MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1523 February, 1995

Recognizing 3D Objects Using Photometric Invariant

Kenji NAGAO **W. Eric. L. GRIMSON**

This publication can be retrieved by anonymous ftp to publication.ai.mit.edu.

Abstract

In this paper we describe a new efficient algorithm for recognizing 3D objects by combining photometric and geometric invariants. Some photometric properties are derived, that are invariant to the changes of illumination and to relative object motion with respect to the camera and/or the lighting source in 3D space. We argue that conventional color constancy algorithms can not be used in the recognition of 3D objects. Further we show recognition does not require a full constancy of colors, rather, it only needs something that remains unchanged under the varying light conditions and poses of the objects. Combining the derived color invariants and the spatial constraints on the object surfaces, we identify corresponding positions in the model and the data space coordinates, using centroid invariance of corresponding groups of feature positions. Tests are given to show the stability and efficiency of our approach to 3D object recognition.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-91-J-4038. WELG was also supported by NSF grant IRI-8900267.

1 Introduction

A typical approach to model-based object recognition[14] matches stored geometric models against features extracted from an image, where the features are typically localized geometric events, such as vertices. Objects are considered to have undergone a transformation in space to yield a novel view for the image. To solve for this transformation explicitly, recognition methods use matches of features to hypothesize a transformation, which is used to align the model with the image and select the best-fit pair of transformation and model. While this approach to recognition has achieved considerable success, there still remain practical problems to be solved.

One such problem is the computational complexity of the method. For example, even with popular algorithms (e.g.[20, 32]), to recognize an object with m features from an image with n features, we must examine m^3n^3 combinations of hypotheses where m and n can be easily on the order of several hundreds in natural pictures. A second problem is the tolerance of the algorithm to scene clutter. To verify the hypothesized transformation, object recognition algorithms have to collect evidence of actual correspondences characterized by that transformation. This is usually done by looking for nearest image features around the transformed model features, or equivalently by casting votes to a hash table of parameters, such as affine invariant parameters, leading to a correspondence (e.g.[24]). In either case, when features are extracted from the image with perturbations, and if the image is cluttered so that the feature distribution is too dense, it is difficult to tell whether an image feature thus detected is the one actually corresponding to the model feature or if it just happened to fall close to the transformed model feature. This issue has been extensively analyzed, both theoretically and empirically, giving arguments about the limitations of geometric feature based approaches to recognition (e.g.[15, 1, 14]).

Considering the limitations of conventional approaches to recognition which depend solely on geometrical features, it is natural to start using other cues than simple local geometric features. One such candidate is photometric information like color, because we know that color often characterizes objects well and it is almost invariant to the change of views and lighting conditions. In parallel with geometry, color properties of the object surface should be a strong key to the perception of the surface. However, most authors who have exploited color in recognition used it simply for segmentation, e.g., [5, 29, 16], mostly because color is considered to be more contributive in building up salient features on the object surface than in giving precise information on the location and the poses of the objects. Exceptions include Swain[27, 28] and Nayar et. al. [26] who have used photometric information more directly for recognition, respectively for indexing and matching processes. At the same time, however, they abandoned the use of local geometric features, which still is very useful in predicting the locations and the poses of the objects. Swain used only a color histogram for representing objects and matched it over the image to identify the object included and localize its presence in the image. Nayar et al. proposed a photometric invariant for matching regions with consistent colors given the partitioned model and image derived by some other color properties. Therefore, it requires a preliminary segmentation of the image into regions having consistent colors.

In this paper, we attempt to exploit both geometric and photometric cues to recognize 3D objects, by combining them more tightly. Our goal is to develop an efficient and reliable algorithm for recognition by taking advantage of the merits of both geometric and color cues: the ability of color to generate larger and thus more salient features reliably, as well as of adding more selectivity to features, which enables more efficient and reliable object recognition, and the rich information carried by the set of local geometric features that is useful in accurately recovering the transformation that generated the image from the model. To realize this, we have developed new photometric invariants which are suitable for this approach. Then, we combine the proposed photometric properties with the Centroid Alignment approach of corresponding geometric feature groups in the model and the image, that we have recently proposed [25]. This strategy gives an efficient and reliable algorithm for recognizing 3D objects. In our testing, it took only 0.2 seconds to derive corresponding positions in the model and the image for natural pictures.

2 Some photometric invariants

In this section, we develop some photometric invariants that can be used as strong cues in the recognition of 3D objects. The invariant is related to the notion of *color constancy*, that is — whether in human or machine vision — the perceptual ability to determine the surface reflectance property of the target objects given the reflected light from the object surface in the receptive field. If a color constancy algorithm could perform sufficiently well, we could use it for object recognition because it would provide a unique property of the object itself. Unfortunately, however, color constancy is generally difficult to compute in practice, so we can not use it by itself. The invariant property to be presented here is efficiently computed from the segmented/non-segmented images at the same time as the geometrical features are extracted.

2.1 Unavailability of color constancy

Color constancy is an underconstrained problem, as we will see in the following. Let $S(\mathbf{x}, \lambda)$ be the spectral reflectance function of the object surface at $\mathbf{x}$, that is the property we have to recover, let $E(\mathbf{x}, \lambda)$ be the spectral power distribution of the ambient light, and let $R_k(\lambda)$ be the spectral sensitivity of the kth sensor, then $\rho_k(\mathbf{x})$, the scalar response of the kth sensor channel to be observed, is described as

$$\rho_k(\mathbf{x}) = \int S(\mathbf{x}, \lambda)E(\mathbf{x}, \lambda)R_k(\lambda)d\lambda \tag{1}$$

where, generally, S is a function describing geometric and spectral properties of the surface at $\mathbf{x}$ that can be an arbitrary function and E could also be an arbitrary function of $\mathbf{x}$ and λ. The integral is taken over the visible spectrum(usually from 380 to 800 nm). The geometric factor of the object surface, that is usually considered to include the surface normal and the relative angle of the incident and reflecting light direction with respect to the surface normal, is very crucial in the 3D world[18]. In addition, there are also other confounding factors such as specularities and mutual reflections on the surface. With these complexities, to perform color constancy, that is to recover $S(\mathbf{x}, \lambda)$, we need to limit the world to which it is applied. To get a simple intuition of this, for example, we might insert an arbitrary scalar function $C(\mathbf{x})$ in (1) so that we have[33],

$$\rho_k(\mathbf{x}) = \int \{S(\mathbf{x}, \lambda)C(\mathbf{x})\}\{E(\mathbf{x}, \lambda)/C(\mathbf{x})\}R_k(\lambda)d\lambda. \quad (2)$$

Clearly, when S with E is a solution for (1), $S' = SC$ with $E' = E/C$ is also a solution for any function C. To turn this into a well-posed problem, almost all authors have addressed problems in a strongly constrained world like Mondarian space [19, 13, 33, 31, 12, 9]: a 2D space composed of several matte patches overlapping each other. Then, based on the observation that both the ambient light and the surface reflectance for planar surfaces can be approximated by linear combinations of a small number of fixed basis functions[7, 21], they can deal with the problem at a fairly feasible level[13, 33, 31, 12, 10, 9]. A good mathematical analysis is given in [10]. However, all of those results are for a 2D world. This two-dimensionality assumption takes away any chance of conventional color constancy being used in recognizing a 3D world. Therefore, we can not employ conventional color constancy algorithms as presented.

2.2 Some color invariants

Knowing that color constancy is not easily attainable for any plausible 3D world, we propose a photometric invariant property for use in the recognition of 3D objects.

Since it is known that a spectrum distribution of the surface reflectance of many materials depends very little on the surface geometry[23], we may break up the surface reflectance function into the product of geometry $G(\mathbf{x})$ and spectrum property $L(\mathbf{x}, \lambda)$ such that $S(\mathbf{x}, \lambda) = G(\mathbf{x})L(\mathbf{x}, \lambda)$. Then, the equation (1) becomes:

$$\begin{aligned} \rho_k(\mathbf{x}) &= \int G(\mathbf{x})L(\mathbf{x}, \lambda)E(\mathbf{x}, \lambda)R_k(\lambda)d\lambda \\ &= G(\mathbf{x}) \int L(\mathbf{x}, \lambda)E(\mathbf{x}, \lambda)R_k(\lambda)d\lambda \qquad (3) \end{aligned}$$

[Constant ambient light assumption over the entire surface]

If we assume that the ambient light spectrum distribution is constant over the entire surface of the objects, E becomes simply a function of wavelength λ. This assumption is justified when the lighting source is sufficiently far away from the object relative to the size of the object surface, and mutual illumination and shadowing are not significant. This yields

$$\rho_k(\mathbf{x}) = G(\mathbf{x}) \int L(\mathbf{x}, \lambda)E(\lambda)R_k(\lambda)d\lambda \qquad (4)$$

Taking the ratios between the two i, j channel responses eliminates the geometric factor $G(\mathbf{x})$ which depends on the relative orientation of the object surface with respect to the camera and/or the lighting source,

$$\frac{\rho_i(\mathbf{x})}{\rho_j(\mathbf{x})} = \frac{\int L(\mathbf{x}, \lambda)E(\lambda)R_i(\lambda)d\lambda}{\int L(\mathbf{x}, \lambda)E(\lambda)R_j(\lambda)d\lambda} \qquad (5)$$

By the same reasoning, we have a similar form after the motion of the object with respect to the camera and/or the lighting source,

$$\frac{\rho'_i(\mathbf{x}')}{\rho'_j(\mathbf{x}')} = \frac{\int L'(\mathbf{x}', \lambda)E'(\lambda)R_i(\lambda)d\lambda}{\int L'(\mathbf{x}', \lambda)E'(\lambda)R_j(\lambda)d\lambda} \qquad (6)$$

where primes show the function after the motion, and this prime notation applies to any symbol expressing some quantity after the motion of the object in the rest of this paper unless otherwise described. Note that $L(\mathbf{x}, \lambda) = L'(\mathbf{x}', \lambda)$, because the spectrum property of the surface reflectance would not be affected by the object motion. When we approximate the spectral absorption functions R by narrow band filters such that $R_i(\lambda) \approx s_i\delta(\lambda_i - \lambda)$, where s_i is the channel sensitivity and the λ_i is the peak of the spectral sensitivity of the ith channel, we obtain ratios from (5) and (6):

$$\begin{aligned} \gamma_{ij}(\mathbf{x}) &\equiv \frac{\rho_i(\mathbf{x})}{\rho_j(\mathbf{x})} \\ &\approx \frac{s_i L(\mathbf{x}, \lambda_i)E(\lambda_i)}{s_j L(\mathbf{x}, \lambda_j)E(\lambda_j)} \qquad (7) \\ \gamma'_{ij}(\mathbf{x}') &\equiv \frac{\rho'_i(\mathbf{x}')}{\rho'_j(\mathbf{x}')} \\ &\approx \frac{s_i L(\mathbf{x}, \lambda_i)E'(\lambda_i)}{s_j L(\mathbf{x}, \lambda_j)E'(\lambda_j)} \qquad (8) \end{aligned}$$

Since the band width over which a real camera sensor responds varies from camera to camera, and the standard ones may not be too narrow, this is only an approximation. However, experiments show that this assumption is not unrealistic for the normal cameras. Taking the ratio of γ's before and after the motion and/or the change of lighting conditions yields,

$$\frac{\gamma_{ij}(\mathbf{x})}{\gamma'_{ij}(\mathbf{x}')} \approx \epsilon_{ij} \qquad (9)$$

where

$$\epsilon_{ij} = \{\frac{E(\lambda_i)}{E(\lambda_j)}\}/\{\frac{E'(\lambda_i)}{E'(\lambda_j)}\} \qquad (10)$$

Since ϵ_{ij} is apparently independent of the position on the surface, $\gamma_{ij}(\mathbf{x})$ can be regarded as approximately invariant to the changes of illuminant conditions and to the motions of the object within a consistent scale factor over the object surface. Note that ϵ_{ij} depends only on the ratios of spectrum distribution of the ambient light before and after the motion of the object.

