

Laboratório VISGRAF

Instituto de Matemática Pura e Aplicada

Classificacao de sinais com Scattering Operators

Leandro Cruz

Luiz Velho (orientador)

Technical Report TR-2012-04 Relatório Técnico

October - 2012 - Outubro

The contents of this report are the sole responsibility of the authors.
O conteúdo do presente relatório é de única responsabilidade dos autores.



Instituto de Matemática Pura e Aplicada – IMPA
Doutorado em Matemática

Classificação de sinais com *Scattering Operators*

Leandro Cruz

Rio de Janeiro
Outubro de 2012

Sumário

1	Introdução	2
2	Descritores	4
2.1	Descritor Invariante por Translações Globais	7
2.2	SIFT, DAISY e Scattering	10
3	Classificação de Áudio usando Operadores de Scattering	11
3.1	Representação por Scattering	12
3.2	Scattering Wavelets	17
3.3	Decomposição e Reconstrução	19
3.4	Cosine Log-Scattering - CLS	21
3.5	Classificação	22
4	Classificação de Sinais bidimensionais com Operadores de Scattering	23
4.1	Transformação de Scattering	23
4.2	Transformação de Scattering e Convolution Network	26
4.3	Reconstrução a partir dos coeficientes de Scattering	27
4.4	Propriedades da Transformada de Scattering	28
4.5	Viabilidade Computacional	29
4.6	Classificação	31
5	Experimentos	35

5.1	Experimento 1: Wavelets de Gabor	35
5.2	Experimento 2: Filter Bank	36
5.3	Experimento 3: Transformação Scattering	37
5.4	Experimento 4: Transformação Scattering - áudios	38
5.5	Experimento 5: Coeficientes de Scattering - exemplares de terreno	41
5.6	Experimento 6: Transformação Scattering - exemplares de terreno	43
5.7	Experimento 7: Transformação Scattering - Variando a escala dos coeficientes de scattering em exemplares de terreno	46
6	Resultados e Conclusões	48
	Referências Bibliográficas	49

Capítulo 1

Introdução

O problema de classificação consiste na escolha de um descritor de um sinal combinado com um classificador. A escolha do descritor é baseada de modo que esse destaque certas características de um sinal. O classificador deve ser capaz de representar bem o espaço das representações, conseguindo discriminar objetos de classes distintas. De um modo geral, quanto mais simples é o descritor, mais complexo deve ser o classificador. Isso ocorre, para que a combinação garanta a discriminabilidade entre as diferentes classes de sinais.

Ao longo desse relatório discutiremos o problema de definição de bons descritores para sinais de dimensão um ou dois, combinados com técnicas de classificação. Os sinais discutidos referem-se principalmente à áudio (1D), imagens (tipicamente texturas) e terrenos (modelos de elevação - 2D).



Entre os descritores de áudio podemos citar o MFSC, MFCC e o Delta MFCC. No caso de imagens, podemos citar como exemplo de descritores o SIFT, SURF, Daisy, módulo da Transformada de Fourier, entre outros. Além desses descritores, podemos citar os coeficientes de scattering, introduzido por Mallat et. al. em um conjunto de trabalhos que serão citados nos capítulos seguintes.

Entre os classificadores podemos citar PCA, SVM, Convolution Network, Adaboost, Cadeias de Markov, Redes Neurais, etc.

Ao longo desse relatório apresentaremos como utilizar os coeficientes de scattering como descritores de sinais de dimensão um e dois. Esse descritor é invariante a pequenas deformações e se mostrou eficiente, mesmo quando combinado com classificadores simples como PCA e SVM.

No Capítulo 2 apresentaremos alguns descritores que podem ser usados no processo de classificação. Compararemos esses descritores com a representação por coeficientes de scattering, que será apresentado nos capítulos seguintes. No Capítulo 3 apresentaremos como obter a Transformada de Scattering para sinais unidimensionais e citaremos alguns resultados no processo de classificação usando essa representação. No Capítulo 4 apresentaremos a Transformada de Scattering para sinais bidimensionais. Nesse capítulo falaremos de algumas propriedades desse operador e também apresentaremos alguns resultados no processo de classificação de dígitos e texturas. Finalmente no Capítulo 5 apresentaremos alguns experimentos realizados pelo autor desse relatório. O objetivo desses experimentos é verificar alguns conceitos apresentados ao longo desse relatório e realizar alguns testes aplicando essa teoria no processo de classificação de amostras de terrenos.

O principal objetivo deste trabalho é entender os conceitos do processo de classificação de sons e imagens usando *Scattering Operators* e pensar como aplicá-los em pesquisas referente a síntese de texturas e terrenos.

