

Lecture Scribe (4/27)

Matting, compositing and Environment Matting

B90902003 張譽馨

B90902096 曾又亭

R93922120 方壯雄

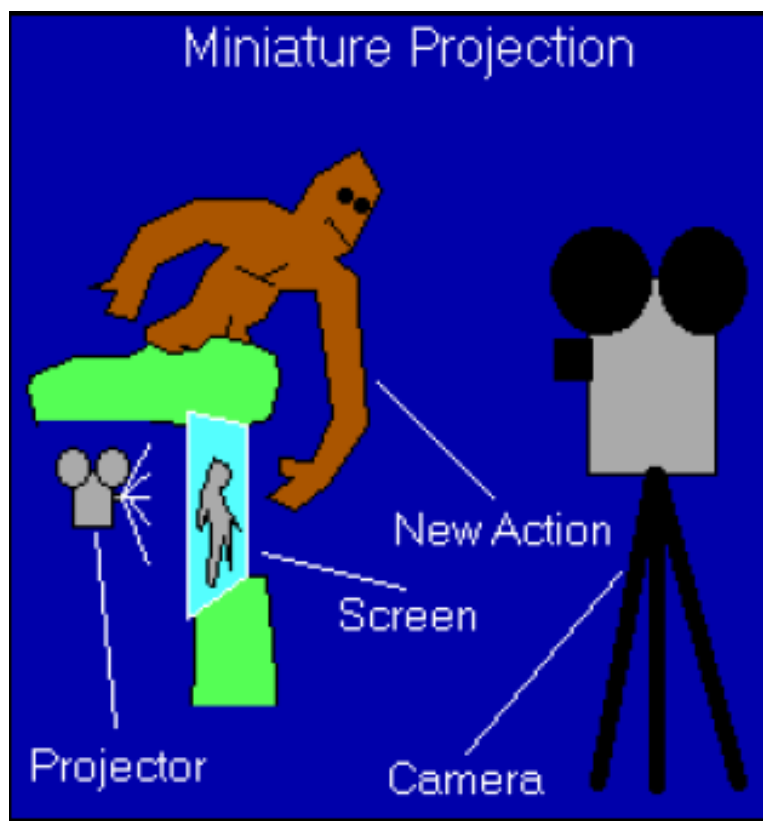
Traditional Matting and Compositing

Matting: 將影像中所需的 element 取出。

Compositing: 將背景與 element 組合起來。

The great train Robbery(1903) matte shot: 分開拍攝影片，利用不同的曝光度來達成效果，再將兩段影片合起來。

King Kong: 先將人的影像拍起來，將其投影到毛玻璃上與金剛玩偶一起拍攝，利用 stop motion 和 optical compositing 的技術。

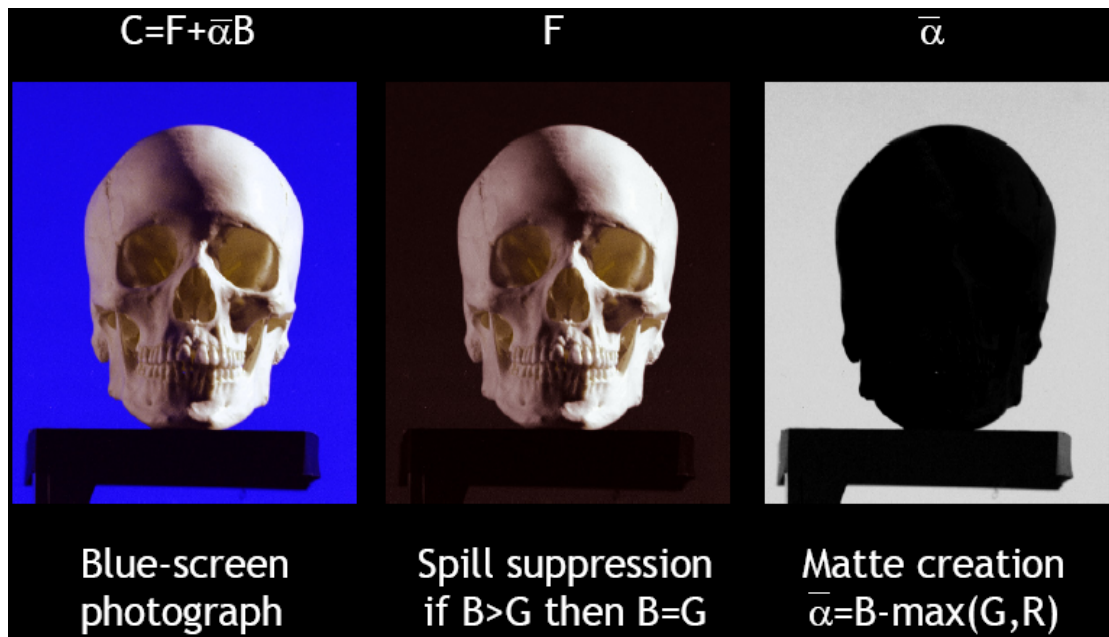


The lost world: (1925) 恐龍是利用玩偶拍攝。

(1997) 恐龍是利用動畫產生。



Color Difference Method (Ultimate)



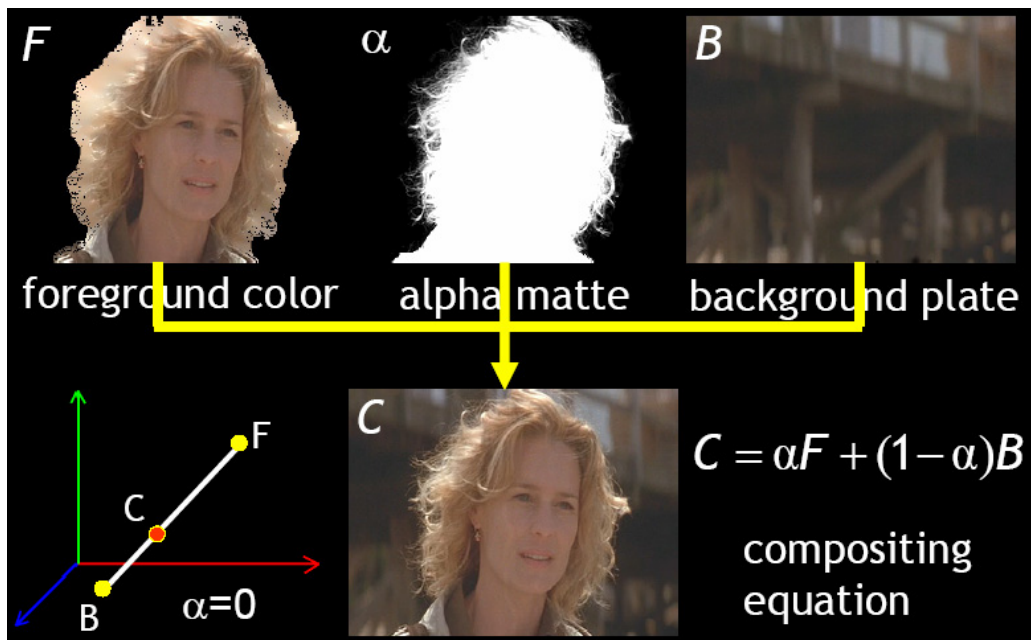
F: 前景 B: 背景

前提: 背景的藍色必須明亮且均勻。

Primatte Method

用兩個 128 面的多面體把 color space 分成三區，中間的區域的 color 的 alpha: $0 < \alpha < 1$ 。

Compositing



α 為透明度，是一個 undetermined 的 parameter。

Estimate the right part from the left part, this is an under constraint problem。

There are 7 undetermined variables。

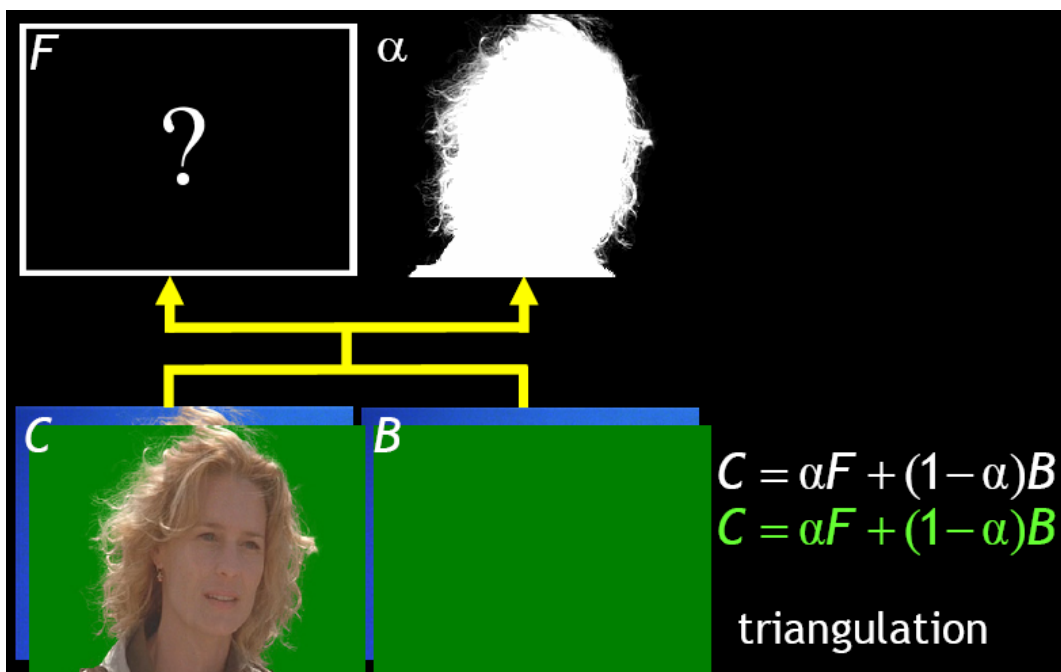
Resolve this equation 有三個方法:

1. reduce unknowns:

 difference matting: (when the background is known), 需要多個 thresholds.

