

# Using Specularities in Comparing 3D Models and 2D Images<sup>\*</sup>

Margarita Osadchy<sup>1,\*</sup> David Jacobs<sup>2</sup> Ravi Ramamoorthi<sup>3</sup>  
David Tucker<sup>2</sup>

---

## Abstract

We aim to create systems that identify and locate objects by comparing known, 3D shapes to intensity images that they have produced. To do this we focus on verification methods that determine whether a known model in a specific pose is consistent with an image. We build on prior work that has done this successfully for Lambertian objects, to handle a much broader class of shiny objects that produce specular highlights. Our core contribution is a novel method for determining whether a known 3D shape is consistent with the 2D shape of a possible highlight found in an image. We do this using only a qualitative description of highlight formation that is consistent with most models of specular reflection, so no specific knowledge of an object's specular reflectance properties is needed. This allows us to treat non-Lambertian image effects as a positive source of information about object identity, rather than treating them as a potential source of noise. We then show how to integrate information about highlights into a system that also checks the consistency of Lambertian reflectance effects. Also, we show how to model Lambertian reflectance using a reference image, rather than albedos, which can be difficult to measure in shiny objects. We test each aspect of our approach using several different data sets. We demonstrate the potential value of our method of handling specular highlights by building a system that can locate shiny, transparent objects, such as glassware, on table tops. We demonstrate our hybrid methods on pottery, and our use of reference images with face recognition experiments.

---

<sup>\*</sup> A more preliminary version of this work appears as: [36]

<sup>\*</sup> Corresponding author.

<sup>1</sup> Computer Science Department, University of Haifa, Mount Carmel, Haifa 31905, Israel. e-mail: rita@cs.haifa.ac.il

<sup>2</sup> Computer Science Department, University of Maryland, College Park, Maryland. e-mail: djacobs@cs.umd.edu and dntucker@wam.umd.edu

<sup>3</sup> Computer Science Department, Columbia University, 450 Computer Science Bldg 500 W 120 St New York, NY 10027. e-mail: ravir@cs.columbia.edu



Fig. 1. Images of real objects contain effects due to specularities, transparency, and interreflections.

## 1 Introduction

An object's appearance varies significantly with viewing conditions. Lighting is a chief source of this variability. Current methods have made substantial progress in accounting for the effects of lighting on objects with Lambertian reflectance. However recognition of objects made of more complex materials, such as shiny or specular objects remains a challenging problem. When the specular effects are minor, existing Lambertian methods treat them as noise, and produce satisfactory results. However many real objects (some examples are shown in Figure 1) are very specular and their appearance changes dramatically with even a minor change in lighting (Figure 2).

We focus on modelling specular effects instead of treating them as noise. Such an approach has multiple advantages: 1) More accurate and realistic appearance models will improve the performance of object recognition systems. 2) Specularity appearance depends on surface normals, thus it contains infor-

mation about the shape of an object. Though local it provides discriminative information that is essential for recognition.

Our contributions are to design:

- A simple model for specular highlights that allows us to use them as a positive source of information about object identity and location.
- A model-based identification method that can identify smooth and textured objects that have mixed Lambertian and specular reflection. This method doesn't require explicit knowledge of Lambertian albedo, which is difficult to measure in shiny objects.
- A novel method for recognition of challenging objects made from specular and transparent materials such as glass, a problem that has not been previously addressed.

We use 3D models to allow us to account for the effects that lighting variation can have on the appearance of specular objects. 3D models can be acquired using stereo or structured light systems, if shiny objects are first covered with powder or paint to reduce their shininess. This however can result in noisy 3D models of shiny objects; consequently an important feature of our approach is its robustness to reasonable levels of noise in the model. We demonstrate this by using models built with commercial structured light systems.

In our current work we assume that the object is convex to prevent self-shadowing and interreflections. The object is illuminated primarily by a distant compact light source. This is true for many indoor settings and outdoors on a sunny day (our next step will be generalizing our method to multiple light sources). The camera and a compact light source are distant from the object. When dealing with glass objects we also assume that glass is thin to

avoid interreflections. Currently we do not explicitly model interreflections, cast shadowing, and occlusions. However when these effects are minor, we treat them as noise. Using these assumptions, we derive an algorithm that we then test on real objects, achieving good results even when these assumptions do not precisely hold.

This paper focuses on *verification*, which we consider to be the core component of recognition and localization systems. This is the ability to judge whether a hypothesized position of a 3D object of known shape is consistent with a corresponding portion of a 2D image. Our algorithms focus on making this judgement. Verification tells us when we have a correct hypothesis, but it is not possible to test all hypotheses about pose. Recognition and localization systems will require some method to generate a reasonable number of hypotheses for testing. For example, one can often hypothesize poses using a few, distinctive features of an object, such as eyes and mouth in a face or distinctive surface markings for a soda can. These features may be insufficient to allow certain judgements about pose and identity, but may be used to generate a limited number of hypotheses that can be evaluated using more extensive image information. As another example, in many localization applications the number of possible poses is inherently small, allowing a more brute-force search. For instance, if we want to monitor a highway, we know that cars will have a constrained position relative to the road. If we want a household robot that can put away dishes, we may know to look for glasses or pottery on a table. Though the focus of this paper is on verification, in order to make our work concrete we also describe a system that uses these judgements to prune hypotheses and select the location of shiny, transparent wine glasses on a table top. These glasses have known shape and only two degrees

of freedom in their position, but we know of no other system that localizes such challenging objects (McHenry et al. [26] propose a system for detecting glass. However, their method doesn't recognize objects, it only locates glass surfaces). The system proposed in this paper also recognizes specific objects based on their highlights.

The paper is organized as follows. We first focus on understanding what information specular highlights make available for recognition. To do this, in Section 2 we propose a simple, qualitative reflectance model that captures the properties of existing models (eg., Torrance-Sparrow[10], Phong[38], Ward[46], ...) over any reasonable choice of parameters. For this model, we show that we can perform verification by using 3D shape to map highlight pixels to the Gaussian sphere, and then checking whether they form a disk there. This test allows us to sort through possible highlight pixels to find ones that could match our model. To refine this test we require that the light direction implied by the highlight will be consistent with some prior knowledge of lighting direction.

Our method also determines the sharpness of highlights (or surface roughness ie., the degree to which specular reflectance is mirror-like or more diffuse), so our model of specularity requires no previously known parameters.

Next (Section 3), we show how to integrate knowledge about highlights with the Lambertian method of Basri and Jacobs [1]. This allows us to use highlights in recognizing objects that are part Lambertian, part specular, such as pottery (top of Figure 2). Using [1], we can determine lighting conditions from Lambertian reflectance. This allows us to remove the assumption that we have rough knowledge of lighting a priori, which we needed to identify glassware.



Fig. 2. Strong appearance variation of specular objects. Top – objects that are part Lambertian, part specular under varying illumination. Bottom transparent specular objects with different backgrounds.

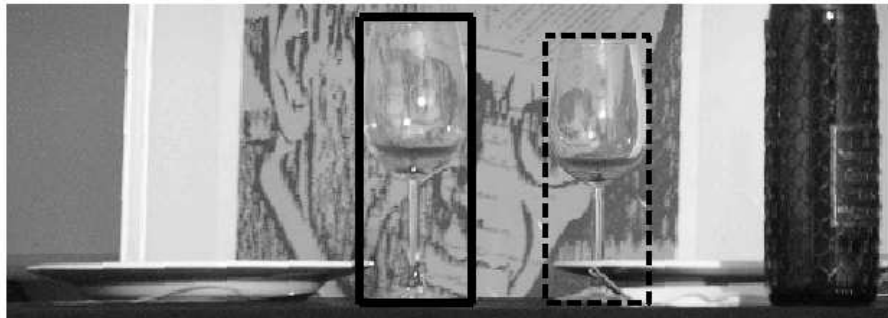


Fig. 3. Using specularity, our algorithm locates two glasses in this image. Then, we can identify highlights consistent with a specific model, and refine the Lambertian results using knowledge of these highlights. This provides us with a comparison method that can be used on objects of diverse materials with mixed properties.

We show how to model Lambertian reflectance using a reference image, rather than albedos, which can be difficult to measure in shiny objects. This approach

is based on the assumption that irradiance (the image that would be produced by the illumination if the object had uniform, white albedo) is dominated by low frequency components of lighting, thus it can be estimated by accounting for as much of the image as possible with low frequency variation in lighting. The texture is more dominant in higher frequencies. Similar ideas are used in homomorphic filtering (Gonzalez and Woods[19]) methods and the Retinex theory of lightness constancy (Land and McCann[25]).

In the experimental section (Section 4) we first evaluate the effectiveness of the specular matching method. We experiment using transparent objects such as glassware that produce specular highlights, but have no Lambertian reflectance. This allows us to verify the specular component of our comparison method before integrating it into a system capable of handling general reflectance. Our experiments show that specular highlights alone can provide useful information in identifying and localizing these purely specular objects. Figure 3 shows an example of our system's ability to localize glasses with a known shape on a table top.

Next we evaluate the ability of the integrated method to identify objects that have both Lambertian and specular components. We run our test on pottery objects that are shiny and have very little texture, because these objects present the greatest challenges, since their appearance depends so much on illumination. We test the ability of our approach to handle unknown Lambertian albedos as a separate recognition method for Lambertian objects, and we test it on faces.

In all tests our methods showed very good results compared to Lambertian methods that treat specular effects as noise.

## 1.1 Previous Work

Most of the work in illumination insensitive recognition can be divided into two main approaches: 2D methods and model-based methods that use object shape information in some way.

2D methods recognize objects by comparing image representations that are insensitive to lighting variations. Osadchy et al.[37] reviews a number of 2D methods, such as normalized correlation, the use of the gradient direction (Chen et al.[9]) or Gabor jets (Lades et al.[24]), and discusses the relationship between them. In general, these approaches work best for highly textured, Lambertian objects, or objects with discontinuities in shape. For these objects, variations in intensity can often be compensated for by normalizing intensities or by using gradient-based representations. Ratios of color or intensity have also been used to achieve lighting invariance ([15,31]). It seems more difficult to extract lighting insensitive image representations for non-textured, 3D objects whose appearance is largely due to shading from lighting variation, or for specular objects.

Model-based methods account for illumination effects by building representations of the set of images that a 3D object can produce under a large range of lighting conditions. This can be done by condensing a large number of training images into a low-dimensional representation (eg., Hallinan[21], Epstein et al.[13], and Murase and Nayar[30]) or by using analytic insights to generalize from a small number of training images (eg., Shashua[43], Moses[28], Belhumeur and Kriegman[2], Georghiades et al. [17,18]). Basri and Jacobs[1] and Ramamoorthi and Hanrahan[40] build on these results to show how to ana-

lytically derive a 9D representation of an object's image set from a 3D model. Ramamoorthi[39] shows that a lower-dimensional 5D or 6D subspace often suffices. [2,1] provide further discussion of the large literature on this topic. Subspace methods cannot be directly extended to handle specular objects, because their appearance cannot be well approximated by a low-dimensional linear sub-space. First, specular objects reflect higher frequency light ([45]), so modeling their appearance requires many more dimensions. Second, in higher dimensions, ignoring the non-negativity constraint on the lighting leads to significant errors in the approximation of an object's appearance [44]. It is particularly hard to see how to apply any of these prior methods to transparent specular objects.

