SYSTEM AND METHODS OF AMPLITUDE-MODULATION FREQUENCY-MODULATION (AM-FM) DEMODULATION FOR IMAGE AND VIDEO PROCESSING

Inventors: Victor Manuel Murray Herrera, Albuquerque, NM (US); Marios Stephanou Pattichis, Albuquerque, NM (US); Peter Soliz, Albuquerque, NM (US); Carla Paola Aguro Rios, Albuquerque, NM (US); Herbert T. Davis, III, Corrales, NM (US)

Assignee: STC, UNM, Albuquerque, NM (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1006 days.

Related U.S. Application Data
Provisional application No. 61/098,165, filed on Sep. 18, 2008.

Int. Cl.
G06K 9/40 (2006.01)

U.S. Cl.
USPC 382/260; 382/100

Field of Classification Search
USPC 382/100, 260

See application file for complete search history.

References Cited
U.S. PATENT DOCUMENTS

Other Publications

Abstract
Image and video processing using multi-scale amplitude-modulation frequency-modulation ("AM-FM") demodulation where a multi-scale filterbank with bandpass filters that correspond to each scale are used to calculate estimates for instantaneous amplitude, instantaneous phase, and instantaneous frequency. The image and video are reconstructed using the instantaneous amplitude and instantaneous frequency estimates and variable-spacing local linear phase and multi-scale least square reconstruction techniques. AM-FM demodulation is applicable in imaging modalities such as electron microscopy, spectral and hyperspectral devices, ultrasound, magnetic resonance imaging ("MRI"), positron emission tomography ("PET"), histology, color and monochrome images, molecular imaging, radiographs ("X-rays"), computer tomography ("CT"), and others. Specific applications include fingerprint identification, detection and diagnosis of retinal disease, malignant cancer tumors, cardiac image segmentation, atherosclerosis characterization, brain function, histopathology specimen classification, characterization of anatomical structure such as carotid artery walls and plaques or cardiac motion and as the basis for computer-aided diagnosis to name a few.

18 Claims, 6 Drawing Sheets
References Cited

OTHER PUBLICATIONS


FIG. 2
\[ I_0 \sim k_2 k_n \]

\[ I_{AS_{filtered}}(k_1, k_2) \xrightarrow{\text{abs}} \hat{a}(k_1, k_2) \]
cally illustrating estimation of a single AM-FM component from a single scale. For purposes of this application, the term “scale” is defined as a collection of bandpass filters. As an example, scales can be categorized into low, medium, and high based on frequency magnitude. As shown in FIG. 1, an input image \( I(x_t, y_t) \) is provided at 102. It is also contemplated that an input video may be provided as described further below. A two-dimensional extended analytic signal \( I_{e}(x_t, y_t) \) of the input image is computed at 104 by applying the Hilbert transform to form a 2D extension of the 1D analytic signal. For either an input image or input video, a multi-scale filterbank is selected at 106. A multi-scale filterbank is selected by defining the bandpass filters at 108 that correspond to each scale as described in further detail with respect to FIG. 2. The extended analytic signal is processed through a multi-scale filterbank as shown by 110, 112. Estimates for instantaneous amplitude, instantaneous phase, and instantaneous frequency are calculated at 114. The dominant AM-FM component is selected at 116 using the maximum amplitude at every pixel. Thus, the estimates with the maximum response are selected. Instantaneous amplitude, instantaneous phase and instantaneous frequency estimates are obtained at 118. These estimates are obtained by the QEA method described above. Instantaneous amplitude and instantaneous frequency estimates are used to reconstruct the image such that it can be displayed on a display unit.

The processing for a discrete video is similar to that described above. A discrete video \( I(x_t, y_t, z_t) \) is provided. A Hilbert transform is applied to form a 3D extension \( I_{e}(x_t, y_t, z_t) \) of the 1D analytic signal. A multi-scale filterbank is selected by defining the bandpass filters that correspond to each scale as described in further detail with respect to FIG. 3. The extended analytic signal is processed through the multi-scale filterbank and the instantaneous frequencies are estimated in all \( x, y \), and \( t \) directions. For each pixel, the AM-FM demodulation estimates are selected from the processing block that gives the largest instantaneous amplitude estimate. Hence, the algorithm adaptively selects the estimates from the bandpass filter with the maximum response. This approach does not assume spatial continuity, and allows the model to quickly adapt to singularities in the input signal. Instantaneous amplitude and instantaneous frequency estimates are used to reconstruct the video such that it can be displayed on a display unit.