Capítulo 2

Descritores

Um descritor consiste em um operador que representa um sinal. Em geral classes de sinais são invariantes a certas propriedades. Essas propriedades podem ser globais ou locais. Entre elas podemos citar invariância à translações, rotações ou escalas. Um descritor é invariante a uma certa propriedade quando o operador associado também é. A invariância a uma determinada propriedade reduz a dimensionalidade da representação do sinal. Isso ajuda no processo de aprendizagem, porém diminui a distância entre classes diferentes, ou seja, piora a discriminabilidade dos sinais.

Dado $c \in \mathbb{R}^n$, a translação de um sinal x por c pode ser definida pela transformação $L_c x = x(\cdot - c)$. Um operador ϕ invariante por translações globais deve satisfazer a propriedade $\phi(L_c x) = \phi(x)$, para todo $c \in \mathbb{R}^n$.

Um exemplo de operador invariante por translações globais é $\phi(x) = x(u + a(x))$. A função $a(x)$, conhecida como *Anchor Point*, associa à cada sinal um vetor de translação. Essa função deve satisfazer a seguinte propriedade: $a(L_c x) = a(x) + c$. Neste caso:

$$\begin{aligned}\phi(L_c x)(u) &= L_c x(u + a(L_c x)) = L_c x(u + (a(x) + c)) \\ &= x(u + a(x) + c - c) = x(u + a(x)) \\ &= \phi(x)\end{aligned}$$

Neste caso, temos que ϕ é um operador invariante por translação global. Um exemplo de função anchor point é $a(x) = \arg \max_u \|(x * h)(u)\|$. Essa função satisfaz a condição $a(L_c x) = a(x) + c$. Com efeito:

$$\begin{aligned}a(L_c x) &= \arg \max_u \|(L_c x * h)(u)\| \\ &= \arg \max_u \|(x(\cdot - c) * h)(u)\| \\ &= \arg \max_u \|(x * h)(u - c)\| \\ &= a(x) + c\end{aligned}$$

Outro exemplo de operador invariante por translações globais é o Módulo da Transformada de Fourier $\phi(f) = \|\hat{f}\|$. Com efeito

$$\begin{aligned}\widehat{L_c x}(w) &= \int x(t-c)e^{-itw} dt \\ &= \int x(u)e^{-i(u+c)w} du \\ &= e^{-icw} \int x(u)e^{-iuw} du \\ &= e^{-icw} \hat{x}(w)\end{aligned}$$

Logo, $\|\widehat{L_c x}(w)\| = \|e^{-icw} \hat{x}(w)\| = |e^{-icw}| \|\hat{x}(w)\| = \|\hat{x}(w)\|$, sendo portanto invariante por translação global, como queríamos concluir.

As translações globais $(\{f(x-\tau)\}_{\tau \in \mathbb{R}^d})$ formam uma classe de equivalência que pode ser pensada como uma variedade d-dimensional em $L^2(\mathbb{R}^d)$. Esse é um tipo fraco de invariância. Uma classe mais forte de invariância são as pequenas deformações locais.

Seja $\tau : \mathbb{R}^d \rightarrow \mathbb{R}^d$, uma deformação local pode ser escrita como $D_\tau x(u) = x(u-\tau(u))$. Observe que, diferentemente do Anchor Point que dependia apenas do sinal, agora τ também depende do parâmetro u . Essa classe de operadores também é conhecido como deformações elásticas.

A norma do tensor do gradiente de τ mede a amplitude da deformação do sinal, criada por L_τ em u . Uma deformação é uma transformação invertível quando $\|\nabla\tau(u)\| < 1$.

Um operador de deformação elástica pode ser aproximado por uma translação na vizinhança de u quando $\|\nabla\tau(u)\|$ é pequena. Nesse caso, uma deformação linear induzida pelo tensor do gradiente é dada por $L_\beta x(u) = x(\beta u)$, onde $\beta = I - \tau(u) \in GL(\mathbb{R}^n)$ (matrizes invertíveis).

Para que L_β aproxime bem D_τ é necessário que a representação ϕ do sinal tenha pequenas variações sempre que $\|\nabla\tau(u)\| = \|I - \beta\|$ (onde $I \in \mathbb{R}(n, n)$ é a matriz identidade) for pequeno. Esta propriedade pode ser reformulada exigindo que ϕ satisfaça a condição de regularidade de Lipschitz, ie, $\forall \beta \in GL(\mathbb{R}^n)$ na vizinhança de I (ie, $\beta = I - \tau$, tara τ tal que $\|\nabla\tau(u)\|$ seja pequena) existe $C > 0$ tal que:

$$\begin{aligned}\|\phi(x) - \phi(L_\beta x)\| &\leq C\|x - L_\beta x\| \\ &\leq C\|x\|\|I - \beta\| \\ &= C\|x\|\|\tau\| \\ &\leq C\|x\|\sup_u \|\tau(u)\|\end{aligned}$$

Logo, se $\|\nabla\tau(u)\|$ é pequena o segundo membro da inequação é pequeno, logo, as representações de x e de $L_\beta x$ são parecidas. Além disso, como nesse caso L_β é uma boa aproximação de D_τ , também queremos que $\|\phi(x) - \phi(D_\tau x)\| \leq C\|x\|\sup_u \|\tau(u)\|$.