A few previous works have addressed the recognition of specular objects. In pioneering work on the recognition of shiny objects (Birk et al.[4], Fukada et al.[14]) the main theme has been to look for specularities with specific shapes that are likely to appear on objects of interest. For example, cylinders were detected using elongated specularities (Birk et al.[4]).

Another direction has been to use specularities to infer the 3D shape of the corresponding region and then match these region to a known 3D model. Such an approach was taken by Koshikawa and Shirai [23] in their recognition system that uses polarized illumination to estimate the surface normals of an object. Georghiades [16] used the Torrance-Sparrow model to recover the shape and the non-Lambertian reflectance properties of faces. These were used to model the appearance of each face under variable lighting and viewpoint. Each representation comprises a collection of (linear subspace approximations) of illumination cones in the space of images.

Sato et al.[41] emphasize that specularities should not be treated as noise, but should serve as a positive clue for recognition. Unlike us, they employ a physics-based simulator to predict specularities from which a set of aspects of the object is generated. For each specular aspect they prepare deformable matching templates. At runtime, an input image is first classified into candidate aspects, then the deformable templates are used to refine matching. The method proposed in Gremban et al.[20] also operates on specular features, but the method mainly focuses on using multiple observations from different viewpoints to resolve an ambiguity in the recognition of specular objects. [41] also gives an overview of earlier attempts in military applications to recognize specular objects.

There has been much additional work aimed at analyzing specular effects in images. Some of this can be found in [11,42,32,34,7,29,12]. Of particular relevance is the work of Nishino et al.[34]. Using a known 3D model, and multiple images, they disentangle diffuse and specular reflectance, and recover lighting. We perform somewhat similar operations in Section 3. Our approach and assumptions are somewhat different, because we wish to perform recognition with a single query image, while they attempt to reconstruct object properties using a larger (though still small) number of images.

In summary, recognition of non-textured, specular objects is still an open problem and there is no existing *general* method that recognizes both homogenous and textured specular objects in a common framework.

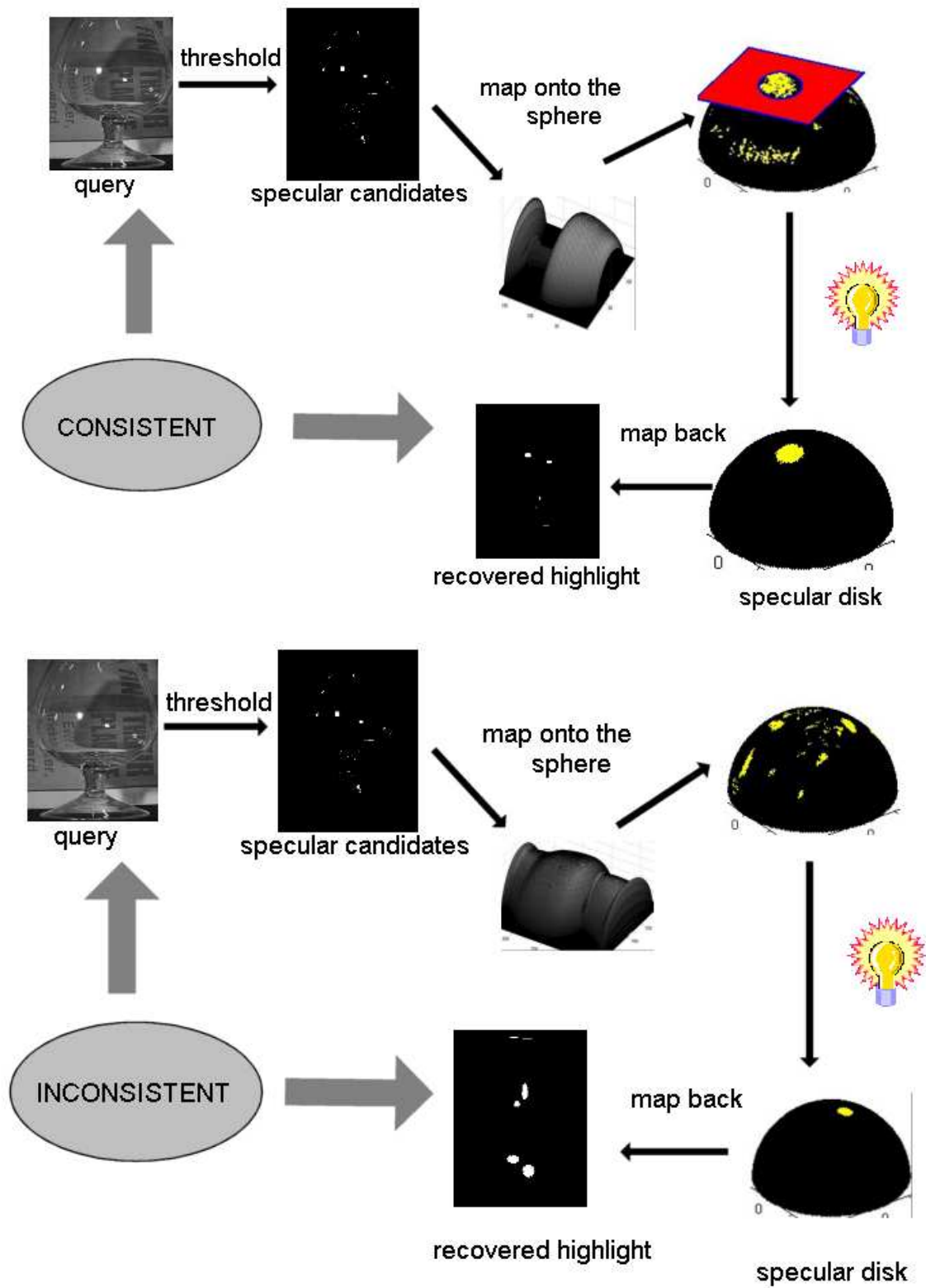


Fig. 4. Identification by matching specularity: top – correct object, bottom – incorrect object. The process begins at the query image.

## 2 Identification Using Specularities

In this section we will focus on the problem of verifying the compatibility of the 3D shape of an object and the 2D shape of a hypothesized specular highlight. To demonstrate the potential of this method we also show how to use specularity to perform verification given a rough knowledge of lighting. We will explain this method using the example of highly specular objects, such as glass. Later we will show that when objects have Lambertian and specular reflectance, we can use Lambertian effects also, in particular automatically determining the lighting direction to allow us to analyze specular highlights.

The algorithm is shown schematically in Figure 4. Our treatment of specularity has two stages. First we must locate pixels that are candidate highlights. This process need not be too accurate, since we refine these candidates in the second stage. Many different cues have been used to detect specularities, such as color, polarization, stereo and multiple views (see Oren and Nayar[35] for a survey). The best choice of a method will depend on the particular application and its constraints. However, determining candidate highlights is not the focus of our research, so we test our method using a simple, heuristic method. When the scene contains a dominant, compact light source, shiny objects such as glass produce highlights that are quite bright (Brelstaff and Blake[8]). For these objects it is sufficient to simply threshold the image based on intensity. This simple method produces many false positive candidates. However, this is not too important because our proposed method is able to reject these in a second stage, and we do not observe errors caused by false positive specular candidates. Therefore, our performance using this simple method of finding specular candidates is sufficient to demonstrate the potential

of our method. We consider the problem of identifying specular candidates for less shiny objects in the next section.

Next, we determine which of these candidates are consistent with a known 3D object in a hypothesized pose. We do this using a simple, qualitative model of specular reflection that captures the properties of many standard models. Let  $\hat{v}$  be a vector giving the direction from the object to the viewer. Let  $\hat{l}$  give the direction from the object to the center of a compact light source. We will assume that the object is reasonably distant from the viewer and the lighting, so that  $\hat{v}$  and  $\hat{l}$  are constant throughout the scene. Next we define  $\hat{n}_p$  to be the surface normal coplanar with  $\hat{l}$  and  $\hat{v}$ , and halfway between them. That is,  $\hat{n}_p$  is the unit vector in the direction of  $\hat{l} + \hat{v}$ . Our model first assumes that specular reflections are most intense at surface normals  $\hat{n}_p$  (eg., [6]). We also assume that the specular reflection produced by a surface normal,  $\hat{n}$ , that is close to  $\hat{n}_p$ , only depends on the angle between  $\hat{n}$  and  $\hat{n}_p$ . That is, the intensity of specular reflectance is rotationally symmetric about  $\hat{n}_p$ . Finally, we assume that reflectance will be a monotonically decreasing function of  $\hat{n}_p \cdot \hat{n}$ . This model holds strictly for a number of models of specularity, such as Phong [38] and Ward[46]. For other models, such as Torrance-Sparrow[10] and Beckmann-Spizzichino [3] it is true under the conditions listed in Ikeuchi and Sato[22], which usually hold. This model is helpful because it implies that if we threshold specularities based on intensity, the surface normals that produce specular points will form a disk on the Gaussian sphere.

Therefore, we proceed by selecting candidate specularities consistent with such a disk. Our hypothesized pose allows us to match image pixels to corresponding surface normals on the model. We map each pixel to a point on the sphere having the same surface normal, noting the pixels that are specular candidates

(Figure 4). There we find a plane that separates the specular pixels from the other pixels with a minimal number of misclassifications. The plane will intersect the sphere in a disk. The plane  $(\hat{w}, b)$ , defined by  $\hat{w} \cdot v - b = 0$  is found by minimizing (1)

$$\min_{\hat{w}, b} \left\{ \sum_i (\text{sign}(\hat{w} \cdot n_i - b) - y(n_i)) \right\} \quad (1)$$

$$y(n_i) = \begin{cases} +1 & n_i \text{ is specular} \\ -1 & \text{otherwise} \end{cases}$$

where  $\{n_i\}$  are the surface normals. Finding a good linear separator is a well-studied problem. Since our problem is low dimensional, we solve it simply, with a brute-force search over many possible values of  $\hat{w}$  and  $b$ . As possible values of  $\hat{w}$  we use the surface normal associated with every candidate specularly observed in the image, provided that the lighting direction implied by this choice of  $\hat{w}$  is within ten degrees of the approximately known lighting direction. This constraint on lighting direction is needed to eliminate false positive disks. Then, for each of these values of  $\hat{w}$ , we consider values of the disk radius,  $r$ , ranging from .05 to .15 (relative to the unit sphere), in increments of .01. These values are empirically determined to work well. For each disk radius we compute the offset  $b = \sqrt{1 - r^2}$ . The pair  $\hat{w}, b$  that minimizes Eq.1 over all disk sizes is chosen to be the hyperplane that defines the specular disk.

Nishino et al.[34] have used a similar representation to recover lighting. Their techniques are different, because they use multiple images to obtain a more certain estimate of which pixels are highlights. They then map these highlights, from multiple images, to the Gaussian sphere, in the same way that we have, and use these, and a Torrance-Sparrow model to estimate the lighting. Because

we only have a coarse estimate of which pixels are highlights, we make use of a more robust estimation of only a single, compact light source.

In summary, this algorithm finds the maximal set of specular points consistent with our knowledge of the object's shape and of the lighting. We can use the degree to which an object can explain possible specular points in the image as a measure of how well it fits the image. In Section 4.1 we use this information to identify objects having no Lambertian reflectance.