Dominant component analysis is applied over each scale to produce a single AM-FM component from each scale. Estimates are adaptively selected from the bandpass filter with the maximum response. This approach does not assume spatial continuity and allows the model to quickly adapt to singularities in the image or video. Although any collection of filters for each scale is contemplated, in one embodiment the corresponding collection of filters for each scale is shown below by Table 1:

<table>
<thead>
<tr>
<th>Scale</th>
<th>Single-scale</th>
<th>Two-scale</th>
<th>Three-scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPF</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>High</td>
<td>2, 3, 4, 5, 6, 7</td>
<td>2, 3, 4, 5, 6, 7</td>
<td>2, 3, 4, 5, 6, 7</td>
</tr>
<tr>
<td>Medium frequencies</td>
<td>NA*</td>
<td>8, 9, 10, 11, 12, 13</td>
<td>8, 9, 10, 11, 12, 13</td>
</tr>
<tr>
<td>Low frequencies</td>
<td>NA*</td>
<td>14, 15, 16, 17, 18, 19</td>
<td></td>
</tr>
</tbody>
</table>

Multi-scale filterbanks covering the whole frequency spectrum are used for images as shown in FIG. 2. For real-valued bandpass filters, each separable channel filter has support over four quadrants. In order to maintain support over only two quadrants, Fast Fourier Transform (“FFT”) pre-filtering is used to remove support in two quadrants such as the two left or two right quadrants. Thus, each bandpass filter has frequency support in only two quadrants of the frequency spectrum so that, in effect, each channel filter operates over a single quadrant. The filters were designed using a min-max, equiripple approach. In one embodiment, passband ripple is set at 0.017 dB and the stopband attenuation is set to 66.02 dB.

The design of an efficient filterbank to be used in 3D methods and applications is shown in FIG. 3 and FIG. 4. FIG. 3 shows the extended 2D filterbank design to generate a 3D multi-scale filterbank illustrating the numbering of bandpass filters for the time variable for 3D single-scale, 3D two-scale, and 3D three-scale filterbanks. The same equiripple design, with the same specifications, is used. The third dimension, time, will increase the total number of bandpass filters. The bandpass filters always have frequency support in only half of the spectrum. For the filters through time, the following notation is used: filter 1 is the low pass filter (“LPF”), filters 2 and 3 are high frequency filters, filters 4 and 5 are medium frequency filters, and filters 6 and 7 are low frequency filters.

As an example, in embodiments where 2D applications have designed filterbanks with 7, 13 and 19 bandpass filters for a single-, two- and three-scale filterbank, respectively, the 3D filterbanks have 21, 65 and 133 3D bandpass filters, respectively.

FIG. 4(a) illustrates frequency spectrum decomposition for a 3D two-scale filterbank, but with the added frequencies associated with the time variable. FIG. 4(b) shows the 3D frequency-domain decomposition for a 3D two-scale filterbank.

For the transition bandwidth, it is required that it remains lower than the passband bandwidth and also that it remains sufficiently large so that the passband and stopband requirements can be met with a reasonable number of digital filtering coefficients. Transition widths are relatively less important for the high frequencies since they also come with filters of larger passband bandwidths. In contrast, low-frequencies require relatively larger transition widths since images contain larger, low-frequency components and the transitions occur over smaller bandwidths. Transition bandwidths are fixed to \( \pi/10 \). Based on this approach, the unit gain over the passband eliminated the need for amplitude correction as required by a Gabor filterbank approach.

In order to develop robust multi-channel techniques, it must be ensured that no intermediate computation step results in the loss of significance, particularly avoiding excessive relative error.

The relative error is defined in the approximation as given by \( b_{\text{error}} = |x_{\text{approx}} - 1|/|x_{\text{true}}| \) where the number \( x_{\text{approx}} \) is used as an approximation of the true value \( x_{\text{true}} \). For one-dimensional functions, the notion of relative error to relative perturbation is generalized as \((f(x+h) - f(x))/f(x)\), which measures the relative change in evaluating \( f \) at \( x \), for relative change in \( x \) given by \( h/x \). A large relative perturbation for small relative change in \( x \) implies instability. A well-known estimate of relative perturbation is based on the condition number of the function as summarized below.

Starting from the Taylor theorem with remainder \( f(x+h) = \frac{1}{h!} f^{(n)}(\alpha) \) for \( f(x+h) = f(x)R(x) \), where \( |R(x)| \leq \max_{x \in [x+a,h]} |f^{(n)}(t)|h^n/2 \), the relative perturbation is given by.

