SUPPLEMENTARY MATERIALS

APPENDIX 1: Experimental Apparatus.

The main body of the set-up is a wood chamber of dimensions 1400 mm x 1050 mm x 1500 mm (Figure 1 of the main article or Figure A1), placed on a wheeled steel platform (not in figure). The roof of the chamber slopes from a height of 1050 mm on the viewer side to a low point of 610 mm at the far end of the chamber. The apparatus contains: two sets of three fluorescent lights (red, green, and blue), a data projector, a frame, a pair of sliding doors, a set of movable panels, a side door, and a rotating system. The latter allows rapid changes in the scene during the experiment and consists of three steel boards placed in vertical position forming a equilateral triangle on a circular wood platform. Each board’s dimensions are 810 mm x 610 mm; each is coated with white matt paint. The observer sits in a rectangular inlet (400 mm x 810 mm x 650 mm) of the chamber and looks inside the chamber through a viewing hole (400 mm x 250 mm). The stimuli are displayed in front of the observer on one of the vertical boards of the rotating system at a distance of 1200 mm. Inside the chamber, on each side of the inlet, a triplet of fluorescent lamps (Osram T5-FQ39W-220 red, -230 green, and -240 blue) is placed behind a diffusing filter (416 Lee uniform three-quarter white diffusion filter film). Each pair of same-color lamps is digitally controlled by a double dimmable ballast (Osram Quicktronick; FQ 2x39/230-240 DIM) and connected through a custom-made interface to a terminal block via a 68-pin SCSI female connector (DIN-68S). This in turn is connected to an 8-channel digital/analog output PCI multi-function board (DAQ 2502). Two adjustable panels, located 250 mm to either side of the viewing hole inside the chamber, restrict the direction of the light. A frame, placed at 800mm from the viewing hole, blocks interaction between the two sources of light. All surfaces visible to the observer inside the chamber are covered with ultra-white matt paper, while all other interior surfaces are covered with deep black matt velvet to minimise mutual reflections. A computer-driven LCOS data projector (Canon Xeed SX6) is placed inside the chamber on a platform above the observer’s seat and hidden from the observer’s view. The projector and side-lamps are controlled independently from a PC running Windows XP.
The geometry of the chamber, i.e. the slope of the ceiling, the light directing panels and the shape of the frame, as well as the paint of the interior surface, ensure that less than 1% of the light produced by the side lamps hit the main screen. The maximum luminance contributed from the side lamps light occurred in the centre of the screen, and was measured at 6 cd/m$^2$, which is less than half of the luminance produced by the data projector there when it is on and set to minimum output (black; 14 cd/m$^2$), and the average luminance from the side lamps in the centre of the screen was 2.8 cd/m$^2$. The average difference in measured luminance for 100 projected colors between side-lamps-on and side-lamps-off was $\sim$1 $\Delta E_{uv}$ units. Therefore, projected images and side-lamp illumination are practically independent. The side-lamp outputs were computed using DIALux software (DIAL GmbH; www.dialux.com).

Small magnetic buttons (Electronics-bonded Neodymium-Iron-Boron discs: 2 mm diameter and 1 mm thickness), glued to the experimental objects (3 to 10 per object), held them in place on the rotating backboard system.
APPENDIX 2: Image generation

Figure A2 shows the general framework of the image generation described in the paper.

Figure A2 - Flowchart of the general framework of the image generation.
A2.1 Camera characterization and conversion model to XYZ coordinates of the photographs

To characterize a Nikon D70 SLR camera with a 18-70mm kit lens, we employed a Verivide Colour Assessment cabinet, containing independent, stable illumination sources, and a SpectraScan PR 650 spectroradiometer. The Verivide cabinet simulates the actual spectrum of three different light sources, among which is the CIE standard D65 illuminant that we used for the characterization. The general idea of the characterization process is to measure the CIE XYZ tristimulus values and the corresponding camera RGB values of a group of paper patches placed in a predetermined region of the Verivide cabinet and construct a characterisation model which fits the relationship between the two sets of values. Specifically, a third order polynomial regression was used to obtain the coefficients of the linear model mapping one set into the other.

