**通过相位一致性检测特征**

不是假设图像应被压缩成一组边缘，而是特征检测的相位一致性模型假定压缩图像格式应该是高*信息* （或*熵*），并且冗余度低。因此，代替寻找那里有强度的急剧变化的点，这种模式搜索模式*顺序*在傅立叶变换的相位分量变换。之所以选择相位，是因为Oppenheim和Lim [ [9](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#opp) ]的实验证明它对视觉特征的感知至关重要。进一步的生理证据[ [6](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#mandb)]表示人类视觉系统对相位信息高度有序的图像中的点作出强烈响应。因此，相位一致性模型将*特征*定义为具有高相序的图像中的点。

相位一致性模型是基于*频率*的视觉处理模型。假设视觉系统不是在空间上处理视觉数据，而是能够使用信号中各个频率分量的相位和幅度来执行计算。因此，底层计算工具是傅立叶变换，或其等价物之一。为此，让我们假设我们可以在傅里叶域中表示我们的图像信号。为了简化演示，我们将假设一个简单的一维信号，表示通过图像的一维切片。这样的信号，比如*f*（*x*），是通过傅立叶变换重建的

\ begin {displaymath} f（x）= \ int _ { -  \ infty} ^ {\ infty} a _ {\ omega} \ cos（T \ omega x + \ phi _ {\ omega}）d \ omega，\ end {displaymath }

其中，对于每个频率$ \ $欧米茄，美元_ {\欧米茄} $是余弦波的幅度，并且$ T \ omega x + \ phi _ {\ omega} $是该波的相位偏移。术语*T*与图像窗口的大小有关，从现在开始我们假设它是1。

例如，如果图像信号是简单的阶梯边缘，那么

\ begin {displaymath} f（x）= \ frac {-4} {\ pi} \ int _ { -  \ infty} ^ {\ infty} \ frac {1} {2 \ omega + 1} \ cos（\ omega x + \ pi / 2）d \ omega，\ end {displaymath}

并且，在阶梯边缘（*x* = 0）处，所有相位项都对齐$ \ PI / 2 $。这是信号中唯一一个相位值*一致的*地方; 在所有其他点，各个频率分量的相位值假设0和0之间的不同值$ 2 \ pi $。

|  |
| --- |
| **图1：** 周期性阶梯边缘的所有傅里叶项在每个步骤点都是同相的。 |
| \ begin {figure} \ par \ centerline {\ psfig {figure = step.ps，angle = -90，height = 2in，width = 4in}} \ par \ end {figure} |

|  |
| --- |
| **图2：** 周期性条形特征的所有傅里叶项在每个峰和谷处都是同相的。 |
| \ begin {figure} \ par \ centerline {\ psfig {figure = tri.ps，angle = -90，height = 2in，width = 4in}} \ par \ end {figure} |

|  |
| --- |
| **图3：** 周期梯形波形的所有傅里叶项在切线不连续点处具有最大一致性，其中可以看到马赫数。 |
| \ begin {figure} \ par \ centerline {\ psfig {figure = trap.ps，angle = -90，height = 2in，width = 4in}} \ par \ end {figure} |

如果图像信号是三角波形，表示相切的不连续性，并且通常在任何信号中存在局部最大一致性或相位值中的阶数的点恰好是人类感知特征的点，则会出现类似的相位一致性。[6](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#mandb) ]。也就是说，如果要求人类绘制图像的草图，精确定位场景中看到的感兴趣的边缘或标记，则所选择的点将是频率的相位分量中存在最大次序的点 - 基于信号的表示。

我们需要精确地确定相位一致性的含义。这是通过定义的完成*相位一致性功能*，*PC*（*X*），在每个点*X*在信号。我们有

\ begin {displaymath} PC（x）= \ max _ {\ theta \ in [0,2 \ pi}} \ frac {\ int a _ {\ omega} \ c ... ... x + \ phi _ {\ omega }  -  \ theta）d \ omega} {\ int a _ {\ omega} d \ omega}。 \ {端} displaymath