TABLE 1

<table>
<thead>
<tr>
<th>Scale</th>
<th>Single-scale</th>
<th>Two-scale</th>
<th>Three-scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPF</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>High</td>
<td>2, 3, 4, 5, 6, 7</td>
<td>2, 3, 4, 5, 6, 7</td>
<td>2, 3, 4, 5, 6, 7</td>
</tr>
<tr>
<td>Medium frequencies</td>
<td>NA*</td>
<td>8, 9, 10, 11, 12, 13</td>
<td>8, 9, 10, 11, 12, 13</td>
</tr>
<tr>
<td>Low frequencies</td>
<td>NA*</td>
<td>14, 15, 16, 17, 18, 19</td>
<td></td>
</tr>
</tbody>
</table>
The error in the relative error approximation is bounded by \( R(x) = f'(x) \max \left( \frac{|f'(x)|}{f(x)} \right) \), noting that the relative perturbation in \( x \) (given by \( h(x) \)) is amplified by the condition number given by:

\[
\text{cond } f(x) = \frac{|x| \cdot f'(x)}{|f(x)|}
\]

More generally, in evaluating functions, the condition numbers of all intermediate functions used in the computation need to be considered. The computation is unstable and prone to error if any one intermediate function has a large condition number. For robust methods, it is sought to develop numerically stable algorithms that are characterized by the smallest possible condition numbers.

In the development of AM-FM demodulation methods, three inverse trigonometric functions are encountered: arcsin, arccos, and arctan with their condition numbers given by:

\[
\text{cond } \arccos(x) = \frac{|x|}{\arccos(x) \sqrt{1 - x^2}}
\]

\[
\text{cond } \arcsin(x) = \frac{|x|}{\arcsin(x) \sqrt{1 + x^2}}
\]

\[
\text{cond } \arctan(x) = \frac{|x|}{\arctan(x) (1 + x^2)}
\]

Again, the approximation error of the relative error is bounded above by \( h^2/(2f(x)) \max \left( \frac{|f'(x)|}{f(x)} \right) \). Therefore the error for the arccos is bounded above by:

\[
\frac{h^2}{2 \arccos(x) \max \left( \frac{|f'(x)|}{f(x)} \right)} \frac{1}{(1 - 2x^2)^3}
\]

Similarly, for the arccos and arctan, the errors are bounded above by:

\[
\frac{h^2}{2 \arcsin(x) \max \left( \frac{|f'(x)|}{f(x)} \right)} \frac{1}{(1 - 2x^2)^3}
\]

and

\[
\frac{h^2}{2 \arctan(x) \max \left( \frac{|f'(x)|}{f(x)} \right)} \frac{1}{(1 - 2x^2)^3}
\]

For the arccos, it is interesting to note that the approximation error remains low around \( x = 0 \).

The condition numbers become infinite for both the arccos and the arcsin for \( x = \pm 1 \) as shown by Figs. 5. Otherwise, finite condition numbers exist for all values of \( x \), including the cases when \( \arccos(x) = 0 \), \( \arcsin(x) = 0 \), or \( \arctan(x) = 0 \).

As shown by Fig. 5, evaluating the arccos is more stable for \( |x| < 0.6 \) than evaluating the arcsin. More specifically, the absolute value of the condition number for the arcsin and the absolute value of the condition number for the arccos is shown. Furthermore, the most stable evaluations occur for small values of \( |x| \), when using the arccos function. Absolute value of the condition number of the arcs (solid line).

Even though the arcsin function provides the most stable evaluations, the arcsin function must also be used to estimate the proper instantaneous frequency quadrant, although, there are also significant problems in accurate estimation at very high frequencies.

According to the present invention, a robust instantaneous frequency estimation is termed herein as Variable Spacing, Linear Local Phase (VS-LLP) described below.