Before the measurements, we optimised the camera’s settings by optimising the image histograms of a Macbeth Digital ColorCheck SG chart placed within the viewing cabinet. We measured two sets of color patches: a training set (87 Munsell paper patches, 10 Macbeth Digital ColorChecker SG chips, 7 coloured art paper pieces, and 1 Ocean Optics WS-1 Diffuse Reflectance Standard, which is >98% reflective from 250-1500nm) and a testing set (37 Munsell paper patches). Table A1 lists the tristimulus values of the 7 art paper pieces and Table A2 lists the patches used in each set. The patches were chosen so as to approximately uniformly sample the color space in saturation, hue and luminance.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>2B</td>
<td>240.22</td>
<td>283.37</td>
<td>363.04</td>
</tr>
<tr>
<td>3B</td>
<td>107.86</td>
<td>140.83</td>
<td>290.13</td>
</tr>
<tr>
<td>1G</td>
<td>420.57</td>
<td>231.47</td>
<td>221.68</td>
</tr>
<tr>
<td>1O</td>
<td>197.62</td>
<td>382.66</td>
<td>70.917</td>
</tr>
<tr>
<td>2O</td>
<td>439.09</td>
<td>260.74</td>
<td>40.517</td>
</tr>
<tr>
<td>3V</td>
<td>508.1</td>
<td>342.91</td>
<td>59.992</td>
</tr>
<tr>
<td>1Y</td>
<td>402.81</td>
<td>567.22</td>
<td>90.781</td>
</tr>
</tbody>
</table>

Table A1 - CIE XYZ values of the 7 colored art paper pieces used to characterize the Nikon D70 SLR camera (D65)
Training Set: 105 samples

| Munsell 5YR: 3/2–5/2 – 5/6–7/4–7/12 – 8/5–8/5–8.5/12 – 9/8 |
| Munsell 5Y: 3/2–5/2 – 5/6–7/4–7/8–7/12 – 8.5/2–8.5/6–8.5/10 |

COLOUR CHECKER: 2A - 5B - 6B - 7C - 2D - 5E - 4H - 6L - 5M - 6M

NON-STANDARD PAPERS: 2B – 3B – 1G – 1O – 2O – 3V – 1Y

REFLECTANCE STANDARD

Testing set: 37 samples

| Munsell 10R: 2.5/2–5/8 – 8/6 |
| Munsell 10YR: 3/2–6/8–8/10 |
| Munsell 10Y: 4/2–7/4–8.5/10 |
| Munsell 10GY: 5/6–8/6–9/4 |
| Munsell 10BG: 3/4–6/4–9/2 |
| Munsell 10G: 4/8–6/4–9/2 |
| Munsell 10BG: 3/4-6/6–8/2 |
| Munsell 10B: 3/6–6/10–9/2 |
| Munsell 10PB: 2.5/6–4/6–7/8 |
| Munsell 10P: 3/4–5/10–7/6 |
| Munsell 10RP: 5/8–7/4–9/2 |
| Munsell N: 2.5–3.5–4.5–5.5–6.5–7.5–8.5 |

Table A2 - Training and testing data sets description for the camera characterization

The main criterion of the setup is to ensure that the camera and the spectroradiometer measure the same signal. Thus, we designed a removable stand allowing each color patch to be positioned within the Verivide cabinet exactly in the same location. The surface reflectance of each color patch was first measured by the spectroradiometer and then the camera recorded its RGB values from exactly the same location. Next, the two sets of recordings were used in a third order polynomial regression model to obtain the coefficient of the best-fitting polynomial for the RGB to XYZ transformation. The final formula of the model is expressed in Equation A1; R, G, and B are the
RGB camera values from which we have subtracted the camera black point obtained by recording an image with the lens covered.