为了理解这个定义，我们可以想象我们的信号*f*（*x*），在图像中的任何点*x*，由不同幅度和相位角的各种正弦波的总和组成，我们在矢量图上绘制。

|  |
| --- |
| **图4：** 构成信号的各个向量。 |
| \ begin {figure} \ par \ centerline {\ psfig {figure = figure61.ps}} \ par \ end {figure} |

如果我们找到平均相位角然后计算所有这些相位角的标准偏差，我们将得到一个相位一致性的度量，但不是一个好的，因为这样的度量给出了355 *o*与1 *o*的大偏差。，实际上它很小。为了克服这个问题，相位一致性功能*PC* 需要被如上，其中该余弦项捕获355的接近度定义*Ò* 1 *Ò*。的$ \ $ THETA最大化该表达式为*PC*代表的加权平均相位角和自由泰勒定理，我们有$ \ cos（x）\约1  -  \ frac {x ^ 2} {2} $对于小*X*，我们看到，*PC*是的度量*方差* 是信号的相位值。当*PC*等于1时，相位项都是相等的，就像阶梯函数中的不连续情况一样。否则，*PC*会在0和1之间取值。

尽管*PC*的定义准确地捕获了我们想要测量的内容，但实现起来却是一个尴尬的功能。幸运的是，一些简单的三角函数操作足以证明*PC*与生物视觉中的众所周知的计算成比例，即信号中的*局部能量*。

信号的局部能量根据信号及其*希尔伯特变换来定义*。希尔伯特变换在空间域中具有稍微复杂的定义，即

\ begin {displaymath} h（x）= \ int _ { -  \ infty} ^ {\ infty} \ frac {f（y）} {xy} dy。 \ {端} displaymath

但这对应于频域中的简单相移。具体地，正频率术语相移90 *Ó*和负频率的条件是相-90移*ö* ; 零频率或直流项设置为零，因为相移在此没有意义。因此，在频域中，

\ begin {displaymath} {\ cal F}（h）（\ omega）= i \ mbox {sgn}（\ omega）{\ cal F}（f）（\ omega），\ end {displaymath}

其中sgn是返回其参数符号的函数。因此，如果

\ begin {displaymath} f（x）= \ int _ { -  \ infty} ^ {\ infty} a _ {\ omega} \ cos（\ omega x + \ phi _ {\ omega}）d \ omega，\ end {displaymath}

然后

\ begin {displaymath} h（x）=  -  \ int _ { -  \ infty} ^ {\ infty} a _ {\ omega} \ sin（\ omega x + \ phi _ {\ omega}）d \ omega。 \ {端} displaymath

我们可以定义一个向量**Ë**在点*X*通过

\ begin {displaymath} {\ bf E}（x）= f（x）{\ hat {\ bf i}} + h（x）{\ hat {\ bf j}}，\ end {displaymath}

其中$ {\ hat {\ bf i}} $和$ {\ hat {\ bf j}} $是沿*x*和*y*轴的单位向量。在信号中的任何特定点*x处*，**E**是傅里叶项的矢量和（或积分）。第*n*个分量是长度为*a n*的矢量，*其*$ nx + \ phi_n $与*x*轴成一角度。

矢量**E的**大小称为 信号的*局部能量*（有时也称为信号的*包络*），它被定义为

\ begin {displaymath} \ Vert {\ bf E} \ Vert = \ sqrt {f（x）^ 2 + h（x）^ 2}。 \ {端} displaymath

局部能量函数中的局部峰值对应于相位一致性函数中的局部峰值。

争论

\ begin {displaymath} Arg（x）= \ mbox {atan2}（\ frac {h（x）} {f（x）}）\ end {displaymath}

给出相位一致性发生的角度，并可用于定义特征类型。

如前所述，人类视觉系统具有通过正交的奇数和偶数对称滤波器来模拟卷积的能力。该过滤器是正交不仅意味着它们形成奇数和偶数对称对（即，由一个过滤器的输出卷积的是一个90 *ø*其它的输出的相移），也表明它们都具有一个零均值和相同的平方和值。更具体地说，如果*M e*代表偶数滤波器而 *M o*代表奇数滤波器，那么