In using γ for object recognition, we might need to normalize its distribution because generally it is invariant only within a scale factor. When we are provided with the sets of γ from corresponding positions over different views, this could be done by applying a normalization process to the original sets:

$$\hat{\gamma_{ij}} = \sigma_{ij}^{-\frac{1}{2}}\gamma_{ij} \tag{11}$$

where σ_{ij} is the variance of the given γ_{ij} distribution. Note that when the ambient light has not been changed, $\epsilon_{ij} = 1$, so that $\gamma_{ij}(\mathbf{x}) = \gamma'_{ij}(\mathbf{x}')$, thus normalization process is not needed.

[Only locally constant ambient light assumption] Now, let us assume only a locally constant ambient light spectrum distribution, instead of the globally constant one over the object surface: $E(\mathbf{x}_l, \lambda) = E(\mathbf{x}_m, \lambda)$ for nearby positions $\mathbf{x}_l, \mathbf{x}_m$. Then, eqs. (7) and (8) must be modified respectively as:

$$\begin{aligned} \gamma_{ij}(\mathbf{x}) &\equiv \frac{\rho_i(\mathbf{x})}{\rho_j(\mathbf{x})} \\ &\approx \frac{s_i L(\mathbf{x}, \lambda_i) E(\mathbf{x}, \lambda_i)}{s_j L(\mathbf{x}, \lambda_j) E(\mathbf{x}, \lambda_j)} \end{aligned} \tag{12}$$

$$\begin{aligned} \gamma'_{ij}(\mathbf{x}) &\equiv \frac{\rho'_i(\mathbf{x}')}{\rho'_j(\mathbf{x}')} \\ &\approx \frac{s_i L(\mathbf{x}, \lambda_i) E'(\mathbf{x}', \lambda_i)}{s_j L(\mathbf{x}, \lambda_j) E'(\mathbf{x}', \lambda_j)} \end{aligned} \tag{13}$$

Incorporating the assumption, that is, $E(\mathbf{x}_l, \lambda) = E(\mathbf{x}_m, \lambda)$, and $E'(\mathbf{x}'_l, \lambda) = E'(\mathbf{x}'_m, \lambda)$, we again have an invariant ψ_{ij}^{lm}:

$$\begin{aligned} \psi_{ij}^{lm} &\equiv \frac{\gamma_{ij}(\mathbf{x}_l)}{\gamma_{ij}(\mathbf{x}_m)} \\ &\approx \{\frac{L(\mathbf{x}_l, \lambda_i)}{L(\mathbf{x}_l, \lambda_j)}\}/\{\frac{L(\mathbf{x}_m, \lambda_i)}{L(\mathbf{x}_m, \lambda_j)}\} \end{aligned} \tag{14}$$

thus, apparently, $\psi_{ij}^{lm} \approx \psi_{ij}^{lm\,\prime}$. However, ψ_{ij}^{lm} is obviously sensitive to perturbations contained in the image signals especially when one makes the values of $\gamma_{ij}(\mathbf{x}_m)$ (the denominator in (14)) close to zero. To stabilize this, we adopt a normalized measure in place of ψ itself:

$$\varphi_{ij}^{lm} \equiv \frac{\gamma_{ij}(\mathbf{x}_l)}{\gamma_{ij}(\mathbf{x}_m) + \gamma_{ij}(\mathbf{x}_l)} \tag{15}$$

It is easy to see $\varphi \approx \varphi'$, that is, φ is approximately invariant to the change of illumination conditions and of orientations of the object surfaces.
Note that for γ_{ij} we can not derive this kind of normalized invariant formula. A very important thing to remember here is that in order to make φ useful, the surface reflectance properties associated with two nearby positions $\mathbf{x}_l, \mathbf{x}_m$ to be picked up must be sufficiently different from each other. Otherwise, even if an invariant of φ in (15) holds true, as the γ's tend to have the same value for $\mathbf{x}_l, \mathbf{x}_m$, the φ's always return values that are close to 0.5, so that it does not provide any useful information involved in their color properties. Fortunately, as we describe later when color properties are picked up from different sides of the brightness boundaries, this situation may often be avoided.

2.3 Related photometric invariants

A related invariant to our photometric invariants was proposed earlier based on an *opponent color model* by Faugeras for image processing applications[8]. The opponent color model was first introduced by Hering[17] to describe the mechanism of human color sensation. He advocated that the three pairs Red-Green, Blue-Yellow, White-Black form the basis of human color perception. A simple mathematical formulation of this[3], which is a linear transformation of R, G, B was used as a color invariant in [27, 28] for indexing 3D objects: $[\text{R-G,Bl-Y,W-Bk}]^T = \text{L}[\text{R,G,B}]^T$, where L is a linear transformation. A similar formalization of an opponent color model was also used for the correspondence process in color stereopsis [5]. However, there are no theoretical explanations of the linear transformation model for the full 3D object surfaces, because, as we noted in the derivation of our invariants, the surface orientation in 3D space with respect to the lighting source and the camera is an unignorable factor (see also [18]) in deriving invariants for a 3D world, and it is never removed by any linear transformation.

Unlike this linear transformation case, Faugeras's form is the logarithm of the ratios between different channel responses for a chromatic model, so is similar to ours, and the logarithm of the products of three of R, G, B responses but with a low-pass filtering accounting for lateral inhibition for achromatic responses.

In [4] a unique illuminant-invariant was proposed which, assuming the existence of at least four local distinct color surfaces, uses the volumetric ratio invariant of the parallelepiped generated by the responses of the three receptors. It seems to us, however, that the assumption of four local distinct color surfaces is demanding too much in practice.

Recently, a new photometric invariant was proposed for object recognition[26]. Limiting its application to only geometrically continuous smooth surfaces, it used as an invariant the ratio between the brightnesses of two adjacent regions each with consistent and different surface spectral reflectance. Therefore, it requires a preliminary complete segmentation of the image into regions having the same colors. Other assumptions introduced in its derivation are almost the same to ours (locally constant ambient illuminant case) except for the additional continuous smooth surface constraint over the boundary of two surfaces with different spectral reflectance.

2.4 Experiments

Experiments were conducted to examine the accuracy of the proposed photometric invariants. Figure 1 shows pictures of a man-made convex polyhedron composed of 6 planar surfaces each with a different surface orientation. The left picture is a front view of the polyhedron, hereafter pose P_A, while in the right picture the object is rotated around the vertical axis (y-axis) by about 30 degrees, hereafter pose P_B. On each side of the boundary of adjacent surfaces, several matte patches with different colors were pasted. Then, we picked up corresponding positions manually within each colored patch in the pictures for the poses (P_A, P_B). The selected positions within patches are depicted by crosses in the pictures. To test the accuracy of the proposed invariants γ, φ under varying illuminant conditions and surface orientations of the object with respect to the illuminant and the camera, we took three pictures: the first at the pose P_A under the usual lighting conditions ($P_A \& L_U$), the second at the pose P_B under a greenish light ($P_B \& L_G$), and the third at the pose P_B but under a bluish light ($P_B \& L_B$). To change the source light spectrum, i.e., to get greenish or bluish light, we covered a tungsten halogen lamp with cellophane of colors green and blue. For φ, the surface positions within planar patches facing over the boundaries of planar surfaces were used as neighboring positions to satisfy the requirement of (locally) constant ambient light. To compute the invariants in practice, we used the ratios $G/R, B/R$ for γ and $\varphi_1 = (G^1/R^1)/(G^1/R^1 + G^2/R^2)$, $\varphi_2 = (B^1/R^1)/(B^1/R^1 + B^2/R^2)$ for φ, where R, G, B are the outputs from the sensor channels respectively of Red, Green, Blue, and the indices attached to R, G, B shows the sides of the surfaces used for computing φ's with respect to their boundaries. As described previously, in theory, when we use the RGB channel outputs to compute invariants, instead of outputs through the exact narrow band filters, they might be only pseudo-invariants. But, the following results confirm that the values of γ and φ computed using RGB are fairly invariant to the changes of the illumination conditions as well as the surface orientations. In Table 1, the correlation coefficients between the sets of values for each invariant measure computed at corresponding positions in different pictures are given, that are measured by the following formula:

$$\sqrt{\frac{C^2_{\alpha\alpha'}}{C_{\alpha\alpha}C_{\alpha'\alpha'}}} \tag{16}$$

where C_{ab}'s ($a, b \in \{\alpha, \alpha'\}$) are the covariances between the sets of the values of the measure α (e.g., γ) before (α) and after (α') the motion of the objects or the changes of the lighting conditions, which is defined by:

