massachusetts institute of technology — artificial intelligence laboratory

Properties and Applications of Shape Recipes

Antonio Torralba and William T. Freeman

AI Memo 2002-019 December 2002

1 Abstract

In low-level vision, the representation of scene properties such as shape, albedo, etc., are very high dimensional as they have to describe complicated structures. The approach proposed here is to let the image itself bear as much of the representational burden as possible. In many situations, scene and image are closely related and it is possible to find a functional relationship between them. The scene information can be represented in reference to the image where the functional specifies how to translate the image into the associated scene. We illustrate the use of this representation for encoding shape information. We show how this representation has appealing properties such as locality and slow variation across space and scale. These properties provide a way of improving shape estimates coming from other sources of information like stereo. [1]

[1]This work was supported in part by a grant from NTT.

1. Introduction

The visual world offers a tremendous amount of data that the human visual sytem processes quickly and continuously. People can quickly infer low-level scene properties such as shape, motion, lightness and occlusion boundaries. We would like to design machines to do similar processing. The scene quantities can be very high dimensional, driven by the rich complexity of the visual world. Such high-dimensional quantities can be difficult to estimate quickly. We propose not to estimate directly the high-dimensional scenes of interest, but instead to estimate a low-dimensional *function* that transforms the incoming image data into the scene quantities of interest. For the examples considered in this paper, we study simple formulas to convert image data into shape data; we call these *shape recipes*.

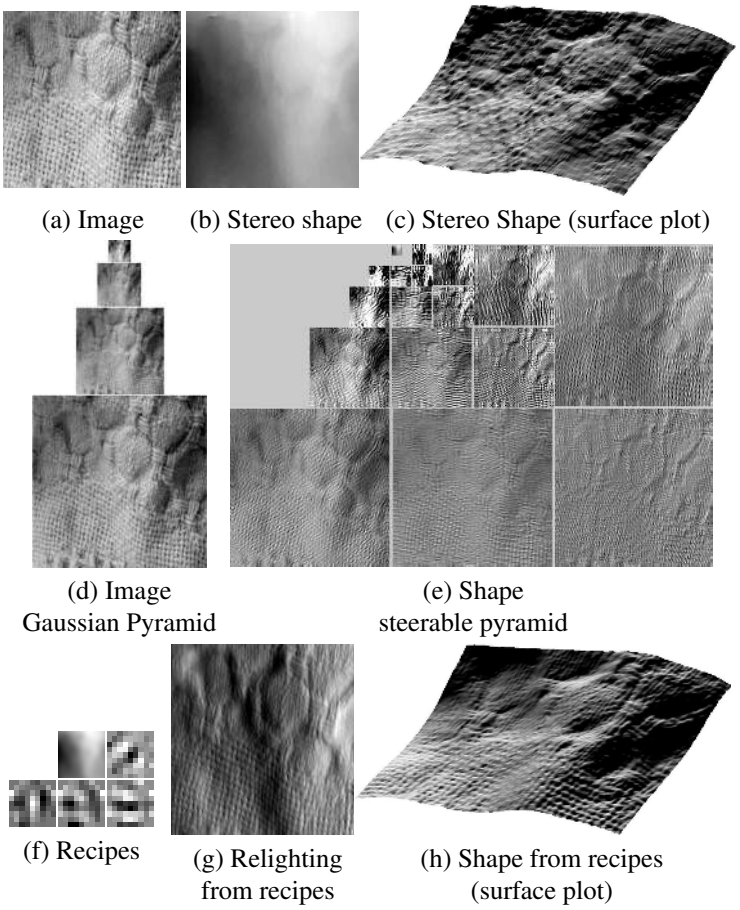
The shape recipes are not general purpose shape-from-shading algorithms: they are faster and weaker. They need to be suitable for online conversion of image data into shape, and therefore they need to be very fast to compute, and only require a local neighborhood of image data as input[1]. This approach to shape estimation separates the complexity of the shape to be estimated into two parts, (a) that described by the image data itself, and (b) that described by the shape recipes, which operate on the image to produce shape. Since we know the image and don't need to estimate it, this separation lets us focus our resources on the lower-dimensional problem of finding the shape recipes. The recipes implicitly encode local lighting and material properties that apply over a given region of a given image.

The low-dimensional shape recipe should be easier to estimate and store than a conventional 3-d shape estimate. It may change slowly over time, even though the image information itself may be changing quite quickly with time. The shape recipe may also change slowly over spatial scale, a point we discuss below and exploit for shape estimation.

Gilchrist [5] and Adelson [1] have proposed the notion of local "atmospheres" to describe simple formulas which transform from the image intensity domain to the perceived lightness domain. In the context of scene recipes, we would call those formulas "lightness recipes". Our learning-based approach relates to Leung and Malik's learning 3-d textons [6], and builds on Pentland's linear shading work [8].

In our first report about shape recipes [9], we described linear shape recipes, using a different multi-scale representation. In this paper, we present a range of shape recipes, and show applications. First, we show a simple example of a shape recipe. Then we discuss how to achieve the locality property of shape recipes. In Section 2, we build up a repertoire of shape recipes, from linear, to non-linear, to color. In Section 4, we discuss where these methods break down. We then show their application for improving shape estimates: we use an initial low-resolution shape estimate to learn a shape recipe, then use those learned recipes to improve the shape estimate (by applying the recipe at high-resolution). We show methods for applying these methods in practice:

[1] In order to meet the locality requirement, we will only compute estimates of *bandpassed* shape.



