system recovers the failure with user interaction. Then the system improves the color transformation with an observed color pair of the recognized target object. By repeating this process, the system adapts to the new environment. Experiments using real-world refrigerator scenes are shown.

1. Introduction

This paper describes a learning method of color transformation for interactive object recognition under various lighting conditions for a service robot which finds a user-specified object and bring it to the user. Nowadays, many service robots using vision are developed such as a robot which brings daily objects [1] and a robot which recognizes apples and books through verbal and gestural interaction [2]. For the vision-based service robots, adaptation to lighting condition change is one of important problems.

There are a variety of researches on color constancy using surface reflectance of objects. Nadimi et al. [3] apply it to shadow detection for moving object based on the fact that shadow regions and illuminated regions have the same surface reflectance.

In addition, problems of color constancy are treated as those of color transformation between different lighting conditions. Some researchers propose linear color transformation [4] and diagonal color transformation (independent transformation of each RGB channel), which are derived from the physics-based color model. Drew et al. [5] propose a voting scheme with pairs of colors of corresponding features to get candidates of diagonal color transformation.

On the other hand, there are statistics-based approaches. Miller et al. [6] propose a method of non-linear color transformation using color eigenflows learned from multiple pairs of images of a same scene under different lighting conditions. These two methods [5][6] need multiple pairs of reference colors for estimating color transformation. It is however difficult for the robot vision system to get multiple reference colors in unknown lighting conditions.

Our method [7] needs only one reference color for estimating color transformation. First we introduce a color model and color transformation based on a light reflection model, then formulate color transformation estimation with an observed color of the reference object. In learning step, parameters for color transformation estimation are calculated using an observed color of an object and the reference object under various lighting conditions. In object recognition, the system observes a color of the reference object and estimates color transformation with the learned parameters.

When the system has enough learning sets of color pairs, the above method works well. However, if the system is used in a new environment, it has no learning sets because the reference object is different. Therefore the system needs to learn the parameters online. The system first initializes a color transformation with standard parameters and tries to recognize objects. When the system fails in recognition due to the imperfect color transformation, the system asks the user to help object recognition. Then the system updates the parameters with correspondent color pairs obtained from the correct object recognition. Every time the system recognizes an object, it repeats this update process. As a result, the system improves color transformation and adapts to the new environment. This adaptation improves object recognition performance and reduces user's burdens.

The outline of this paper is as follows. In sec. 2, we give overview of our recognition system. In sec. 3, we introduce a physics-based color model and formulate relation of color transformation and the observed color of the reference object. In sec. 4, we describe a learning method of the color transformation. We show experiments of color transformation using objects in a real-life refrigerator in sec. 5 and give a conclusion in sec. 6. We also deal with a dialogue system via voice and an object manipulation system, and give details of these topics in [8] and [9] respectively.

2. Overview

First, the system constructs object models in advance. The model is composed of a texture image of the object (see Fig. 1(b)). Then a user asks the system to bring an object via voice. Because lighting conditions in model construction and object recognition are different, the system estimates a current lighting condition (see the following sections for details) and transforms the color of the object model (see Fig. 1(c)). Last the system recognizes the object with the transformed model as automatically as possible (see Fig. 1(d)). If the system fails in recognition, it recovers the result by user interaction. Details of these recognition processed are described in [10] and [11].

(a) Original image (b) Original object model (c) Transformed object model (d) Recognition result