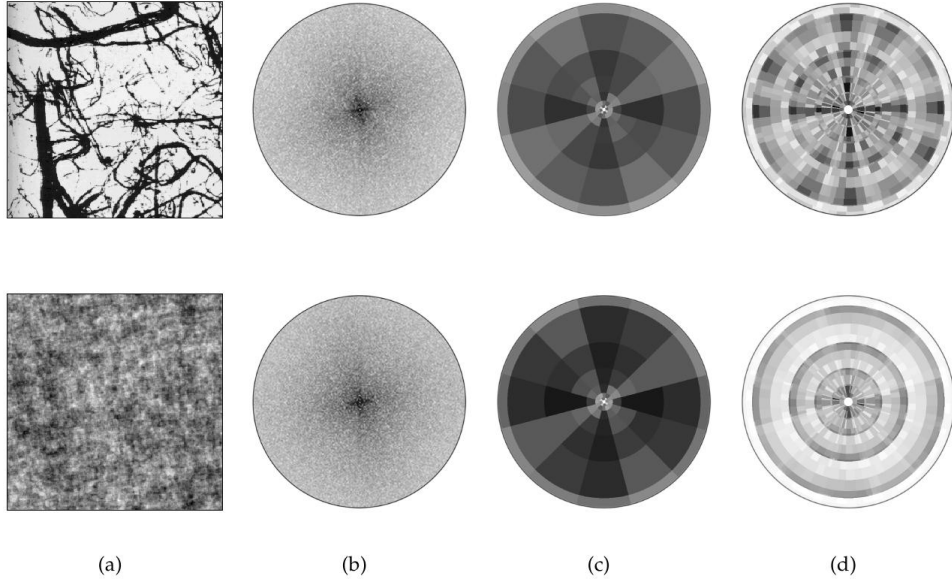


Figura 2.1: (a) exemplar de textura; (b) módulo da transformada de fourier; (c) coeficientes de scattering de primeira e (d) de segunda ordem.

As translações globais, quando $\|\nabla\tau(u)\| = 0$, o que implica que $\forall u$ temos que $\tau(u) = c$, são um tipo de deformação global. Porém, como facilmente se observa, esse é um exemplo muito restritivo.

O Operador de Módulo de Fourier é invariante a translações globais, mas não satisfaz a condição de regularidade de Lipschitz. Mesmo que $\|\nabla\tau(u)\|$ seja pequeno, para altas frequências a desigualdade pode não ser válida. Isso é um resultado esperado, pois deformações locais podem afetar muito altas frequências. Outro ponto negativo desse operador é a perda da fase do sinal. Isso reduz a discriminabilidade dos sinais e inviabiliza reconstruir o sinal a partir da sua representação.

O Operador de Scattering satisfaz a propriedade de regularidade de Lipschitz, e portanto é invariante a pequenas deformações. No próximos capítulos abordaremos como definir esse operador. Por ora, faremos uma pequena análise comparando-o com outros descritores.

Vale destacar que esse operador é capaz de capturar as frequências relacionadas com as waveletes usadas no filter bank, mas também capta as co-ocorrências dessas frequências. Isso torna essa representação, embora mais redundante, mais discriminante.

A Figura 2.1 mostra um exemplo de duas texturas visualmente muito distintas, cujo módulo da transformada de fourier é o mesmo. Por outro lado, os coeficientes de scattering, principalmente os de segunda ordem, são bem distintos, mostrando uma eficiente representação. A eficiência dos coeficientes de scattering nesse caso deve-se ao fato destes incorporarem à representação, além dos momentos de segunda ordem, também capturado