### 3 Recognition of Hybrid Objects

In this section we show how to incorporate knowledge about highlights into a more general scheme for recognition of objects that have diffuse and specular components. In this method, we do not assume prior knowledge of lighting direction. We also do not assume knowledge of the relative magnitudes of Lambertian and specular reflectance. It seems impossible to recognize a Lambertian object without some knowledge of its Lambertian albedo, since an object with unknown albedo might have any image painted on it. We first present our algorithm assuming we know the Lambertian albedo. Since this can be tedious to measure for shiny objects, we will later present a method that only requires a reference image of the object, registered with the 3D model, but taken with unknown lighting.

We will describe each step of the algorithm in a separate subsection. Overall, the algorithm proceeds as follows:

- (1) Find an approximation to the lighting using a purely Lambertian model.  
This also gives us an estimate of the luminance (light reaching the cam-

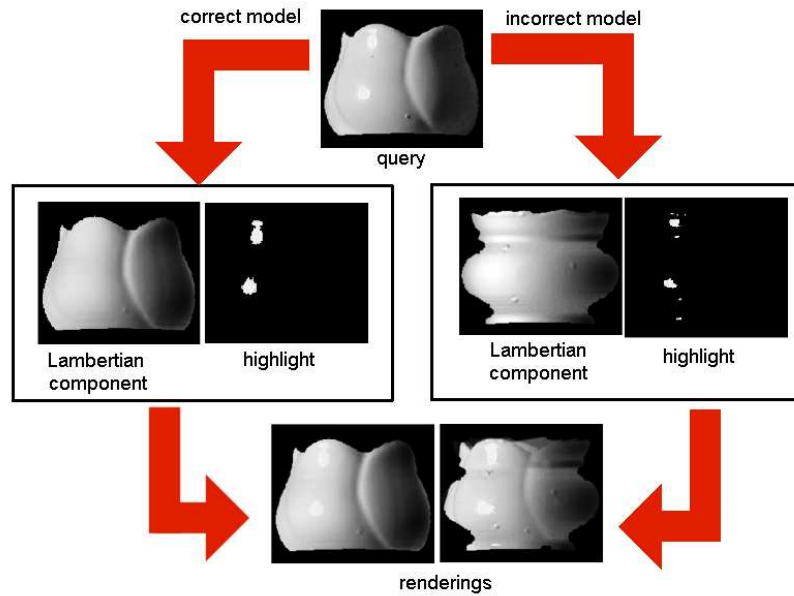


Fig. 5. Recognition of Hybrid Objects

- era) that is due to purely Lambertian reflectance.
- (2) Use this estimate of luminance to locate candidates for specular highlights.
  - (3) Find specularities using the method in Section 2.
  - (4) Re-estimate the lighting and Lambertian luminance excluding specularities.
  - (5) (Optional) At this stage, since we have separated Lambertian and specular reflectance in the image, if we desire we can recover the intensity profile of the specularity, and the Lambertian albedo within highlights if the albedo is unknown.
  - (6) Finally, we compare the image to the specular and Lambertian luminance produced by the resulting model and recovered lighting.

### 3.1 Recovering Lighting and Lambertian Appearance

In the first step of our algorithm, we use a Lambertian model to recover information about lighting in the scene. To do this, we apply the model proposed by Basri and Jacobs[1] and Ramamoorthi and Hanrahan[40].

This model approximates the set of images an object produces under varying illumination by a 9D linear subspace that is computed from the 3D model. The dimensions of this subspace are low-degree polynomial functions that are spherical harmonics (MacRobert[27]) of the surface normals of the object in a specific position, scaled by albedo, ie.:

$$b_{nm}(x, y) = \lambda(x, y)h_{nm}(\theta(x, y), \phi(x, y)), \quad (2)$$

where  $b_{nm}$  are basis functions in the image for  $(0 \leq n \leq 2, -n \leq m \leq n)$ , each of which describes how the object will look under a different, low-frequency lighting condition.  $\lambda(x, y)$  is the albedo at pixel  $(x, y)$  and  $h_{nm}$  is the spherical harmonic evaluated at the surface normal  $(\theta, \phi)$  corresponding to pixel  $(x, y)$ . For example,  $b_{00}$  represents the appearance of the object when lighting intensity is constant from all directions.  $b_{10}$  describes the object under light that is brightest from above, and falls off in intensity with the cosine of its angle to the  $z$  axis. [1,40] show that nine basis images accurately approximate all images of a convex, Lambertian object. Given an image  $I$ , [1] seeks a vector  $a$  that minimizes  $\|Ba - I\|$ .  $B$  denotes the basis images, arranged as a  $p \times 9$  matrix, where  $p$  is the number of points in the image. Every column of  $B$  contains one harmonic image  $b_{nm}$ , as per equation 2. The vector  $Ba$  corresponds to the rendered image that can be produced by the model with low frequency lighting that best fits  $I$ .

After solving this linear minimization problem, we can derive the low frequency components of the lighting from the coefficients  $a$  (see [1] for details). As [40] point out, it is not possible to use a Lambertian object to accurately determine the high frequency components of lighting. In Section 3.3 we discuss how to use this low frequency information to determine the direction of a dominant, compact light source.

### 3.2 Finding Highlight Candidates

Very shiny objects, like glass, produce bright specularities that we can find by simple thresholding. Lambertian plus specular objects may produce highlights that are less bright, because there is relatively less specular reflection. Brighter pixels can arise from Lambertian shading, or from light albedos. To account for this, we consider  $I_{diff} = I - Ba$ , the difference between the real image and the image we render using Lambertian reflectance. This is the intensity that we cannot explain with a Lambertian model. Where this difference is due to specular highlights, the rendered image will be dimmer than the true image. We therefore threshold  $I_{diff}$ , designating the brightest 5% of the pixels as candidates for specular highlights. We set the threshold to get many false positives and fewer false negatives, since we don't want to miss the specular highlight. The next phase of the algorithm is capable of rejecting candidates that do not correspond to specular highlights.

### 3.3 Finding the Highlights

Using these candidate highlights, we can find the ones most consistent with a hypothesized model using the method in Section 2. The only variation for partly Lambertian objects is that we now insist that the lighting direction that we find be consistent with the computed, low-frequency lighting, rather than a known lighting direction. We estimate the light source direction with the vector  $v$  that minimizes  $\|E - Nv\|$ , where  $N$  is a  $p \times 3$  matrix of the model's surface normals, and  $E$  is a rendering of these normals under low-frequency lighting.

Of course, specular reflections can cause errors in our estimates of lighting direction, which uses a Lambertian model. However, several factors tend to make these errors small. First, consider the worst case of a purely specular object (ie., a mirror), such as a sphere. Let  $\vec{v}$  denote a vector from the center of the sphere to the viewer, and let  $\vec{l}$  denote a vector from the center of the sphere to the light. If the angle between  $\vec{v}$  and  $\vec{l}$  is  $\theta$ , then there will be a highlight half between the two. Using a Lambertian model, we will recover lighting in the direction of the highlight, producing an error of  $\frac{\theta}{2}$ . Second, we have found that even quite shiny objects have much more diffuse than specular reflection. In this case, the recovered light will be a compromise in which errors due to specularity play a smaller role. Third, while most Lambertian reflection is due to low frequency components of lighting ([40,1]), most specular reflection is due to high frequency components of lighting that do not effect our lighting estimate. Therefore, in practice we find that we encounter much less than the maximum possible error of  $\frac{\theta}{2}$ . As a consequence, our method produces sufficiently accurate results to identify highlights.

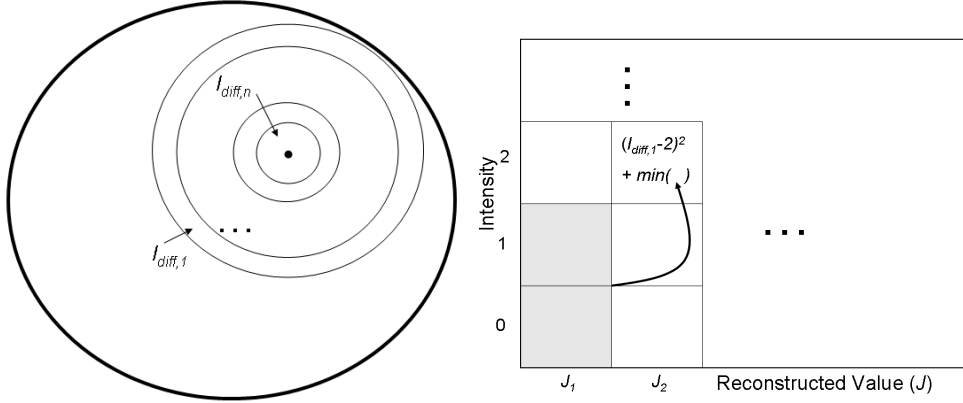


Fig. 6. Left: We map the difference between observed intensities and those predicted by a Lambertian model to the Gaussian sphere, and group these into concentric circles. Right: We can use dynamic programming to find the monotonically increasing values that best match the mean intensities in these circles. The cost of assigning a specific value to one circle depends on the mean intensity of pixels in that circle and the minimum cost that can be obtained by assigning a smaller intensity to the previous circle.

### 3.4 *Recomputing the Lambertian Component*

In this step, we simply recompute the lighting (as described in Section 3.1 or in Section 3.7, Equation 3), excluding the pixels we have identified as specular. This improves our estimate of luminance from Lambertian reflectance.

### 3.5 *Computing the Specular Profile*

For the case of glossy objects, we can then recover the intensity profile of the specular highlight. If we allow this profile to be arbitrary, then we could perfectly fit the image by setting the profile to be  $I_{diff}$ . Alternately, we can constrain the specular intensities at a surface normal to be a monotonically decreasing function of the angle between that normal and the normal at the

specular peak. In practice we find that we can only recover the specular profile when our 3D models are fairly accurate.

In experiments reported in this paper, we fit the specular profile in a simple way (Figure 6 Left). We divide  $I_{diff}$  into concentric rings (on a Gaussian sphere) so that each ring contains pixels with surface normals at roughly the same angle relative to the specular peak. Then we choose the value of each ring in the specular profile to be the mean of the corresponding values in  $I_{diff}$ . In our experiments, these values happen to obey the monotonicity constraint.

In general, this need not be the case. However, we can efficiently find the optimal specular profile that does obey the monotonicity constraint. To explain this, we index the pixels in  $I_{diff}$  as  $I_{diff,i}$ , arranging them in a 1D list. Let  $J_i$  denote the value assigned in the specular profile to the point  $I_{diff,i}$ . We order the pixels so that if  $i < j$  then the monotonicity constraint requires  $J_i \leq J_j$ . Then, our goal is to find  $J_1 \dots J_n$  that minimizes  $\Sigma(I_{diff,i} - J_i)^2$  subject to the constraint that  $J_i \leq J_{i+1}$ .

It is straightforward to solve this using dynamic programming (Figure 6 Right). To see this, note that once we have assigned an intensity  $J_i$  to match pixel  $i$  in the difference image, the cost of the best matching of  $J_{i+1} \dots J_n$  to  $I_{diff_{i+1}}, \dots, I_{diff_n}$  is independent of all  $J_1 \dots J_{i-1}$ . So we can use dynamic programming to fill in a table in which each cell indicates the lowest cost matching in which monotonicity is preserved and pixel  $i$  is assigned value  $J_i$ . If there are  $m$  possible intensity values, this table has  $nm$  entries. Filling in a cell of this table only requires looking at all cells representing values of  $J_{i-1}$ . The total cost of this approach is  $O(m^2n)$ . Other tricks can speed this up further, but we omit these since they have not proven necessary.