$$C_{ab} = \sum P(a, b)(a - \bar{a})(b - \bar{b}) \tag{17}$$

where $\bar{x}$ is the average of the measure x, $P(a, b)$ is the probability density function, and the sum is taken over all corresponding values of the measures a, b. A high correlation, that gives a value close to 1, shows that the proposed invariant measures remained unchanged within a consistent scale over the set of positions between the two picture, while a low correlation, that gives a value close to 0, means that the values of the measures changed in a irregular manner. For comparison, other color properties including raw (R, G, B), (H, S, V) (hue,saturation,value), and a linear-transformation implementation of the opponent color model[3] are also included. In these tests, R, G, B, $R - G, B - Y$, $\gamma = G/R, B/R$, are almost equally good, though γ is the best among them on average, that mean those properties have been changed but only within a consistent scale between the different pictures (recall the property of γ being invariant within a scale factor). The reason why R is very good is probably just that we did not happen to change the intensity of the red light spectrum. The values of H, S, V is unexpectedly quite unstable. The measure φ is extremely stable. To see how far the color properties remained unchanged in addition to the correlative relation, in Figure 2 the actual distribution of the color properties are shown, where the horizontal axes are the values for the pose P_A, while the vertical axes are those for the pose P_B. If the color measures remained unchanged between the two pictures before and after the motions of the object and/or the changes of the light conditions, the distributions should present linear shapes, and their slopes should be close to 1. Indeed, the measure φ is certainly found to remain almost unchanged under varying light conditions, while other color properties H, S, and $\gamma = G/R, B/R$ included for comparison are found not. The biases of the slopes of γ either toward the horizontal or vertical axes indicate that the light spectrum has been changed between the two compared pictures. Figure 3 shows the performance of γ constancy against the change of the object pose, under the same lighting conditions. In other words, unlike in the last experiments, this time the ambient light has not been changed for both of the two pictures, and only the object pose has been changed. For comparison, the performance of $B - Y$ (linear-trans implementation for blue vs. yellow, the second figure from the left) as well as raw B (blue, the first one) are also shown. Note that what should be observed here is how the slopes of the distributions are close to 1. Except for the two samples in the upper area in the figure (the fourth picture), $\gamma = B/R$ is found to be almost unchanged between the two pictures. The two exceptional samples were from patches with almost saturated blue channel in the picture at pose P_B. The performance of $\gamma = G/R$ (the third figure) is almost perfect. On the other hand, $B - Y$ and B are perturbed around the slope of 1, which is probably caused by the perturbed orientations of the patches. This suggests that γ may be used for object recognition without applying any normalization process, so that extracting object regions might not be a prerequisite, as long as the lighting conditions are not changed.

Similarly, in Table 2 the results of the same tests as above but on a natural object, a doll which is shown in Figure 4, are given, for which both the ambient light and the object pose were changed. We refer to the pose of the doll similarly to the above tests on the Test-Object: left pose P_A, right pose P_B. The first picture was taken under a usual lighting conditions from the oblique angle($P_A \& L_U$), the second and third were taken respectively under a greenish and a bluish light from the front angle($P_B \& L_G$, $P_B \& L_B$). Corresponding positions were picked up manually as done in the previous tests. As the surface colors varied smoothly, we can not expect that we could pick up corresponding points accurately. Thus, unwanted errors could be introduced in this operation. This time for φ, two positions which are closest to each other among the selected points are used. In this tests, R, G, B and S, V performed poorly, though somehow H was very good. The linear model $R-G, B-Y$ and $\gamma = G/R, B/R$ performed well again, though γ was better. The measure φ is quite stable again. Unlike the results on the Test-Object, however, since the surface of the doll, especially in the body parts, had similar surface colors in near positions, the distribution of φ — $\varphi_1 = (G^1/R^1)/(G^1/R^1 + G^2/R^2)$, $\varphi_2 = (B^1/R^1)/(B^1/R^1 + B^2/R^2)$ — did not spread very well, thus having a weak selectivity photometrically, as seen in Figure 5. Therefore, when picking up two nearby positions for φ for object recognition, it is important that they have different spectral reflectance. For comparison, the values of H, S, and γ are also plotted in Figure 5.

2.5 Sensing limitations

As we note in the examination above, the invariant properties are sometimes perturbed around the ideal values which support our theories. This is caused mainly by the limited dynamic range of the sensors of the camera. These effects include *Color Clipping* and *Blooming* as argued carefully in [23]. When the incident light is too strong and exceeds the dynamic range of the sensor, the sensor can not respond to that much input and thus clips the upper level beyond the range. This means the sensor does not correctly reflect the intensity of the light any more. Note that this is very serious for our invariants, because both γ and φ are ratio invariant, and a basis of their theory is, whether locally or globally, the consistency of the amount of light falling onto the concerning positions on the object surfaces. Here, our natural and important assumption is that this consistency is correctly reflected in the responses of the sensors. Therefore, if the sensor response does not meet this assumption, our theory no longer holds. The same arguments also hold for the blooming effect. When the incoming light is too strong to be received by the sensor element of the CCD camera, the overloaded charge will travel to the nearby pixels, thus crippling the responses of such pixels.

3 Combining photometric and geometric constraints for 3D object recognition

In this section, we describe how we can exploit the photometric invariant developed in the preceding section for recognizing 3D objects. The basic idea is to combine it with the Centroid Alignment approach we have recently proposed in [25].

3.1 Centroid invariant of geometric feature groups

We argued in [25] that when an object undergoes a linear transformation caused by its motion, the centroid of a group of 3D surface points is transformed by the same linear transformation. Thus, it was shown that under an orthographic projection model, centroids of 2D image geometric features always correspond over different views regardless of the pose of the object in space. This is true for any object surfaces (without self-occlusion). Note that this property is very useful, because if we have some way to obtain corresponding feature groups over different views, we can replace simple local features used for defining alignment in conventional methods by those groups, thereby reducing computational cost. We demonstrated the effectiveness of this approach to object recognition on natural as well as simulation data [25].

3.2 Grouping by photometric and geometric constraints

To obtain corresponding groups of 2D geometric features, we can use the proposed photometric invariant measures associated with each feature.

In [25], to obtain corresponding geometric feature groups, a clustering operation, in which the criterion was rotationally invariant, was applied in the coordinates which had been normalized up to a rotation prior to a clustering. This time, we again use a clustering technique to obtain corresponding geometric feature groups in different views. Our intention is to yield corresponding cluster configurations using a criterion incorporating spatial proximity constraints of geometric features and the invariance of their associated photometric invariants we have proposed. Therefore, we assume that surface colors (surface spectral reflectance) vary mostly from place to place. In other words, within some local areas surface colors are almost consistent. Note that this assumption should be justified for most object surfaces, because otherwise we must always be seeing diffused colors over the surface and thus always having difficulty in trying to distinguish surfaces. We also normalize the geometric feature distributions by the linear transformation we presented in [25]. This transformation has been confirmed, both mathematically and empirically, to generate a unique distribution up to a rotation, for feature sets from a planar surface on the object, regardless of the surface orientations in 3D space. We note that even 3D object surfaces often tend to become planar in their visible surfaces, thus justifying the use of our transformation for 3D object surface. This will be seen later in the experiments.

3.3 Implementation

We employ the *Kmean* clustering algorithm, in which the criterion is rotationally invariant, to obtain corresponding feature groups in the feature set from different views. The feature vector f used in clustering is the extended feature (from local geometrical feature) which is defined by the following vector:

$$f = [f_g^T, sf_p^T]^T \tag{18}$$

where f_g is the 2D geometric feature composed of spatial coordinates $f_g = (x, y)^T$ of a feature point in the xy image plane, and f_p is the vector of photometric invariant properties we proposed in the preceding sections, and s is a balancing parameter. Note that what we ultimately need here is simply the configuration of geometric features, that is f_g, in the clustering results, and photometric invariant is used only as a cue in performing clustering.

After the clustering, an alignment process starts by using centroids of clusters so derived to recover the transformation which generated a novel view, the image data, from the model. It is known that only 3 point correspondences suffice to recover the transformation either by using Linear Combination of the models[32] or a full 3D object model[20]. Therefore, we examine every possible combination of triples of cluster centroids of model and data that are generated by clustering, and select the best-fit transformation to generate the data from the model in terms of their match. In our testing, which we will see later, this number of clusters could be suppressed to less than 10. Further, we should note that we only need to consider the combination of model and data cluster centroids which have compatible values of γ or φ. This means that adding photometric properties contribute not only to the clustering but also to the selectivity of the features (cluster centroids). Therefore, considering the computational complexity of conventional alignment approach to recognition, this should bring a noticeable computational improvement.

4 Empirical results

In this section, we show experimental results of our algorithm for identifying corresponding positions in different views. Tests were conducted on natural pictures including 3D objects to be recognized.

4.1 Preliminaries

Geometric features used for our algorithm can be extracted as follows:
(Step 1) Use an edge detector[6] after preliminary smoothing to obtain edge points from the original gray level images.
(Step 2) Link individual edge points to form edge curve contours.
(Step 3) Using local curvatures along the contours, identify features as corners and inflection points respectively by detecting high curvature points and zero crossings based on the method described in [20]. Before actually detecting such features, we smooth the curvatures along the curves [2].
In obtaining color attributes from corresponding positions we should note that the positions of the geometric features thus extracted in different views do not always correspond exactly in discrete image coordinate space. This is not only due to quantization error, but also because edges detected to derive feature points can shift to the other side of the surface beyond the boundary under a object rotation within a image plane. Note that this is serious because the occurrences of gray level edges often tend to coincide with color edges[5]. So, we can not simply use the color attributes of the geometrical feature points derived from gray level edges. To solve this problem, we picked up color values from two positions over the gray level boundary, which are away from the geometric feature positions in the opposite directions along the local normals of the contours. Then, we used two color values from both of two positions. As we do not know which sides of an edge in one picture correspond to which in another, the distance metric between the photometric invariant vectors associated to two different feature positions should be independent of the correspondences of those sides of the surfaces. Thus, the actual measure used for photometric invariant vector f_p and the distance metric for two of those (that are used for computing the values for clustering criterion) are designed such that they support the symmetry on the sides of the surfaces over the boundaries: $f_p = [{f_p^1}^T, {f_p^2}^T]^T$, where $f_p^i = (G^i/R^i, B^i/R^i)$ for γ and $f_p^i = ((G^i/R^i)/(G^i/R^i + G^j/R^j), (G^j/R^j)/(G^i/R^i + G^j/R^j), (B^i/R^i)/(B^i/R^i + B^j/R^j), (B^j/R^j)/(B^i/R^i + B^j/R^j))$ for φ, and indices $(i,j) \in \{(1,2),(2,1)\}$ show the sides of the surfaces with respect to their boundaries, and the distance metric between f_{p1} and f_{p2} for geometric feature positions 1, 2 is:

$$|f_{p1} - f_{p2}|^2 = min\{||f_{p1}^1 - f_{p2}^1||^2 + ||f_{p1}^2 - f_{p2}^2||^2, \\ ||f_{p1}^1 - f_{p2}^2||^2 + ||f_{p1}^2 - f_{p2}^1||^2\} \quad (19)$$

where $||\cdot||$ denotes Euclidean distance. This apparently supports the symmetry on the sides of the surfaces over the boundaries of the gray level, and is invariant to the rotation of the objects within a image plane. The following experiments test our algorithm with both of the proposed invariants γ, φ. For each feature position, the associated invariant φ was computed using color attributes of those two points mentioned above, that is, two points a little away from the geometrical feature points along the contour normals in the opposite directions. As described earlier, since gray level edges tend to coincide with color edges, the color values collected from those two positions facing across the gray level edges are usually quite different, thereby producing φ distributions that spread over the feature space. To satisfy the requirement for γ, that is to be provided with the corresponding sets of points between the model and the data views, the object regions were extracted prior to the application of our algorithm. This was done manually though we expect that this could be done automatically using several cues such as motion, color, texture, (see e.g.,[30, 29, 27, 28].) Note that, however, in using φ this process, i.e., region extraction, is not necessarily required, as long as the background in the picture happened to have different colors from object ones. This is because φ is a complete invariant, unlike γ which needs further normalization to remove scale factors as we have argued. This is also true for γ when the ambient light has not been changed before and after the motion of the objects. Hereafter, we refer to $\hat{\gamma}$, the normalized measure, as simply γ.