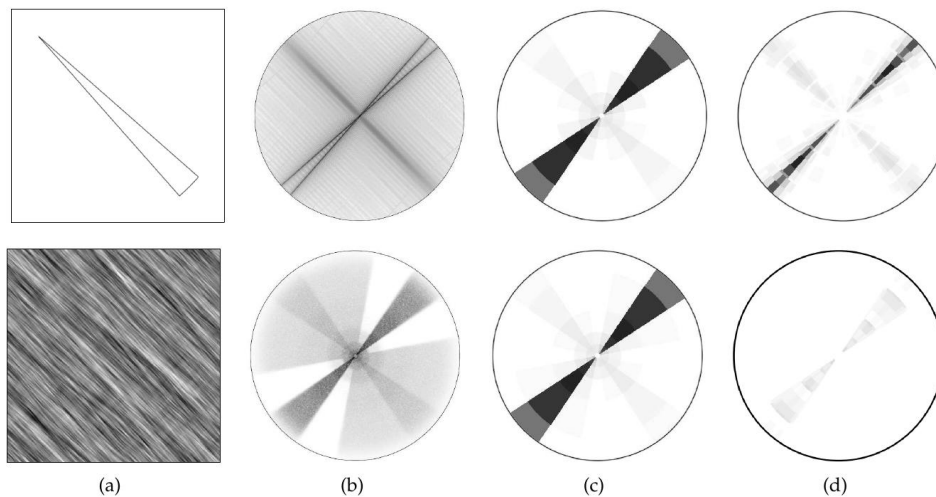


Figura 2.2: (a) exemplar de textura; (b) módulo da transformada de fourier; (c) coeficientes de scattering de primeira e (d) de segunda ordem.

pelo módulo da transformada de fourier, momentos de ordem maior. Isso torna a representação mais discriminante.

Vale destacar que há casos em que os coeficientes de scattering de primeira ordem não são suficientes para discriminar certos sinais. A Figura 2.2 apresenta um desses exemplos. As imagens apresentadas são visualmente bem distintas, apresentam módulo da transformada de fourier distintos, mas os coeficientes de scattering de primeira ordem são iguais. Porém, novamente os coeficientes de segunda ordem são bem distintos, sendo ainda uma representação satisfatória.

2.1 Descritor Invariante por Translações Globais

Nessa seção descreveremos alguns experimentos relacionados com classificação usando um descritor invariante por translações globais, implementado pelo autor desse relatório junto com Francisco Ganacim. Nesse exemplo comparou-se amostras de um determinado tipo de terreno (o Grand Canyon, nos EUA), contra terrenos de outras três classes: Flat Head, Nevada e Chest Nut Ridge (outras regiões topograficamente distintas, também nos EUA). Aplicou-se um processo de classificação binária, comparando os exemplos do Grand Canyon contra os demais. A Figura 2.3 mostra um exemplo de algumas amostras dessas regiões.

Nesse experimento, tomou-se um conjunto de amostras de terrenos (no formato de elevação). A representação consiste apenas no conjunto de alturas, independente da localização, e portanto invariante a translações globais. Essa representação é semelhante a usar um *anchor point*, uma vez que trabalhando apenas com a imagem da função de altura, podemos aplicar translações no domínio que a representação ficará inalterada.

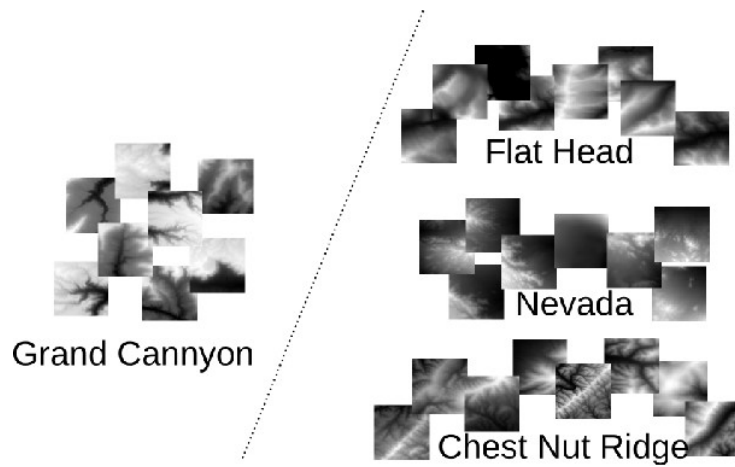


Figura 2.3: Classificação com descritor invariante a translação global

A Figura 2.4 mostra um experimento onde utilizou-se 128 exemplares do Grand Canyon contra 128 exemplares de cada uma das outras três classes para treinar uma SVM (Support Vector Machine). A Figura 2.4a mostra os exemplares de treinamento pertencentes ao Grand Canyon. Ao testar a SVM com esses exemplares apenas um foi classificado errado (exemplar em vermelho). Nas demais imagens os exemplares azuis referem-se aos exemplares de treinamento, os verdes foram classificados como parte do Grand Canyon e os Vermelhos como não pertencentes.