Due to the unit gain over the passbands, a plug-in rule is used to produce a new signal \( \hat{f}_s \) with unit instantaneous amplitude using:

\[
\hat{f}_s(k_1, k_2) = \frac{\hat{I}_s(k_1, k_2)}{\tilde{p}_s(k_1, k_2)} = \frac{n(k_1, k_2) \exp(j\Phi(k_1, k_2))}{m(k_1, k_2)} = \exp(j\Phi(k_1, k_2)).
\]

A linear approximation is considered for the estimated phase \( \Phi(p_1, p_2) \) of order \( \|p_1 - k_1, k_2 - k_2\| \), given by \( \Phi(p_1, p_2) = \Phi(k_1, k_2) + \Phi(k_1, k_2) \| p_1 - k_1, p_2 - k_2 \| \), to get:

\[
\hat{I}_s(k_1 + n_1, k_2) + \hat{I}_s(k_1 - n_1, k_2) = \frac{2\hat{I}_s(k_1, k_2)}{2\hat{I}_s(k_1, k_2)} = \frac{\exp(j\Phi(k_1 + n_1, k_2)) + \exp(-j\Phi(k_1 - n_1, k_2))}{2\exp(j\Phi(k_1, k_2))}
\]

This gives an arcsine expression for estimating the instantaneous frequency component using:

\[
\frac{\partial \hat{\Phi}}{\partial x}(k_1, k_2) = \frac{1}{n_1} \arccos \left( \frac{\hat{I}_s(k_1 + n_1, k_2) + \hat{I}_s(k_1 - n_1, k_2)}{2\hat{I}_s(k_1, k_2)} \right)
\]

The analysis for \( \nabla \hat{\Phi} \) is similar.

For low instantaneous frequency magnitude, it is clear that the local linear phase approximation will hold over a larger range of \( n_1, n_2 \). For larger instantaneous frequency magnitude, the phase must be modulated down to lower frequencies as described below.

The arccosine function is evaluated at different possible values for \( n_1 \). For stable function evaluations, the argument to the arccosine function is considered:

\[
\gamma_{\arccos}(n_1) = \frac{\hat{I}_s(k_1 + n_1, k_2) + \hat{I}_s(k_1 - n_1, k_2)}{2\hat{I}_s(k_1, k_2)}
\]

Integer values for \( n_1 \) so as to have \( \gamma_{\arccos}(n_1) \) as close to zero as possible are considered (see FIG. 5). Four possible values: \( n_1 = 1, 2, 3 \) and 4 are considered—increasing the value of \( n_1 \) could lead to an unstable zone. Additionally, only integer values are considered since non-integer values require image interpolation and possible additional errors due to the additional interpolation step.

To establish the limits of this approach, in order for the argument of \( \gamma_{\arccos}(n_1) \) to be zero, it is required that:

\[
\hat{I}_s(k_1 - n_1, k_2) - \hat{I}_s(k_1 + n_1, k_2) = 0.
\]
where \( s \) is the number of scales used, \( d \) is a global DC image estimate, \( G(k_x,k_y) \) is the low-pass filter output, \( a_1 \cos \phi_1 \) is the high-frequency scale AM-FM component, \( a_2 \cos \phi_2 \) is the medium-frequency scale AM-FM component and \( a_3 \cos \phi_3 \) is the low-frequency scale AM-FM component. The AM-FM multi-scale coefficients \( c_n, n=0, 1, \ldots, s \) are computed so that \( \hat{l}(k_x,k_y) \) is a least-squares estimate of \( l(k_x,k_y) \).

Video is reconstructed using its AM-FM components by extending the application presented for 2D signals to get 3D versions of the three methods discussed above. The 3D multi-scale filterbank is used to extend the 2D reconstruction to reconstruct videos using 3D Least-Squares Reconstructions using AM-FM harmonics (3D-LESHA), 3D multi-scale least-squares reconstructions (3D-MULTILES) and 3D Least-Squares Reconstructions using AM-FM harmonics and the DCA (3D-LESCA). 3D-MULTILES will be discussed since the application of 3D-LESCA and 3D-LESHA are similar to the 2D methods.

3D-MULTILES is based on the scales of the filterbanks designed by defining \( d \) as the global DC image estimate, \( G(k_x,k_y) \) as the low pass filter output, \( a_1 \cos \phi_1 \) as the high-frequency scale AM-FM component, \( a_2 \cos \phi_2 \) as the medium frequency scale AM-FM component, and \( a_3 \cos \phi_3 \) as the low-frequency scale AM-FM component. In this case, least squares reconstructions is given by:

\[
\hat{l}(k_x, k_y, k_z) = d + \sum_{n=1}^{s} c_n G(k_x, k_y, k_z) \cos(\phi_n(k_x, k_y, k_z)),
\]

where \( s \) is the number of scales used.

The AM-FM multi-scale coefficients \( c_n, n=0, 1, \ldots, s \) are computed, so that \( \hat{l}(k_x,k_y,k_z) \) is a least-squares estimate of \( l(k_x,k_y,k_z) \). An orthonormal basis is also computed over the space of the AM-FM estimations scale by scale.