(a) Image (b) Stereo shape (c) Stereo Shape (surface plot)

(d) Image Gaussian Pyramid (e) Shape steerable pyramid

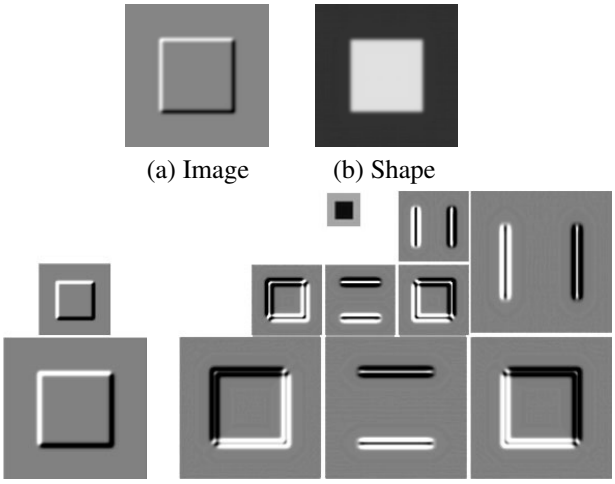
(f) Recipes (g) Relighting from recipes (h) Shape from recipes (surface plot)

Figure 1: Example of shape recipes and its use in representing and improving shape estimates. (c) shows the shape estimate by a stereo algorithm. (h) shows the shape improved by the shape recipes (f). The new shape allows relighting (g) the original image (a).

dealing with paint and occlusion.

2 Shape recipe algorithms

Figure 1 shows an example of a shape recipe. (a) is an image, one of a stereo pair. Using a downloadable stereo algorithm [14], we computed a range map, (b), from the stereo pair. From our experience, we expect that the stereo shape will be accurate at the low spatial frequencies, but inaccurate at the high frequencies. Using linear regression, we learn a convolution kernel that, when applied to the low-resolution image data, yields the low-resolution stereo bandpass filtered data (the coefficients of a steerable pyramid [11]). These kernels are the shape recipies, instructions for how to transform image data into bandpassed shape data that applies for this particular material under these particular lighting conditions. In general, shape recipes vary slowly across scale (Sect. 3.1). We can improve the shape reconstruction of the stereo by learning the shape recipes at a low-resolution, assuming the recipes are constant over scale, and applying the learned recipes at the high resolu-



(a) Image (b) Shape

(c) Image gaussian pyramid (d) Shape steerable pyramid

Figure 2: The dependency of the shape (b) with respect to the image (a) is non-local. When decomposing the shape in a bandpass pyramide (d), local shape variations are related to local image properties at each scale (c).

tion pyramid levels to reconstruct improved high resolution shape estimates, shown in Figure 1 (h). Note the improvement over the initial shape estimate from stereo.

2.1 Bandpass shapes and local recipes

The linear shading [8] approximation holds well for Lambertian surfaces under oblique illumination, or for extended light sources. These conditions are simple to analyze mathematically, yet provide insight to more general problems.

Under linear shading, the relationship between image, i, and shape is given by: $i \simeq 1 + l_1 p + l_2 q$ where, $p = dz/dx$ and $q = dz/dy$ and z is the range data. Without taking special care, the relationship between shape and image cannot be a local one. This is illustrated in Fig. 2. Fig. 2 (a) shows the image obtained by rendering the shape of Fig. 2 (b). When inferring the shape from the image, the shape values depend on non-local image values: the same local image intensities can give very different shape values, a function of image data far away. However, it is possible to have a local relationship between local image data and a *bandpass* version of the shape. In 1D, a bandpass filtered shape is obtained as $z_b = z * b$ where b is a compact wavelet of zero mean. Therefore, $z_b = (\int i\,dx) * b = i * (\int b\,dx) = i * r_b$, with $r_b = \int b\,dx$. Bandpass shape can be obtained by applying a local linear filter to the image, r_b. As shown in Fig. 2 (c-d), local image variations are in correspondence with local bandpass shape variations at different scales. Here, the 2D image is decomposed with a gaussian pyramid and the shape with a steerable pyramid [11].

Under some non-generic conditions, even under the approximation of linear shading, we can encounter shape changes that do not produce contrast changes in the image [4], leading to non-local shape recipes. We treat such special cases in Section 5.2.

2.2. Linear recipes

The simplest shape recipe is a linear filter. In 2D, for each scale k and each orientation n, the shape subband $z_{k,n}$ can be estimated by linear filtering:

$$\hat{z}_{k,n} = i_k * r_{k,n} \tag{1}$$

where $\hat{z}_{k,n}$ is the approximation of the shape subband $z_{k,n}$, i_k is the k band of the gaussian pyramid [2] of the image and $r_{k,n}$ is the local regression kernel. These regression kernels are shape recipes, local formulas that specify how to translate image information into each shape subband. We can represent the shape in reference to the image by encoding the recipes $r_{k,n}$ together with the image. This provides a compact way of encoding all the complexities of the shape variations by letting the image itself describe the various details of the shape. Assuming that shape and image are available, then the recipes are obtained by minimizing $\sum_x (z_{k,n} - \hat{z}_{k,n})^2$. We discuss later how to learn the recipes when only a coarse approximation of the shape is available (for instance, given by a stereo algorithm).