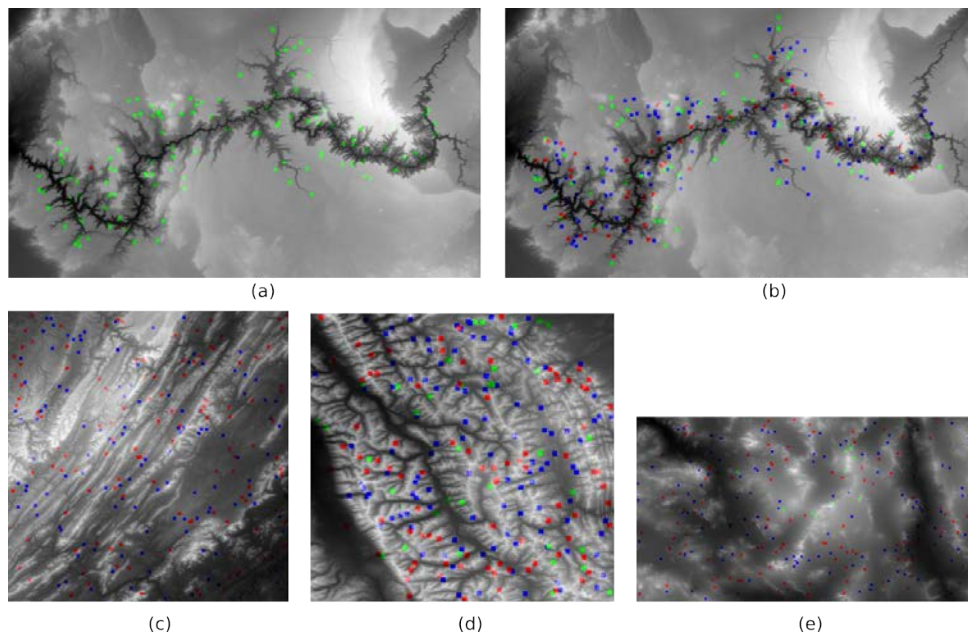


Figura 2.4: Resultado da classificação: pertencente ao Grand Canyon (verde), não pertencente ao Grand Canyon (vermelho), exemplares de treinamento (azul)

Outro experimento realizado com esses exemplares do Grand Canyon consiste na Análise das Componentes Principais (PCA) desses dados. A Figura 2.5 mostra os dados e algumas componentes principais (autovetores da matriz de covariância dos dados).

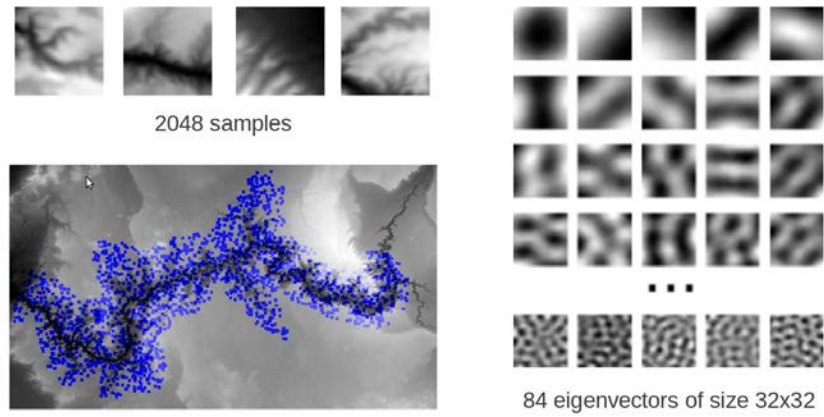


Figura 2.5: PCA calculada com exemplares do Grand Canyon.

Como supracitado, outro descritor invariante a translações globais é o Módulo da Transformada de Fourier. Esse descritor foi utilizado em um terceiro experimento. Este experimento consiste em criar uma base de dados de relevos artificiais (tipicamente morros), e tomar de cada relevo curvas de silhueta observadas de diferentes pontos de vista. Para cada uma dessas curvas calculou e armazenou-se o módulo da Transformada de Fourier.

As informações obtida com esse descritor para cada relevo é comparada com o resultado desse descritor aplicado a uma dada curva (tipicamente introduzido utilizando sketches). A aplicação recupera os relevos que apresentam alguma silhueta parecida com a curva introduzida. A Figura 2.6 mostra um exemplo do uso dessa aplicação.