### 3.6 Verification

At this point, we have computed the lighting, luminance due to this lighting and Lambertian reflectance, and the position and intensity of specular highlights that are consistent with the lighting and object geometry. We identify an object by doing this for all models and measuring the residual error between the query image and the rendered models. We find the one that best fits the image by minimizing the sum of square differences (SSD). For correct objects, we can fit the Lambertian and specular parts of the image they produce; for incorrect models incorrect specular computation leads to extra errors, plus a failure to account for true specularities. Figure 5 shows an example of matching a query image to correct and incorrect models. <sup>4</sup>

### 3.7 Handling Unknown Lambertian Albedos

When Lambertian albedo is unknown we can use a reference image of the object instead. This image is aligned with the 3D model and taken under unknown lighting conditions. Let  $I_1$  and  $I_2$  be the reference image and the query image. Let  $E$  represent the irradiance, the image that would be produced by the illumination if the object had uniform, white albedo. The image is the pointwise product of  $E$  and the albedo. Since  $E$  is dominated by low frequency components of lighting, we estimate it by accounting for as much of the image as possible with low frequency variation in lighting. Similar ideas are used in

---

<sup>4</sup> Figure 5 shows images from our experiments for which we do not measure the exact Lambertian albedos. Rather we estimated them using an algorithm explained in the next section. The comparison step is the same either way.

homomorphic filtering (Gonzalez and Woods[19]) methods and the Retinex theory of lightness constancy (Land and McCann[25]). Specifically, if  $I$  is the image (a column vector of  $p$  elements) and  $H$  is a  $p \times 9$  matrix of the nine spherical harmonic basis vectors  $h_{nm}$  on the image, then by solving the least squares problem

$$\min_a \|Ha - I\|, \quad (3)$$

we find the coefficients  $a$ , which approximate the incident irradiance or illumination effects,

$$E \approx Ha. \quad (4)$$

This approximation will be exact when the albedo contains no low frequency components, except for a zero order, DC component. In practice, we find that we obtain good results for a variety of real objects in which this is not strictly true.

We estimate the low-frequency incident irradiance or illumination effects  $E_1$  and  $E_2$  separately for both images, as per equations 3 and 4 excluding the pixels inside the highlights that we found using the method from Section 3.3.

At each pixel, for both images we now know the effects of illumination. We pick an albedo  $\lambda$  to scale these intensities to minimize the sum of square error in both images.

$$\min_{\lambda} \|I_1 - \lambda E_1\| + \|I_2 - \lambda E_2\| \quad (5)$$

where  $\lambda E_1$  and  $\lambda E_2$  are pointwise multiplications.

For pixel  $j$ , this is done by choosing  $\lambda_j$  so that:

$$\lambda_j = \frac{E_{1,j}I_{1,j} + E_{2,j}I_{2,j}}{(E_{1,j})^2 + (E_{2,j})^2}. \quad (6)$$

Now we have an initial approximation of albedo and irradiance functions in both images. One could use an iterative process where the current estimation of albedo is used for re-estimation of irradiance. However, we found that this iterative process didn't improve the initial estimates. We emphasize that we neither require nor recover accurate parametric models, since we only use them to tell whether the images might be consistent with the 3D model, not to render new images. In our empirical tests, we have seen that the initial estimates suffice for the purposes of recognition.

The above process approximates the albedos in pixels outside the highlights. We can also estimate the albedos inside the highlights, by combining the above method with the specular profile estimation. Specifically, we consider the set of pixels belonging to the specular highlight in the query image  $I_1$ , and the corresponding set of nonspecular pixels in the reference image  $I_2$ . (Here, we make an assumption that specular highlights in  $I_1$  and  $I_2$  do not overlap.) We want to explain these intensities as a combination of Lambertian reflectance in both images, and a specular highlight in image  $I_1$  whose intensity is rotationally symmetric and monotonically decreasing about the specular peak.

We can express this by saying we want to minimize:

$$\sum_{j \in \mathcal{J}} (I_{1,j} - (\lambda_j E_{1,j} + s(N_j \cdot P)))^2 + (I_{2,j} - \lambda_j E_{2,j})^2 \quad (7)$$

Here  $P$  is a vector giving the direction of the specular peak.  $N_j$  gives the surface normal corresponding to  $I_{1,j}$ .  $s$  is a function expressing the specular profile in  $I$  as a monotonically decreasing function, and  $\mathcal{J}$  contains the indices of all pixels in the specular highlight.

We can efficiently find  $\lambda$  and  $s$  to minimize this expression. We divide the highlight into a series of rings as we did in Section 3.5. Beginning with the outer ring, we solve for the albedos and the value of the specularity in the ring; this is a straightforward minimization of a quadratic. Again, it is possible to enforce the monotonicity constraint by using dynamic programming, as described in Section 3.5.

A single reference image will provide information about the albedos over only a portion of an object. We will experiment with this approach for face recognition, in which we only expect to deal with intensities produced by the front of a face. To apply our method to general 3D objects that might be viewed from any direction, we will need reference information about all portions of the object. This can be obtained by taking multiple images of a fixed object in which we do not alter the lighting. In this case, intensities can be mapped onto the 3D structure of the object, and used to generate a virtual reference image from any viewpoint, for verification of that hypothesized viewpoint. This essentially amounts to texture mapping the reference image onto the object and projecting them into an image, except we do not treat the projected intensities as true albedos, but as a reference image produced under some unknown lighting condition. Note that this approach requires us to take images that together cover all areas of the object; in general there will not be too many.

## 4 Experiments

We designed the experiments to test each component of our approach in a setting that would be challenging for it. First we evaluate the specular component of our comparison method described in Section 2. In this test we identify the transparent objects shown in Figure 7, bottom. These objects have no Lambertian reflectance, and so they allow us to test the value of purely specular cues to identity. Note that model-based methods that assume Lambertian reflectance will obviously fail on these objects, because they have no Lambertian reflectance. Fitting 3D to 2D based on edge information seems very difficult, since edges mostly belong to the background. Consequently, the edge maps of different objects against the same background are more similar than the edge maps of the same object against different backgrounds.

Next we test this specular comparison method in more general setting, where the pose is unknown. In this experiment the goal is to locate transparent wine glasses on a table top. These glasses have known shape and only two degrees of freedom in their position, but we know of no other system that localizes such challenging objects.

We next test the method described in Section 3 on more general objects that have both specular and Lambertian components. We evaluate the ability of this comparison method to identify pottery. Although our method can make use of albedo information, identification is relatively easy between sets of objects that have very different albedos. So to test the effectiveness of our method on a moderately sized data set, we have chosen most of the models with very similar albedos. It is easiest to ensure that our models have similar albedos

if we choose models of nearly uniform Lambertian albedo. For these models, recognition depends mainly on cues from specularity and smooth shading. To show that the algorithms do not have any prior knowledge that the albedos are nearly uniform we added models with variation in texture.

In this experiment we estimated albedos using the method described in Section 3.7. We also tested the method for albedo recovery using a reference image of the object in a more challenging task of identification of textured objects with similar appearance, specifically faces.

#### 4.1 Purely Specular Objects

To test the method described in Section 2 we identify the transparent objects shown in Figure 7, bottom. Given a query image,  $I$ , we produce a binary image  $I_{bin}$  by thresholding  $I$  to find the brightest pixels, which we consider to be candidates for specular highlights. Since the object is transparent, two normals help produce every pixel in the query image: one for the front surface and one for the back, and either of the surfaces may produce specularity. Consequently, we map each candidate highlight to two points on the sphere. As a result, specular highlights map to a disk in the correct location, and also in another place on the sphere. The mapping of specular pixels to the second location simply acts as more noise, similar to that produced by spurious specular candidates. The disk found is then mapped back and the resulting binary image  $I'_{bin}$  is matched to image candidates  $I_{bin}$  by computing the overlap between them:

$$overlap = \frac{size(I_{bin} \wedge I'_{bin})}{size(I_{bin} \vee I'_{bin})} \quad (8)$$

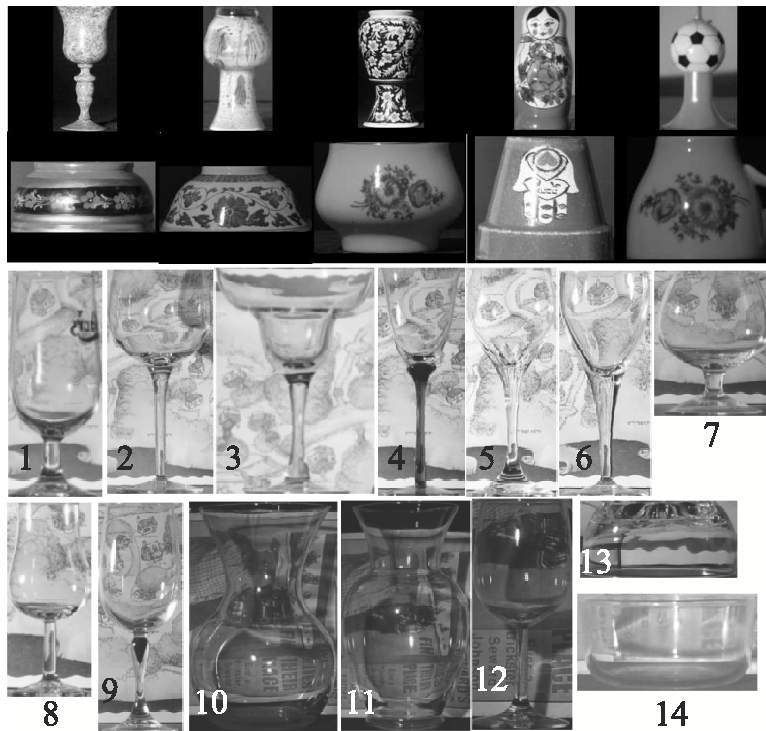


Fig. 7. Objects tested in the experiments. Top – smooth objects for the Lambertian plus Specular test; middle – textured objects for the Lambertian plus Specular test; bottom – transparent specular objects. The numerical labels on the transparent objects are added for the confusion table below. Test images have no labels.

The best match corresponds to the biggest overlap. If the image and the model belong to the same object, then the two methods will produce similar segmentations, otherwise the mapping method that uses the wrong model will detect erroneous specularities that don't match the segmentation produced by



Fig. 8. Typical viewing conditions in the glassware experiment: top – different backgrounds, bottom – an example of changing lighting directions

thresholding.

Figure 4 shows an example in which we attempt to explain the specular highlights in the query image. The top figure shows the flow of the algorithm when the correct model is used. The algorithm correctly finds specularity ( $I'_{bin}$ ) using the disk model and the corresponding geometry. The bottom figure shows the results of the algorithm using the wrong model. The specularity then is inconsistent with the specular highlights in the query image.