\ begin {displaymath} \ int M_e（x）dx = \ int M_o（x）dx = 0，\ end {displaymath}

和

\ begin {displaymath} \ int M_e ^ 2（x）dx = \ int M_o ^ 2（x）dx。 \ {端} displaymath

此外，人类视觉系统具有计算卷积输出与奇数和偶数对称滤波器的平方和的能力，即，通过组合简单和复杂细胞响应的输出，它计算*局部能量*。信号。通过定义使其更加精确

\开始{displaymath} E（X）= \ Vert的{\ BFë} \韦尔= \ SQRT {{（M_E * F（X）}）^ 2 + {（M_o * F（X））} ^ 2}。 \ {端} displaymath

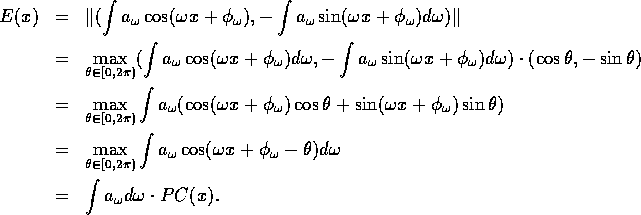
过滤器*M e*和*M o*需要仔细设计。选择均匀对称滤波器，使其覆盖尽可能多的频谱，同时消除直流项。奇对称滤波器是，不只是一个90 *Ô*的相移滤波器甚至过滤。从而

\开始{displaymath} M_E * F（X）\约\ INT _ { -  \ infty} ^ {\ infty}一个_ {\欧米加} \ COS（\欧米加X + \披_ {\欧米加}）d \欧米加\端{ displaymath}

和

\开始{displaymath} M_o * F（X）\约 -  \ INT _ { -  \ infty} ^ {\ infty}一个_ {\欧米加} \罪（\欧米加X + \披_ {\欧米加}）d \欧米加。 \ {端} displaymath

现在是简单明白为什么$ PC \ propto E $，对于



因此，为了在相位一致性函数中搜索局部最大值，等效地搜索局部能量函数中的局部最大值。这些局部最大值将出现在奇偶校验（向上或向下），线条和条形边缘的步进边缘，以及其他类型的特征，例如虚幻的马赫带[ [8](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#nature) ]。

为了说明这是如何工作的，图5显示了一个简单的测试图像，其中包含不同对比的各种功能。图6显示了一个简单的基于梯度的边缘检测器（此处为Sobel算子）的输出。请注意，输出取决于边缘的相对对比度，并且线要素的输出是*两条*边，一条线位于线的两侧。图7显示了本地能量（或相位一致性）检测器的输出。在这里，我们注意到输出是一个统一的响应，无论所涉及的功能的类型或对比如何。

|  |
| --- |
| **图5：** 一个简单的测试图像。 |
| \ begin {figure} \ par \ centerline {\ psfig {figure = test.ps}} \ par \ end {figure} |

|  |
| --- |
| **图6：** Sobel运算符的输出。 |
| \ begin {figure} \ par \ centerline {\ psfig {figure = sobel.ps}} \ par \ end {figure} |

|  |
| --- |
| **图7：** 测试图像的相位一致性图。 |
| \ begin {figure} \ par \ centerline {\ psfig {figure = phase.ps}} \ par \ end {figure} |

局部能量不会对它试图检测的特征的形状做出任何假设，因为它只是寻找局部最大相位一致性的点。注意，所有其他空间边缘检测方法必须对要检测的特征的形状进行假设。Perona和Malik [ [10](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#p&m) ]已经注意到，真实图像中的大多数特征由阶梯，屋顶和坡道剖面的组合组成，并且没有*线性* 特征检测方案可以检测到这种组合。然而，他们证明了二次方案，例如局部能量方案，足以检测所有这些特征。