Figure 1. Overveiw of object recognition

3. Color Estimation under Unknown Lighting Conditions

In this section, we describe color estimation method briefly. Details are described in [7]. First, we formulate a color transformation model with a physics-based light reflection model. According to a well-known finite-dimensional linear model [12], an observed color $\boldsymbol{c}^{OL} = [c_1^{OL}, c_2^{OL}, c_3^{OL}, c_4^{OL}]^T = [R, G, B, 1]^T$ of an object O under a lighting condition L is expressed as

$$c_p^{OL} = \sum_{k=1}^{4} \sum_{l=1}^{4} f_{pkl}\, s_k^O\, e_l^L \qquad (1)$$

where s_k^O and e_l^L represent coefficients for surface reflectance of the object O and received light under the lighting condition L respectively. Then we introduce color transformation between two colors $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ under lighting condition L_1 and L_2 as follows,

$$\boldsymbol{c}^{OL_2} = A^{L_2 L_1} \boldsymbol{c}^{OL_1} \qquad (2)$$

where $A^{L_2 L_1}$ is a color transformation matrix.

Next, in order to estimate the color $\boldsymbol{c}^o$ of a target object under an unknown lighting condition in object recognition, we use a color of a reference object O_r (the refrigerator's door or frame in this case). An outline of this problem is

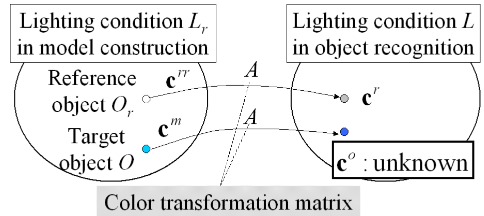

Figure 2. Problem setting of color estimation

estimating $\boldsymbol{c}^o$ given a model color $\boldsymbol{c}^m$ of the target object and an observed color $\boldsymbol{c}^r$ of the reference object under the unknown lighting condition in object recognition (see Fig. 2). We assume that the color transformation matrix A is expressed with $\boldsymbol{c}^r$. Then $\boldsymbol{c}^m$ is transformed into the estimated color $\hat{\boldsymbol{c}}^o$ in object recognition with the color transformation matrix A. These relations are expressed as the following two equations.

$$\hat{\boldsymbol{c}}^o = A(\boldsymbol{c}^r; \boldsymbol{\alpha})\, \boldsymbol{c}^m \qquad (3)$$

$$A(\boldsymbol{c}^r; \boldsymbol{\alpha})_{ij} = \sum_k \alpha_{ijk} c_k^r \qquad (4)$$

where $\boldsymbol{\alpha}$ is a vector including all the parameters α_{ijk} for color transformation estimation. The former means the model color $\boldsymbol{c}^m$ of target object is transformed linearly with transformation matrix $A(\boldsymbol{c}^r; \boldsymbol{\alpha})$ and the latter means each component of the transformation matrix $A(\boldsymbol{c}^r; \boldsymbol{\alpha})$ is written as a linear combination of the observed color $\boldsymbol{c}^r$ of the reference object. Because the parameters $\boldsymbol{\alpha}$ vary for each environments, we need to learn $\boldsymbol{\alpha}$ for the environment. We describe the learning in the following section.

4 Learning of Parameters for Color Transformation Estimation

When we have enough learning sets of color pair samples for a target environment, we calculate the parameters $\boldsymbol{\alpha}$ by the least square method of color transformation errors under the lighting condition in model registration. In addition, we normalize the errors for each object because observed color variations are different due to the difference of their surface reflectance. As a result, we obtain the parameters $\boldsymbol{\alpha}$ by minimizing the following objective function.

$$S = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \boldsymbol{\Delta}_{ij}^T \Sigma_i^{-1} \boldsymbol{\Delta}_{ij} \qquad (5)$$

$$\boldsymbol{\Delta}_{ij} = \boldsymbol{c}_i^m - A(\boldsymbol{c}_{ij}^r; \boldsymbol{\alpha})^{-1} \boldsymbol{c}_{ij}^o \qquad (6)$$

where N is the number of objects, n_i is the number of observation times for each object i, Σ_i is a covariance matrix of color transformation error for each i.