It is important to note that adding decomposition levels also reduces the total amount of video signal energy that is captured by the decomposition. For a single scale decomposition, video signal energy is captured by the low-pass filter component and the dominant high-frequency components, selected from the high frequency 3D bandpass filters. Then, in two-scale decompositions, the 3D spectrum captured by the low-pass filter is further decomposed into two new scales. The dominant components are found in this second scale while the lowest frequency components are captured by the new low-pass filters. Similarly, for three-scales, decompose the frequency spectrum of the 3D low-pass filter is decomposed.

The extracted dominant components from each scale provide decompositions using an independent AM-FM component per scale. Furthermore, the corresponding dominant channel filters allow the extraction of local spatiotemporal content over each pixel. This approach allows the re-formulation of the classical motion estimation problem with several independent equations over each scale. It is also important to note that the AM-FM decomposition also track both continuous and discontinuous motions since at each pixel three different dominant channels from three different scales can be associated.

For the 3D-LESHA reconstruction method, reconstructing the input video using AM-FM harmonics is considered:

\[
\hat{l}(k_x, k_y, k_z) = d + \sum_{n=1}^{s} c_n G(k_x, k_y, k_z) \cos(\phi_n(k_x, k_y, k_z)).
\]

where \( d \) is a scalar and \( h \) is the maximum number of AM-FM harmonics to use. For 3D-LESCA, \( G(k_x,k_y,k_z) \) (the LPF output) is also used. Thus, least squares video reconstructions is considered using:

\[
\hat{l}(k_x, k_y, k_z) = d + \sum_{n=1}^{s} c_n G(k_x, k_y, k_z) \cos(\phi_n(k_x, k_y, k_z)).
\]

VS- LLP provides significant improvement in instantaneous frequency estimation, while the least-squares AM-FM decompositions can be used to reconstruct general images.

For both instantaneous amplitude and instantaneous frequency estimations, significant improvements are realized when using the multi-scale filterbanks. Instantaneous frequency estimation does suffer when the instantaneous frequency components are very low. This comes from the requirement that the AM and FM magnitude spectra should remain clearly separated.

For instantaneous frequency estimation, VS- LLP is consistently shown to improve estimation over all other methods for single-component signals, without the use of filterbanks. Similarly, for the same signal, significant improvements are obtained over regularized AM-FM demodulation. Dramatic improvements in instantaneous frequency estimation are seen when using VS- LLP with modulation to a lower frequency. Here, modulating to a lower frequency has the effect of "slowing-down" the signal, allowing the consideration of instantaneous frequency estimation algorithms with spacings of 1 to 4 pixels. In turn, this "slowing-down" helps the local linear phase model become more applicable. The use of post and pre-filtering with a 3x3 median filter also helps reduce noise. Again, the advantage of the median filter is that it removes noise without reducing the bandwidth (for single AM-FM components only).

It is also interesting to note the 70 dB improvement over QEA, when the Quadrature signal was provided to both VS- LLP and the QEA, which suggests that most of the error comes from estimating the Quadrature signal. When the Quadrature signal is provided, the VS- LLP is directly compared to the standard QEA, without accounting for any pre-processing.

To analyze the results for image reconstructions, it is important to note the significant role played by the channel filters in each scale. For example, for a single-scale, the spectral support for the LPF is the largest. Similarly, for three-scale reconstructions, the LPF support is the smallest. Thus, in considering the quality of the reconstructions, given the fact that most image energy is concentrated around the low-frequency components, the LPF provides a very significant contribution. This also applies to LESA reconstructions since the low-frequency components estimated over the LPF will dominate. For LESA, the use of AM-FM harmonics did not contribute much to the reconstruction. This result is attributed to the fact that low-frequency components, estimated over the LPF, dominate the reconstruction.

For LESA and MULTILES, AM-FM demodulation is not allowed over the LPF. This allows the measurement of the contribution of higher-frequency AM-FM components. It is clear that the use of AM-FM components extracted from
multiple scales provided for the most significant AM-FM contributions towards the reconstruction.

The success of the VS-IIP algorithm comes from the fact that it adaptively selects accurate instantaneous frequency estimates at every pixel. Here, the basic idea is to look at instantaneous frequency estimates coming from different spacings between the samples, and then select the estimate that also produces the lowest condition number estimate. It has been verified experimentally that the instantaneous frequency methods that produced lower condition number estimates have also produced more accurate demodulation results. Overall, this resulted from the fact that numerical stability of the instantaneous frequency estimate was considered a function of frequency. It thus makes sense to consider samples separated by larger distances for lower frequencies as opposed to using smaller distances for higher frequencies.