4.2 Experiments

We tested our algorithm to see how accurately it can identify corresponding positions over different pictures taken under varying light conditions and poses of the objects to be recognized. It would not be hard to see that identifying corresponding positions perfectly is not an easy task, because in doing that we must fight against two different kind of instabilities: one in extracting geometric features, most serious one of which is the missing of features, and the other substantially contained in photometric properties of the image, such as the ones described in the arguments for sensing limitations. Remember that, however, for our ultimate objective, that is recognizing objects using the identified positions, only three correspondences are sufficient under orthographic projection model[32] or weak perspective projection model[20]. Therefore, what have to be observed in the following results are whether our algorithm could identify at least this minimum number of correspondences or not. First, the results of using γ as photometric invariant are shown.

[With γ for photometric invariant]
Figure 6 shows the results of obtaining feature group centroids on Band-Aid-Box pictures, which includes characters of some different colors on a white base on the surface. All the pictures were taken to involve the same three surfaces of the box, which are to be used for the recognition. The figures in the first row from the top show the edge maps with extracted geometric features superimposed on them with small closed circles. The first from the left (hereafter first) picture was taken under a usual light conditions. The second from the left (hereafter second) and third from the left (hereafter third) pictures were taken respectively under a greenish and a bluish light at a different pose from the first one. Throughout the rest of the paper, we refer to the figures by the order they are presented from the left as above. The lighting conditions were changed by the same way used in the experiments presented in section 2.4. The figures in the second and the third rows show the respective original and normalized distributions of γ. The horizontal axes of the figures are for G/R while the vertical axes are for B/R. These figures show how the invariant property γ remained unchanged between the different pictures. When it performs well, the original distributions of γ should show the similar shape over different views except for some scale change along the axes. Then, those scale distortion (e.g., dilation) should be corrected by the normalization of the distribution, thus ideally showing linear distributions of slop 1. Note that even if the shape of the distributions are distorted in addition to the dilation, we can not conclude that the proposed invariants performed poorly. This is because unstable results of the geometrical feature extraction will also distort the shape of the distribution of the photometric properties. The intermediate results of clustering are shown in the fourth row in their normalized coordinate of the geometric features. In the figures of the first row, identified corresponding positions using our algorithm are superimposed by large closed circles. Therein, the accuracy of our algorithm are found to be fairly good. Apparently perturbations of identified positions were caused partly by the unstable results of feature extraction, e.g., missing features, rather than by clustering errors or incompleteness of the proposed photometric invariant.

In Figures 7 results on Spaghetti-Box pictures taken in the same way as the Band-Aid-Box pictures are given. The surfaces of this box include some textures including large/small characters. This is a little cluttered texture compared with the Band-Aid-Box surface. The first row shows the edges with extracted geometric features superimposed on them. The first picture was taken under a usual light condition. The second and the third pictures were taken respectively under a greenish and a bluish light at different poses. The second and the third row figures show the respective original and normalized distribution of γ. The algorithm could perform identification of the corresponding positions fairly accurately as we see in the top figures.

Similarly, in Figure 8 the results on Doll (the same one as the one used in the section 2.4) pictures are presented. Unlike the last two example, the surface of this doll does not have man-made texture such as characters, but only has color/brightness changes partly due to the changes of materials and partly due to depth variations. The surface is mostly smooth except for some parts including hair, face, and finger parts. The pictures in the first row show the edges with extracted geometric features superimposed on them. The first and second pictures were taken under a usual light conditions, but at different poses of the doll. The third picture was taken under a moderate greenish light plus usual room light. For the fourth picture, we used an extremely strong tungsten halogen lamp with a bluish cellophane covering it. The second and the third row figures show the respective original and normalized distributions of γ. Comparing the shapes of original and normalized distributions of γ for the first and the second pictures, we can confirm that when the light conditions have not been changed the distributions of γ are not affected by the change of pose of the object. The algorithm could perform identification of the corresponding positions fairly accurately as we see in the pictures.

[With φ for photometric invariant]
The results of using φ as a photometric invariant on the same pictures used for γ are shown. Figure 9 presents the results on Band-Aid-Box pictures. The first row shows the edge maps with extracted geometric features superimposed on them with closed circles. In the second row, respective distributions of φ are shown. The horizontal axes are for $(G^i/R^i)/(G^i/R^i + G^j/R^j)$, while the vertical axes are for $(B^i/R^i)/(B^i/R^i + B^j/R^j)$ where $(i, j) \in \{(1, 2), (2, 1)\}$. As described already, since we do not know the correspondences of the sides of the surface over the edges (contours), we included properties from both sides of the edges. Consequently, we had 2-fold symmetric distributions of φ around its centroid as noted in the second row figures (see eq. (15)). When φ performs well as an invariant, this distribution should remain unchanged over different pictures. Thus, the second row figures demonstrate a fairly good performance of it for this picture. The intermediate results of clustering are given in the third row figures in their normalized coordinate of the geometric features. In the figures of the first row, identified corresponding positions using our algorithm are also superimposed by large closed circles. Thus, the accuracy of our algorithm are found to be fairly good.

In Figures 10 the results with φ on Spaghetti-Box are given. The first row shows the extracted geometric features. The second row shows the distributions of φ. The performance of φ is almost perfect. As we see in the pictures, the algorithm with φ could perform identification of the corresponding positions very well.

Figure 11 presents the results on Doll pictures. In the first row, the edge maps with extracted geometric features superimposed on them are shown. The second row shows the the respective distributions of φ. Since for the fourth picture we used extremely intensive blue light, the blue channel of many pixels were saturated. As a consequence, the distribution of φ was shrunk in the vertical direction as noted in the fourth picture of the second row. For these doll pictures, generally, the results of identifying corresponding positions with φ were not as good as those with γ, though not very bad. This is probably because as the surface colors of the doll varies quite smoothly in most parts, the distribution of φ did not spread well, so that it did not work so well as to separate clusters in terms of colors.

5 Discussions and conclusion

We argued that by combining the proposed photometric invariants with geometric constraints tightly, we can realize very efficient and reliable recognition of 3D objects. Specifically, we conducted the experiments of identifying the corresponding feature positions over the different views taken under different conditions. Although we did not include the demonstrations of the actual recognition process, as described, by connecting the presented method for identifying features using photometric invariants with the popular recognition algorithms, such as the full 3D model method[20] or the Linear Combination of the model[32], we can perform object recognition quite efficiently. This may be demonstrated somewhere. In the experiments, we showed that our methods could tolerate perturbations both in color and geometric properties, and could provide at least minimum number of correspondences of positions necessary for object recognitions. Although we extracted the object regions manually in the experiments this is sometimes easily done from sequences of images, from the simple background, or may be performed by using color segmentations. In addition, we stress again that as long as the background has different colors from the object ones, we can use φ without any preliminary processing for region extraction. This also holds true for γ when the ambient light has remained unchanged. The weakness of φ comes out when the discontinuities of gray level do not coincide with the ones of colors. In this case, the distribution of φ does not spread very well. This emerged in the body parts of the doll. Compared with the conventional approaches of matching local features of which the number is of the order of several hundreds, the computational cost of our approach for recognizing 3D objects should be very small. The time for identifying (about 10) corresponding feature positions, i.e., cluster centroids, was around 0.2 sec for pictures with several hundreds features. In addition, we can use the invariant photometric values in searching for the correspondences between the derived feature points in the model and the image, so that needless searches could be further suppressed.

The advantages of our approach compared with Nayar's are as follows. Their method uses invariant photometric properties derived for regions each with a consistent and different color, so that the color segmentation is a prerequisite. In our view, this color segmentation is an essential process to reduce the size of the search space for correspondences, and the photometric invariant was used only for further limiting possible matches between the model and the data regions. Unfortunately, however, achieving complete color segmentation is often quite hard and time consuming[29]. Of course, it can still contribute to reduce the computational cost, since in general the number of color regions included in the entire image could still be on the order of some tens. But, it appears to be less of a contribution than color segmentation to the reduction of computational cost. Contrary to their approach, since our photometric invariant can be computed only locally, we do not necessarily need color segmentation as mentioned above, so is less demanding. In addition, since the color properties are passed to the following clustering plus feature centroid alignment process, our method can tolerate many confounding factors, such as inaccuracies of region and/or feature extraction, happening in the application to the real world. The clustering plus feature centroid alignment process is very suitable for compensating those uncertainties. We should also point out that, to be theoretical, region centroids which they used for matching can not be used for 3D surfaces, while our feature centroids can.

An alternative way of using the proposed photometric invariant in recognition is just to incorporate it into the conventional framework of recognition. For example, in selecting features to form hypothesized corresponding triples of features between the mode and the data, photometric properties can be used to limit the possible matches between the model and the data features, trimming a bunch of needless combinations in the search space, thereby effectively reducing the computational cost. This kind of idea has been used in [26] for matching corresponding regions.

Acknowledgments

Kenji Nagao is thankful to Berthold Horn for his useful comments on this research. He also thanks to many other colleagues in the MIT AI LAB for the discussions on his research.