We tested the proposed algorithm on 54 images of 14 objects taken with different lighting directions and against different backgrounds. Typical backgrounds are shown in Figure 8 top. We moved the light source manually, so the lighting directions were not precisely known. In general, they can be characterized as left, frontal, and right illuminations (Figure 8 bottom shows an example of the light sources directions). This introduced significant noise in our prior

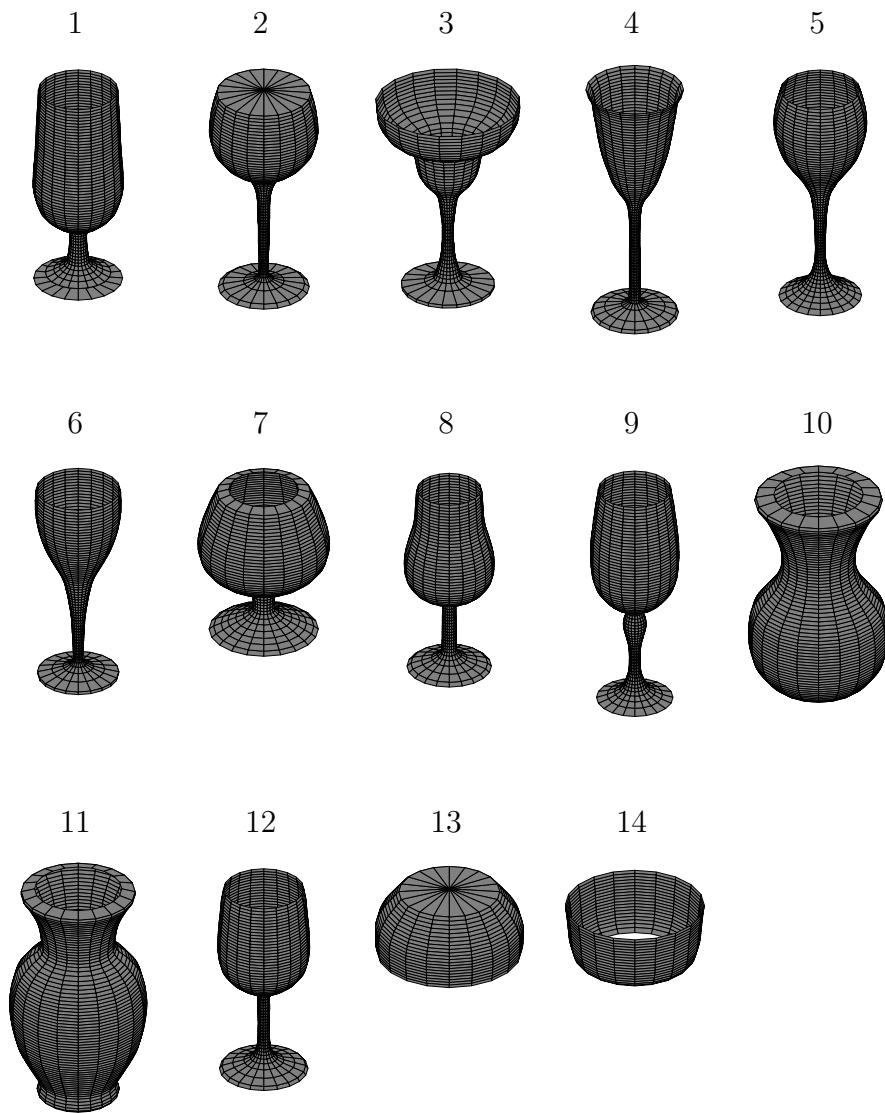


Fig. 9. 3D models of the transparent object.

knowledge of lighting direction. However our method showed robustness to such noise.

The 3D models (shown in Figure 9) were computed from 2D images assuming rotational symmetry. To do this we took an image of each object filled with an opaque liquid. Then the image was rotated so that the rotational symmetry axis of the object coincided with the vertical axis of the image. The boundaries of the object were manually extracted from the image and some smoothing

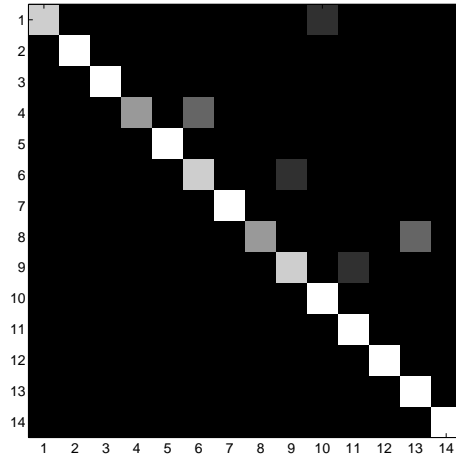


Fig. 10. The confusion matrix of the transparent objects. White corresponds to 100% recognition, black to 0%. The columns correspond to the models.

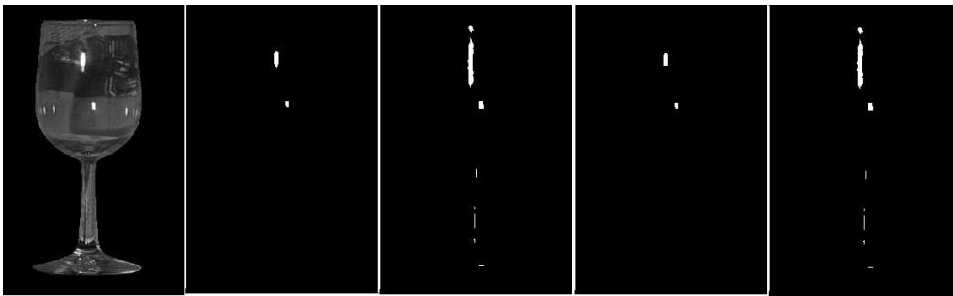


Fig. 11. An example of errors in specular localization as the result of errors in the 3D model. From left to right: image masked by a model, specular candidates obtained by thresholding, recovered specularity (with errors), “AND” between thresholded and recovered specularity, and “OR” between thresholded and recovered specularity. These images show that when the 3D model has large mistakes the predicted specularities poorly matches the thresholded highlight.

was done prior to computing the normals. Then 2D normals to the boundaries were calculated at each point. 3D model normals were calculated as follows: for each  $y$  on the symmetry axis of the object we define a radius of the object  $R_y$  as the distance between the symmetry axis and a boundary point at the height  $y$ . The 3D normals for the  $x$  going from the left to the right boundary

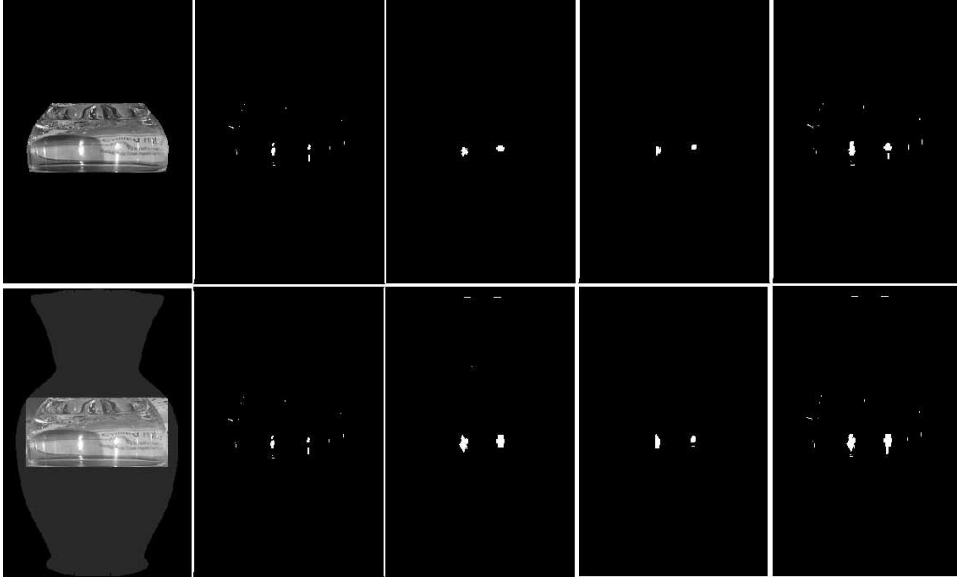


Fig. 12. An example of misclassification because two shapes can produce similar highlights. First row – correct model, second row – incorrect model. From left to right: tested image superimposed on the model silhouette, specular candidates obtained by thresholding, recovered specularity, “AND” between thresholded and recovered specularity, and “OR” between thresholded and recovered specularity. In corresponding top and bottom images the specularity is almost the same, because the correct and incorrect models have similar surface normals inside the specular highlight.

points are computed as:

$$N_x = \frac{x}{R_y} \cos \theta_y \quad N_y = \sin \theta_y \quad N_z = \frac{\sqrt{R_y^2 - x^2}}{R_y} \cos \theta_y$$

where  $\theta_y$  is an angle between the 2D normal at the right boundary point and the  $x$  axis. We emphasize that our method is not restricted to rotationally symmetric objects. We run our tests on such objects only to simplify the construction of 3D models.

Each test image was compared to 14 models. The one that produced the biggest overlap between the specular candidates (obtained by thresholding)

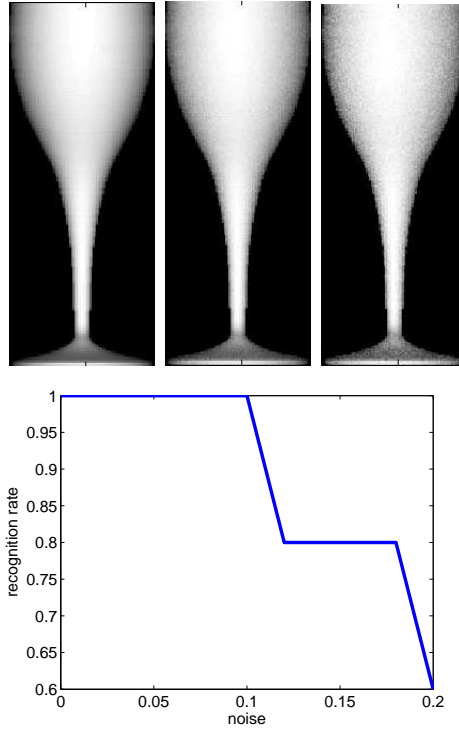


Fig. 13. Noisy 3D models: Top - renderings of a model with different levels of noise in normals. From left to write: no noise, uniform noise 0.1, uniform noise 0.2. Bottom - recognition rates as a function of noise level injected to normals.

and the specularity obtained from the disk was chosen as the correct object. 87% of images were recognized correctly using only consistency between the geometry and the specular highlights in a query image. Figure 10 shows the confusion matrix between the tested objects.

The sources of errors in our system are:

- (1) Large errors in some 3D models. All models used in our experiment are quite noisy. However large errors in curvature can result in large errors in specularity prediction. An example of such an error is shown in Figure 11. The real object is more curved than the manually created model. Consequently, the predicted specularity is significantly larger than the real one.