The accuracy of the multi-scale QEA amplitude estimates are attributed to the relatively flat spectral magnitude response of the designed digital filters. The use of non-flat magnitude response will clearly affect the AM-FM component spectrum dramatically. Apparently, it is far more efficient to use digital filters with flat responses, instead of attempting to fix for the non-flat response afterwards.

The use of a robust least-squares approach for combining AM-FM components is an important step that directly leads to accurate reconstructions of general images. In other words, the standard use of dominant component analysis does produce AM-FM components that are (generally) far from orthogonal. Use of a least-squares approach relaxes the assumption that the demodulated AM-FM components should be orthogonal.

In sum, the first step in developing the AM-FM methods according to the present invention is to design a new multi-scale filterbank. This design allows correct instantaneous frequency component sign estimation. The almost flat response in the bandpass frequency of the 1D filters eliminates errors due to the use of an amplitude correction as in the case of using Gabor filterbanks. The use of these filters in the AM-FM demodulation problem produces big improvements in the instantaneous amplitude estimations and instantaneous frequency estimations. For noisy signals, VS-IIP produces better results for robust instantaneous frequency estimation than previous methods such as QEA or QLM.

Images can be reconstructed based on their instantaneous amplitude and instantaneous phase information. The least squares methods MULTILES, LESSHA and LESCA produced good image reconstructions and MULTILES showed that AM-FM components from different scales of the frequency spectrum contain important information that can be used to improve image quality in the reconstruction. Overall, the present invention provides for accurate reconstructions of general images and the extraction of AM-FM component parameters. Clearly, all prior applications that were based on AM-FM demodulation can benefit from using the new filterbanks presented.

As for 3D, the first step was to design new 3D multi-scale filterbanks with support in four octants of the frequency spectrum. Similar to the 2D case, the flat response in the bandpass frequency of the 1D filters eliminates errors due to the use of an amplitude correction as in the case of using Gabor filterbanks. Basically, the 2D AM-FM formulation and theory is extended to 3D.

Video reconstructions can be obtained using AM-FM from multiple-scales in three different forms based on the 2D methods. These AM-FM methods can be used for motion estimation, allowing the estimation of pixel motion with up to three equations per pixel per scale (AM, FM, and continuity equations). The new filterbanks cover the entire frequency spectrum providing dense motion estimates.

One application of the present invention will now be discussed with respect to retinal image analysis using AM-FM methods. For retinal image analysis, a four-scale filterbank was designed as shown by FIG. 8. In FIG. 8, filter I corresponds to a low pass filter ("LPF") with frequency support [−π/16, π/16] for both x and y directions. For all the other filters, the bigger the label number of the filter, the lower the frequency support that it has. The filters in the highest frequency, such as filters from 2 to 7, have a bandwidth of π/2 for both x and y directions. The bandwidth is decreased by a factor of 0.5 for each added scale. FIG. 8(a) illustrates a complete frequency spectrum of the filterbank and FIG. 8(b) illustrates the zoom on the low frequency bandpass filters.

AM-FM components are extracted from different scales. Table 3 describes the correspondence between the scales and bandpass filters according to one embodiment:

<table>
<thead>
<tr>
<th>TABLE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandpass filters used for estimating AM-FM in a four-scale filterbank</td>
</tr>
<tr>
<td>Scales</td>
</tr>
<tr>
<td>LPF</td>
</tr>
<tr>
<td>VL</td>
</tr>
<tr>
<td>L</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

As shown by Table 4 below, nine different cases of extracting dominant AM-FM component from different scales is considered for retinal applications:

<table>
<thead>
<tr>
<th>TABLE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scales used for the nine cases in retinal image analysis</td>
</tr>
<tr>
<td>Case #</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
</tbody>
</table>

From each case i, i=1, ..., 9, histograms of both the instantaneous amplitude \( a_i \) and the magnitude of the instantaneous frequency \( \| \hat{\Phi} \| \) of the image are features used. It should be noted that high-frequency bandpass filters (filters from 2 to 7 of FIG. 8) are not used for the analysis because the information contained therein is the high-frequency noise of the image.

The histograms of instantaneous amplitude and the magnitude of the instantaneous frequency, \( |\hat{f}| \), are used to create a feature vector for detection of diabetic retinopathy ("DR"). Using histograms at different scales (see Table 4) the information extracted with AM-FM can be analyzed to find differences among retinal images with DR and healthy images. A region containing micro-aneurysms, hemorrhages, and exudates will have different estimates for instantaneous amplitude than a region lacking these features. Using these histograms, it can be found if a certain frequency component that encodes a feature is present in the image.