2.3. Non-linear recipes

Linear shape recipes can represent a surprisingly large range of illuminations and material properties. However, under more general conditions, for instance in presence of strong specularities, we need to introduce a non-linear relationship between local image data and the local shape subbands. More general shape recipes than the linear ones can be obtained by using Volterra filters [7]. For instance, a second order Volterra filter can be written as $\hat{z}_{k,n} = \sum_{m,l} r_{k,n}(m,l) i_k(x+m) i_k(x+l)$, but this approximation results in a large number of parameters that need to be learned. We propose to use a simpler non-linear model, applying first a point-wise non-linearity to the image and then a linear kernel:

$$\hat{z}_{k,n} = f_{k,n}(i_k; \mathbf{w}_{k,n}) * r_{k,n} = \left(\sum_m w_{k,n,m} f_m(i_k) \right) * r_{k,n} \tag{2}$$

where the non-linearity for each shape subband is parameterized by $\mathbf{w}_{k,n}$. The non-linear function is obtained by a linear combination of a basis of functions. We use here the functions $f_m(x) = x^m$ with $m = 0, ..., M$. Eq. (1) and (2) are special cases of the Volterra filter.

In order to learn the recipes from image and shape data, we can rewrite $\hat{z}_{k,n}$ as a bilinear equation: $\hat{z}_{k,n}(x) = \mathbf{w}_{k,n}^T \mathbf{F}_x \mathbf{r}_{k,n}$. We want the recipes (parameters of the non-linearity and linear kernel) that minimize $\sum_x (z_{k,n} - \mathbf{w}_{k,n}^T \mathbf{F}_x \mathbf{r}_{k,n})^2$. The minimization is done by iteratively solving linear equations for each of the two variables of the bilinear form, until the error does not decrease. Fig. 3 show the results of the iterations for real image data. The object has a strong specular component. The iteration begins with a linear recipe (c). Fig. 3 (g) shows the desired shape subband (after removing occlusion edges, see Sect. 5.2). With

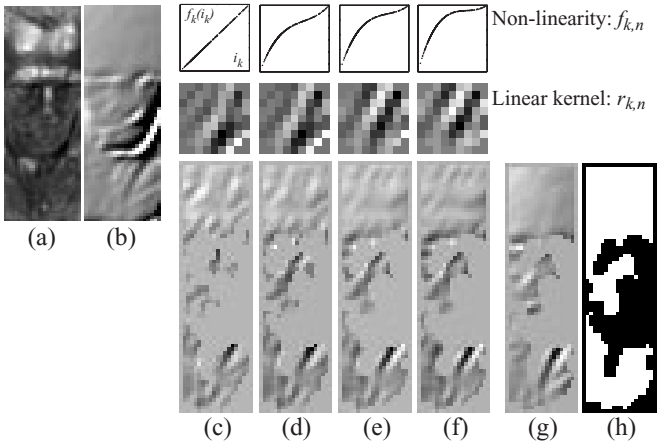


Figure 3: Learning the non-linear shape recipes for one subband. (a) and (b) are the image and shape at one scale. The object has some specular reflections. Figures (c) to (f) show the results at iterations 1, 2, 4 and 8. Figure (g) shows the desired output. Occlusion edges (black pixels in (h)) in the shape have been removed from the shape data (b) to give (g), as they cannot be inferred from the image in this case. The shape was obtained from a laser scan of the object.

subsequent iterations, the function $f_{k,n}$ becomes a compressive non-linearity in order to reduce the effect of the specular reflection (top of the images).

To test the range of illumination and material conditions that can be represented using these simple models, we use synthetic shapes rendered using a simplified Phong model in which image intensity corresponds to $i = R(\mathbf{l}^T\mathbf{n})^\alpha$ with $\mathbf{n} = (1, p, q)/\sqrt{1+p^2+q^2}$, $l = (\cos\theta, \sin\theta, 0)$, and $R(x) = x$ for $x > 0$ and $R(x) = 0$ for $x < 0$. Fig. 4 shows spheres rendered under this model. The angle θ controls the elevation of the light source, from lateral (in which a large part of the surface is not illuminated) to frontal illumination. The parameter α controls the specularity of the material. This simplified Phong model, in which only the specular component is included, enhances the effects of the non-linearities and provides a worst-case scenario for testing the recipes. Fig. 4 shows the region of the rendering space for which linear (dashed line) and non-linear recipes (solid line) can represent at least 75% of the energy of the shape contained in the pyramid subbands. Despite these restrictions, the range of renderable conditions is large.

2.4. Shape recipes for color images

The use of color provides additional information to remove non-linear effects in the relationship between image and shape. Including color, the Phong model of a material can be written as:

$$i(R,G,B) = \mathbf{L}[\mathbf{K}_a + \mathbf{K}_d(\mathbf{l}^T\mathbf{n}) + \mathbf{K}_s(\mathbf{h}^T\mathbf{n})^\alpha] \qquad (3)$$

where, $\mathbf{L}$ is a 3x3 diagonal matrix. The elements of the diagonal represent the chromaticity of the illumination. The vectors $\mathbf{K}_a$, $\mathbf{K}_d$ and $\mathbf{K}_s$ are the ambient, diffuse and specular chromatic reflectance of the surface.