When the system tries to recognize objects in a new environment, it has no color pair samples. Then the system first assumes that the observed color of the reference object in model registration is white [255, 255, 255] and estimates a color transformation of only brightness compensation defined as $A_{ij} = \delta_{ij} Y_r/255$, where Y_r is brightness component of the color $\boldsymbol{c}_r$ and δ_{ij} is Kronecker's delta. If this estimation is inaccurate, the system may fail in automatic recognition. Then the system tries to recover the failure with user interaction, that is, the user tells the system the correct object name of the image region by speech (see [11] for details). As a result, the system obtains a true correspondence of the image region and the target object model. Thus the system stores a new observed color pair of the object and also the reference object. Because one color pair alone is not enough to calculate all of the parameters $\boldsymbol{\alpha}$ for color transformation estimation, the system calculates only important parameters and constrains the others to be zero. For the next time, the system tries to recognize an object with the partially constrained parameters and obtains another new observed color pair. Then the system updates the parameters again. If the system continues to use the initial parameters, the color transformation errors increase due to lack of degrees of freedom (DOF). Therefore the system should relax the constraints of the parameters: it calculates new parameters with the new observed color pair.

The main problem is how to relax the constraints. This problem is mainly divided into the following two subproblems.

1. The priority of the constraints to relax
2. Timing to relax each constraint (How many color pairs are required to relax the constraints?)

[The priority of the constraints to relax]

For simplicity, according to the element of the color transformation matrix defined in eq. (4), we first classify the parameters into $n_p = 9$ groups: $G_1\{\alpha_{iii}\}$, $G_2\{\alpha_{ii4}\}$, $G_3\{\alpha_{i44}\}$, $G_4\{\alpha_{i4i}\}$, $G_5\{\alpha_{iji}\}$, $G_6\{\alpha_{iij}\}$, $G_7\{\alpha_{ij4}\}$, $G_8\{\alpha_{i4j}\}$, and $G_9\{\alpha_{ijk}\}$, where $1 \leq i, j, k \leq 3, j \neq i, k \neq i$. The classification is based on properties of the parameters such as diagonal terms, cross terms, and constant terms. We determine the priority of the constraints to relax based on each group's contribution to recognition performances. The recognition performances are judged by adaptation experiments: experiments on object recognition using test color pairs (see sec. 5 for detail) for cases of relaxing the constraints of each group. As a result, the priority turns out to be the same as the above G_i order. We define parameters including only brightness compensation as $\boldsymbol{\alpha}_0$ and parameters whose constraints of groups from G_1 to G_i are relaxed as $\boldsymbol{\alpha}_i$. Thus, using $\boldsymbol{\alpha}_9$ corresponds to using all the parameters. We switch the parameters from $\boldsymbol{\alpha}_i$ to $\boldsymbol{\alpha}_{i+1}$ according to the increase of the number n_c of observed color pairs.

[Timing to relax each constraint]

If we use the parameters $\boldsymbol{\alpha}_i$ calculated with the minimal number of color pairs, the performance is often get worse because of over learning. Therefore we adopt a method to relax the parameter groups when n_c becomes enough large. The problem is how to determine the concrete number of color pairs for relaxation. Unfortunately, the optimal number of color pairs depends on cases. If we use the fixed number for every case, we run the risk of rapid downs of recognition performance as a result of too early relaxation. Therefore, we leave the effect of previous parameters after the relaxation by defining parameters $\boldsymbol{\alpha}$ as a weighted linear combination of each $\boldsymbol{\alpha}_i$ and by changing the weights according to increase of n_c. These are expressed as the following equations,

$$\boldsymbol{\alpha} = \sum_{i=0}^{n_p} w_i \boldsymbol{\alpha}_i, \quad w_i = \frac{d_i}{\sum_{i=0}^{n_p} d_i} \tag{7}$$

$$d_i = \begin{cases} \delta_{i0} & (n_c < (2/3)n_i) \\ 1 - (1-\delta_{i0})\frac{(4/3)n_i - n_c}{(2/3)n_i} & ((2/3)n_i \leq n_c < (4/3)n_i) \\ 1 - (1-\delta_{in_p})\frac{n_c - (4/3)n_i}{(2/3)n_i} & ((4/3)n_i \leq n_c < 2n_i) \\ \delta_{in_p} & (n_c \geq 2n_i) \end{cases} \tag{8}$$