Finally, the performance of the model was assessed by computing the color difference between the measured XYZ values and the model-predicted XYZ values of the training and the testing set. For both sets, the accuracy of the prediction of the model is comparable to 3 JND (testing set on average 2.5 ΔEuv units difference), revealing good performance. This model was then used to convert each point of the distribution of the photographed objects into XYZ values.

\[
\begin{bmatrix}
    X \\
    Y \\
    Z
\end{bmatrix}
=\begin{bmatrix}
    \alpha_1 \\
    \alpha_2 \\
    \vdots \\
    \alpha_n \\
    \beta_1 \\
    \beta_2 \\
    \vdots \\
    \beta_n \\
    \gamma_1 \\
    \gamma_2 \\
    \vdots \\
    \gamma_n
\end{bmatrix} \cdot \begin{bmatrix}
    R \\
    G \\
    B
\end{bmatrix}
+ \begin{bmatrix}
    \alpha_o \\
    \beta_o \\
    \gamma_o
\end{bmatrix}
\]

Equation A1

**A2.2 Texture extraction**

Using the Nikon D70 SLR camera calibrated as described above, we photographed natural familiar objects in the Verivide cabinet under D65 illumination. To minimize specular highlights, the objects were spray-painted with full-matt, totally transparent paint (MARABU Matt Varnish) prior to being photographed. To eliminate luminance shading due to 3D shape, after taking the initial photograph we spray-painted each object matt gray, and photographed it again in exactly the same location and under exactly the same conditions (see right-hand column in Figure 2 of the main article). The luminance channel of the image of this gray-painted object (Lg), containing shading information but not intrinsic color surface variations, was then subtracted from the luminance channel of the original image (Lo), so that the resulting image (Lv = Lo − Lg) contained only the luminance variations due to the veridical texture of the object. Residual highlights (defined as pixels with CIE Lab luminance values greater than or equal to 99) were equalized to the mean of the 8 nearest-neighbour values with luminances below 99. (Note that residual highlights were found on the apple only, for fewer than 0.5% of the pixels.) Figure 4 (main article) shows the result of the
texture extraction. Figure 2 (left-hand column of the main article) illustrates the chromaticity distributions for the three objects after highlight and shading correction in EMG cone contrast space.

A2.3 Geometric calibration and morphing technique

After texture extraction and manipulation, the contour of the projected image must exactly overlap the outline of the real object inside the box. Therefore we developed a geometrical calibration method to align the projected image with the scene from the observer’s viewpoint (Figure A2).

From the observer’s viewpoint, a Canon PowerShot S45 digital camera takes a picture of an object on the main display, in the location in which it will be placed during the experiment. Then the object’s contour is segmented from the background of the picture; we define the segmented part as the stencil. A lookup table is then created to store, for each pixel of the stencil, the corresponding pixel projected by the data projector, creating a 1-to-1 correspondence between object surface position and projector pixel position.

At this point we have two images: the first contains information about the original object (output of the block Texture Manipulation in Figure A2 – containing the texture and shape of the original object) while the second contains information about the visible surface of the test object from the observer’s viewpoint (i.e. the stencil, output of the block Stencil Design in Figure A2). Because the final goal is to obtain an image possessing the original object texture within the contour of the stencil, we then “morphed” one image into the other. This process alters the spatial distribution of the points of the original image without corrupting the original texture structure (example in Figure A3). A piecewise linear spatial transformation was employed as follows (Goshtasby, 1986). Initially a number of control points were selected on the original image (arrow-ends on the left of Figure A3) on the contour or another salient locations. Then their corresponding locations on the stencil were selected (arrowheads on the right of Figure A3). The transformation was performed using the control points as anchors for the spatial morphing triangulation (modification of Goshtasby, 1986). We implemented the piecewise linear spatial transformation and
selected for each object the number and location of the control pairs based on the following constraints: (1) the fidelity in reproducing the same texture, (2) the overlap of the 2D texture features with the object’s 3D structural features, and (3) the percentage of points of the morphed texture overlapping the stencil (less than 1% outside the stencil contour). Specifically, point (2) requires that the texture’s spatial characteristics align with the 3D object surface features that modulate it; for example, the real banana’s texture possesses distinctive elongated features where its skin bends (indicated by the red arrow in Figure A4-A). Therefore, these features in the projected texture are constrained to fall on the same location of the 3D replica (black arrow in Figure A4-B). The final number of pairs varied between 47 for the banana to 6 for the apple. Note that, the transformation being linear, the chromaticity of each point is not altered, only the position of a point relative to its neighbors. After morphing, the image is converted to the projector RGB color space using the model described in Section A2.4. Examples of the resultant images are given in the main article and also in Figure A6-B.