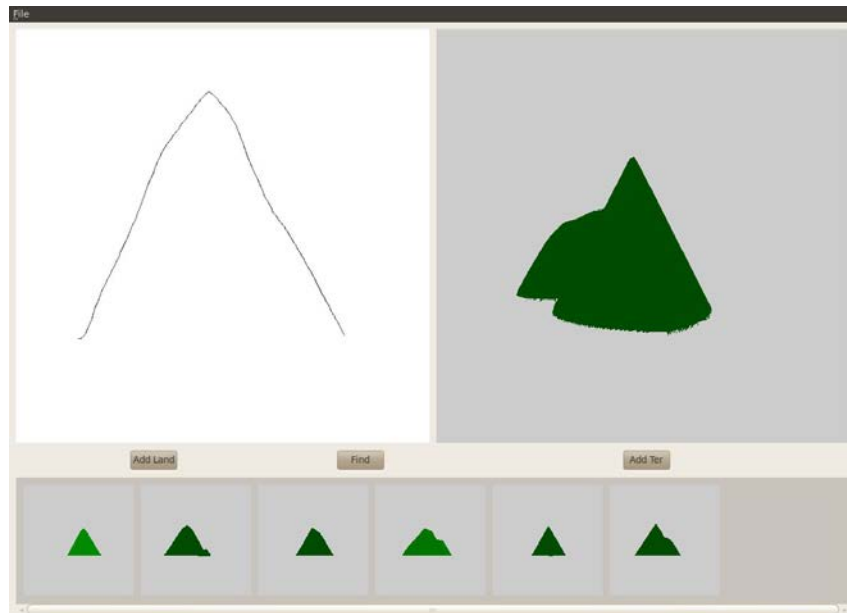


Figura 2.6: Métrica baseada em descritor baseado no Módulo da Transformada de Fourier.

2.2 SIFT, DAISY e Scattering

Dois exemplos de descritores utilizados com imagens são o SIFT [8] e o DAISY [11]. Para wavelets apropriadas, os coeficientes de scattering de primeira ordem são equivalentes a representação do SIFT. Analogamente, a representação obtida com o DAISY é semelhante aos coeficientes de scattering de segunda ordem para certas wavelets (obtidas por derivadas parciais de gaussianas).

Muitos descritores de imagens são definidos como médias de amplitude de coeficientes de wavelets. O SIFT é calculado através de histogramas sobre o gradiente da imagem, em escalas 2^j . Um elemento (neste caso ângulo) $\gamma \in \Gamma$ no histograma soma as amplitudes dos vetores dos gradientes cuja orientação é próxima a γ .

Uma aproximação do SIFT pode ser calculada mais eficientemente através de médias, calculadas usando um filtro passa baixa ϕ_J , das amplitudes dos vetores da derivada parcial de f , em K direções $\gamma \in \Gamma$. Essas médias podem ser calculadas como coeficientes de wavelets $|f * \psi_{j,\gamma}| * \phi_J(x)$, onde $\psi(x) = \frac{\partial g(x)}{\partial x_1}$ e $g(x) = e^{-\frac{|x|^2}{2}}$. Esses coeficientes são quase invariantes por translações ou deformações menores do que 2^J .

As wavelets obtidas por derivadas parciais são boas para obter arestas no sinal. Porém, esses filtros não captam bem certas informações de frequências e direções, que aparecem em sinais mais complexos, como texturas.

Um exemplo de filtro capaz de gerar boas wavelets para análise de textura, pois apresentam melhor localização na frequência, é o filtro de Gabor: $\psi(x) = e^{i\xi x_1} e^{-\frac{|x|^2}{2}}$. Esse filtro foi usado nos experimentos de [3].

Outro exemplo semelhante de filtro é a Wavelet de Morlet: $\psi(x) = C_1(e^{i\langle x, \xi \rangle} - C_2)e^{-\frac{|x|^2}{2\sigma^2}}$, onde C_2 é escolhida de modo que $\int \psi(u) du = 0$. Esse filtro foi usado nos experimentos de [4].

Como dito até este ponto, os coeficientes de scattering são uma boa representação para sinais invariantes a pequenas deformações. Nos próximos capítulos descreveremos como obter esse descritor.

Capítulo 3

Classificação de Áudio usando Operadores de Scattering

Nesse capítulo descreveremos o processo de classificação de sinais unidimensionais, tipicamente áudios. A teoria apresentada foi introduzida por Anden e Mallat no artigo *Multiscale Scattering For Audio Classification* [1].

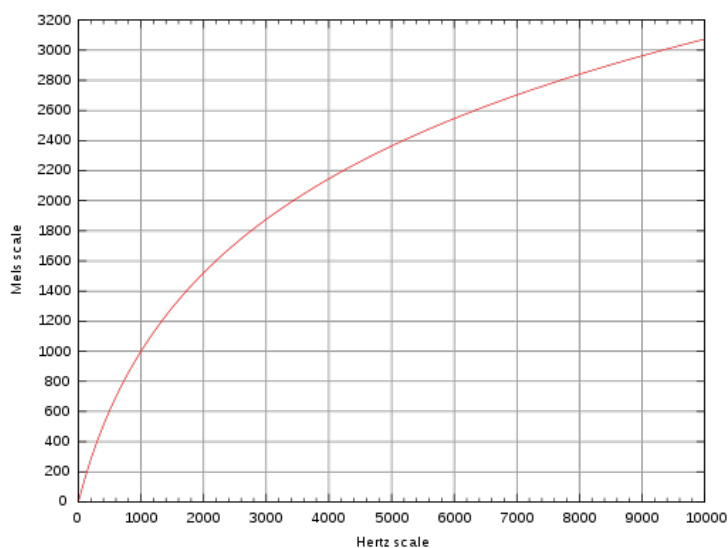


Figura 3.1: Relação entre frequência em Hertz e Mel-frequency

O conceito de *Mel-Scale* é um conceito muito utilizado em processamento de som. O prefixo *Mel* vem da palavra *Melody*. Essa é uma escala perceptual, de medidas de tons em um sinal sonoro. A relação entre a frequência (em hertz) de um sinal e a *Mel-frequency* é dada por:

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right)$$

A Figura 3.1 mostra o gráfico da função de conversão de frequência numa escala em Hertz para a *Mel Scale*.