 Color difference method: (blue screen matting)

2. add observation: 在藍幕、綠幕各拍一次，實際應用不有效。



3. add priors: user 先勾勒出前景；或 Ruzon-Tomasi 方法，由 user 將影像分爲 foreground, background 和 unknown 三個 region。



Bayesian Image Matting

posterior probability

$$\arg \max_{F, B, \alpha} P(F, B, \alpha | C)$$

likelihood priors

$$= \arg \max_{F, B, \alpha} P(C | F, B, \alpha) P(F) P(B) P(\alpha) / P(C)$$

C: observation，已經發生的影像結果。

Likelihood: 在已經發生 F(foreground), B(background), α 的前提下，發生 C 的機率。

Priors: 發生 F(B or α)的機率。

如何有效率解出 optimal solution 是我們的訴求。

$$L(C | F, B, \alpha) = -\|C - \alpha F - (1 - \alpha)B\|^2 / \sigma_C^2$$

(爲 Bilinear equation)

$$C' = \alpha F + (1 - \alpha)B$$

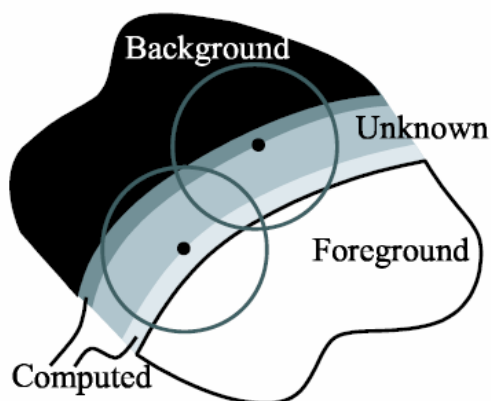
Minimize C 與 C' 的距離，因為會有 noise，所以結果不一定會好。

通常用 smooth Gaussian 當 mode priors。

定 prior:

$$\bar{F} = \frac{1}{W} \sum_{i \in N} w_i F_i$$

$$\Sigma_F = \frac{1}{W} \sum_{i \in N} w_i (F_i - \bar{F})(F_i - \bar{F})^T$$



Unknown region 離 foreground(background)越接近, weight 越高, 與 foreground (background) 越遠, weight 越低。

利用 Gaussian 的好處: 取 log 時, equation 變成 quadratic, 取偏為分時, equation 變成 linear system。所以:

$$L(F) = -(F - \bar{F})^T \Sigma_F^{-1} (F - \bar{F}) / 2$$

(為 Gaussian 的指數部分)

Optimize $L(C | F, B, \alpha)$

$$E = \|C - \alpha F - (1 - \alpha)B\|^2 / \sigma_C^2 - (F - \bar{F})^T \Sigma_F^{-1} (F - \bar{F}) / 2$$

1. fix α , solve F, B (linear)。

$$\begin{bmatrix} \Sigma_F^{-1} + I\alpha^2 / \sigma_C^2 & I\alpha(1 - \alpha) / \sigma_C^2 \\ I\alpha(1 - \alpha) / \sigma_C^2 & \Sigma_B^{-1} + I(1 - \alpha)^2 / \sigma_C^2 \end{bmatrix} \begin{bmatrix} F \\ B \end{bmatrix} = \begin{bmatrix} \Sigma_F^{-1} \bar{F} + C\alpha / \sigma_C^2 \\ \Sigma_B^{-1} \bar{B} + C(1 - \alpha) / \sigma_C^2 \end{bmatrix}$$

2. fix F, B , solve α (linear)。

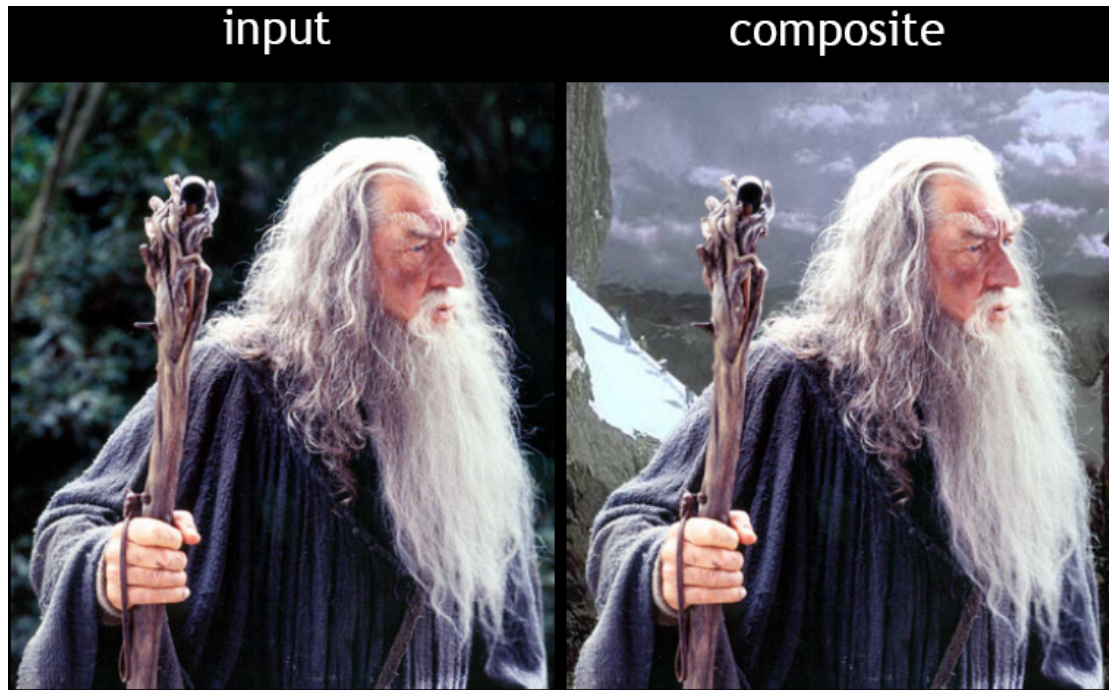
$$\alpha = \frac{(C - B) \cdot (F - B)}{\|F - B\|^2}$$

Repeat until converge

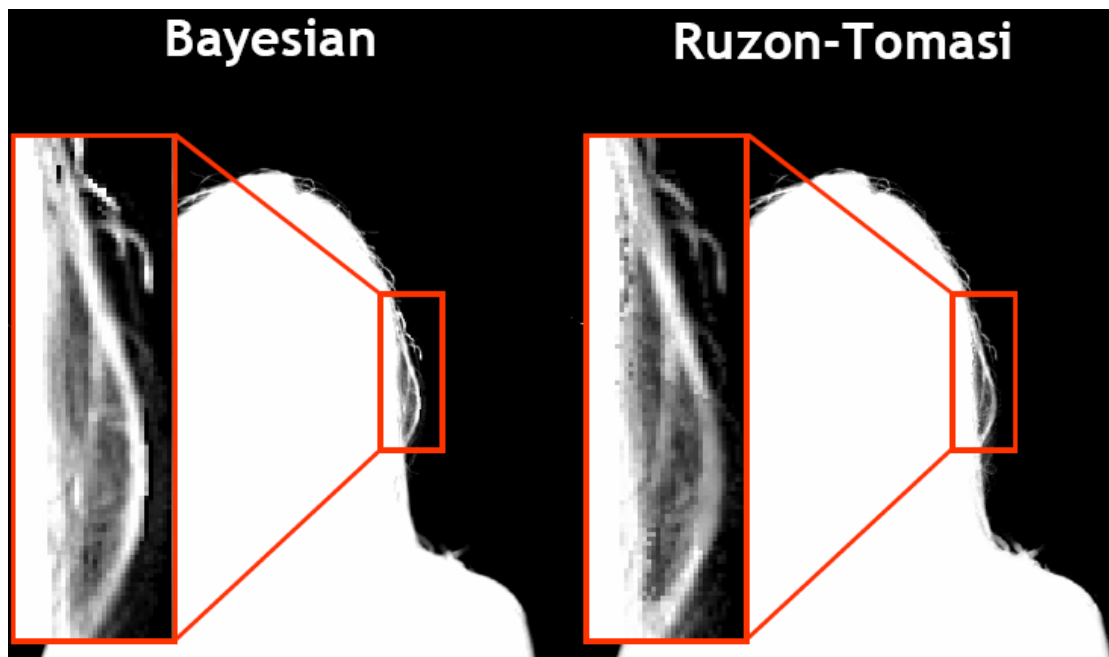
Pixels in the unknown region are inserted into a priority queue. Pixels near FG or BG will be processed first. 計算順序會影響結果, 對 trimap 很 sensitive。

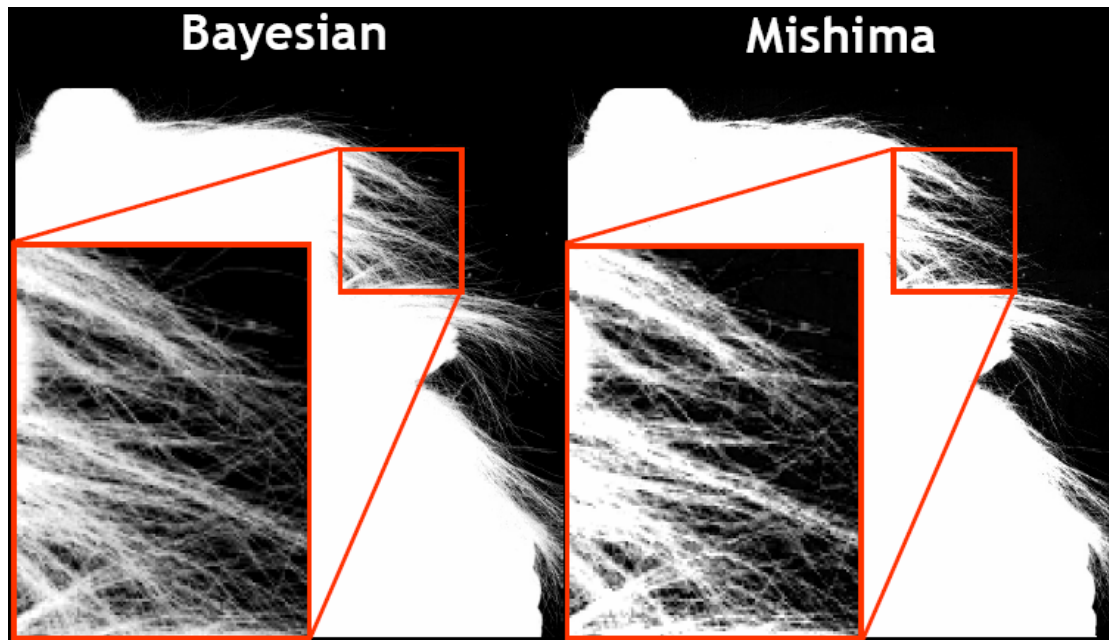
Bayesian Matting is often used in natural image matting, but still performs good in blue screen matting and the result is better than traditional blue screen matting algorithm