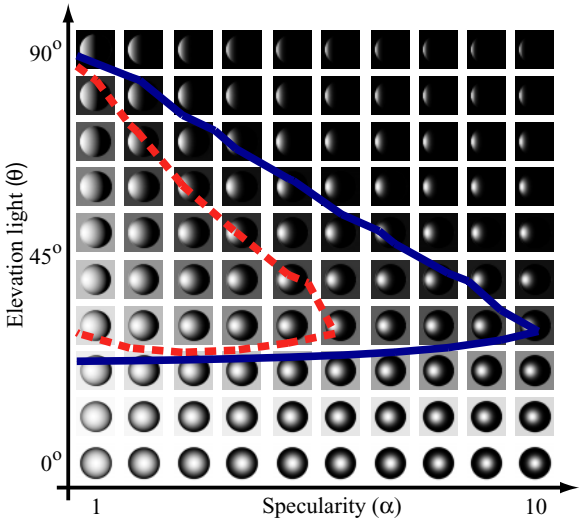


Figure 4: The figure illustrates the range of illumination directions and specularities that allow representing more than 75% of the energy of the shape using the linear (dashed line) and non-linear (solid line) recipes. The meaning of each location in the parameter space is illustrated by an image of a hemisphere rendered under those material and lighting conditions.

If the specular and diffuse components have different chromaticities, then by properly combining the three color channels of the image, we can cancel the non-linear effect of the specular component. Therefore, we can build shape recipes as:

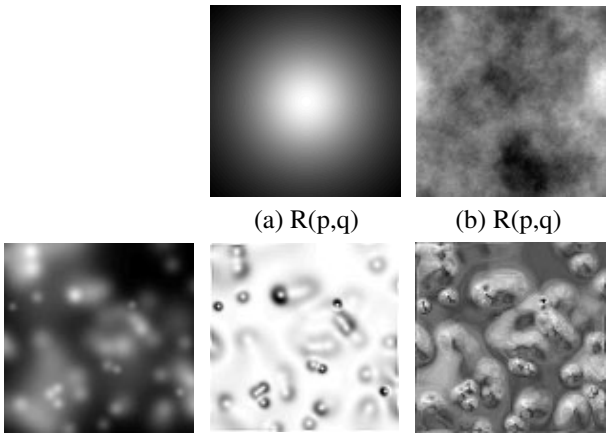
$$z_{k,n} = (a_{k,n}R_k + b_{k,n}G_k + c_{k,n}B_k) * r_{k,n} \qquad (4)$$

when the recipe parameters vector $(a,b,c)_{k,n}$ is orthogonal to the vector $\mathbf{L}\,\mathbf{K}_s$ the relation between shape and the combination of image color channels becomes linear. In such a situation, the color recipes have the same performances as the non-linear recipes for representing the shape.

If the diffuse component is zero or if there is no difference in chromaticity between diffuse and specular components, then we need to use the non-linear recipes proposed in the precedent section. For learning the color shape recipes we can use the same technique as with the non-linear recipes by writing again the relation between shape subband and image using the bilinear equation.

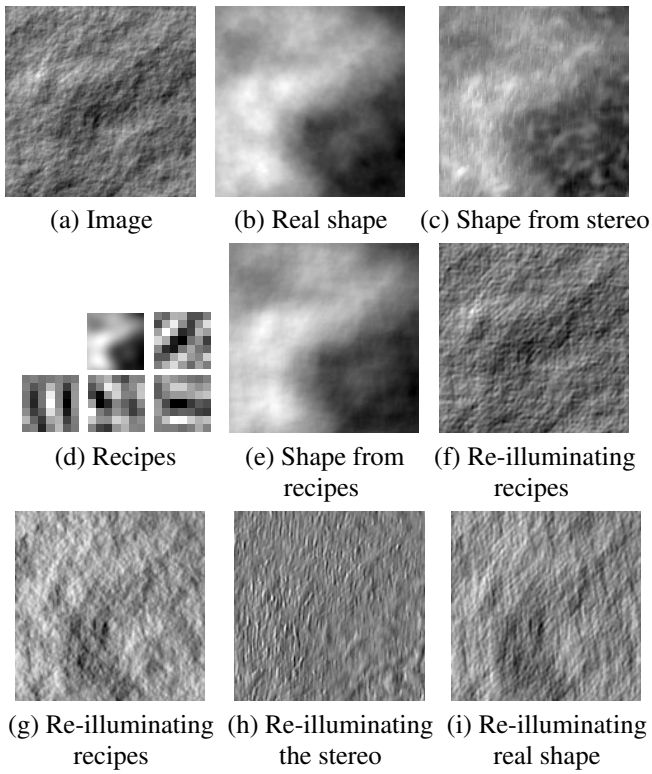
3. Regularities of shape recipes

Until now we have assumed that both shape and image were available and we have studied the material and illumination conditions that can be represented using the shape recipes proposed here. But in general, shape will not be available. We argue here that shape recipes have very appealing properties that make them interesting for estimation. We illustrate this in the next section in which we show how the use of shape recipes allows improving poor shape estimates that come from sources of information as stereo.