where w_i is a weight for $\boldsymbol{\alpha}_i$, n_i is the number of relaxed parameter components in $\boldsymbol{\alpha}_i$, δ_{ij} is Kronecker's delta. The above equations indicate gradual transition of the parameters based on the weights control. We first define the 4 domains for the weights according to n_c. In the first domain (less than $(2/3)n_i$), we let the weight w_i for $\boldsymbol{\alpha}_i$ be zero so as to prevent over learning. In the second interval (from $(2/3)n_i$ to $(4/3)n_i$), we let it be gradually large so as to increase the effect of the parameter $\boldsymbol{\alpha}_i$. In the third domain (from $(4/3)n_i$ to $2n_i$), we let it be gradually small so as to prevent poor representation of color transformation due to lack of DOF. In the forth domain (more than $2n_i$), we let it be zero. Here the domain boundaries $(2/3)n_i$, $(4/3)n_i$, and $2n_i$ are determined by adaptation experiments (see sec. 5 for detail).

5. Experiments

We made adaptation experiments on color transformation estimation in new environments. First we divide stored color pairs into a learning set and a test set. Next, we pick up a color pair of the learning set one by one. Each time a color pair is picked up, the system updates the parameters $\boldsymbol{\alpha}$ and makes recognition tests for the test sets. We made this experiments for 10 times with various learning sets and test sets.

Fig. 3 shows average recognition ratios in a test environment used for determining the priority of the constraints to relax and timing to relax each constraint described in section 4. Here we show recognition ratio of $\boldsymbol{\alpha}_0$, $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\alpha}_5$, $\boldsymbol{\alpha}_8$, $\boldsymbol{\alpha}_9$, and $\boldsymbol{\alpha}$. When we use the parameters with low DOF such as $\boldsymbol{\alpha}_0$, $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$, we obtain relatively high recognition ratios in the early phase, but their performances make