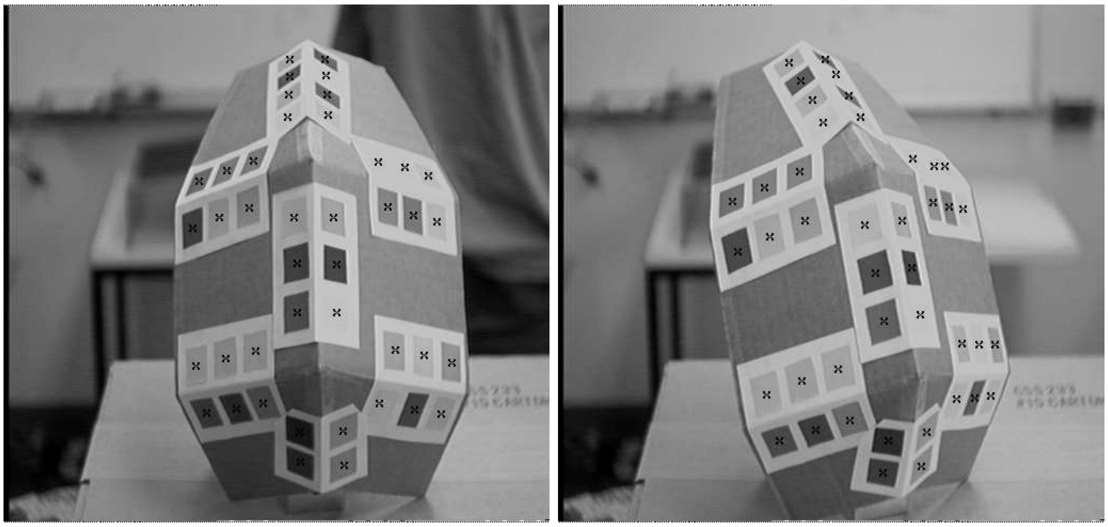

Figure 1: Tests of Invariant on Convex Polyhedron
The pictures show the convex polyhedron in different poses: left pose P_A, right pose P_B. This object is composed of 6 planar surface patches each with different surface orientation. On each side of the boundary of adjacent surfaces, several matte patches with different colors were pasted. Then, we picked up corresponding positions manually within each colored patch in both pictures. The selected positions within patches are depicted by crosses.

	$P_A \& L_U - P_B \& L_G$	$P_A \& L_U - P_B \& L_B$
R	0.988368	0.989877
G	0.967951	0.974081
B	0.946251	0.882816
H	0.724681	0.701377
S	0.914236	0.749529
V	0.945473	0.668672
$R-G$	0.985398	0.985687
$B-Y$	0.935039	0.908867
G/R	0.978163	0.988289
B/R	0.962186	0.907126
$\varphi_1 = (G^1/R^1)/(G^1/R^1 + G^2/R^2)$	0.997766	0.997532
$\varphi_2 = (B^1/R^1)/(B^1/R^1 + B^2/R^2)$	0.991843	0.988893

Table 1: Correlation coefficients between the sets of the values of the color properties from different pictures of Test-Object.
The correlation coefficients between the sets of values of the proposed invariants from pictures taken under different light conditions and at the different poses of the object are given to show how much they remain unchanged within a consistent scale. For comparison, other color properties including (R, G, B), (H, S, V), and a linear-trans implementation of opponent color model[3] are also presented. In these tests, (R, G, B), $(R-G, B-Y)$, $\gamma = (G/R, B/R)$, are almost equally good, though γ is best among them. The reason why R is also fine is probably just that we did not happen to change the intensity of the red light spectrum. The values of (H, S, V) (hue,saturation,value) is unexpectedly unstable. The measure φ is extremely stable.

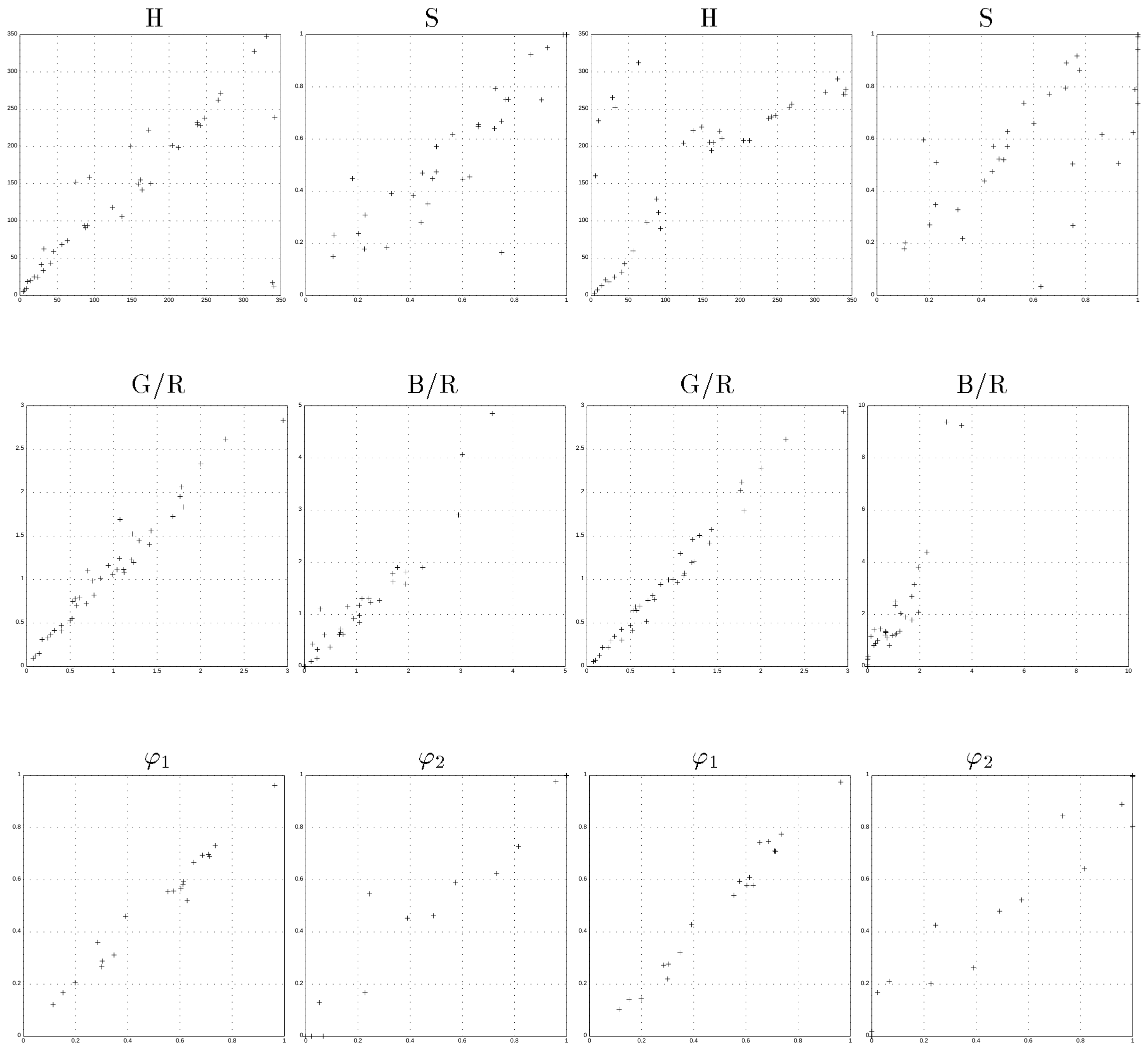

Figure 2: Distributions of invariants on Convex Polyhedron

The left two columns are from pictures taken under $P_A \& L_U$ (horizontal axis) and $P_B \& L_G$ (vertical axis), and the right two columns are from pictures under $P_A \& L_U$ (horizontal axis) and $P_B \& L_B$ (vertical axis). The rows in each two columns are respectively: top left and right: H and S, middle left and right: G/R and B/R, bottom left and right: $\varphi_1 = (G^1/R^1)/(G^1/R^1 + G^2/R^2)$ and $\varphi_2 = (B^1/R^1)/(B^1/R^1 + B^2/R^2)$.

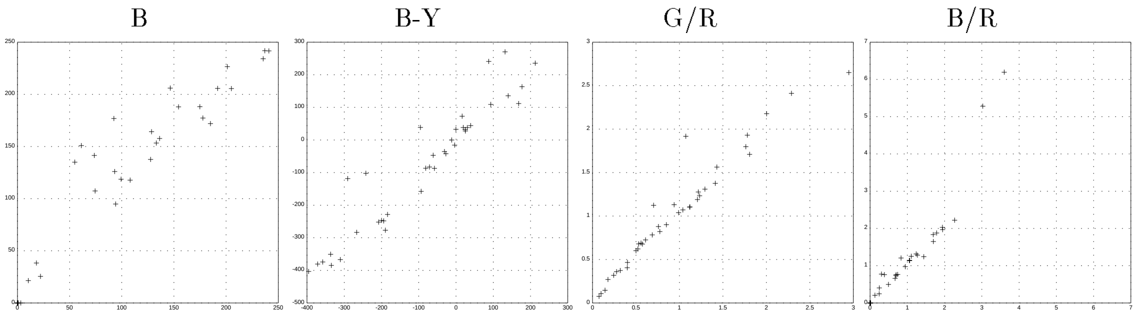

Figure 3: Tests of γ at different poses of object but under the same illuminant conditions
The first from the left: distribution of Blue, the second : $B - Y$ (Blue vs. Yellow), the third: G/R, the fourth: B/R. The horizontal axis is for the pose P_A and the vertical axis is for the pose P_B. Except for the two samples in the upper right area of the distribution, $\gamma = B/R$ is found to be almost unchanged in both of the pictures because the slope is almost 1, while $B - Y$ and B are perturbed around the slope of 1. Those two exceptional samples were from patches with almost saturated blue channel in the picture at pose P_B. The distribution of $\gamma = G/R$ is almost perfect. This gives the evidence that γ may be used for object recognition without applying any normalization process, so that extracting object regions might not be a prerequisite, as long as the lighting conditions are not changed.

Figure 4: Tests of Invariant on natural pictures
The pictures show a doll at different poses: left pose A, right pose B. We picked up corresponding positions in both views. The selected positions are depicted by crosses.