Result:



Comparisons:





Mishima: 每一點獨立去解。
 Bayesian: smooth computing。

Video Matting



對每個 frame 畫 trimap 相當費時，要求使用者對 keyframe 畫 trimap 就好。
 但要解決相鄰 frame 不相容的問題。



得到 keyframe 的 trimap，用 optical flow 算其他的 trimap，並且算出正逆方向的 optical flow，再 warp 出其他的 trimap，將兩個 trimap combination。



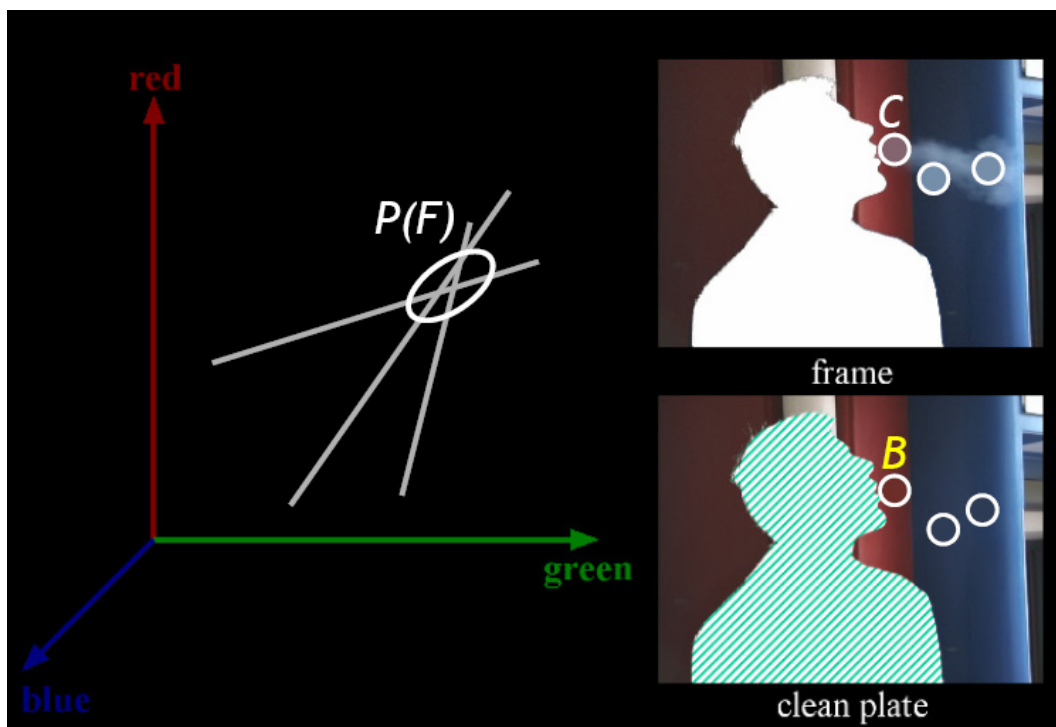
Optimize: 可將背景先算出來。

Result:



Smoke:

1. Assume clean plane is known.
2. Remove foreground.
3. Compare the color on the foreground removed image and clean plane.
4. Solve the color of smoke (least square).

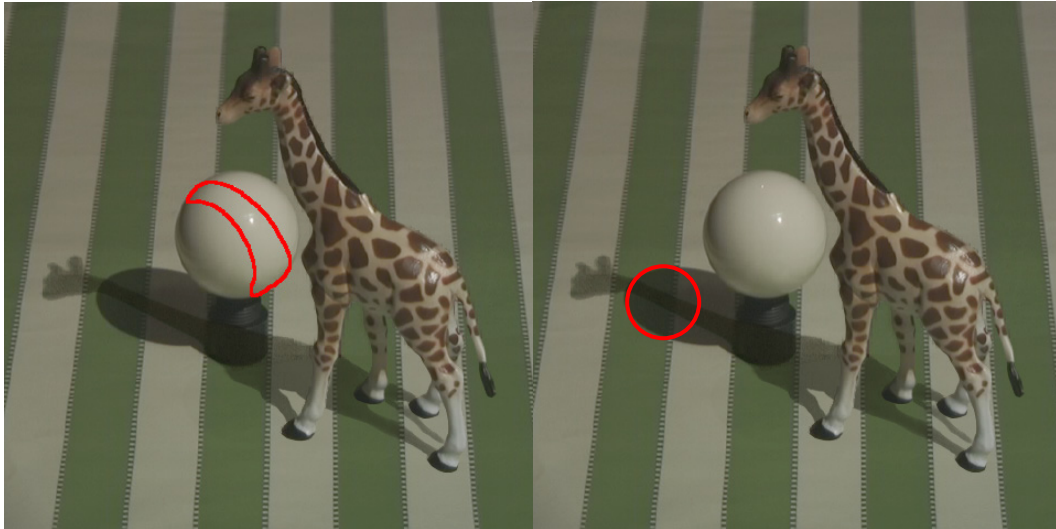


Shadow Matting

constrain: image 的 light 需相近。

Shadow on Traditional Matting

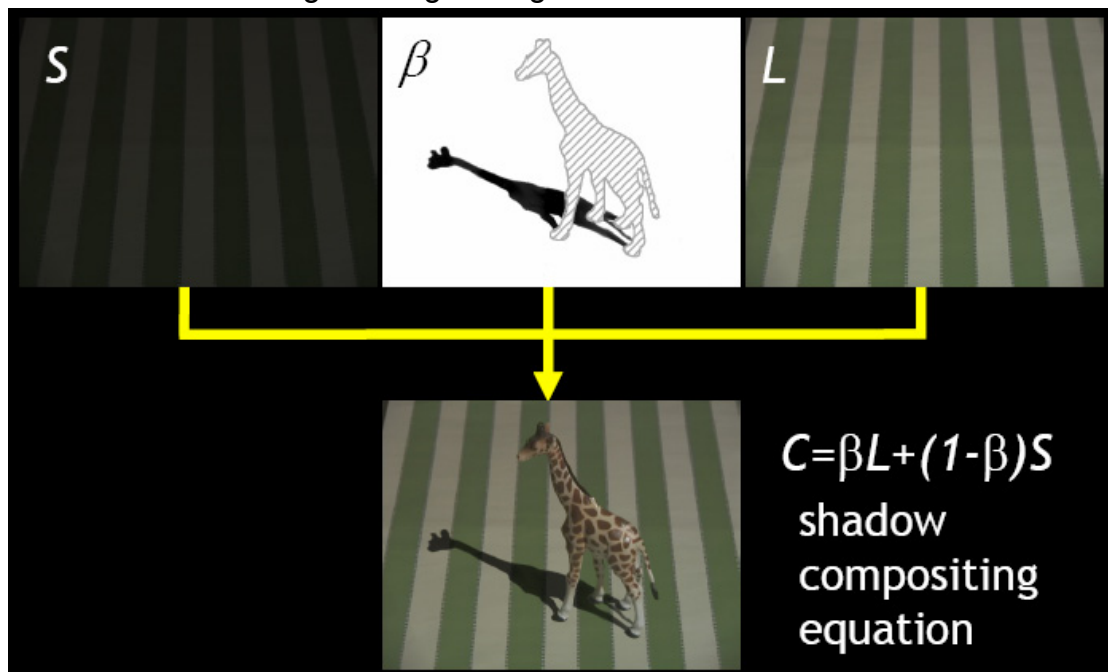
1. geometric errors
2. double shadow errors(photometric errors)



Shadow Matting Equation

$$C = \beta * L + (1-\beta) * S$$

Assume shadow image and light image is known



Estimate shadow image and light image from a video sequence

1. Remove foreground.

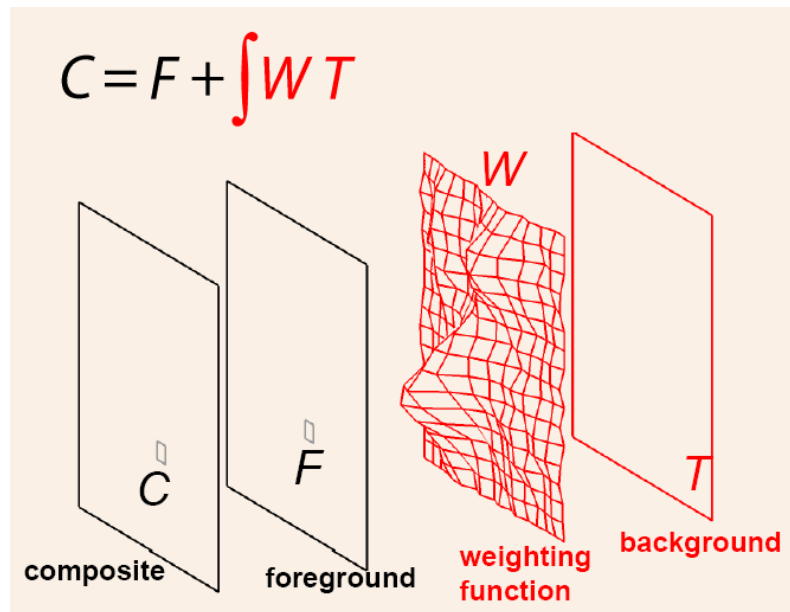
2. Construct a video volume.
3. The brightest pixel → lit image, The darkest pixel → shadow image