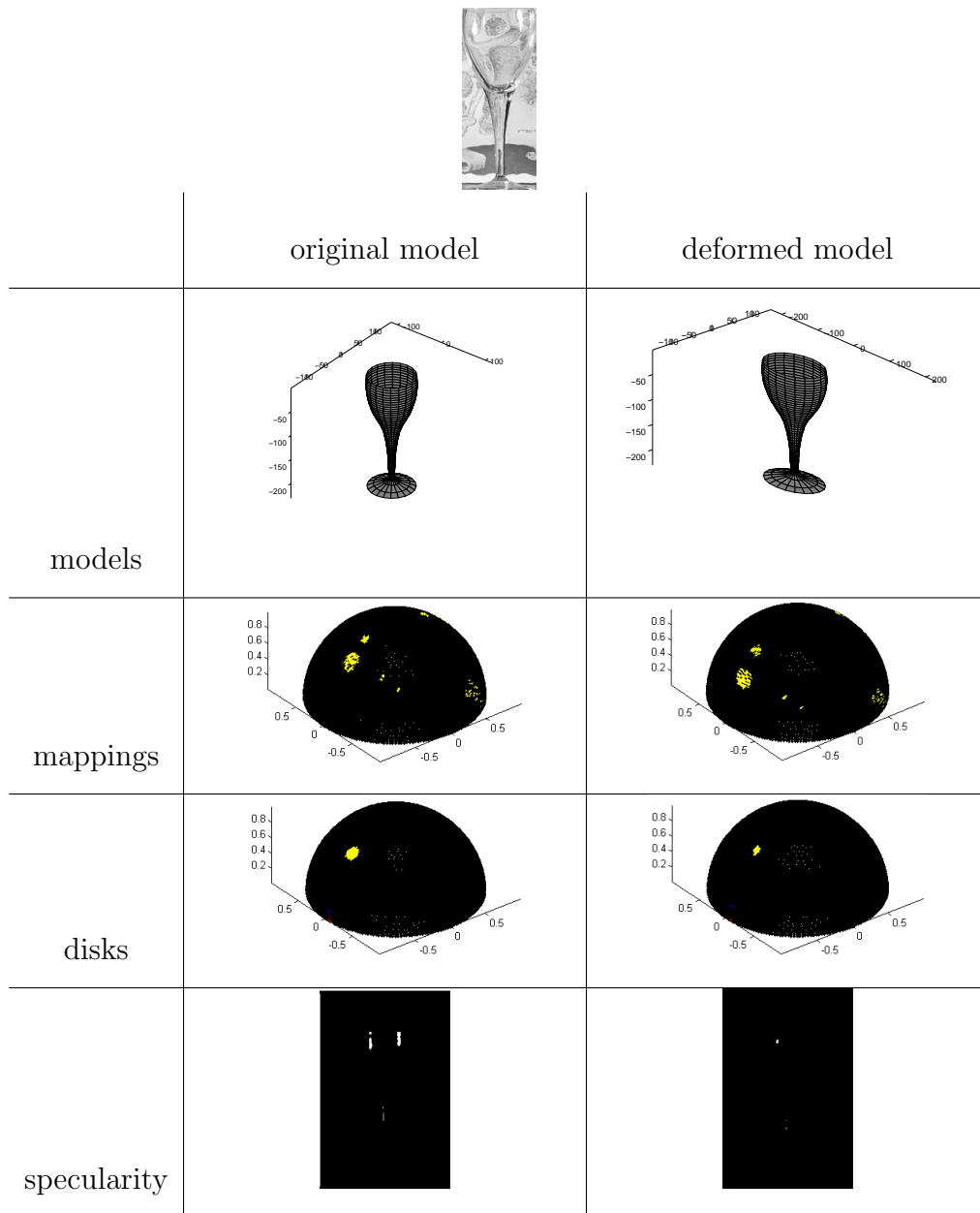


Fig. 14. Example of deformation in 3D model: First row - test image. The table compares the stages of the method between the original and significantly deformed model.

- (2) Two different objects may have very similar shape in the area that produces highlights (Figure 12). This example demonstrates that even though specularities present a positive cue in recognition, using all available information is essential in designing robust recognition models.

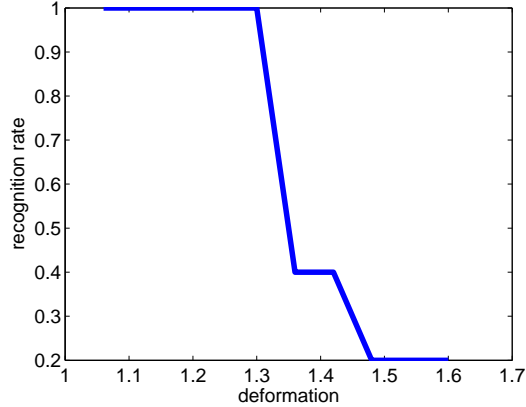


Fig. 15. 3D deformation test: Recognition rate as a function of model deformation. We tested the robustness of our approach to two types errors in 3D models: random noise in normals and smooth deformation of 3D shape. In the noise test we injected uniform noise in each coordinate of the normals, with the maximum noise varying between 0 and 0.2. The examples of noisy normals and the recognition degradation as a function of noise are shown in Figure 13. This test shows that our method is quite robust to the random noise in normals.

In the deformation test we gradually deform the model from circular cross-section to elliptical. One of the radii (for each height) is kept constant and the other is changed in the range  $[1.1r, 1.6r]$ , where  $r$  is the original radius. This way we change the 3D model and thus the prediction of the specularities, but we do not change the occluding contours - thus the registration between the model and the images is kept unchanged. Figure 14 compares the mappings onto the Gaussian sphere and the computed secularities of the original and highly deformed (60%) models. This example shows that large deformation of 3D shape corrupts the prediction of specularities and damages the recognition results. However our experiments show (Figure 15) that the method is robust to moderate (up to 30%) deformation of shape.

We have not methodically considered the effect of pose on recognition accuracy. Since our paper focuses on lighting effects, we have in many cases limited our experiments to a single pose. In the case of the glass-finding system, described below, pose variations do not effect appearance much, and have not effected recognition performance.

#### *4.2 An Automatic System to Find Glass Objects on a Table.*

While this paper focuses on verification, we now provide one example of how to use this in the inner loop of a system that recovers the position of glasses on a table top. We also demonstrate that highlights can be used to locate a known object in a scene. In these experiments, we placed wineglasses on a table with a camera level with the center of the height of the glass. We measured the approximate position of a single light bulb illuminating the scene. This, and knowledge of the table position constrained the angle between the light, glass, and camera, to a small range. We used the center of this range as our estimate of lighting angle and of the angle from the object to the camera. We keep this estimate constant for all object positions, allowing for some uncertainties in these estimates in our algorithm. In this setup, different positions of the glass correspond to translation in the  $x$  direction of the image, and to scaling. Knowledge of the table geometry was used to constrain these values. Still, recognition requires some search over these two degrees of freedom in the glasses' positions.

To speed up this search, we made use of a coarse upper bound on the measure of fit between the model and image given in Equation 8. Given a hypothesized position and scale of the glass, we can compute the number of specular pixels

that intersect a mask representing the glass' shape (this is  $I_{bin}$ ). From the model, we generate a second mask to represent all of the pixels within the glass that correspond to a normal that can reflect light to within 10 degrees of the direction to the camera. This mask does not depend on the glass' position. The intersection of this mask with the image specularities will produce an upper bound for the possible image specularities that the model could generate in this position. We call this  $I'_{bin,up}$ . Using this upper bound on  $I_{bin}$ ,  $\frac{I_{bin} \wedge I'_{bin,up}}{I_{bin}}$  provides an upper bound on the value of *overlap*.

While the full computation of *overlap* described in previous sections takes .5 seconds per position in our implementation, the computation of this coarse upper bound takes only .005 seconds. So we compute this upper bound for all possible positions of the glass. Then we sort these positions by this bound, and compute the full cost function on those with the highest upper bound. In our experiments, when we found five positions whose match score was higher than the upper bound on all remaining positions, we terminated the search. On average, this allowed us to avoid evaluating the full cost function on about 95% of the poses. For example, in the case of the scene at the top of Figure 16, specularities were fully rendered for 60 positions, while over 2500 were skipped. Total run time was less than one minute on a standard hardware. Results for the four scenes are shown in Figures 3 and 16. Figure 17 shows the highlight candidates for one image. This demonstrates that our method can localize glasses even in images with other highlights and light pixels. Together, these experiments demonstrate that the shape and position of highlights are distinctive enough to allow us to find the location of a known object in a moderately complex scene.

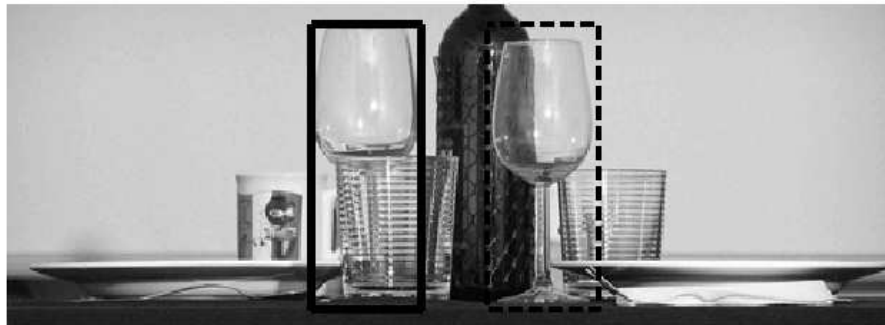
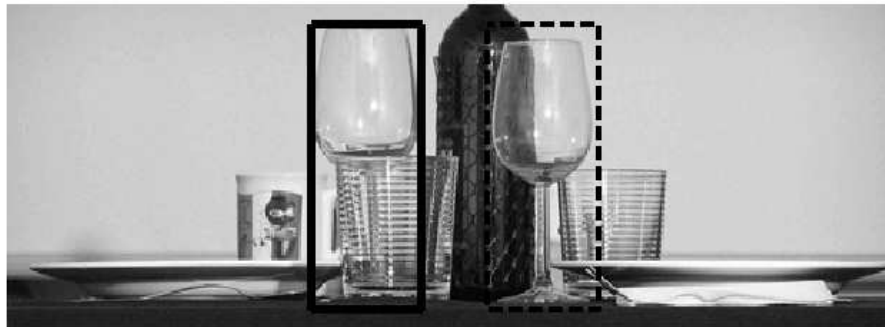
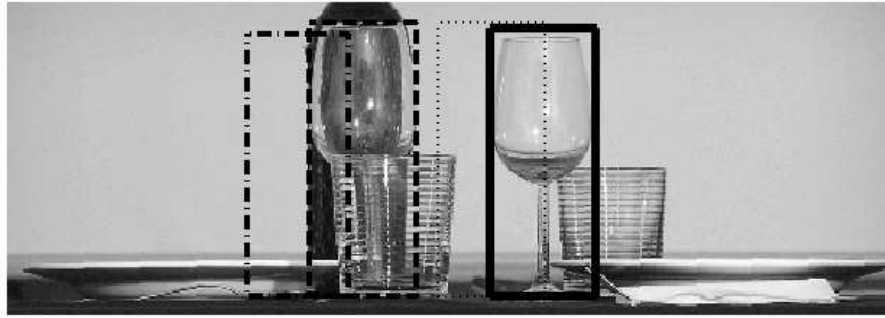


Fig. 16. Three scenes and the positions found by our algorithm that best fit the glass model. The highest scoring position is shown as the thickest, solid line. The second position is the next thickest, dashed line. In every image, the correct positions are the top two, although the scale of the identified glass is slightly off in a couple of cases.