(a) R(p,q) (b) R(p,q) (c) Shape (d) Image (e) Image

Figure 5: Examples of illumination maps that cannot be represented by the recipes proposed here. The images (d) and (e) are produced by rendering the shape (c) using the $R(p, q)$ maps shown in (a) and (b). (a) corresponds to frontal illumination of a lambertian material. (b) corresponds to a complex illumination for a mirror like material. The recipes can account for 25% of the energy of the bandpass shape with frontal illumination. In (b), despite the complexity, the recipes still can represent 60% of the energy of the subbands of the shape.



(a) Image (b) Real shape (c) Shape from stereo (d) Recipes (e) Shape from recipes (f) Re-illuminating recipes (g) Re-illuminating recipes (h) Re-illuminating the stereo (i) Re-illuminating real shape

Figure 6: Experiments with a synthetic shape. Stereo has an error of 9% of the energy of the real shape. Using the recipes the error goes down to 1.5%. When rendering the stereo estimate, it produces a very poor image (h) compared to the recipes (g).

3.1. Scale invariance of shape recipes

In the 1D linear lambertian case, $i = dz/dx$. At the scale k, where i_k and z_k are the image and shape subsampled by a factor of 2^k, we have $i_k = 2^{-k} d(z_k)/dx$ (assuming that there is no alliasing when subsampling the signals). Then, the output of a shape subband is obtained by filtering the subsampled shape z_k with a wavelet b: $z_{k,b} = z_k * b$, where b is a compact kernel of zero mean. Therefore, $z_{k,b} = (\int 2^k i_k \, dx) * b = i_k * (2^k \int b \, dx) = i_k * r_k$ where the recipes are $r_k = 2^k \int b \, dx$. The recipes are also compact with a support equal or smaller that the one of the kernel b. They are constant across scale with only an amplitude factor. Therefore, in this case, we only need to store the shape recipes computed at one level of the pyramid and then propagate the kernel, with the amplitude factor, to recover the shape at the other resolutions. This is interesting because, in general, stereo shape estimates provide a good approximation to the shape at low spatial frequencies but it is noisy at the high spatial frequencies [3].

We expect that under more general illumination and material conditions than the lambertian or Phong model, the shape recipes will vary slowly across scales.

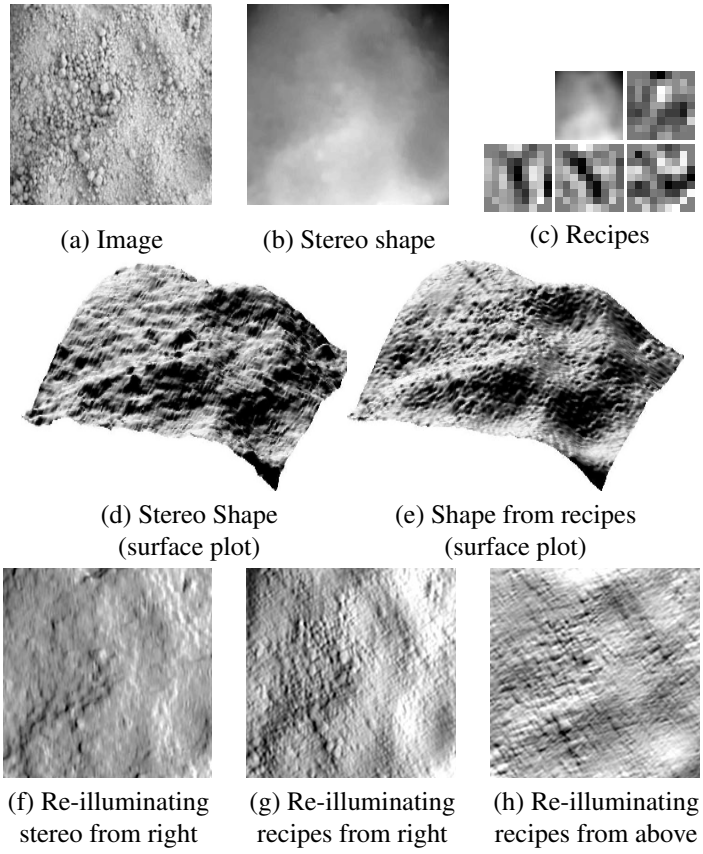
3.2. Constant across space

Even when a shape may have complex variations like the cloth example in fig. 1 (a), the relationship between image and shape remains constant across large image regions. This is interesting because it allows to learn the recipes in a region of the image were shape is known accurately and then use them to recover shape information in regions were shape is poorly known. Within each subband, we can write the shape z_k as a mixture of recipes: $\hat{z}_k = \sum_{n=1}^{N} w_n \, f_{k,n}(i_k)$. The weights w_n, which will be a function of location, will specify which recipe has to be used within each region and, therefore, will provide a segmentation of the image [9].

4. Limitations of shape recipes

There are situations in which shape recipes breakdown. In the case of frontal illumination (fig. 5.a), the image is given by $i \simeq 1/\sqrt{1 + p^2 + q^2}$. Both concavities and convexities produce the same image. The recipes proposed here cannot solve this ambiguity. In the example shown in fig. 5 (d), the recipes can only represent 25% of the energy of the pyramid subbands of the shape (c). This is also illustrated in fig. 4 in which is shown that both linear and non-linear fail to represent the shape in reference to the image for materials under frontal illumination.