Both histograms, of \( a_i \) and \( \| \hat{\Phi} \| \), for i=1, 2, ..., or 9, are computed using forty bins, leading to one histogram of eighty
For example, in the detection of Risk 1 in the data containing Risk 1 as the only level of retinopathy and Risk 0 (normals), cases 4 and 9 from Table 4 provided the best model for the detection of the Risk 1 patients. These cases appear to indicate that the spatial information, as encoded by the associated scales, serve to differentiate the two classes of images, i.e., normal versus Risk 0. Similarly, one can select appropriate combinations of scale from Table 4 for the detection of the Risk 2 and Risk 3. Likewise, detecting certain lesion types, such as Neovascularization/New Vessels of the Disc (“NV/ NVD”), will be performed using specific scales. Based on this factor-based analysis, the next step is to use the information to produce a matrix of independent variables that is parsimonious, well conditioned and robust. To accomplish this, the data for each case with 10 or fewer factors is used in a PLS model using the optimal number of factors as shown in the tables above. For the normal's versus Risk 1 retinopathy, for example, case 1 was fitted with a PLS model using 9 factors producing a T matrix of t-scores, T1. The case 2, which requires 16 factors, is not fitted since the number of factors is greater than 10. Finally, case 9 is fitted using 9 factors producing a t-score matrix T9. From this, a matrix of independent variables is constructed as:

\[ X = [T_1 T_2 T_3 T_4 T_5 T_9] \]

From a regression model, X is constructed from the T matrices. This model was fitted using PLS with 2 factors and jackknifed predictions of the diagnoses obtained. The resulting ROC curves and AUC's estimated from these predictions is discussed below.

The application of AM-FM feature extraction and PLS classification to the four categories of DR severity (0–none; 1=few microaneurysms (“MAs”); and 2=MA's and hemorrhages present, and 3=extensive MAs, hemorrhages, possible macular edema (“MEs”) and neovascularization). The sensitivity, specificity, and area under the ROC curve are given for both the testing processes described above. First, the ability to correctly detect those images is determined with signs of DR in a set of images composed of Risk 0 (normal N=92) and Risk 1 (N=71). A total of 163 images are selected from the total available (N=265). AM-FM features are calculated for the nine cases (see Table 4). The PLS-based classifier is tested using all combinations to determine the best model (as measured by AUC).

Table 5-continued

| Number of factors in normal versus risk (1,2, 3) retinopathy |
|-----------------|-----------------|-----------------|
| Case | # Factors | Case | # Factors | Case | # Factors |
| 4    | 18        | 4    | 18        | 4    | 18        |
| 5    | 5         | 5    | 2         | 6    | 8         |
| 6    | 2         | 6    | 4         | 7    | 4         |
| 7    | 2         | 8    | 7         | 8    | 10        |
| 9    | 9         | 9    | 3         | 7    | 9         |

Fig. 9(b) shows the ROC curve for detecting, classifying and comparing Risk 0 versus Risk 2 images. As shown by Fig. 9(b), for a sensitivity of 99%, the specificity is 90%. Area under the ROC is 0.95. Fig. 9(c) shows the ROC curve for detecting, classifying and comparing Risk 0 versus Risk 3 images. Fig. 9(c) shows a sensitivity of 100%, and a specificity of 82%. Area under the ROC is 0.973. Finally, Fig. 9(d) shows the ROC curve for detecting, classifying and comparing Risk 0 versus all patients with any form of DR, which shows a sensitivity of 100%, specificity of 82% and an area under the ROC equal to 0.95.

Now the application of AM-FM/PLS processing to images sets with vascular abnormalities and risk for macular edema (DR level 3 and macular edema level 2) is considered. Also, the effects of image quality are addressed. Table 6 shows the distribution of image quality for 265 test images with higher values indicating better image quality:

<table>
<thead>
<tr>
<th>Risk</th>
<th>Total</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>92</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>43</td>
<td>39</td>
</tr>
<tr>
<td>1</td>
<td>71</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>37</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>52</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>36</td>
<td>7</td>
</tr>
</tbody>
</table>

Here, 36 images are combined from Risks 2 and 3 with risk of macular edema and 11 images with vascular abnormalities are separated from them resulting in 100% of these images being correctly classified.