	$P_A\&L_U$–$P_B\&L_G$	$P_A\&L_U$–$P_B\&L_B$
R	0.764343	0.819267
G	0.588161	0.881416
B	0.936572	0.843604
H	0.951843	0.923887
S	0.934587	0.490994
V	0.398850	0.459425
$R-G$	0.764240	0.939152
$B-Y$	0.948642	0.877519
G/R	0.779377	0.944164
B/R	0.962186	0.895180
$\varphi_1 = (G^1/R^1)/(G^1/R^1 + G^2/R^2)$	0.996245	0.998781
$\varphi_2 = (B^1/R^1)/(B^1/R^1 + B^2/R^2)$	0.988840	0.983675

Table 2: Correlation coefficients between the sets of the values of the color properties from different pictures of the Doll.

The results on natural object, a doll, are given. The first picture was taken under a usual lighting conditions from the oblique angle($P_A\&L_U$), the second and third were taken respectively under a greenish and a bluish light from the front angle($P_B\&L_G$, $P_B\&L_B$). This time for φ — $\varphi_1 = (G^1/R^1)/(G^1/R^1 + G^2/R^2)$, $\varphi_2 = (B^1/R^1)/(B^1/R^1 + B^2/R^2)$ — two positions which are closest to each other are used. In this tests, R, G, B and H, S, V were very unstable. The linear model $R-G$, $B-Y$, $\gamma = G/R$, B/R did perform well again, though γ was better. The measure φ is quite stable again.

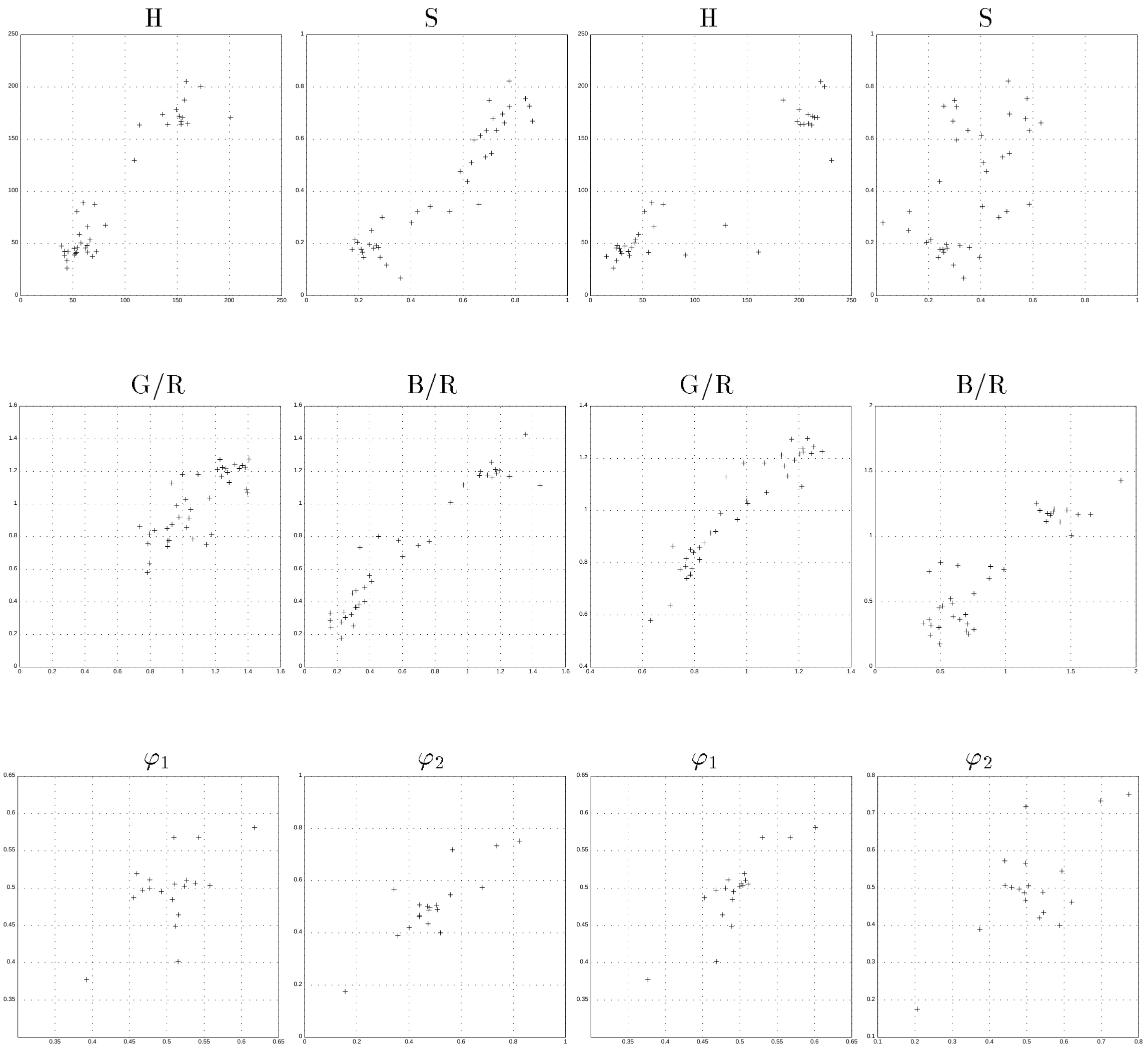

Figure 5: The distributions of invariant measures on Doll pictures.
The left two columns are from pictures taken under $P_A\&L_U$ (horizontal axis) and $P_B\&L_G$ (vertical axis), and the right two columns are from pictures under $P_A\&L_U$ (horizontal axis) and $P_B\&L_B$ (vertical axis). The rows in each two columns are respectively: top left and right: H and S, middle left and right: G/R and B/R, bottom left and right: $\varphi_1 = (G^1/R^1)/(G^1/R^1 + G^2/R^2)$ and $\varphi_2 = (B^1/R^1)/(B^1/R^1 + B^2/R^2)$.

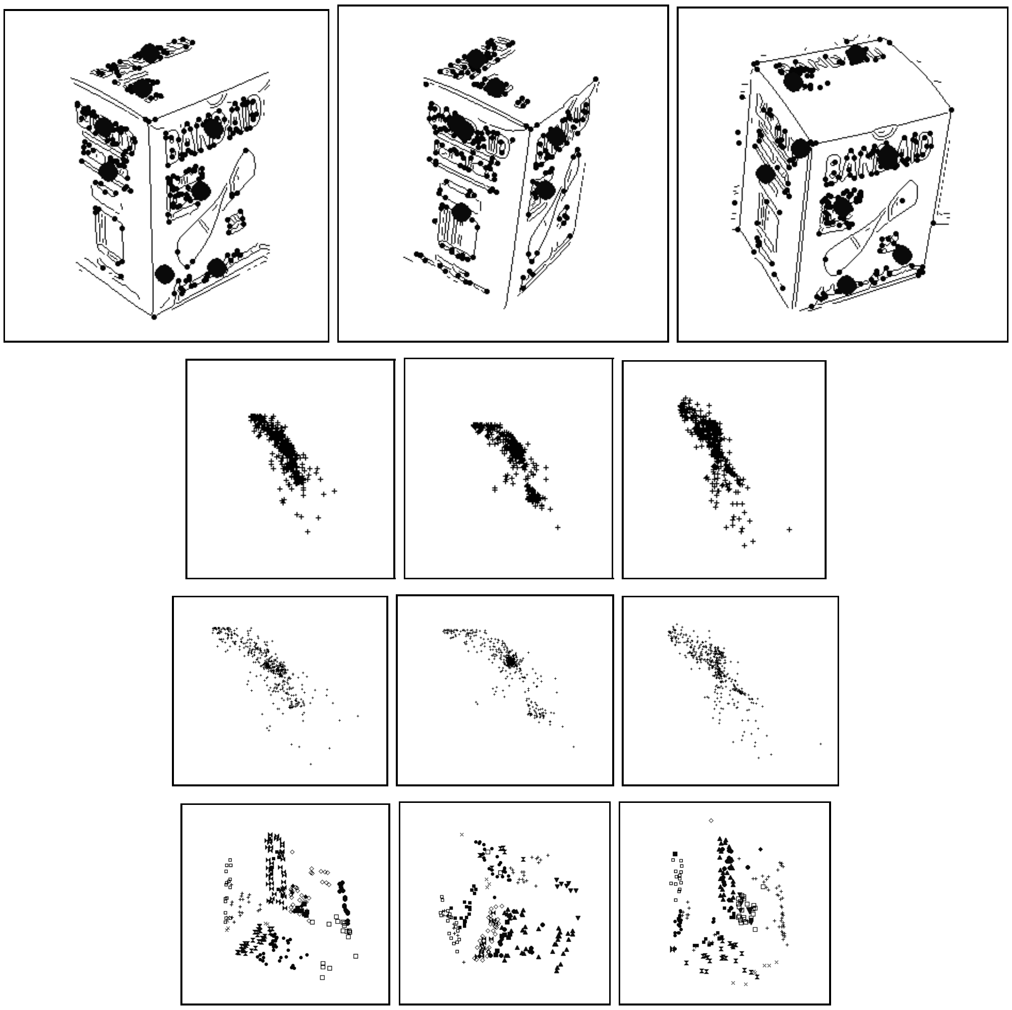

Figure 6: Tests with γ on Band-Aid-Box picture.

Edge maps are shown with extracted geometric features superimposed on them in the first row. The first picture (from the left) was taken under a usual light conditions. The second and third pictures were taken respectively under a greenish and a bluish light at a different pose. Identified corresponding positions using our algorithm are also superimposed by large closed circles. The figures in the second and third rows show the respective original and normalized distributions of γ. The intermediate results of clustering are shown in the fourth row figures in their normalized coordinate of the geometric features.