The lambertian and Phong models are particular cases of illumination and material models. In general, the image may be a complex function of the shape derivatives p, and q. For instance, fig. 5 (b) shows a complex reflectance map and the image (e) produced by rendering the shape (c) under that map. The resulting image looks like a mirror like material under a complex pattern of illumination.



(a) Image (b) Stereo shape (c) Recipes

(d) Stereo Shape (surface plot) (e) Shape from recipes (surface plot)

(f) Re-illuminating stereo from right (g) Re-illuminating recipes from right (h) Re-illuminating recipes from above

Figure 7: Example of real image and the improvement in the shape estimate using the shape recipes. (a) shows one image of the stereo pair and (b) shows the shape obtained by the stereo algorithm. (c) shows the recipes leant at the lowest resolution of the pyramids.

5. Application: improving shape estimates

In this section we discuss the application of shape recipes to improve shape estimates. We focus on stereo estimates but the same technique could be applied to other sources like the visual hull.

5.1. Experiments with stereo

We generated a synthetic shape as a fractal surface (fig. 6 (b)). From the shape we generated a stereo pair of images rendered using a lambertian function. Fig. 6 (a) shows the left image of the stereo pair. First we run the stereo algorithm by Zitnick and Kanade [14] to have an estimate of the shape (c). The shape (c) has the overall correct structure at low spatial frequencies but appears noisy at the high spatial frequencies (the error was of 9% of the energy of the shape). We learnt the recipes by decomposing the image and the shape using a gaussian and a steerable pyramid respectively with 4 scales. We learnt the recipes at the lowest resolution (we used linear recipes, with a kernel size of 9x9 pixels). By applying the scale invariance of the recipes, we apply the same recipes at the levels of the pyramid with higher resolution for which the shape information was noisy. A new shape is obtained by collapsing the new shape steerable pyramid. The obtained shape is shown in fig. 6 (e). The new shape reduced significantly the error in the shape estimation to 1.5%.

Fig. 6 (f) shows the image produced by rendering the estimated shape (e) using the same lighting conditions than in (a). The resulting image looks very close to the original one. Figs. 6 (g), h and i show the shape illuminated from a direction different than the one used for the training. The stereo shape provides a very poor rendering.

Figures 1, 7 and 10 show examples of real surfaces and the shapes obtained by applying the shape recipes in order to improve the stereo estimates.

5.2. Occlusions

When learning the relationship between image and shape, special attention has to be paid to occlusions and paint boundaries. Not all image and shape variations are directly related. For instance, an image can be decomposed into paint and shade variations [12]. Shade variations are directly produced by the shape, however, the paint will be related to changes in the reflectance function of the material and not directly related to the shape. In the same vein, the shape can be decomposed into undulations that will generate image variations (due to lighting) and occlusions with may or not render into the image (consider the extreme case of a random dot stereogram in which the shape boundary does not produce an edge in the image). Shape recipes, as presented here, learn the relationship between undulations in the shape and the shade image.

Although edges in the image and the location of edges in the shape are not independent (occlusion correlate with changes of material and therefore with changes in paint), some occlusion boundaries may not render in the image. For instance, some of the occlusions in fig 10 (b) or in fig 11 do not produce sharp discontinuities in the image. In such cases, shape recipes will tend to provide smooth shape variations were an occlusion edge is expected. This difficulty can be circumvented by encoding occlusion edges by a separate mechanism. In this section, we propose a statistical framework for combining the shape given by the recipes with the occlusion edges detected by a stereo algorithm [14].

Within a shape subband, considering that stereo s_k and image intensity i_k are conditionally independent with respect to the real shape z_k, we can write:

$$p(s_k, i_k|z_k) \simeq p(s_k|z_k)p(i_k|z_k) \tag{5}$$

s_k, z_k and i_k refer to the outputs of the subband k.

The distributions $p(s_k|z_k)$ and $p(i_k|z_k)$ will depend on the noise models for the depth from stereo and for the shape recipes respectively. Assuming that the stereo algorithm is able to correctly detect most of the occlusion edges, we model the error as having a gaussian distribution:

$$p(s_k|z_k) = p_s(z_k - s_k) = \frac{e^{-|z_k - s_k|^2/\sigma_s^2}}{(2\pi)^{1/2}\sigma_s} \tag{6}$$

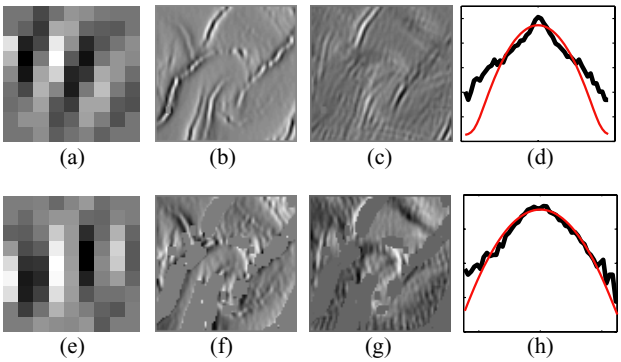


Figure 8: Distribution of the error before (d) and after (h) removing occlusion edges for one of the subbands of the shape shown in fig 10. The distributions are plot on a semi-logarithmic axis. Superimposed with the error distribution is shown the best gaussian fit.