Fig. 10 illustrates an exemplary computer system 200, or network architecture, that may be used to implement the methods according to the present invention. One or more computer systems 200 may carry out the methods presented herein as computer code. One or more processors, such as processor 204, which may be a special purpose or a general-purpose digital signal processor, is connected to a communications infrastructure 206 such as a bus or network. Computer system 200 may further include a display interface 202, also connected to communications infrastructure 206, which forwards information such as graphics, text, and data, from the communication infrastructure 206 or from a frame buffer (not shown) to display unit 230. Computer system 200 also includes a main memory 205, for example random access memory (“RAM”), read-only memory (“ROM”), mass storage device, or any combination thereof. Computer system 200 may also include a secondary memory 210 such as a hard disk drive 212, a removable storage drive 214, an interface 220, or any combination thereof. Computer system 200 may also include a communications interface 224, for example, a modem, a network interface (such as an Ethernet card), a communications port, a PCMCIA slot and card, wired or wireless systems, etc.

It is contemplated that the main memory 205, secondary memory 210, communications interface 224, or a combination thereof function as a computer usable storage medium, otherwise referred to as a computer readable storage medium, to store and/or access computer software and/or instructions.

Removable storage drive 214 reads from and/or writes to a removable storage unit 215. Removable storage drive 214 and removable storage unit 215 may indicate, respectively, a floppy disk drive, magnetic tape drive, optical disk drive, and a floppy disk, magnetic tape, optical disk, to name a few.
In alternative embodiments, secondary memory 210 may include other similar means for allowing computer programs or other instructions to be loaded into the computer system 200, for example, an interface 220 and a removable storage unit 222. Removable storage units 222 and interfaces 220 allow software and instructions to be transferred from the particular embodiments disclosed, but on the contrary, the intention is to cover all modifications, equivalents, and alternatives falling within the scope of the disclosure as defined by the appended claims.
10. A computer system method for modeling video content comprising the steps of:
   providing an input video;
   computing a three-dimensional extended analytic signal of
   the input video;
   selecting a multi-scale filterbank by defining one or more
   bandpass filters that correspond to each scale of the
   multi-scale filterbank;
   processing the three-dimensional extended analytic signal
   through the multi-scale filterbank;
   calculating estimates for instantaneous amplitude, instan-
   taneous phase, and instantaneous frequency, said calculat-
   ing step further comprising the step of applying a
   Variable Spacing Local Linear Phase Model;
   selecting the instantaneous amplitude estimates with the
   maximum amplitude response, the instantaneous phase
   estimate with the maximum phase response, and the
   instantaneous frequency estimate with the maximum
   frequency response from said calculating step;
   reconstructing the input video using the instantaneous
   amplitude estimate and instantaneous frequency esti-
   mate from said selecting step to obtain a reconstructed
   video; and
   displaying the reconstructed video on a display unit.
11. The computer system method for modeling video content of claim 10 wherein said applying step further comprises
   the steps of:
   selecting the instantaneous frequency estimates with the
   minimum value for a condition number;
   pre-filtering the instantaneous frequency estimate with a
   median filter;
   post-filtering the instantaneous frequency estimate with
   the median filter; and
   modulating high-frequency outputs to baseband, wherein
   the Variable Spacing Local Linear Phase Model is
directly applicable.
12. The computer system method for modeling video content of claim 10 wherein the multi-scale filterbank includes
   the one or more bandpass filters that have been optimally
designed using min-max criteria.
13. The computer system method for modeling video content of claim 10 wherein said reconstructing step further
   includes the step of:
   figuring the instantaneous amplitude estimate and the
   instantaneous frequency estimate scale by for each scale
   of the multi-scale filterbank to obtain computed esti-
   mates.
14. The computer system method for modeling video content of claim 10 wherein the method is applied to images
   identifying disease at different stages.
15. The computer system method for modeling video content of claim 10 wherein the method is applied to retinal
   image analysis.
16. The computer system method for modeling video content of claim 15 wherein the retinal image analysis further
   includes diabetic retinopathy classification.
17. The computer system method for modeling video content of claim 10 wherein the method is applied to processing
   X-ray images.
18. The computer system method for modeling video content of claim 10 wherein the method is applied to describing
   images featuring atherosclerotic plaque.
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

In the Specification

Column 1, line 16, “FA9453-06-C-0211 awarded by DARPA” should read -- FA945306-C-0211 awarded by the United States Air Force --

Signed and Sealed this
Twenty-first Day of April, 2015

Michelle K. Lee
Director of the United States Patent and Trademark Office