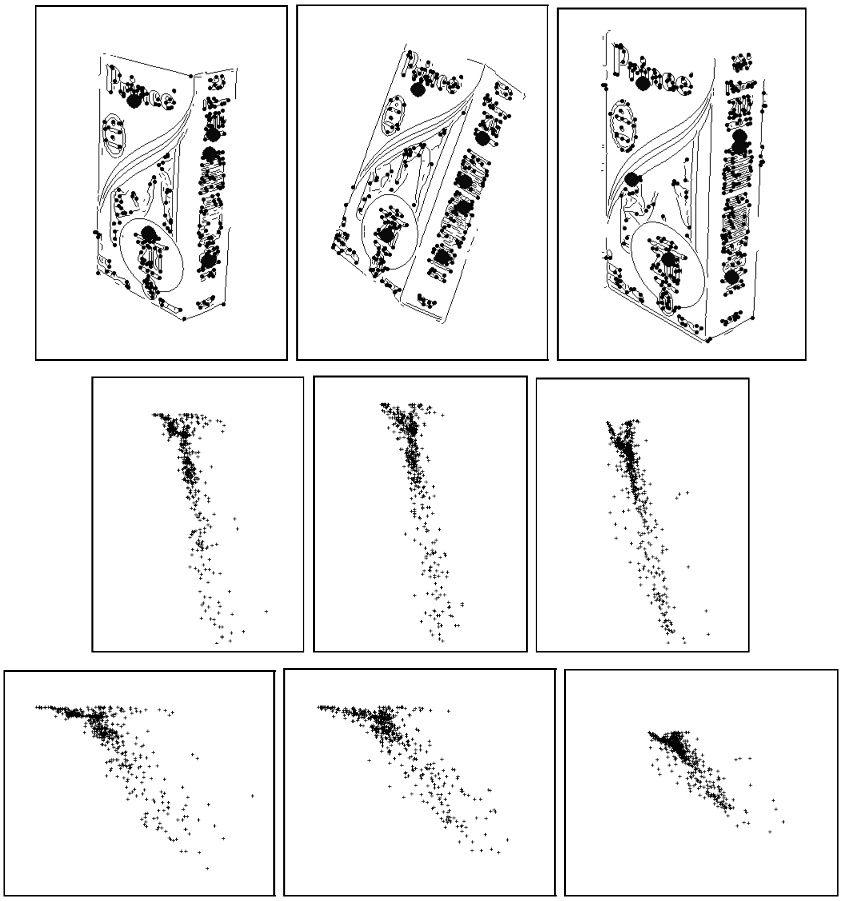

Figure 7: Tests with γ on Spaghetti-Box pictures

The surface of this boxes include some colored textures including large/small characters. The pictures in the first row show the edges with extracted geometric features superimposed on it. The first picture (from the left) was taken under a usual light conditions. The second and third pictures were taken respectively under a greenish and a bluish light at a different pose from the first one. The second and third rows show the respective original and normalized distributions of γ. The identified positions are depicted by large closed circles in the figures of the first row. The algorithm could perform identification of the corresponding positions fairly accurately as we see in the upper figures.

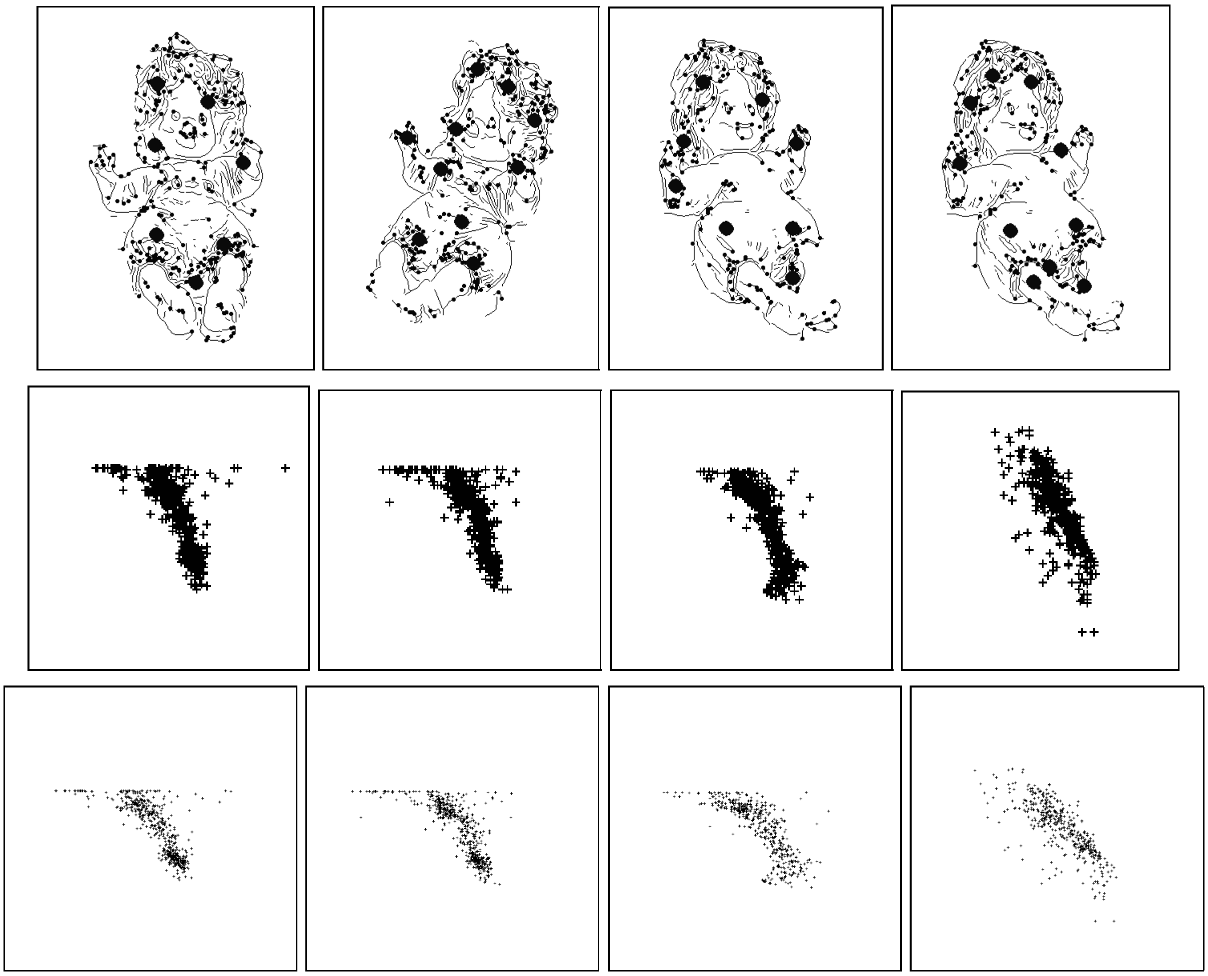

Figure 8: Tests with γ on Doll pictures

The surface of this doll does not have man-made texture like characters, but only has color/brightness variation partly due to the changes of materials and partly due to depth variations. The surface is mostly smooth except for some parts including hairs, face, and finger parts. The first row shows the edge maps with the extracted geometrical features superimposed on it with small closed circles. The first and second pictures (from the left) were taken under a usual light conditions, but at different poses of the doll. The third picture was taken under a moderate greenish light plus usual room light. For the fourth picture, we used a extremely strong tungsten halogen lamp with a bluish cellophane covering it. The second and the third rows show the respective original and normalized distributions of γ. The identified positions are depicted by large closed circles in the figures of the first row. The algorithm could perform identification of the corresponding positions fairly accurately as we see in the figures.

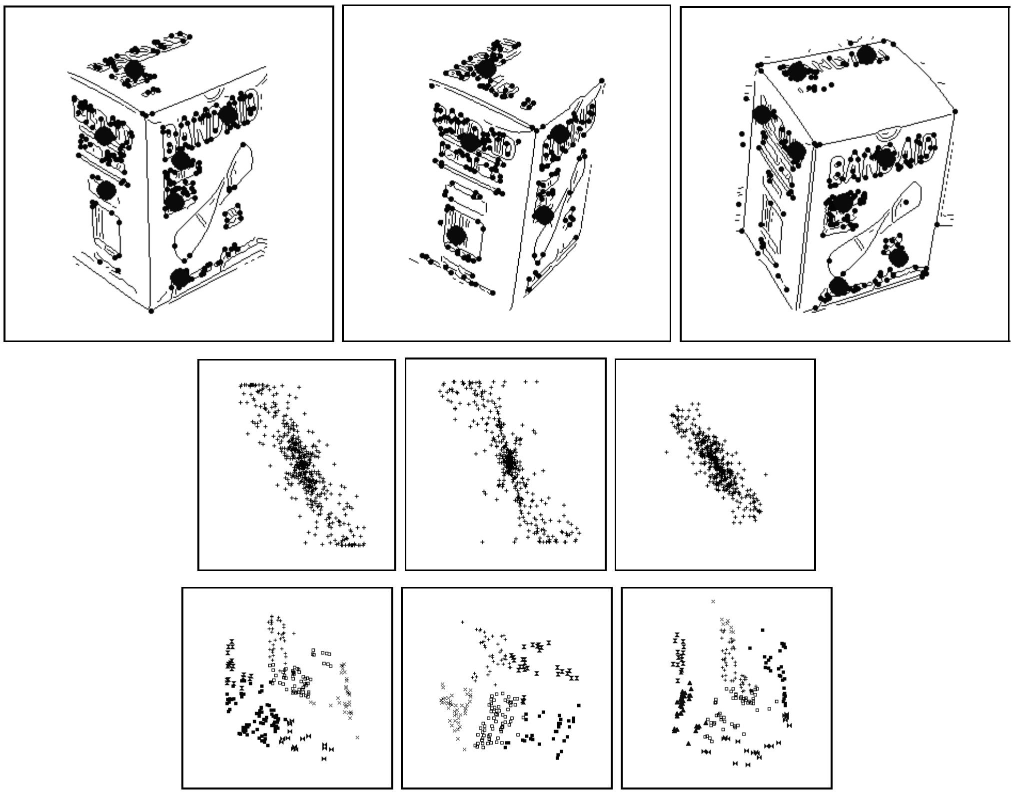

Figure 9: Tests with φ on Band-Aid-Box pictures

The pictures in the upper row show the edge maps with extracted geometric features superimposed on them. The first picture (from the left) was taken under a usual light conditions. The second and third pictures were taken respectively under a greenish and a bluish light at a different pose from the first one. The second row figures show the respective distributions of φ. The third row figures show the intermediate results of the clustering. The identified positions are depicted by large closed circles in the figures in the first row.