图8和图9示出了局部能量图和用于心轴图像的检测特征。

|  |
| --- |
| **图8：** mandrill图像的相位一致性图。 |
| \ begin {figure} \ par \ centerline {\ psfig {figure = mandrillpcmap.ps}} \ par \ end {figure} |

|  |
| --- |
| **图9：** 非最大值抑制和阈值处理后的特征。 |
| \ begin {figure} \ par \ centerline {\ psfig {figure = mandrillpcedge.ps}} \ par \ end {figure} |

原文：

**Feature detection via phase congruency**

Rather than assuming that what an image should be compressed into is a set of edges, the phase congruency model of feature detection assumes that the compressed image format should be high in *information* (or *entropy*), and low in redundancy. Thus, instead of searching for points where there are sharp changes in intensity, this model searches for patterns of *order* in the phase component of the Fourier transform. Phase is chosen because the experiments of Oppenheim and Lim [[9](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#opp)] demonstrated that it is crucial to the perception of visual features. Further physiological evidence [[6](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#mandb)] indicates that the human visual system responds strongly to points in an image where the phase information is highly ordered. Thus the phase congruency model defines *features* as points in an image with high phase order.

The phase congruency model is a *frequency-based* model of visual processing. It supposes that, instead of processing visual data spatially, the visual system is capable of performing calculations using the phase and amplitude of the individual frequency components in a signal. Thus, the underlying computational tool is the Fourier transform, or one of its equivalents. To this end, let us suppose that we can represent our image signal in the Fourier domain. To simplify the presentation, we will assume a simple one-dimensional signal, representing a 1D slice through an image. Such a signal, say *f*(*x*), is reconstructed from its Fourier transform by

\begin{displaymath}
f(x) = \int_{- \infty}^{\infty} a_{\omega} \cos(T \omega x + \phi_{\omega})d \omega, \end{displaymath}

where, for each frequency $\omega$, $a_{\omega}$ is the amplitude of the cosine wave and $T \omega x + \phi_{\omega}$ is the phase offset of that wave. The term *T* is related to the size of the image window, and from now on we will assume it is 1.

For example, if the image signal were a simple step edge, then

\begin{displaymath}
f(x) = \frac{-4}{\pi} \int_{- \infty}^{\infty} \frac{1}{2\omega + 1} \cos(\omega x + \pi/2)d \omega, \end{displaymath}

and, at the point of the step edge (*x* = 0), all the phase terms are aligned at $\pi/2$. This is the only place in the signal where there is *congruency* in the phase values; at all other points, the phase values of individual frequency components assume differing values between 0 and $2 \pi$.

|  |
| --- |
| **Figure 1:** All the Fourier terms for a periodic step edge are in phase at each step point. |
| \begin{figure} \par \centerline{ \psfig {figure=step.ps,angle=-90,height=2in,width=4in} } \par\end{figure} |

|  |
| --- |
| **Figure 2:** All the Fourier terms for a periodic bar feature are in phase at each peak and trough. |
| \begin{figure} \par \centerline{ \psfig {figure=tri.ps,angle=-90,height=2in,width=4in} } \par\end{figure} |

|  |
| --- |
| **Figure 3:** All the Fourier terms for a periodic trapezoid wave form have maximal congruency at the point of tangent discontinuity where the Mach bands are seen. |
| \begin{figure} \par \centerline{ \psfig {figure=trap.ps,angle=-90,height=2in,width=4in} } \par\end{figure} |

A similar congruency of phase values occurs if the image signal is a triangular waveform, representing a tangent discontinuity and, in general, points in any signal where there is local maximal congruency or order in the phase values are precisely those points where humans perceive features [[6](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#mandb)]. That is, if a human were asked to draw a sketch of the image, localising precisely the edges or markings of interest as seen in the scene, then the points chosen would be those were there is maximal order in the phase components of a frequency-based representation of the signal.