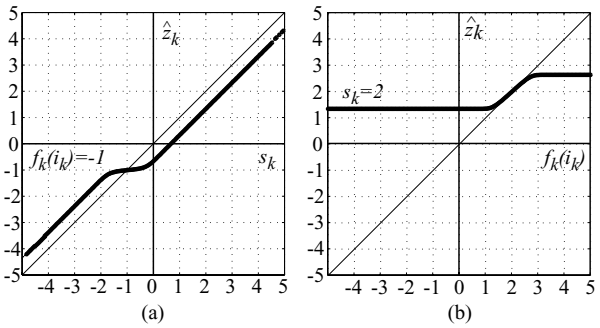


Figure 9: Least square estimate for the shape subband z_k given both stereo s_k and shape recipes $f_k(i_k)$. The graph (a) shows the final shape estimate as a function of the shape estimated by stereo. When the s_k is close to the value given by the recipes $f_k(i_k)$, then, the final shape z_k does not vary with variations in the stereo estimation. The graph (b) shows z_k as a function of $f_k(i_k)$ for a fix value of s_k. z_k is equal to $f_k(i_k)$ only when both recipes and stereo give similar estimations.

However, this model of noise is not appropriate for the shape estimated by the shape recipes. Fig. 8 (d) shows the distribution of the difference between the shape subband estimated by the recipes (c) and the shape estimated from stereo (b). The shape recipe fails to represent some of the edges. As the subband shown correspond to low spatial resolution we assume that $z_k \simeq s_k$. The distribution of the error $s_k - f_k(i_k)$ is better fit by a Laplacian distribution:

$$p(i_k|z_k) = p(z_k - f_k(i_k)) = \frac{e^{-|z_k - f_k(i_k)|^\lambda / \sigma_i^\lambda}}{2\sigma_i/\lambda\Gamma(1/\lambda)} \qquad (7)$$

with $\lambda \simeq 1$. Fig 8.b shows the shape subband given by stereo and 8.c the subband estimated by the linear recipe (8.a).

The least square estimate for the shape subband z_k given both stereo and image data, is: $\hat{z}_k = \int z_k p(z_k|s_k, i_k)\, dz_k$. When $\lambda = 2$, then the LSE estimation is a weighted linear combination of the shape from stereo and shape recipes.

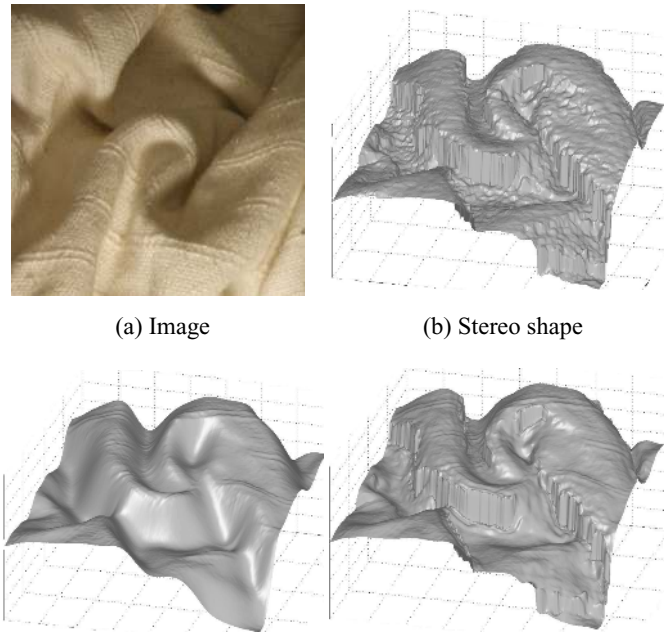


Figure 10: Direct application of shape recipe across occlusion boundary misses the shape discontinuity. Stereo algorithm catches that discontinuity, but misses other shape details. Probabilistic combination of the two shape estimates captures the desirable details of both.

However, with $\lambda \simeq 1$ this problem is similar to the one of image denoising [10] providing a non-linear combination of stereo and shape recipes (fig. 9).

Note that as suggested by the laplacian model of eq. (7) the recipes should be learn by minimizing the sum of absolute differences instead of a least square error. We use a different procedure, we look for the recipes that minimize the weighted least squares:

$$\sum_x u_x \left(z_{k,n} - \mathbf{w}_{k,n}^T \mathbf{F}_x \mathbf{r}_{k,n}\right)^2 \qquad (8)$$

where u_x are the weights. We set $u_x = 0$ in the image regions in which the error is large as it might contain an occlusion edge and we set $u_x = 1$ in the rest of the image. Fig. 8 (f) shows the stereo shape s_k with the regions with large difference with $f_k(i_k)$ suppressed. Fig 8 (g) shows the shape estimated by the recipes $f_k(i_k)$. Now the recipe is trying to follow the variations responsible for the shade image instead of being affected by the occlusions. The new recipe (fig. 8 (e)) differs from fig. 8 (a). In fig. 8 (h) we show that once the occlusion boundaries are removed, the error follows more closely a gaussian distribution.