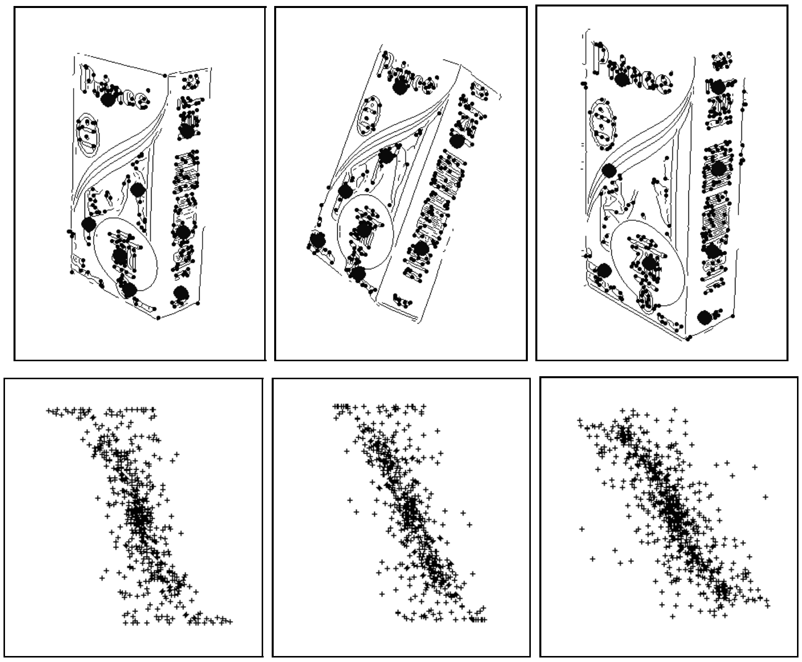

Figure 10: Tests with φ on Spaghetti-Box pictures
The surface of this box include some colored textures including large/small characters. Upper pictures show the edges with extracted geometric features superimposed on it. The first picture was taken under a usual light conditions. The second and third pictures were taken respectively under a greenish and a bluish light and at a different pose. The lower figures show the respective distributions of φ. The identified positions are depicted by large closed circles in the figures of the upper row. The algorithm could perform identification of the corresponding positions fairly accurately as we see in the upper figures.

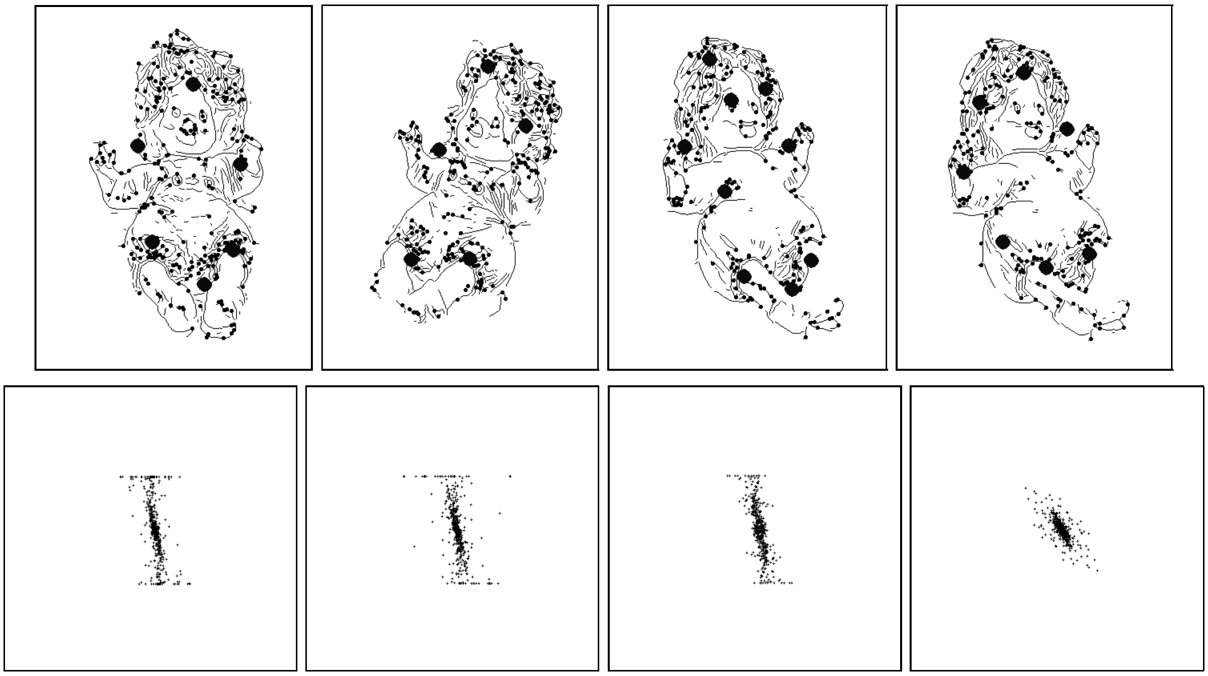

Figure 11: Tests with φ on Doll pictures

The surface of this doll does not have man-made texture like characters, but only has color/brightness variation due to the change of material. The surface is mostly smooth except for some parts including hairs, face, and finger parts. The pictures in the upper row show the edges with extracted geometric features superimposed on it. The first and second pictures were taken under a usual light conditions, but at different poses of the doll. The third picture was taken under a moderate greenish light and fourth pictures was taken under an extremely bright bluish light. The lower figures show the respective distributions of φ. The identified positions are depicted by large closed circles in the figures of the upper row. The algorithm could perform identification of the corresponding positions fairly well as we see in the pictures.

References

[1] T. Alter, W. Eric L. Grimson, "Fast and Robust 3D Recognition by Alignment" *In Proc. ICCV 93*, pp. 113–120, 1993.

[2] H. Asada, M. Brady, "Curvature Primal Sketch" *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-8, pp.2–14, 1986.

[3] D. H. Ballard, C. M. Brown, *Computer Vision*, pp. 31–35, Prentice Hall, 1982.

[4] M. H. Brill, "A Device Performing Illuminant-Invariant Assessment of Chromatic Relations", *J. Theor. Biol.*, pp. 473–478, 1978.

[5] D. C. Brockelbank, Y. H. Yang, "An Experimental Investigation in the Use of Color in Computational Stereopsis", *IEEE Transaction on Systems, Man, and Cybernetics*, vol. PAMI-19, NO. 6, pp. 1365–1383, 1989.

[6] J. F. Canny, "A Computational Approach to Edge Detection", *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-8, pp.34–43, 1986.

[7] J. Cohen, "Dependency of the spectral reflectance curves of the Munsell color chips", *Psychon. Sci.*, vol. 1, pp. 369–370, 1964.

[8] O. D. Faugeras, "Digital Color Image Processing Within the Framework of a Human Visual Model", *IEEE Transaction on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, NO. 4, pp. 380–393, 1979.

[9] G. D. Finlayson, M. S. Drew, B. V. Funt, "Diagonal Transforms Suffice for Color Constancy", In Proc. ICCV, pp. 164–170, 1993.

[10] D. A. Forsyth, "A Novel Algorithm for Color Constancy", *Inter. Journ. Comp. Vision*, 5:1, 5–36, 1990.

[11] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press 1972.

[12] B. V. Funt, M. S. Drew, "Color Constancy Computation in Near-Mondarian Scenes using a Finite Dimensional Linear Model", In Proc. CVPR, pp. 544–549, 1988.

[13] R. Gershon, A. D. Jepson, J. K. Tsotsos, "From [R,G,B] to Surface Reflectance: Computing Color Constraint Descriptors in Images", In Proc. 10th Int. Jt. Conf. on Artificial Intelligence, pp. 755-758, 1987.

[14] W. E. L. Grimson, *Object Recognition by Computer*, MIT Press, 1991.

[15] W. E. L. Grimson, "Affine Matching With Bounded Sensor Error: A Study of Geometric Hashing and Alignment", A.I. Memo No. 1250, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, August 1991.

[16] W.E.L. Grimson, A. Lakshmi Ratan, P. A. O'Donnell, G. Klanderman, "An Active Visual Attention System to "Play Where's Waldo"", IEEE CVPR Workshop on Visual Behaviors, Seattle, June 1994.

[17] E. Hering, "Outlines of a theory of the light senses", translated by Leo M. Hurvich and Dorothea. Cambridge, MA:Harvard Univ. Press, 1964.

[18] B. K. P. Horn, *Robot Vision*, Cambridge, MA:The MIT Press, pp. 185–277, 1986.

[19] B. K. P. Horn, "Determining Lightness from an Image", *Comput. Gr. Image Process.*, vol. 3, pp.277–299, 1974.

[20] D. P. Huttenlocher, S. Ullman, "Recognizing Solid Objects by Alignment with an Image", *Inter. Journ. Comp. Vision*, 5:2, pp.195–212, 1990.

[21] D. B. Judd, D. L. MacAdam, and G. Wyszecki, "Spectral distribution of typical daylight as a function of the correlated color temperature", *J. Opt. Soc. Am.*, vol. 54, pp. 1031–1040, August 1964.

[22] G. J. Klinker, "Physical Approach to Color Image Understanding", Ph.D thesis, Carnegie Mellon University, 1988.

[23] G. J. Klinker, Steven A. Shafer and Takeo Kanade, "The Measurement of Highlights in Color Images", *Inter. Journ. Comp. Vision*, pp. 7–32, 1988.

[24] Y. Lamdan, J. T. Schwartz, H. J. Wolfson, "Affine Invariant Model Based Object Recognition", *IEEE Trans. Robotics and Automation* vol. 6, pp. 238–249, 1988.

[25] K. Nagao, W. E. L. Grimson, "Object Recognition by Alignment using Invariant Projections of Planar Surfaces", A.I. Memo No. 1463, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, February 1994, also in Proc. 12th ICPR, pp. 861–864, 1994, and in Proc. DARPA Image Understanding Workshop, November, 1994.

[26] S. K. Nayar, R. M. Bolle, "Reflectance Ratio: A Photometric Invariant for Object Recognition" In Proc. Fourth International Conference on Computer Vision, pp.280–285, 1993.

[27] M. J. Swain, Color Indexing, PhD Thesis, Chapter 3, University of Rochester Technical Report No. 360, November 1990.

[28] M. J. Swain, D. H. Ballard, "Color Indexing", *Inter. Journ. Comp. Vision*, 7:1, pp. 11–32, 1991.

[29] T. F. Syeda-Mahmood, "Data and Model-driven Selection using Color Regions", In Proc. European Conference on Computer Vision, pp.321–327, 1992

[30] W. B. Thompson, K. M. Mutch, V. A. Berzins, "Dynamic Occlusion Analysis in Optical Flow Fields", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. PAMI-7, pp.374–383, 1985.

[31] M. Tsukada, Y. Ohta, "An Approach to Color Constancy Using Multiple Images", In Proc. ICCV, pp. 385-389, 1990.

[32] S. Ullman and R. Basri, "Recognition by Linear Combinations of Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**(10),pp.992–1006, 1991.

[33] B. A. Wandell, "The Synthesis and Analysis of Color Images", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, NO. 1, pp.2–13, 1987.