Environment Matting

Capture 光線的折射和反射

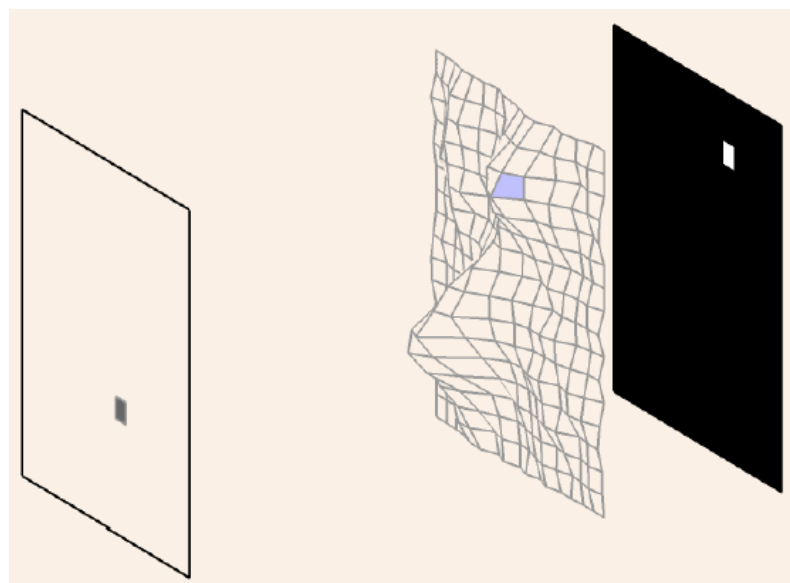
Framework

$C = F + \int W T$



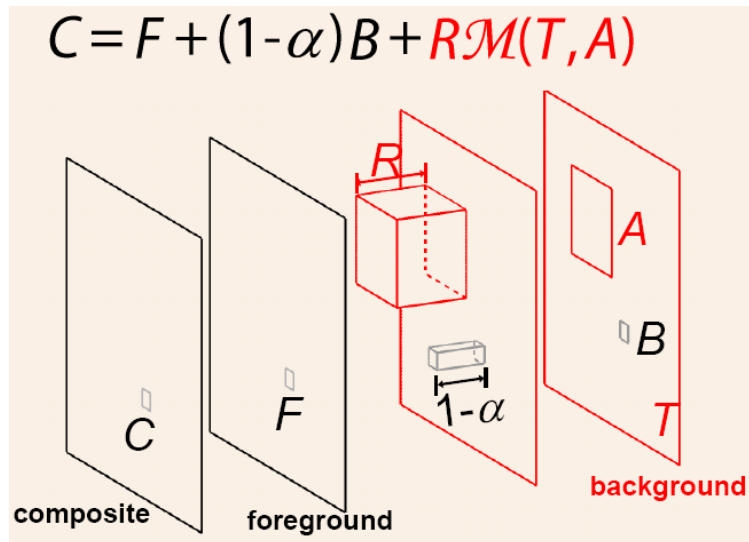
Arbitrary Weighting Function

LCD scanning. Capture $n \times n$ images



Hierarchical environment matting

1. $C = F + (A - \alpha)B + RM(T, A)$
2. Hierarchical backgrounds
3. Divide a 4D optimization problem to two 2D optimization problems.



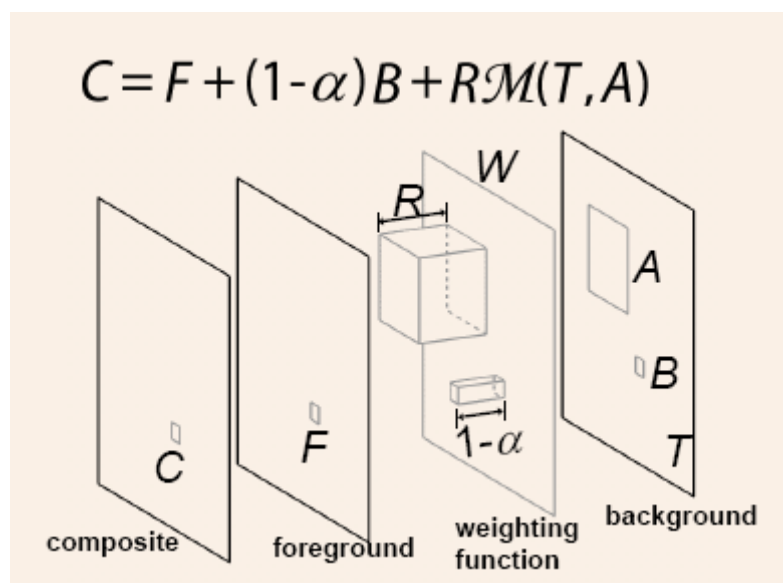
Extension

1. tree side background
2. background on several distance

Disadvantage

1. need many background images
2. multiple mapping
3. glossy surface

Real-time environment matting



$$C = F + (1 - \alpha)B + RM(T, A)$$

The properties of the matte that we would most like to preserve:

- (1) the capacity to refract, reflect, and attenuate the background(A,R)
- (2) smooth blending with the background on silhouettes(α)
- (3) specular highlights due to foreground lighting(F)

我們只專注於保留第一個屬性

$$C = RM(T, A)$$

假設物件是 colorless，那 R 就會變成一個純量 ρ ，與波長無關。

假設物件為 specularly reflective and refractive，意指 neighboring pixels 不需要有 overlapping support in their weighting functions。

Area A 可以被寫成{C,W} C=(Cx, Cy)，area 的中點，W=(Wx, Wy)為 x 軸及 y 軸的寬度。

$$W_x \equiv \frac{\partial}{\partial x} C_x \approx \frac{1}{2} [C_x(x+1, y) - C_x(x-1, y)]$$

$$W_y \equiv \frac{\partial}{\partial y} C_y \approx \frac{1}{2} [C_y(x, y+1) - C_y(x, y-1)]$$

$$M(T, A) \approx T(c)$$

$$\Rightarrow C \approx \rho T(c_x, c_y)$$

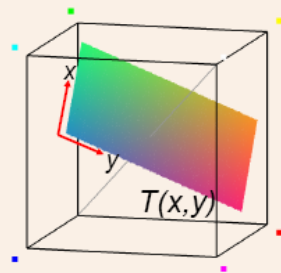
3 observations, 3 variables

然後我們可以解出三個未知數。

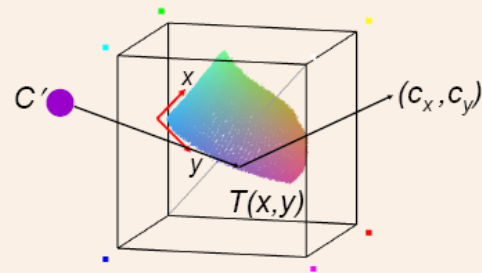
Single image matte recovery

Because we assume that $F=0$ everywhere, we photograph the object in a dark room, lit only by the structured backdrop. The structured background is a smoothly-varying wash of color, in particular, a planar slice through the rgb cube. Due to non-linearities in the system, including crosstalk between the spectra of monitor phosphors and the CCD elements, the gamma of the backdrop display, and processing in the camera's electronics, this plane in color space will be distorted, becoming a curved 2D manifold lying within the rgb cube.

Ideal plane in RGB cube

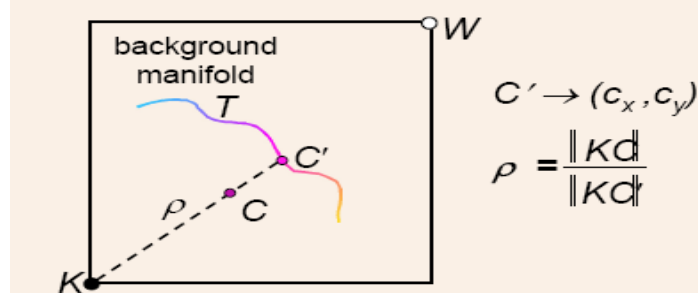


Calibrated manifold in RGB cube



To extract matte parameters at each pixel, we consider the line joining the observed color and the black point in rgb space. The point where this line intersects the background-color manifold gives us the point c , and the fractional distance of the observed color to the manifold gives us ρ .

Estimate c_x, c_y and ρ



To classify pixels, we first take a series of pictures of the background without the object and average them to give us a low noise estimate of the ramp background. Once we begin recording video of the object, we apply a simple difference threshold to each frame, comparing the image of the object to the image of the background alone. This step separates foreground and background pixels.

Difference thresholding



We then use some morphology operations(dilation followed by hole-filling followed by erosion) to clean up this binary map, giving us a reasonably

accurate mask of the pixels covered by the object.

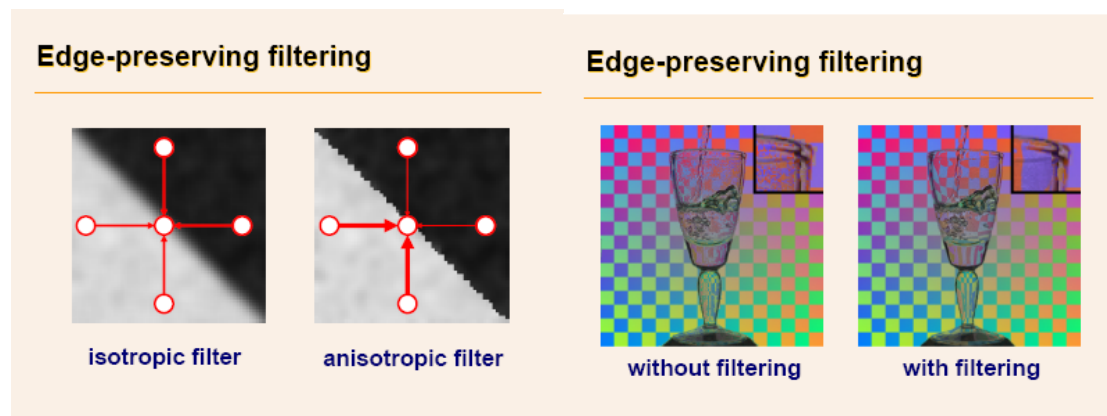


To avoid a sharp discontinuity, we slightly feather the alpha at the boundaries of the object as a post-processing step.