![Figure A3](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932812/)  
**Figure A3** – Example of the morphing between the banana original profile (on the left) and the banana stencil (on the right); e.g. the point A of the original profile is paired with the point A’ inside the object stencil.

![Figure A4-A](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932812/)  
![Figure A4-B](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932812/)  
**Figure A4 – A)** Photograph of the original banana. The red arrow points to a location of the banana chromatic surface texture with a specific direction corresponding to a 3D shape cue.  **B)** Photograph of the banana replica. The black arrow points to the corresponding location in Panel A.

**A2.4 Spatial and spectral calibration**

The projector is driven by a 256MB ATI® Radeon™ HD 3650 graphics card (spatial resolution of 1280X1024 at a refresh rate of 70Hz). Spatial uniformity was tested using a uniform
grey projected image. Recorded chromaticity values were almost constant (0.01% max variability) across the full extent of the image (600 mm by 800 mm; 35 points measured on an even grid of roughly 120 cm$^2$ per unit), while luminance varied up to 7%. Consequently, to ensure colorimetric accuracy, we generated a distinct calibration look-up table for each experimental location and object, as follows. The spectra of the red, green and blue primaries were measured for input levels from 0 to 255 in steps of 5 using a Photo Research PR-650 spectroradiometer and used to compute the CIE 1931 xyY coordinates of each value. After subtraction of the black bias (tristimulus values for RGB=[0 0 0]), the data were used to obtain the transformation matrix of the forward device-response function for each location and object. The inverse of this function maps each device-independent color to the device-dependent RGB color. Model accuracy was tested by comparing the predicted and measured xyY coordinates of a set of 100 randomly-selected RGB values; the mean CIE Luv color difference was 3±0.4 (averaged over all points and objects). Note that we have as many matrices as object-location combinations; for this experiment, 4 distinct objects were used in one central location, and hence we obtained 4 matrices. The chromaticities of the three same-color lamp pairs were measured using the Photo Research PR650 spectroradiometer at four locations, one for each reflecting surface (two sidewalls and two lateral sectors) and found not to vary significantly between these locations. Hence, only one set of look-up tables was used to correct for nonlinearities in the relationship between voltage input and lamp output. In addition, the additivity of the three lamps was verified to hold over the range of intensities used in this experiment.

**A2.5 Gamut mapping technique**

In performing the gamut mapping of the image, we applied a blend of two different techniques: gamut clipping and linear compression. This non-linear compromise between gamut mapping methods is close to the “soft clipping” technique (for review on gamut mapping techniques see Baranczuk, Giesen, Simon, & Zolliker, 2010; Dugay, Farup, & Hardeberg, 2008). This choice was performed considering that the image’s and the projector’s gamuts are rather similar in terms of hue and chroma and differ mostly in the lightness ranges. Furthermore, less than 0.1% of the colours of the points of the original image resulted in out-of-gamut values after lightness scaling, and even
after extreme alteration of the texture’s original hue for certain experiments, no more than the 20% of the colours fell out of the projector’s gamut.