We need to make precise what it is we mean by phase congruency. This is done by defining the *phase congruency function*, *PC*(*x*), at each point *x* in the signal. We have

\begin{displaymath}
PC(x) = \max_{\theta \in [0, 2\pi)} \frac{\int a_{\omega} \c...
 ...x + \phi_{\omega} - \theta)d \omega}{\int a_{\omega}d \omega}. \end{displaymath}

To understand this definition, we can think of our signal *f*(*x*), at any point *x* in the image, as being made up of the sum of various sine waves at different amplitudes and phase angles, which we plot on a vector map.

|  |
| --- |
| **Figure 4:** The individual vectors making up the signal. |
| \begin{figure} \par \centerline{ \psfig {figure=figure61.ps} } \par\end{figure} |

If we find the mean phase angle and then calculate the standard deviation of all these phase angles about this mean we will have a measure of phase congruency, but not a good one, since such a measure gives a large deviation of 355*o* from 1*o*, when in fact it is small. To overcome this problem, the phase congruency function *PC* needs to be defined as above, where the cosine term captures the proximity of 355*o* to 1*o*. The $\theta$ that maximizes this expression for *PC* represents the weighted mean phase angle and since, by Taylor's theorem, we have $\cos(x) \approx 1 - \frac{x^2}{2}$for small *x*, we see that *PC* is a measure of the *variance* is the phase values of the signal. When *PC* is equal to 1, the phase terms are all equal, as is the case at the discontinuity in a step function. Otherwise, *PC* takes on some value between 0 and 1.

Although the definition of *PC* captures precisely what it is we want to measure, it is an awkward function to implement. Luckily, some simple trigonometric manipulations suffice to prove that *PC* is proportional to a well-known computation in biological vision, namely the *local energy* in a signal.

The local energy of a signal is defined in terms of the signal and its *Hilbert transform*. The Hilbert transform has a somewhat complicated definition in the spatial domain, namely

\begin{displaymath}
h(x) = \int_{-\infty}^{\infty} \frac{f(y)}{x-y}dy. \end{displaymath}

But this corresponds to simple phase shifting in the frequency domain. Specifically, the positive frequency terms are phase shifted by 90*o* and the negative frequency terms are phase shifted by -90*o*; the zero frequency, or d.c. term, is set to zero, since phase shifting has no meaning here. Thus, in the frequency domain,

\begin{displaymath}
{\cal F}(h)(\omega) = i \mbox{sgn}(\omega){\cal F}(f)(\omega), \end{displaymath}

where sgn is the function that returns the sign of its argument. Thus, if

\begin{displaymath}
f(x) = \int_{- \infty}^{\infty} a_{\omega} \cos(\omega x + \phi_{\omega})d \omega, \end{displaymath}

then

\begin{displaymath}
h(x) = - \int_{- \infty}^{\infty} a_{\omega} \sin(\omega x + \phi_{\omega})d \omega. \end{displaymath}

We can define a vector **E** at a point *x* by

\begin{displaymath}
{\bf E}(x) = f(x){\hat {\bf i}} + h(x){\hat {\bf j}}, \end{displaymath}

where ${\hat {\bf i}}$ and ${\hat {\bf j}}$ are the unit vectors along the *x* and *y* axes. At any particular point *x* in the signal, **E** is the vector sum (or integral) of the Fourier terms. The *n*th component is a vector of length *an* making an angle of $nx + \phi_n$with the *x* axis.

The magnitude of the vector **E** is called the *local energy* of the signal (sometimes also called the *envelope* of the signal) and it is defined as

\begin{displaymath}
\Vert {\bf E} \Vert = \sqrt{f(x)^2 + h(x)^2}. \end{displaymath}

Local peaks in the local energy function correspond to local peaks in the phase congruency function.

The argument

\begin{displaymath}
Arg(x) = \mbox{atan2}
(\frac{h(x)}{f(x)}) \end{displaymath}

gives the angle at which the phase congruency occurs, and can be used to define the feature type.