Uma representação de sinal muito utilizada em classificação de sons é *Mel-Frequency Cepstrum Coefficients* (MFCC). Estes coeficientes são obtidos aplicando a transformada dos cossenos sobre *Mel-Frequency Spectrum Coefficients* (MFSC). Os MFSCs medem a energia das frequências do sinal em intervalos *mel-frequency* de largura de banda constante, digamos Q . Esse cálculo é obtido a partir de médias do espectro do sinal (obtido pela transformada de Fourier em janela).

Seja $x : \mathbb{R} \rightarrow \mathbb{R}$ um sinal dado. A Transformada de Fourier de x é dada por $\hat{x}(s) = \int_{\mathbb{R}} x(t)e^{-2\pi i s t} dt$. Vamos definir a função $x_{t,T}(u) = x(u)w_T(u-t)$, onde w_T é uma função para localizar x em uma janela do tempo de tamanho T . Sendo assim, a transformada de Fourier em Janela de x é dada por:

$$\hat{x}_{t,T}(s) = \int_{\mathbb{R}} x_{t,T}(u)e^{-2\pi i s t} dt$$

3.1 Representação por Scattering

Sejam $\psi_j(w)$, com $j \in \mathbb{Z}$, funções tais que sua respectiva função de transferência cobre um determinado intervalo no domínio *mel-frequency*. Os MFSCs podem ser calculados como:

$$M_T x(t, j) = \int_{\mathbb{R}} \|\hat{x}_{t,T}(s)\|^2 \|\hat{\psi}_j(s)\|^2 ds$$

Logo, pelo Teorema de Parseval temos que:

$$M_T x(t, j) = \int_{\mathbb{R}} \|(x_{t,T} * \psi_j)(\tau)\|^2 d\tau$$

Podemos interpretar essa relação da seguinte forma: MFSC mede a energia do sinal x numa vizinhança de t , com tamanho T . Essa medida é feita sobre o domínio *mel-frequency*. O valor de $M_T x(t, j)$ consiste numa média do espectrograma de $\|\hat{x}_{t,T}(s)\|^2$ sobre o intervalo *mel-frequency* coberto por ψ_j . Por isso, esse valor não é capaz de determinar medidas de estruturas não estacionárias (ie, estruturas que não variam com o tempo) cujo tamanho é menor do que T .

Seja ψ um filtro, cuja oitava de largura de banda é $1/Q$, podemos construir as $\psi_j(w)$ como dilatações deste filtro:

$$\psi_j(w) = a^{-j} \psi(a^{-j} w)$$

onde $a = 2^{1/Q}$, $j \leq J = \log_2 \left(\frac{f}{2Q} \right)^Q$ e f é a frequência de amostragem. A função ψ é uma wavelet, e desta forma temos que as ψ_j são obtidas a partir de dilatações da wavelet, para altas frequências.

A função ψ deve ser um filtro complexo normalizado (o suporte no domínio do tempo deve ter largura igual a 1s). O suporte da função de transferência de ψ cobre o intervalo de frequência $[2Q\pi - \pi, 2Q\pi + \pi]$. Para $j < J$ o suporte no tempo de ψ_j tem largura menor do que a^j e sua respectiva função de transferência cobre o intervalo de frequência $[2Q\pi a^{-j} - \pi a^{-j}, 2Q\pi a^{-j} + \pi a^{-j}]$. A Figura 3.2 mostra um exemplo dos intervalos cobertos por ψ_j , com $j < J$. Neste exemplo, usamos $Q = 20$ e $J = 30$.