#### 4.3 *Lambertian plus Specular Objects*

We have tested the algorithm for Lambertian plus specular surfaces on a database of 30 objects made from shiny ceramic and wood (Figure 7 top

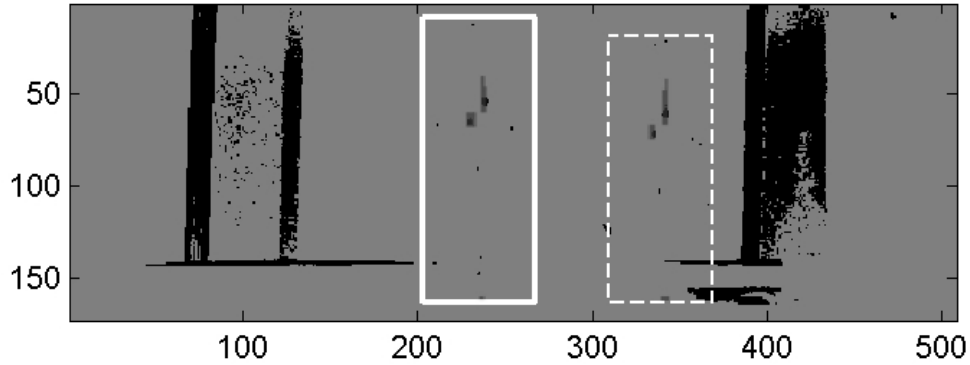


Fig. 17. For the scene in Figure 3 we show the candidate highlights produced by simple thresholding as black, the two best hypotheses are shown as white rectangles, and the rendered highlights for these models are shown as gray.

and middle). The set consists of two groups of objects. The larger group of 20 objects have very similar almost uniform albedo (Figure 7 top). Range images of these objects were obtained using a Cyberware Model 3030 Rapid 3D Scanner. These models were somewhat noisy and inaccurate due the difficulty of scanning shiny objects. We have registered the brightness images with the range images using three feature points, which were enough since there was no out-of-plane rotation.

The other group of 10 objects have textured albedo and are rotationally symmetric (Figure 7 middle). Their models where obtained from occluding contours as with the crystal objects discussed above.

116 intensity images of 30 objects where taken under different illuminations in a range of up to approximately 70 degrees from frontal (Figure 18). The specular highlights were significant in all images. Since the models were very noisy, we didn't recover the specular profile in this test. The highlights where localized using the algorithm described in Section 3.3. Then we masked out the highlights, re-estimated the Lambertian component and approximated the

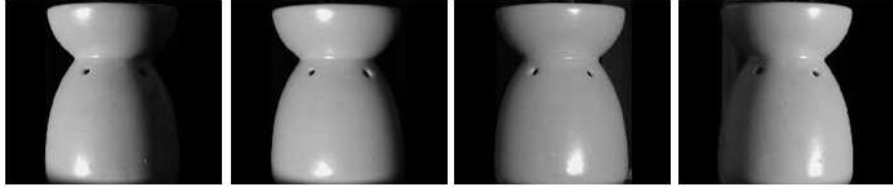


Fig. 18. Illumination directions used in the pottery experiment

albedos using a reference image as outlined in Section 3.7. The object was recognized by doing this for all models and then measuring the residual error (SSD) between the query image and the rendered models, with the specular-ity masked out. Note that excluding the identified specular pixels from the verification step doesn't mean that we discard the information from specular-ites, because when incorrect models are used to explain the image, the masked out areas do not correspond to the true highlights, while the actual highlights remain in the rendered image. This amplifies the error when we attempt to verify incorrect models.

Each object was captured under four different illuminations. Thus we performed four rounds of tests using a different image as a reference at each round and three remaining images as queries (total 360 test images). In this experiment our method showed a 95.2% overall recognition rate.

Figure 19 shows two examples of objects with accurate models where we not only locate the highlights, but also recover the specular profile and the albedo inside the highlight.

Our method is not directly comparable to existing methods. First, the few approaches for recognition of specular objects either work on specific types of objects or a require physics-based simulator or a polarization filter (see Section 1.1). Thus it is hard to perform a fair benchmark; these methods would

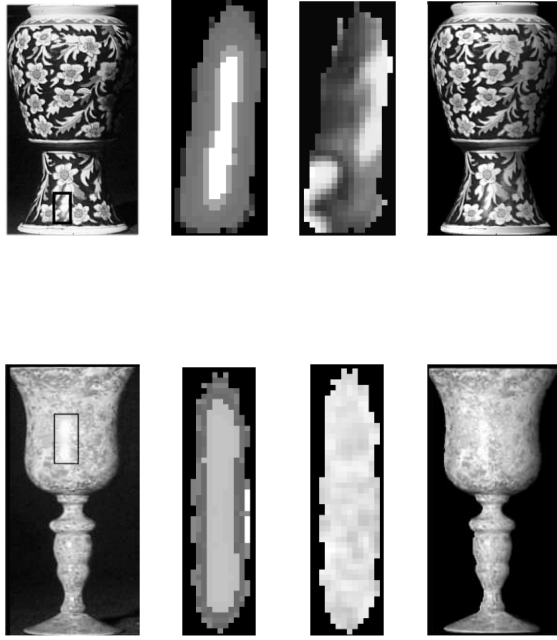


Fig. 19. Nonuniform albedo objects with accurate models. From left to right column: query images, specular profiles, albedos inside specular highlights, rendered model. We zoomed in on the albedos and specular profiles at the areas marked by rectangles in the query images.

	Recognition Rate(%)
Lambertian+Specular with reference image	95.2
Lambertian (Section 3.7)	93.5
Gradient direction	38

Table 1

Pottery Recognition on all objects (20 smooth and 10 textured)

perform poorly or not at all in our setting. Since there has been so little work on the recognition of specular objects and the vast majority of recognition systems treat specularities as noise, we feel that it is appropriate to compare our method to the existing Lambertian methods that ignore specularities. These tests will show the importance of going beyond a Lambertian assumption.

	Recognition Rate(%)
Lambertian+Specular with reference image	98
Lambertian+Specular assuming uniform albedo	98
9D subspace assuming uniform albedo	95

Table 2

Smooth Pottery Recognition (on 20 smooth objects)

First we consider a complete method using all the components we have described, including the use of a reference image and accounting for specular reflection. We call this "L+S". Next, we consider a method that does not account for specularities, which we call the new Lambertian method. This method is primarily different from the 9D method of Basri and Jacobs in using reference images, as described in Section 3.7. We also compare to the direction of gradient method, which is proposed in Chen et al.[9] for illumination insensitive image comparison. This comparison is admittedly unfair to [9] because that method doesn't use 3D information and the query image is directly compared to the reference image. However this does provide a baseline for our tests that involve reference images. Moreover, the comparison is of interest because Chen et al. do report good results for face recognition, even in comparison to 3D methods. Table 1 shows the results of these tests.

As apparent from the results, L+S performs slightly better than the new Lambertian method. Both methods, though, perform quite well because the size of the data set is not extremely large and the objects are quite different (especially textured objects). Since both methods output a distance between the test and the rendered images, we can obtain a finer grained comparison between the methods by investigating the distances they produce. The method

should perform well if the distance between the test image and the rendered image using the correct model is small (we'll call this the *correct distance*) and the distance using an incorrect model is large (*incorrect distance*). The larger is the gap between correct and incorrect distances the more robust recognition will be.

For each object, we measured the difference between the correct distance and the mean of the incorrect distances divided by the standard deviation of the incorrect distances. Averaging this distance over all objects, we found that for L+S, the correct distance is, on average, 2.8 standard deviations below the mean of the incorrect distances, while for the new Lambertian method the correct distance is 2.6 standard deviations from the mean of incorrect distances. To get an idea of the significance of this difference, suppose that the incorrect distances were distributed according to a Gaussian, and the correct distance was always 2.8, or 2.6 standard deviations below the mean of this distribution. This would mean that for L+S, an incorrect model would have a probability of 0.25% of preferring an incorrect model over a correct model, while new Lambertian would have a probability of 0.5% of making such a mistake. In a data base with 30 models, L+S would prefer the correct match over all 29 possible incorrect matches 93% of the time, while new Lambertian would get the right answer 86.5% of the time. This is similar to, but a bit worse than the performance we observe. Extrapolating, in a database of 100 objects, L+S would be right 78% of the time while new Lambertian would be right 61% of the time. This is a very simplified analysis. However, it illustrates that the differences we observe in correct and incorrect distances produced by the two algorithms may be quite important.

Thus we can conclude that even though the recognition results on the test set

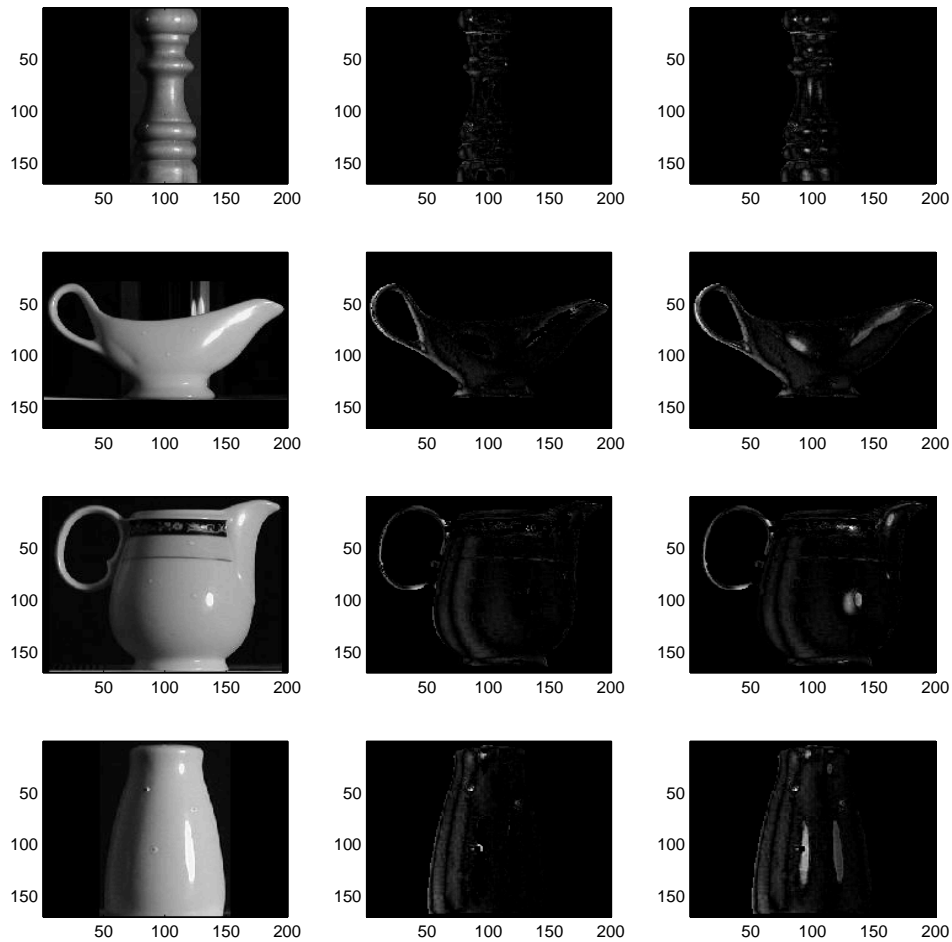


Fig. 20. Errors between test images and their renderings using the correct model. Left column – test image; middle column – error between the test image and the rendering produces by L+S; right – error between the test image and the rendering produced by the Lambertian method. Note the errors at specular highlights in the Lambertian method.

are good in both methods, the method that takes care of highlights is more robust.

Another way to see the advantage of L+S over Lambertian is by looking at the errors between the test images and their renderings. Figure 20 shows that

Lambertian method produces more errors than L+S and these errors are at places of highlights.

Next we compare L+S to the 9D linear subspace algorithm proposed by [1]. The 9D subspace method requires knowledge of Lambertian albedos. We haven't measured the albedos in our experiments. However 20 of the objects have very small albedo variations, so we assumed that albedos are uniform in these objects. We tested L+S also assuming uniform albedo instead of using the reference images. Both methods were tested on 80 images of 20 objects (four images per object). In this new setting L+S was 98% correct and the 9D subspace method showed a 95% recognition rate (Table 2). The L+S method using reference images on this set of 20 smooth objects also produced 98% accuracy. Note that L+S is a more general method that can also handle unknown albedos while the 9D subspace method requires the albedos to be known.