Problem: noisy matte

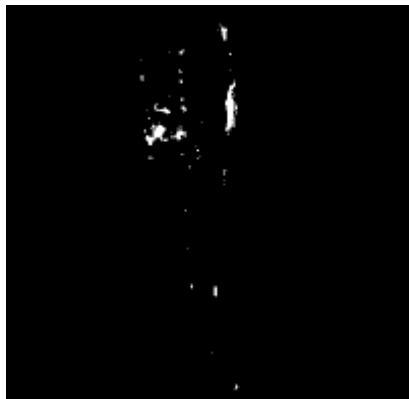
Apply the edge-preserving smoothing operator of Perona and Malik to the Cx and Cy channels. This operator average each pixel with its neighborhood, with unequal contributions are determined by difference of the pixels' values, so that similarly-values pixels affect each other more. This filter smooths out regions with low-to-moderate noise levels while preventing significant energy transfer across sharp edges.



Heuristics for specular highlights

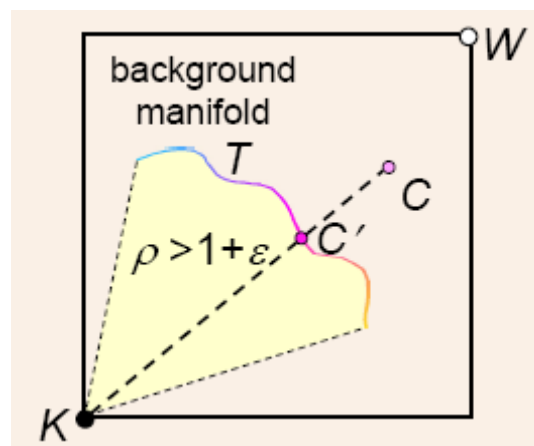
To recover the intensity of the foreground color F , under the restriction that it is white. Thus, $F=fW$ where $W=(1,1,1)$, so that only one additional parameter, f is added to the matting equation. This new single-image environment matting equation then becomes

$$C = fW + (1 - \alpha)B + \rho T(c)$$



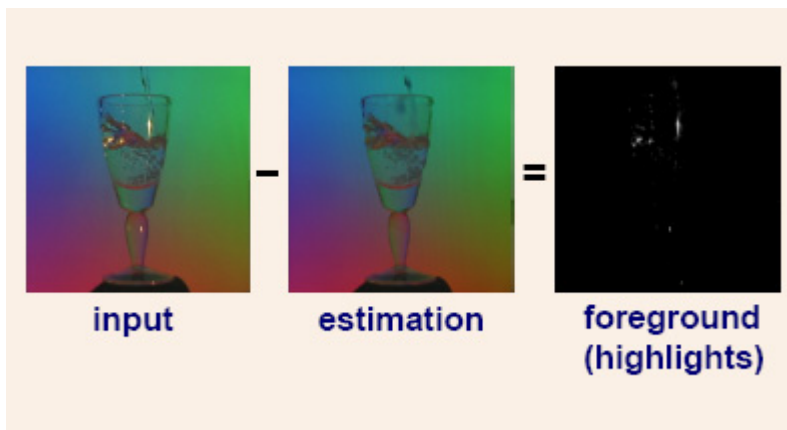
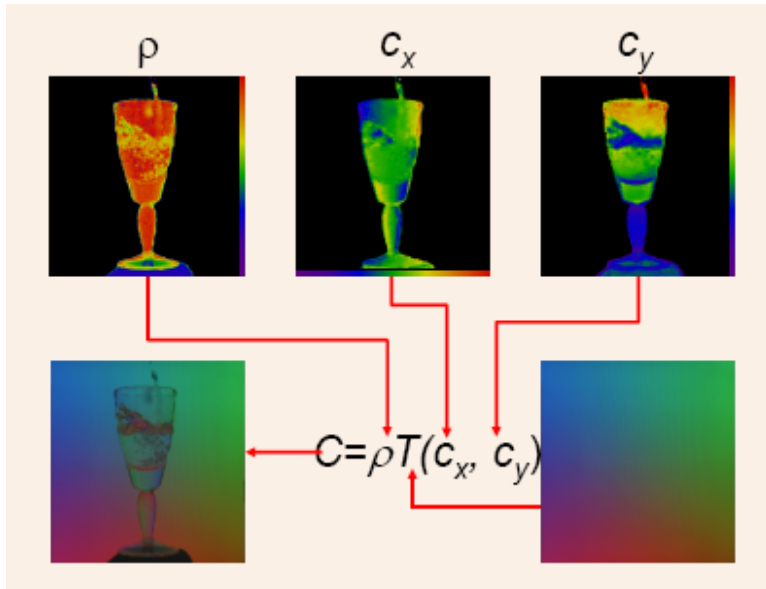
Foreground F

The recovery algorithm of the last section will discover some points where $\rho > 1$, i.e., where the observed color point lies on the side of the background manifold closer to white.

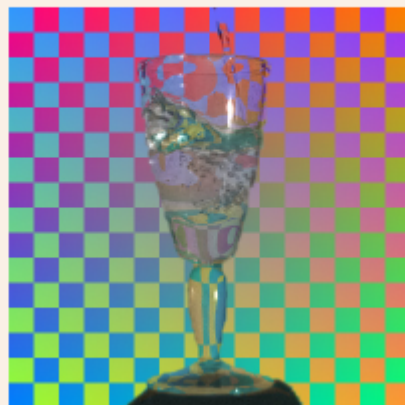


The theory of light transport tells us that this should not happen when the object is lit only by the backdrop, so we assume that wherever ρ exceeds unity there must be some F -term contribution to the pixel.

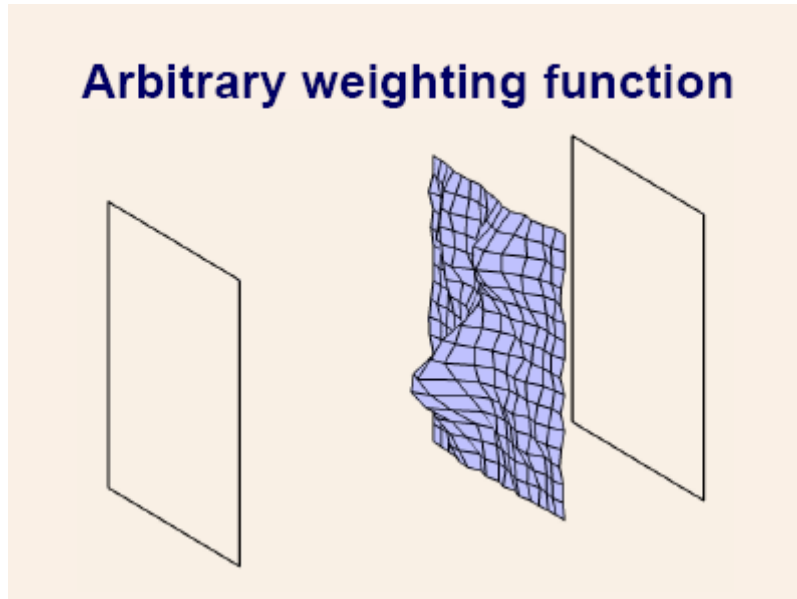
$$f = C - (1 - \alpha)B - \rho T(c)$$



Composite with highlights



Multimodal oriented Gaussian



We generalize arbitrary weighting function to a sum of Gaussian:

$$W(x) = \sum_{i=1}^n R_i G_i(x)$$

Here, R_i is an attenuation factor, and each G_i is the unit-area, elliptical, oriented 2D Gaussian:

$$G_i(x) \equiv G_{2D}(x', c_i, \sigma_i, \theta_i)$$

Where G_{2D} is defined as

$$G_{2D}(x', c, \sigma, \theta) \equiv \frac{1}{2\pi\sigma_u\sigma_v} \exp\left[-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right]$$

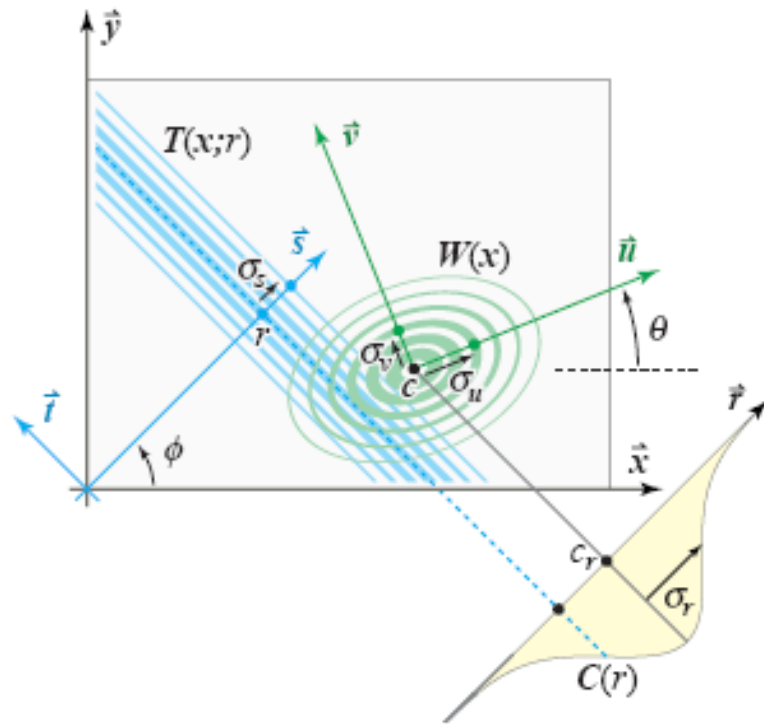
with

$$u = (x - c_x) \cos \theta - (y - c_y) \sin \theta$$

$$v = (x - c_x) \sin \theta + (y - c_y) \cos \theta$$

Here, $x = (x, y)$ are pixel coordinates, $c = (c_x, c_y)$ is the center of each

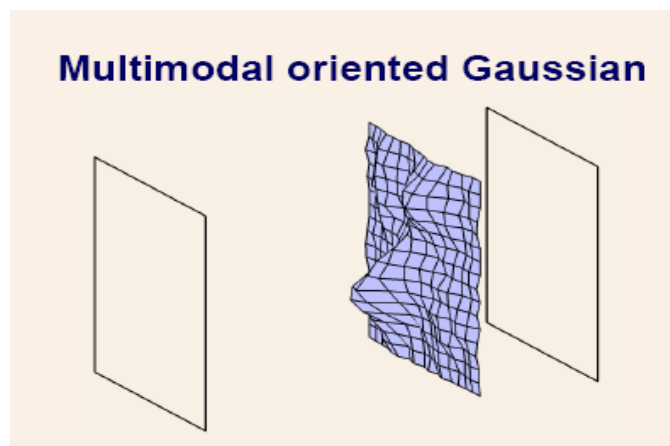
Gaussian, $\sigma = (\sigma_u, \sigma_v)$ are unrotated widths in a local uv-coordinate system, and θ is orientation.