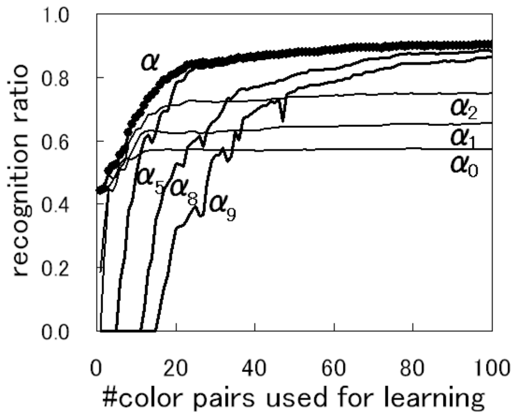

Figure 3. Pretest of learning performance

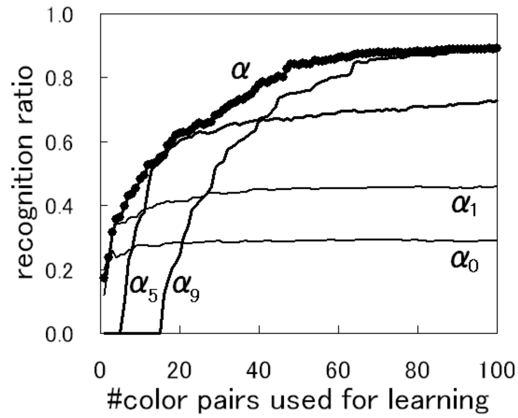

Figure 4. Learning performance in a new environment

little progresses in the latter phase due to increase of color transformation errors. On the other hand, when we use the parameters with high DOF such as $\boldsymbol{\alpha}_5$, $\boldsymbol{\alpha}_8$, and $\boldsymbol{\alpha}_9$, we obtain relatively low recognition ratios in the early phase because of over learning, but their performances make gradual progresses and exceed those of the parameters with low DOF in the latter phase. Thus, when we use $\boldsymbol{\alpha}$, the weighted linear combination of $\boldsymbol{\alpha}_i$ defined in eqs. (7) and (8), we can gradually transit the parameters from those with low DOF in the early phase to those with high DOF in the latter phase. Therefore we achieve as high recognition ratio as possible in each phase (see the bold line in Fig. 3). The reason of approximate 10% recognition failures after enough learning is the effect of local change of lighting condition in the refrigerator. These recognition failures should be recovered by user interaction.

Next we made an experiment for a new environment whose reference color is light green, which is different from white reference in the test environment. Fig. 4 shows learning results of the parameters for color transformation estimation for the environment. As a result, we achieve as high recognition ratio as possible in each phase and can adapt the system to the new environment.

6. Conclusion

We proposed an online learning method of color transformation for interactive object recognition under various lighting conditions. First, we formulated color transformation estimation using the reference object. Next, we proposed online learning in a new environment by relaxing constraints of the parameters for color transformation estimation according to the observed number of color pairs samples. Then we experimentally decided the priority of the constraints to relax and the timing to relax each constraint. Finally, we showed the effectiveness of our method by adaptation experiments with some real-life refrigerator scenes under various lighting conditions.

Future works are as follows.

- Adaptation to local changes of lighting conditions in the refrigerator
- Recognition of specular objects
- Recognition of objects with various shapes such as bottles and fruits.

References

[1] Y. Takahashi et al. Development of the mobile robot system to aid the daily life for physically handicapped. Proc. of Int. Conf. on Machine Automation 2000, 549-554, 2000.

[2] T. Takahashi et al. Human-Robot Interface by Verbal and Nonverbal Communication. Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, 924–924, 1998.

[3] S. Nadimi et al. Moving Shadow Detection Using a Physics-based Approach. Proc. of 16th Int. Conf. on Pattern Recognition, 2:701-706, 2002.

[4] D. H. Marimont and B. A. Wandell. Linear models of surface and illuminant spectra. J. Opt. Soc. Amer., 11, 1992.

[5] M.S. Drew, Z.N. Li, and Z. Tauber. Illumination color covariant locale-based visual object retrieval. Patter Recognition, 35(8):1687-1704, 2002

[6] E. G. Miller et al. Color Eigenflows: Statistical Modeling of Joint Color Changes. Proc. of 8th Int. Conf. on Computer Vision, 1:607-614, 2001.

[7] Y. Makihara, M. Takizawa, Y. Shirai, and N. Shimada. Object Recognition in Various Lighting Conditions. Proc. of 13th Scandinavia Conf. on Image Analysis, 899-906, 2003.

[8] M. Takizawa, Y. Makihara, N. Shimada, J. Miura, and Y. Shirai. A Service Robot with Interactive Vision -Object Recognition Using Dialog with User. Proc. of 1st Int. Workshop on Language Understanding and Agents for Real World Interaction, 16-23, 2003

[9] J. Miura, Y. Shirai, N. Shimada, Y. Makihara, M. Takizawa, and Y. Yano. Development of a Personal Service Robot with User-Friendly Interfaces. Proc. of the 4th Int. Conf. on Field and Service Robotics, 293-298, 2003.

[10] Y. Makihara et al. Object Recognition Supported by User Interaction for Service Robots. Proc. of 16th Int. Conf. on Pattern Recognition, 3:561-564, 2002

[11] Y. Makihara et al. Object Recognition Supported by User Interaction for Service Robots. Proc. of 5th Asian Conf. on Computer Vision, 2:719-724, 2002

[12] D. B. Judd et al. Spectral distribution of typical daylight as a function of correlated color temperature. J. Opt. Soc. Am., 54:1031-1040, 1964.