Figure 10 shows the final result when applying the same procedure to all the subbands. The surface plot shows the shapes estimated using only stereo (b), and using the shape recipes (c) and both (d). The stereo captures the occlusion boundaries but fails in estimating the shape of the details in the image. On the other hand, shape recipes capture the shape variations correlated with image details, but the oc-

clusion boundaries are smoothed. The combination (d) has the details captured by the recipes and the occlusion edges captured by the stereo.

6. Conclusion

We introduced shape recipes, a method to represent scene properties through local, fast operations on image data. The shape and its rendered image can be very complex. We let the image itself describe that complexity and focus on estimating the transformation functions that convert image values into bandpassed shape. We call these functions shape recipes.

Under conditions of low specularity, linear regression fits well We can learn a low-order polynomial point-wise non-linearity to apply to the image intensities for non-linear shape recipies.

We showed, for the two approaches, the range of applicability over a set of material and illumination conditions. For backlit shapes or very specular materials, these models are not appropriate. We generalized to color images.

The application we explored was to use shape recipes to improve initial shape estimates gathered by other means. The idea is to learn the shape recipes at a low-resolution scale, where the initial shape estimates are accurate. We then extrapolated the shape recipes across scale to form high-resolution shape estimates from the high-resolution image data and the shape recipes. In this way, we improved the shape estimates derived from stereo and laser rangefinder point range estimates.

Shape recipes may be useful for general shape estimation algorithms: estimating the shape recipes may be easier than estimating shape. The general approach may apply to the estimation of other scene properties in addition to shape. Then low-level vision becomes an image labelling task–assigning the appropriate labels or transformation functions (recipes) to apply to the different parts of the incoming visual stream.

References

[1] E. H. Adelson. Lightness perception and lightness illusions. In M. Gazzaniga, editor, *The New Cognitive Neurosciences*, pages 339–351. MIT Press, 2000.

[2] P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Trans. Comm.*, 31(4):532–540, 1983.

[3] J. E. Cryer, P. S. Tsai and M. Shah. Integration of shape from shading and stereo. *Pattern Recognition*, 28(7):1033–1043, 1995.

[4] W. T. Freeman. The generic viewpoint assumption in a framework for visual perception. *Nature*, 368(6471):542–545, April 7 1994.

[5] A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou. An anchoring theory of lightness. *Psychological Review*, 106(4):795–834, 1999.

[6] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Intl. J. Comp. Vis.*, 43(1):29–44, 2001.

[7] V. Matthews and G. Sicuranza. *Polynomial signal processing*. John Wiley and Sons, 2000.

[8] A. P. Pentland. Linear shape from shading. *Intl. J. Comp. Vis.*, 1(4):153–162, 1990.

[9] W.T. Freeman and A. Torralba Shape Recipes: Scene Representations that Refer to the Image. *Advances in Neural Information Processing Systems*, 2002. To appear.

[10] E. P. Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf. on Sig., Sys. and Computers*, Pacific Grove, CA, 1997.

[11] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *2nd Annual Intl. Conf. on Image Processing*, Washington, DC, 1995. IEEE.

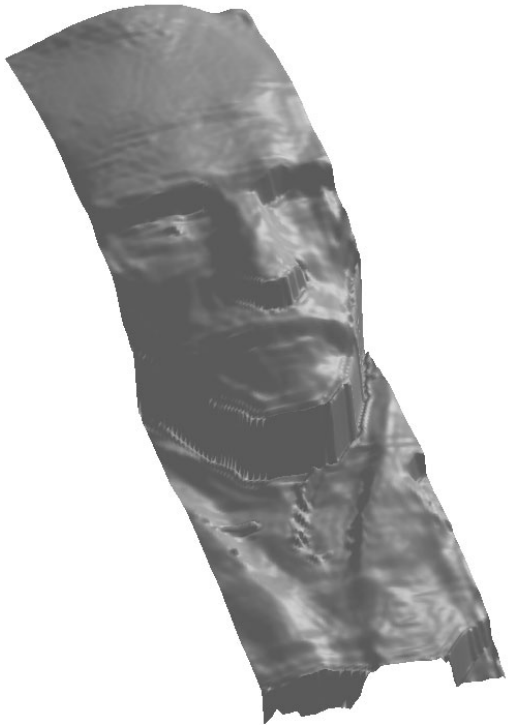
[12] M. F. Tappen, W. T. Freeman and E. H. Adelson. Recovering Intrinsic Images from a Single Image. *Advances in Neural Information Processing Systems*, 2002. To appear.

[13] Y. Weiss. *Bayesian motion estimation and segmentation*. PhD thesis, M.I.T., 1998.

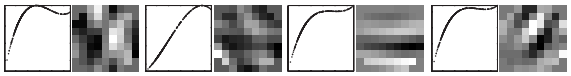
[14] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Pattern Analysis and Machine Intelligence*, 22(7), July 2000. http://www-2.cs.cmu.edu/ clz/stereo.html.

(a) image

(b) Shape data



(c) Shape from recipes



(d) Non-linear shape recipes
(point non-linearity and
convolution kernel for each orientation)

Figure 11: The shape data (b) has been obtained from a laser scan of the object shown in (a). We use the shape recipes to add details to the shape not captured in the original data. We use the non-linear recipes (d) due to the strong specularities.