Thus, our weighting function is some n-modal Gaussian with each term contributing a reflective effect from the object. We arrive at new form of the matting equation:

$$C = F + \sum_{i=1}^n R_i \int G_{2D}(x', c_i, \sigma_i, \theta_i) T(x) dx$$

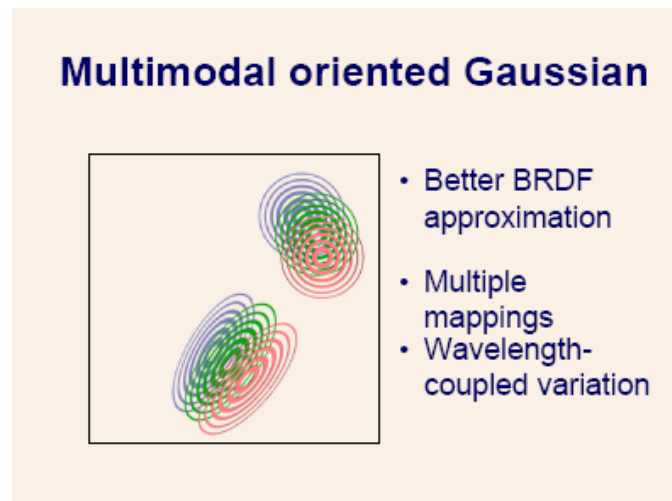
$T(x)$ represent the set of all texture maps.



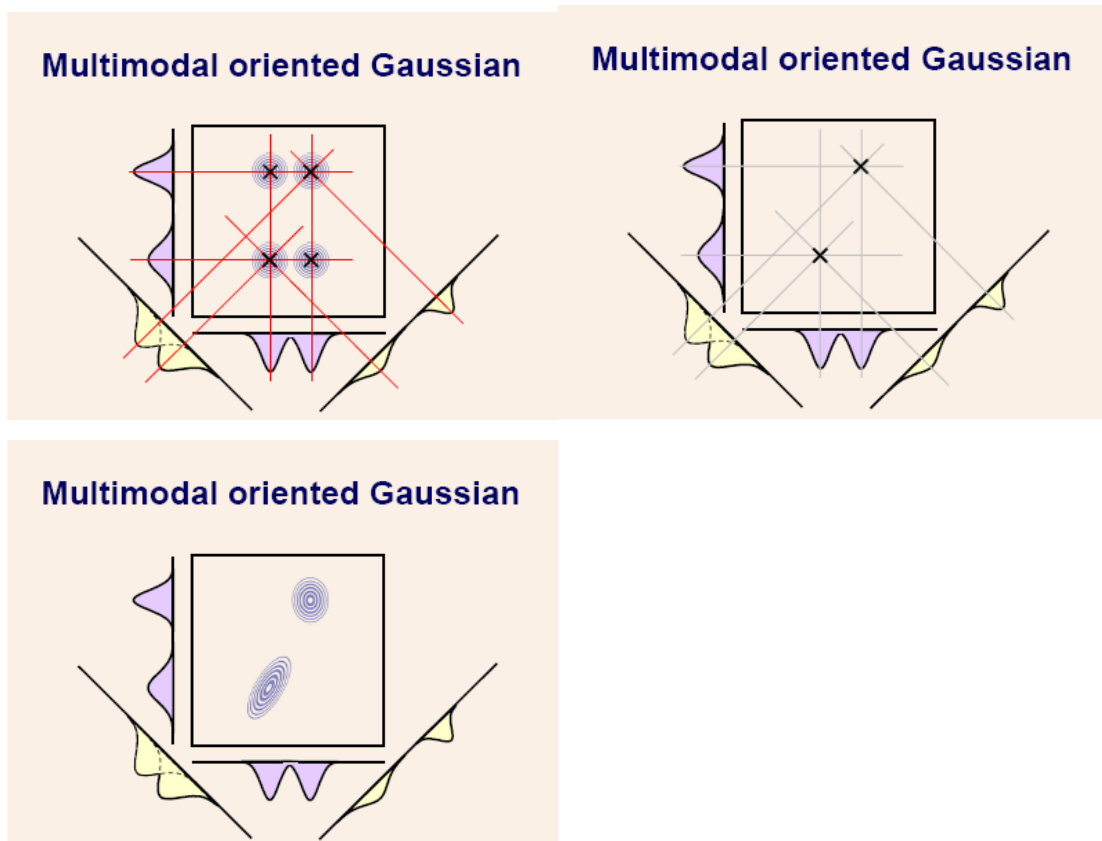
The key advantage of this weighting function:

- (1) the spatial variation can be coupled with wavelength to permit modeling of dispersion.
- (2) it supports multiple mappings to a single texture.

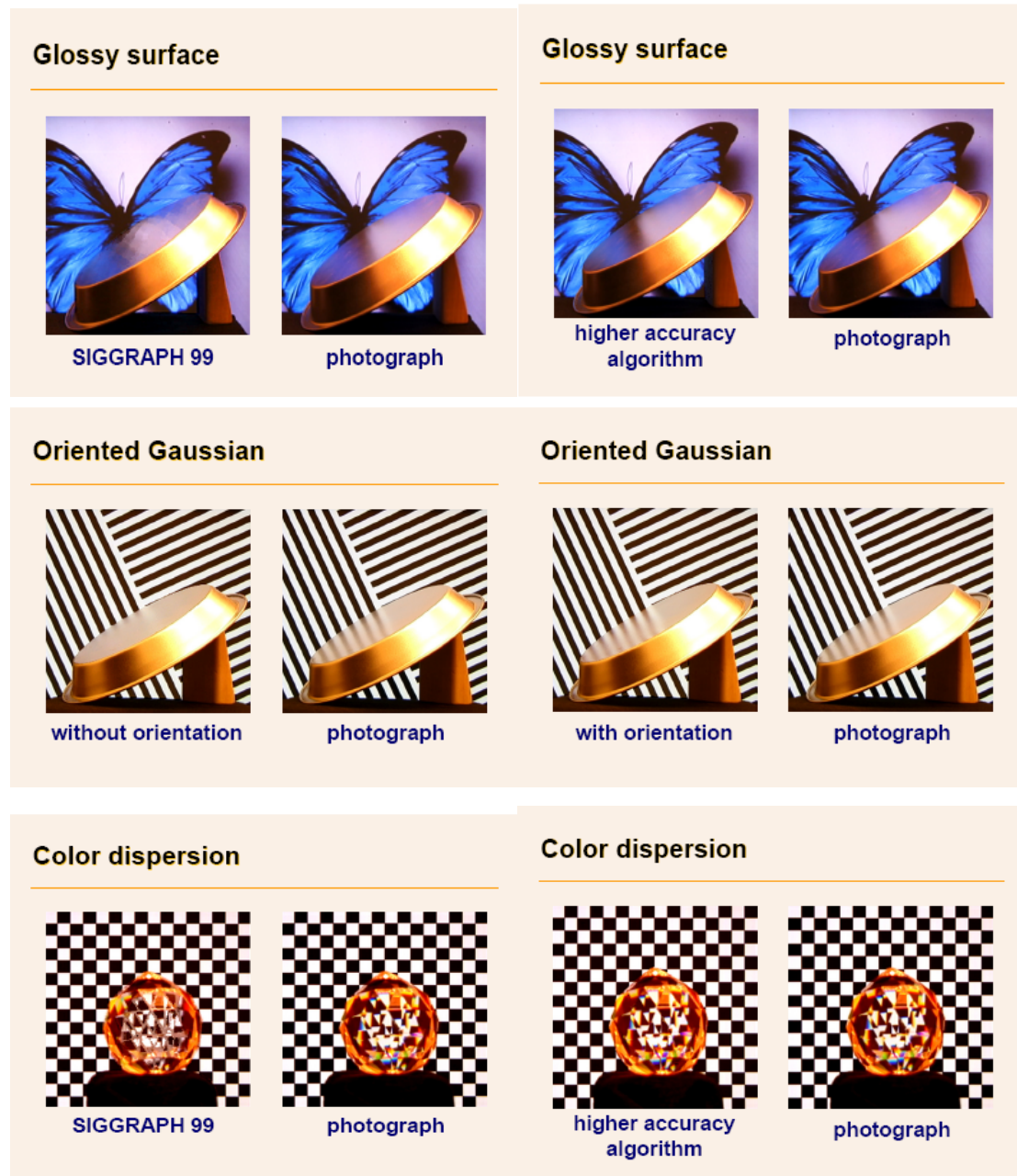
(3) it approximates the behavior of BRDF's more closely.



Horizontally and vertically swept stripes alone are not enough to determine the weighting function, so we introduce two diagonal passes at 45° and -45° .