We've tested our method on glossy objects with a small albedo variation to demonstrate that we could handle objects that previous algorithms find especially difficult. However, 10 of the objects have textured albedo and the algorithm works well for those objects too. Figure 19 shows that our approach can handle nonuniform albedo objects even when the specular highlight is not the brightest part of the image (Figure 19, top row).

#### *4.4 Recognition of Objects with Unknown Albedos*

In this section we test the ability of our method to handle non-uniform albedo objects in a much larger experiment. Since the approximation of unknown albedos using reference images is done after the specular pixels are found and



Fig. 21. Four lighting conditions in the face experiment

excluded from the images, this part of the algorithm can be tested on Lambertian objects as well. We evaluate the method in Section 3.7 in a face recognition experiment, since faces are textured, similar, and there are available databases of faces that include 3D models.

We have experimented with a database containing 3D models of 42 faces. As query images we use 168 images of 42 individuals, taken across four different lighting conditions (shown in Figure 21) with a frontal pose. As a reference image for each model we randomly select one of the three lighting conditions different from the illumination in the query image. This is consistent with the assumption that the reference image is taken under uncontrolled, variable lighting. In our experiments, each of the query images is compared to each model. 98% of the images were recognized correctly. This result suggests that we can successfully use the proposed algorithm to approximate albedos using a reference image instead of measuring them; and measuring albedos of specular objects is a tedious process.

## 5 Conclusions and Future Work

This paper discusses how to use specular highlights as a positive clue for recognition of shiny objects. We have proposed a simple, qualitative reflectance

model that captures the properties of existing models over any reasonable choice of parameters. Using this model, we can efficiently judge the consistency of a known 3D structure and possible specularities in the image. We have demonstrated that this information can be used to successfully identify very challenging objects with no Lambertian reflectance, such as glassware. We have also shown how to integrate knowledge about highlights with Lambertian methods, so that we can use them in recognizing objects that are part Lambertian, part specular. We extended our method so that instead of using Lambertian albedos, which are hard to measure in shiny objects, it uses reference images of the objects under unknown lighting, aligned with the model. Experiments with glossy pottery show that using specular highlights as a source of information improves recognition.

This is only an initial step towards the recognition of complex objects. It has a few limitations that we plan to address in forthcoming research. Specifically:

- The current method assumes a compact light source. Our specular model can be extended to several light sources. However it requires some knowledge of the direction of these sources. Our current method estimates the light direction from the Lambertian component. The next step will be estimating directions of multiple light sources by incorporating cues from Lambertian and specular components. Accomplishing this will take a large step towards making the proposed method work under arbitrary lighting.
- Our current method treats interreflection as noise. In the future, we plan to extend our method to handle interreflections.
- One source of errors in our experiments was large errors in 3D models of the objects. Our current method is robust to moderate noise. However more work should be done to improve its robustness.

- Currently we test our approach using simple heuristics for hypothesizing highlights. A complete system might improve performance by using other cues, such as color or polarization to more accurately determine highlight pixels.
- Finally, we plan to investigate the possibility of the integration of our method into a full system that allows arbitrary pose variation. An iterative techniques similar to 3D Morphable Models [5] can be considered to recover pose parameters together with identity.

## 6 Acknowledgments

This research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No.608/06)

## References

- [1] R. Basri and D. Jacobs. “Lambertian reflectance and linear subspaces”. In *Proc. of ICCV*, vol. II, pp. 283–390, 2001.
- [2] P. N. Belhumeur and D. J.Kriegman, “What is the set of images of an object under all possible lighting conditions?” *IJCV*, 28(3): 245–260, 1998.
- [3] P. Beckmann and A. Spizzochino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, New York: Pergamon, 1963.
- [4] J.W. Birk, R.B. Kelly, and H.A.S. Martines. “An orienting robot for feeding workpieces stored in bins”, In *IEEE Trans. Systems Man Cybernet* SMC-11(2):151–160, 1981.

- [5] V. Blanz and T. Vetter. “Face Recognition Based on Fitting a 3D Morphable Model,” *IEEE Trans. on PAMI* 25(9):1063 – 107,2003.
- [6] J.F. Blinn and M. Newell. “Texture and reflection in computer generated images”, *Comm. of the ACM*, Vol. 19, No. 10., pp. 542-547, 1976.
- [7] T. Bonfort and P. Sturm. “Voxel carving for specular surfaces”. In *Proc. of Int. Conf. of Computer Vision*, 2003 pp. 394-403.
- [8] G.Brelstaff and A. Blake. “Detecting specular reflections using Lambertian constraints”, *Proc. of ICCV*, pp. 297–302, 1988.
- [9] H.F. Chen, P.N. Belhumeur, and D. W. Jacobs. “In search of illumination invariants”. *IEEE Conf. on CVPR*, pp.254–261, June 2000.
- [10] R. L. Cook and K.E. Torrance. “A reflectance model for computer graphics,” *ACM Trans. Graphics*, 1(1):7–24, January 1982.
- [11] A. Criminisi, S. B. Kang, R. Swaminathan, R. Szeliski and P. Anandan. “Extracting Layers and Analyzing their Specular Properties Using Epipolar-Plane-Image Analysis”, *Computer Vision and Image Understanding* No. 1, January 2005, pp. 51-85.
- [12] A. DelPozo, and S. Savarese, “Detecting Specular Surfaces on Natural Images,” *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, June 2007.
- [13] R. Epstein, P. Hallinan, A. Yuille. “ $5 \pm 2$  eigenimages suffice: an empirical investigation of low-dimensional lighting models”, *IEEE Workshop on Physics-Based Vision*: 108-116, 1995.
- [14] Y. Fukada, H. Doi, k. Nagamine, and T.Inari. “Relationships-based recognition of structural industrial parts stacked in bin”, *Robotica*,2 1984, pp.147–154

- [15] B. Funt and G. Finlayson. “Color constant color indexing”, *IEEE Trans. on PAMI*7(5):522-529, 1995.
- [16] A.S. Georghiades. “Incorporating the Torrance and Sparrow Model of Reflectance in Uncalibrated Photometric Stereo,” *Int. Conf. on Computer Vision*,2003,pp. 816-823
- [17] A.S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, “From few to many: Generative Models for Recognition under variable pose and illumination”. *IEEE Trans. on PAMI*.23(6):643–660, 2001
- [18] A.S. Georghiades, D. J. Kriegman, and P. N. Belhumeur, “Illumination cones for recognition under variable lighting: Faces”, *Proc. IEEE Conf. on CVPR.*, 1998.
- [19] R.C. Gonzalez and R.E. Woods. “Digital Image Processing”. Addison-Wesley, Reading, MA, 1992.
- [20] K.D. Gremban and K. Ikeuchi. “Planning Multiple Observations for Specular Object Recognition”, *IEEE Conf. on Robotics and Automation*, Vol. 2, May, 1993, pp. 599–604.
- [21] P. Hallinan. “A low-dimensional representation of human faces for arbitrary lighting conditions”, *IEEE Conf. on CVPR*: 995–999, 1994.
- [22] K. Ikeuchi and K. Sato. “Determining reflectance properties of an object using range and brightness images”, *IEEE Tran. PAMI*, Vol. 13, No. 11, November, 1991, pp. 1139–1153.
- [23] K. Koshikawa and Y. Shirai. “A 3-D modeler for vision research”, in *Proc. International Conference on Advanced Robotics*,pp.185-190, Robotics Society of Japan,1985
- [24] M. Lades, J. C. Vortbruggen, J. Buhmann J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen. “Distortion Invariant Object Recognition in the

Dynamic Link Architecture”, *IEEE Transactions on Computers*, 42:300–311, 1993.

- [25] E. Land and J. McCann. “Lightness and retinex theory”, *Journal of the Optical Society of America*,61(1):1–11, 1971.
- [26] K. McHenry, J. Ponce, and D. Forsyth. “Finding Glass”, *IEEE Conference on CVPR*, pp. 973–979, 2005.
- [27] T. MacRobert. *Spherical harmonics; an elementary treatise on harmonic functions, with applications*, Dover Publications, 1948.
- [28] Y. Moses. *Face recognition: generalization to novel images*, Ph.D. Thesis, Weizmann Institute of Science, 1993.
- [29] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson. “Image statistics and the perception of surface qualities.” *Nature* 2007.
- [30] H. Murase and S. K. Nayar. “Visual learning and recognition of 3d objects from appearance”. *IJCV*, 14(1):5–24, 1995.
- [31] S.K. Nayar and R. Bolle. “Reflectance based object recognition”. *IJCV*, 17(3): 219-240, 1996.
- [32] S.K. Nayar, K. Ikeuchi, T. Kanade. “Surface Reflection: Physical and Geometrical Perspectives”, *IEEE Tran. PAMI*,13(7):611–634,1991.
- [33] S.K. Nayar, K. Ikeuchi, T. Kanade. “Determining shape and reflectance of hybrid surfaces by photometric sampling,” *IEEE Trans. Robotics Automat.*, 6(4):418–431, 1990.
- [34] K. Nishino, Z. Zhang, and K. Ikeuchi. “Determining Reflectance Parameters and Illumination Distribution from a Sparse Set of Images for View-dependent Image Synthesis”, *ICCV01*, I:599-606, 2001.

- [35] M. Oren, S.K. Nayar. “A theory of specular surface geometry”, *IJCV*, 24(2):105–124, 1997.
- [36] M. Osadchy, D. Jacobs, and R. Ramamoorthi, “Using Specularities for Recognition,” *IEEE International Conference on Computer Vision*, Vol. II:1512-1519, 2003.
- [37] M. Osadchy, D. Jacobs, and M. Lindenbaum, ”On the equivalence of common approaches to lighting insensitive recognition,” *IEEE International Conference on Computer Vision*, 2005 (forthcoming).
- [38] B. Phong, “Illumination for computer generated images”, *Comm. ACM* 18, 6(June 1975) 311–317.
- [39] R. Ramamoorthi, “Analytic PCA construction for theoretical analysis of lighting variability in images of a lambertian object”, *IEEE Tran. PAMI*, 24(10):1322–1333, 2002.
- [40] R. Ramamoorthi and P. Hanrahan. “On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object”, *JOSA A*, 18(10):2448–2459, 2001.
- [41] K. Sato, K. Ikeuchi, and T. Kanade. “Model-Based Recognition of Specular Objects Using Sensor Models”, *Proc. IEEE Workshop on Directions in Automated CAD-Based Vision*, June, 1991, pp. 2–10. S.
- [42] S. Savarese, M. Chen and P. Perona. “Local Shape from Mirror Reflections,” *International Journal of Computer Vision*, 64(1), pp.31-67, 2005.
- [43] A. Sashua, *Geometry and Photometry in 3D Visual Recognition*, PhD thesis, MIT, 1992.
- [44] S. Shirdhonkar, D.W. Jacobs. “Non-Negative Lighting and Specular Object Recognition” *IEEE International Conference on Computer Vision*, pp: 1323-1330, 2005.

- [45] K.K Thornber and D.W. Jacobs. “Broadened, specular reflection and linear subspaces”, NEC TR#2001-033, 2001.
- [46] G. J. Ward. “Measuring and modeling anisotropic reflection,” *SIGGRAPH 92*, 26(2):265–272, 1992.