![Figure A5: Gamut of the projector (colored tessellation) and distribution of the colors of the banana (red points) before (A) and after (B) gamut mapping for the illuminant D65.](image)

In general the goal of soft clipping is to compress the highest chroma value (or lowest luminance values) of *all* points and leave unchanged the lowest chroma values (or highest luminance values). In the presented approach the points compressed are only the ones close to the gamut boundary (using a 5th order regression model). Specifically, this method selects the points with \( \Delta E \) difference from the boundary equal or lower than 1 and compresses the lowest luminance values to the lowest boundary of the gamut and the highest luminance values to the highest boundary of the gamut, leaving hue and chroma unchanged. Then, the remaining out-of-gamut points are clipped using the HPmin\( \Delta E \) or the SCLIP algorithm depending on the pleasantness of appearance of the output image (in general three figures were generated for each texture/condition in the following experiments, with an average of 200 images for one object). Figure A5 gives an example of the distribution of colors of the banana under the illuminant D65 before (Figure A5-A) and after (Figure A5-B) gamut mapping, while Figure A6 shows the image of the same banana before (A) and after (B) gamut mapping. Note that Figure 6B in the main article is also a final image of the banana, i.e. the one presented to the observer.
APPENDIX 3: Control experiment

The experiment was divided in two parts with in total 87 participants, aged between 20-29. All participants were undergraduate or postgraduate students in the Psychology course at Newcastle University, and they were all naïve to the purpose of the experiment. None of them had participated in the main experiment of the article.

Two groups of participants were asked to view photographs of objects projected on a white screen in a normally illuminated classroom, recognize the objects and rate their certainty on the object identity on a scale of 1 to 7 (with 1 as very uncertain and 7 as 100% certain). The participants recorded their responses on a questionnaire sheet distributed at the beginning of the experiment. They had no time restriction to perform the choice and the successive photograph was displayed only after all participants had made their choice. It was requested that the subjects did not share their opinion with their neighbors and they were seated in such a way to make this interaction difficult. Each part of the experiment lasted about 15 minutes.

A3.1 Methods

Three photographs of the 3D solid replicas of familiar objects (3DN – Figure A7) used in the main experiment, taken under D65 on a gray background, were displayed to a first group of 45 participants (Group A, 15 male and 30 female, age 20-29).
Figure A7 - Photographs of the 3D replicas presented to the Group A of observers. A) Apple, B) Banana. C) Carrot.

The second group of 40 participants (Group B, 25 female and 15 male, age 20-29) performed the same task and procedure, but a different set of images was shown, as in Figure A8. In this case, the objects are the filled outlines of the images presented to the first group.

A difference in the object recognition performance between objects and between 2D and 3D configurations will help to establish how much information is provided by the contours and surface shading cues of the objects in the main experiment.

Figure A8 - Images of the outlines of the experimental solid objects (3DN) presented to Group B. A) Apple. B) Banana. C) Carrot.

A3.2 Results and Conclusions

For the 3D case (Group A), 100% of the observers recognizes the banana with a correct response certainty mean of 6.82 (out of 7), 97.8% recognize the carrot with a correct response certainty of 5.38 and 15.6% recognized the apple with 4.14 certainty.

For the 2D case (Group B), the banana was identified by 100% of the observers with a mean certainty of 6.81, the carrot by 61.9% with 3.35 mean certainty and the apple by 14.3% of observer with 2.67 mean certainty. Note that the maximum certainty is 7 and the minimum is 1, therefore setting to zero the subject’s “certainty” for a specific object when it is wrongly recognized, we can calculate the object recognition index (RI) as:

\[ RI = A \times C / P \]  

Equation A2
where A is the percent of correct answers, C is the mean certainty and P is the perfect certainty (i.e. 7). Results show that 3D apple, banana, and carrot are significantly more identifiable than their 2D counterparts (p<0.05), with the 3D banana performing best.

In conclusion, even though the fruit and vegetable objects used in this experiment are familiar, their solid shape (3D shape) encompass on average 30% more information on the object identity compared to their contours alone (2D shapes).

REFERENCES