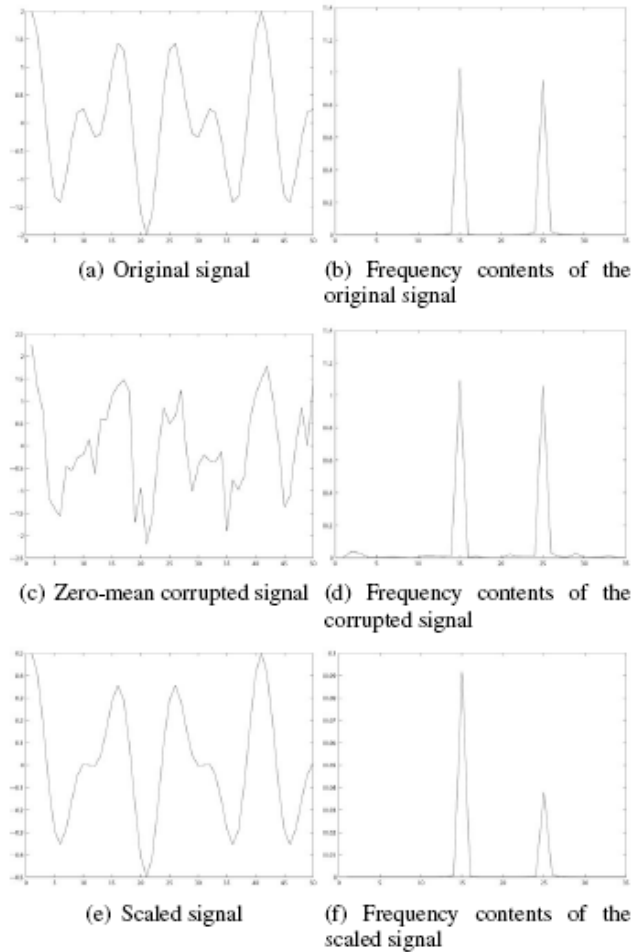


Result

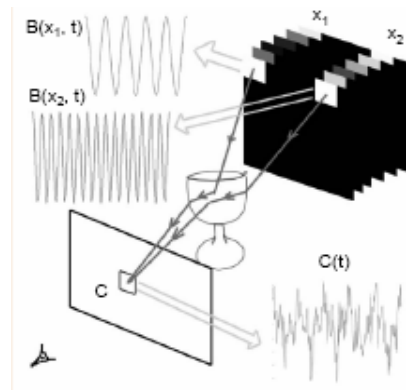


Frequency-based environment matting

Frequency analysis

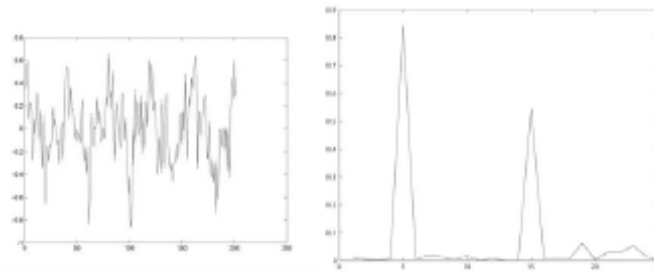


We display the background image B on the CRT monitor. Light rays start from pixels $B(s, t)$ in the background image, interact with the foreground object and finally reach pixel $C(x, y)$ in the recorded image plane. The goal of environment matting is to compute the inverse mapping of this process, i.e. to find the mapping from $C(x, y)$ to $B(s, t)$. It is noteworthy that the mapping can be one-to many.



We regard each pixel in the background image as a signal emitter. Recorded images are stacked in time order and the pixel intensities are interpreted as

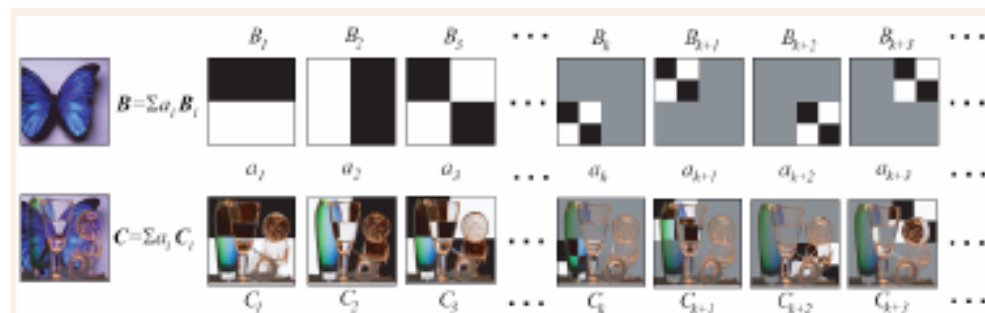
signals in this “time” axis.



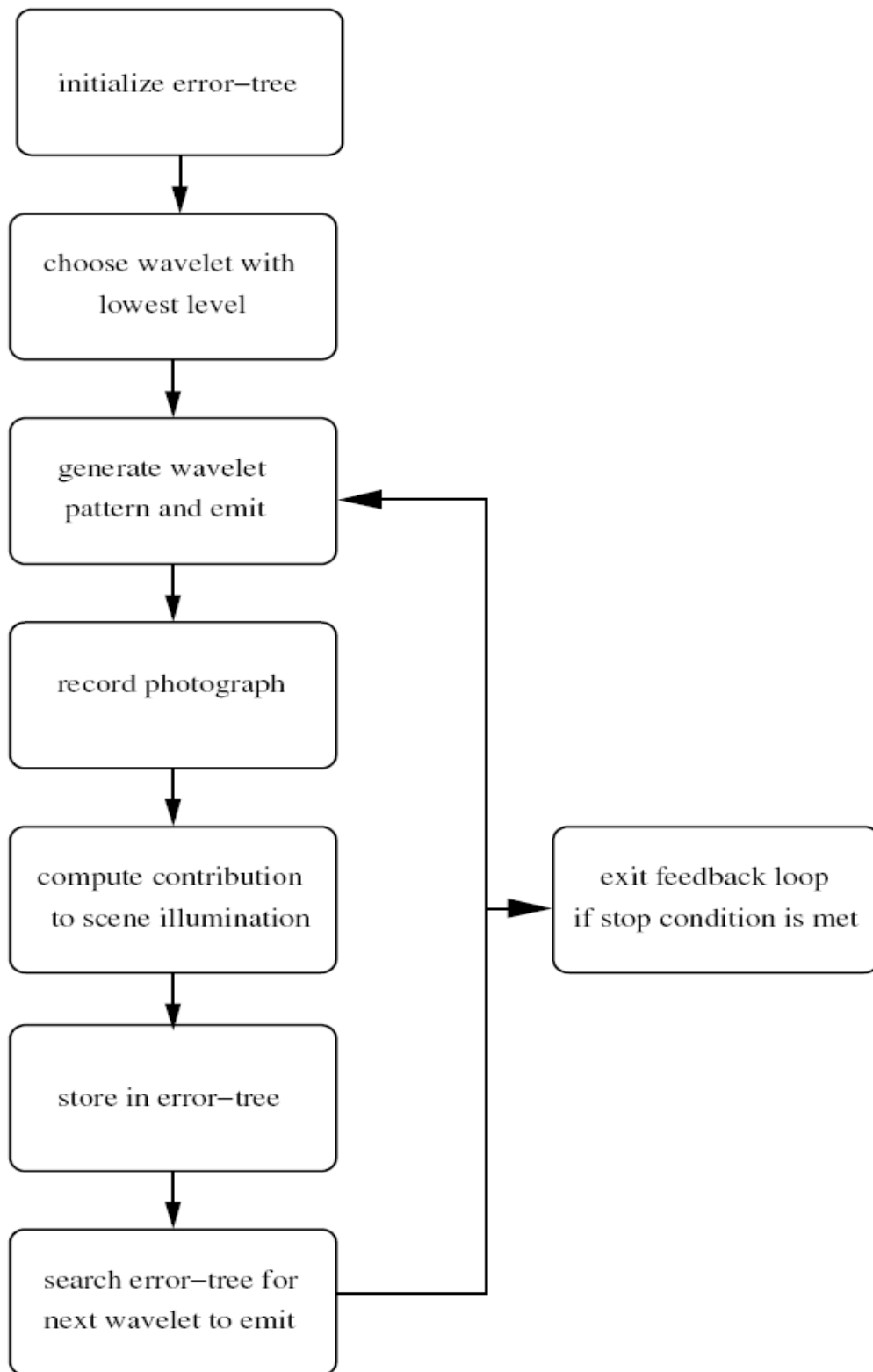
Signal pattern in $C(x,y)$ and its frequency components.

<p>Results: refraction</p> <p>frequency-based environment matting photograph</p>	<p>Results: color dispersion</p> <p>frequency-based environment matting photograph</p>
<p>Results: oriented</p> <p>frequency-based environment matting photograph</p>	<p>Results: oriented</p> <p>frequency-based environment matting photograph</p>

Wavelet environment matting



An overview of the wavelet environment matting algorithm.




$$C = \int wB = \int w \sum a_i B_i = \sum a_i \int w B_i = \sum a_i C_i$$

理論上可以處理任何的 weighting function




a is reference image, b is 1200 basis image.