As was mentioned earlier, the human visual system has the capacity to simulate convolution by odd and even symmetric filters in quadrature. That the filters are in quadrature means not only that they form an odd and even symmetric pair (that is, the output of convolution by one filter is a 90*o* phase shift of the output of the other), but also that they both have a zero mean value and the same sum-of-squares value. More specifically, if *Me* represents the even filter and *Mo* represents the odd filter, then

\begin{displaymath}
\int M_e(x)dx = \int M_o(x)dx = 0, \end{displaymath}

and

\begin{displaymath}
\int M_e^2(x)dx = \int M_o^2(x)dx. \end{displaymath}

Moreover, the human visual system has the capacity to compute the sum of squares of the output from convolution with the odd and even symmetric filters, that is, by combining the output of the simple and complex cell responses, it computes a *local energy* for the signal. This is made more precise by defining

\begin{displaymath}
E(x) = \Vert {\bf E} \Vert = \sqrt{{(M_e * f(x)})^2 + {(M_o * f(x))}^2}.\end{displaymath}

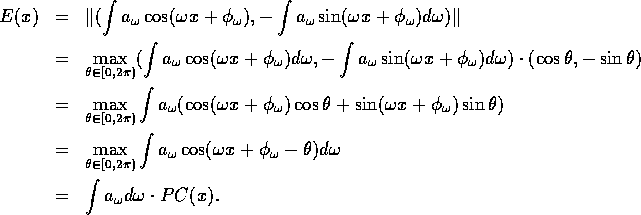
The filters *Me* and *Mo* need to be carefully designed. The even symmetric filter is chosen so that it covers as much of the frequency spectrum as possible, whilst eliminating the d.c. term. The odd symmetric filter is then just a 90*o* phase shift filter of the even filter. Thus

\begin{displaymath}
M_e * f(x) \approx \int_{- \infty}^{\infty} a_{\omega} \cos(\omega x + \phi_{\omega})d \omega \end{displaymath}

and

\begin{displaymath}
M_o * f(x) \approx - \int_{- \infty}^{\infty} a_{\omega} \sin(\omega x + \phi_{\omega})d \omega. \end{displaymath}

It is now simple to see why $PC \propto E$, for



So, in order to search for local maxima in the phase congruency function, one equivalently searches for local maxima in the local energy function. These local maxima will occur at step edges of either parity (up or down), lines and bar edges, and other types of features such as the illusory Mach bands [[8](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#nature)].

To illustrate how this works, figure 5 shows a simple test image that contains a variety of features at different contrasts. Figure 6 shows the output of a simple gradient-based edge detector (here, the Sobel operator). Note that the output depends on the relative contrast of the edge, and that the output for line features is *two* edges, one on either side of the line. Figure 7 shows the output of the local energy (or phase congruency) detector. Here we note that the output is a uniform response, regardless of the type or contrasts of the feature involved.

|  |
| --- |
| **Figure 5:** A simple test image. |
| \begin{figure} \par \centerline{ \psfig {figure=test.ps} } \par\end{figure} |

|  |
| --- |
| **Figure 6:** The output of the Sobel operator. |
| \begin{figure} \par \centerline{ \psfig {figure=sobel.ps} } \par\end{figure} |

|  |
| --- |
| **Figure 7:** The phase congruency map of the test image. |
| \begin{figure} \par \centerline{ \psfig {figure=phase.ps} } \par\end{figure} |

Local energy makes no assumptions about the shape of the features it is trying to detect, as it is only looking for points of local maximum phase congruency. Note that all other spatial edge detection methods have to make assumptions about the shape of the feature to be detected. Perona and Malik [[10](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT7/node6.html#p&m)] have noted that most features in real images are composed of combinations of steps, roofs and ramps profiles, and that no *linear* feature detection scheme can detect such combinations. However, they prove that a quadratic scheme, such as the local energy scheme, is sufficient for the detection of all such features.

Figures 8 and 9 illustrate the local energy map and the detected features for the mandrill image.

|  |
| --- |
| **Figure 8:** The phase congruency map of the mandrill image. |
| \begin{figure} \par \centerline{ \psfig {figure=mandrillpcmap.ps} } \par\end{figure} |

|  |
| --- |
| **Figure 9:** The features after non-maxima suppression and thresholding. |
| \begin{figure} \par \centerline{ \psfig {figure=mandrillpcedge.ps} } \par\end{figure} |