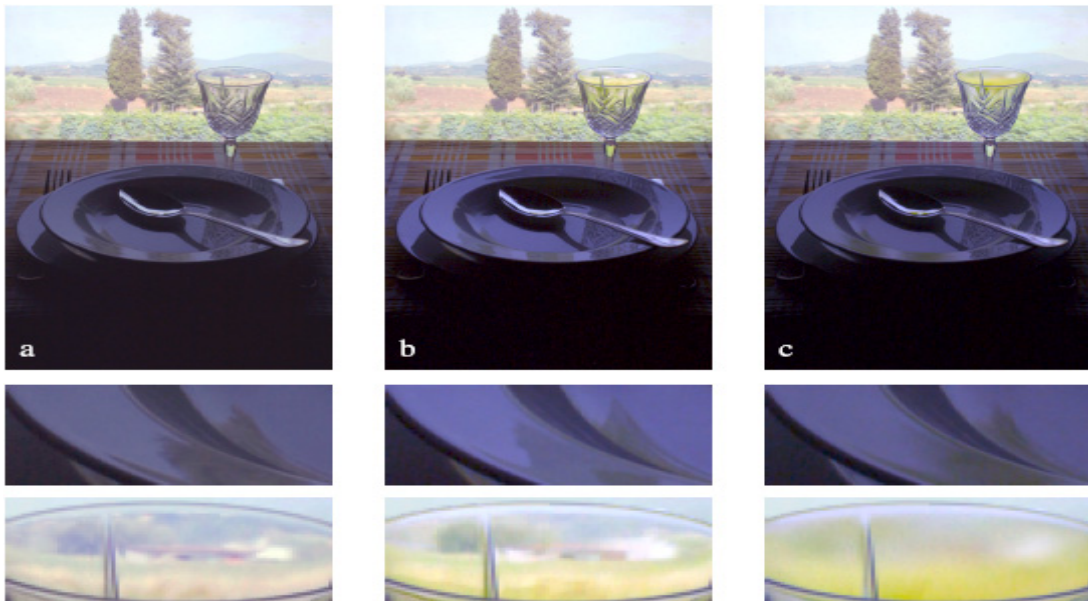
Results: number of basis images



reference image



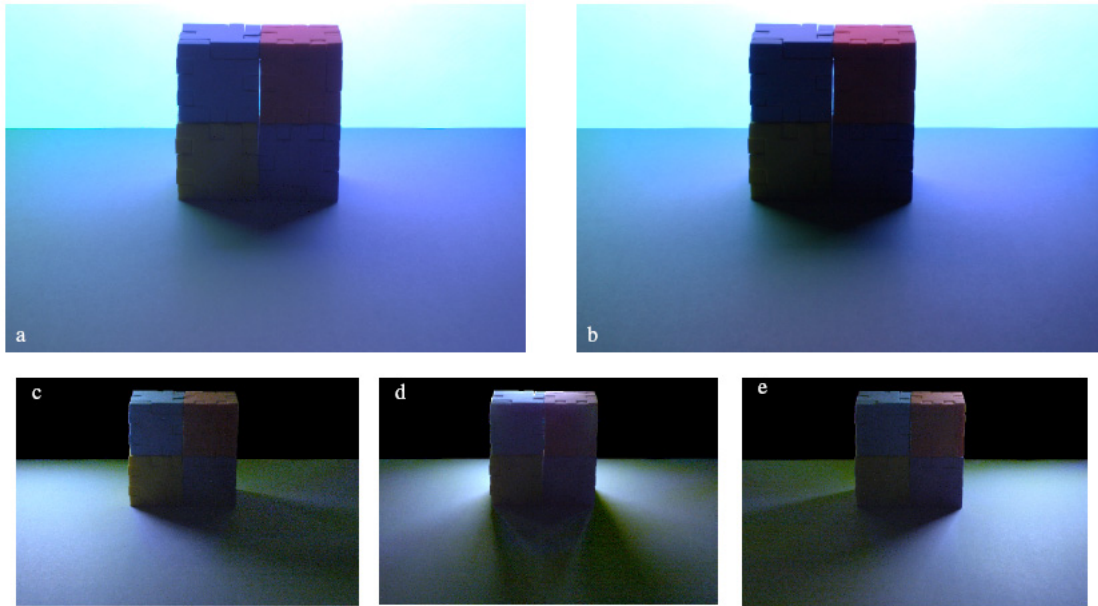
1200 basis images



a is reference image, b is 1000 Haar patterns, c is 1000 Daubechies (9,7)

patterns

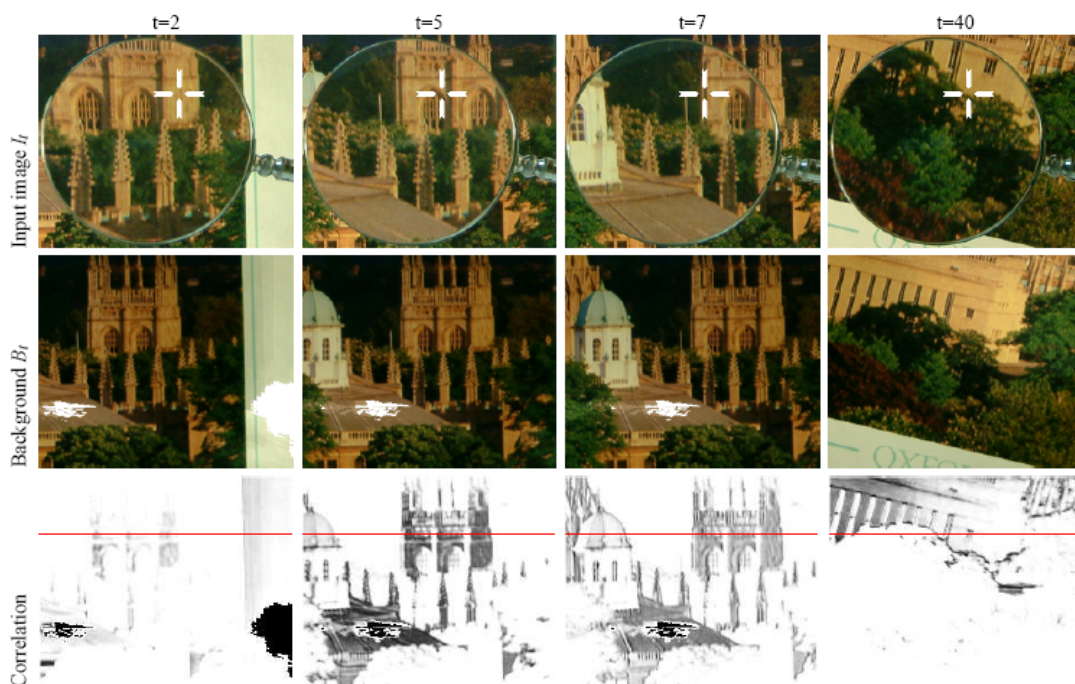
Diffuse material

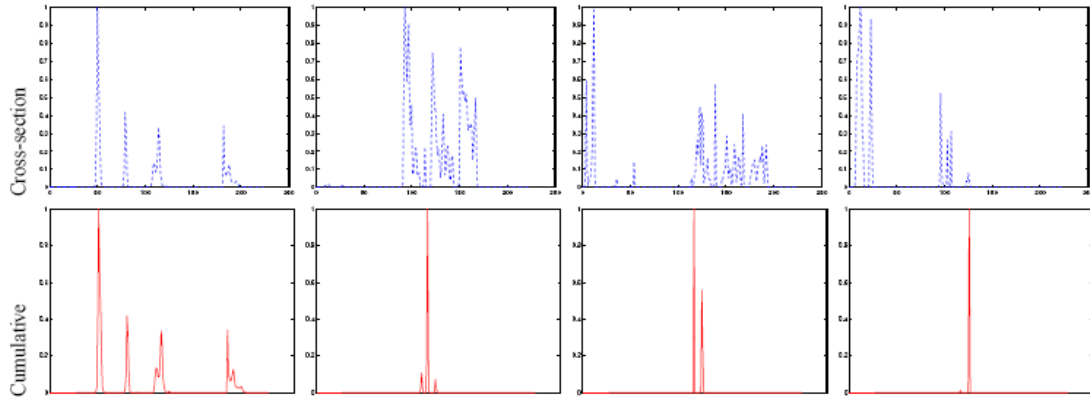


A scene containing colored cubes placed on a diffuse surface. The scene, composited with a low frequency plasma backdrop, is shown in figure b. A reference image is shown in figure a. In figure c, d and f the same scene is composited with different backdrops containing a white square at different locations (respectively located on the left, middle and right).

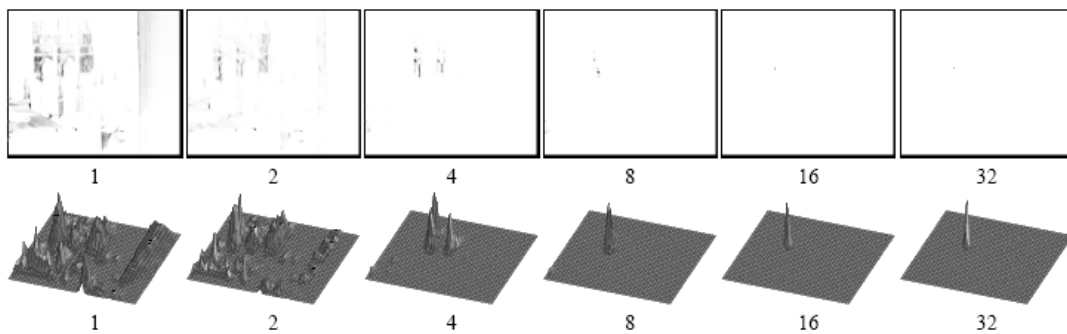
Image-based environment matting

利用之後的 image 消除 false peak





The query pixel is marked with a white cross. Each column corresponds to a new image pair. The first two rows show measured foreground I_t and registered background B_t . The white smudge on the background images is an area where no background color could be computed, as it was always occluded by the magnifying glass. The third row shows the receptive field $r(u, v)$ of the output pixel, computed from just that view pair. The fourth row shows a cross-section through the receptive field. It can be seen that a single image does not constrain $r(u, v)$ very tightly – the curves are far from unimodal. The red curves in the fifth row, on the other hand, show the normalized cumulative products of the preview curves. These represent the integrated receptive fields, and show that the erroneous peaks in the distribution are quickly eroded as more images are added. Furthermore, the finally accepted peak does not necessarily correspond to a maximum in any individual image.



Refining receptive fields for a single output pixel. The receptive field $r(u, v)$ for a single (x, y) pixel as more views are added. After the first pair, the receptive field is far from accurate, with many false maxima (dark regions). As more views are integrated, the estimate is progressively refined. (Top row): intensity map—dark pixels have higher weights. (Bottom row): surface plot.

Results



Results



Comparisons

category	method	asymptotic # of images	typical # of images	weighting function	materials
active	RTEM	1	1	warping function	colorless, specularly refractive
	HEM	$O(\log k)$	20	box filter	refraction, translucency, highly specular, color transparency
	GEM	$O(k)$	600	sum of Gaussians	+color dispersion, multiple mappings and glossy reflection
	FBEM	$O(k)$	1,200	product of two 1D functions	-multiple mappings
	WEM	$O(k^2)$	1,200	object images	+diffuse reflection
passive	IBEM	N/A	40	probability map	colorless, specularly refractive
	ROEM	N/A	200	warping function	colorless, specularly refractive