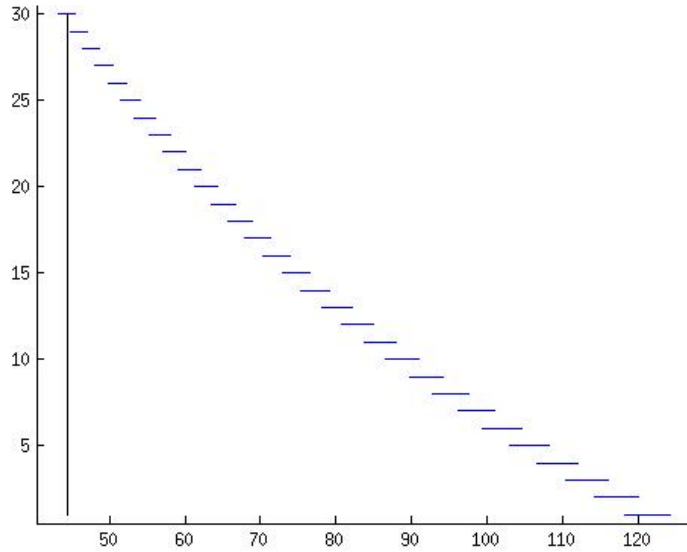


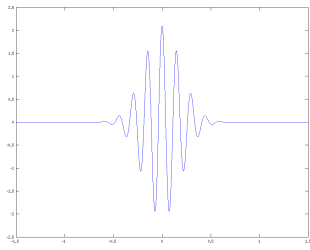
Figura 3.2: Intervalos de suporte de $\hat{\psi}_j$, onde a abscissa representa o tamanho do intervalo e a ordenada o índice j .

As frequências abaixo de $2\pi Q a^{-J}$ são cobertas por P filtros ψ_j , com $J < j < J+P$, que tem a mesma largura de banda (domínio da frequência) de ψ_J que é $2\pi a^{-J}$ e a^J no domínio do tempo. As Figuras 3.3 e 3.4 mostram exemplos de ψ_j variando j e Q .

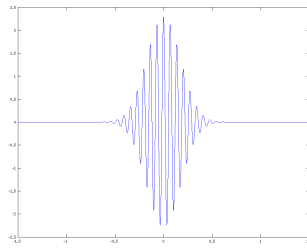
Seja ϕ_J um filtro passa baixa, capaz de cobrir as frequências não cobertas por ψ_j , podemos definir a seguinte Transformação Wavelet:

$$W_J x(t) = \begin{pmatrix} x * \phi_J(t) \\ x * \psi_j(t) \end{pmatrix}_{j < J+P}$$

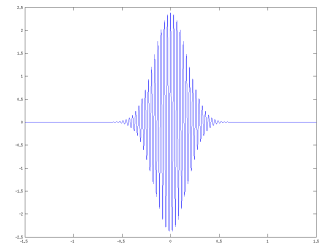
Os filtros wavelets são projetados de modo que existe um $\epsilon \geq 0$, que satisfaz a desigualdade abaixo para toda frequência w (*condição de admissibilidade*):



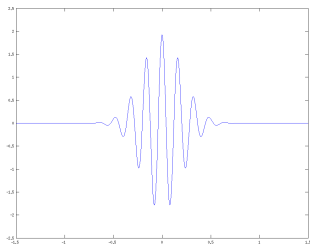
(a) $j=2$ e $Q=8$



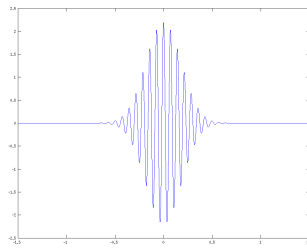
(b) $j=2$ e $Q=16$



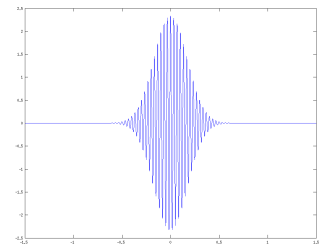
(c) $j=2$ e $Q=32$



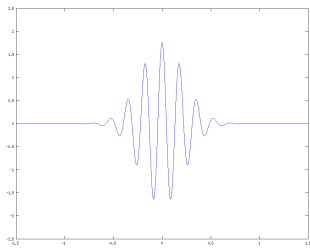
(d) $j=3$ e $Q=8$



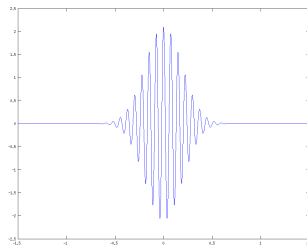
(e) $j=3$ e $Q=16$



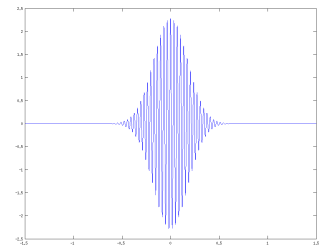
(f) $j=3$ e $Q=32$



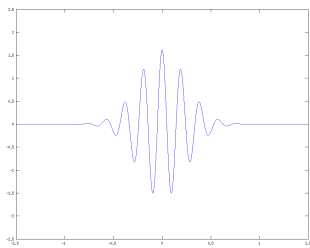
(g) $j=4$ e $Q=8$



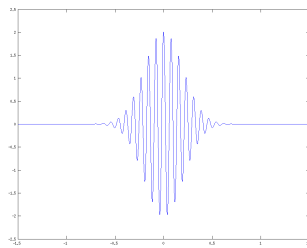
(h) $j=4$ e $Q=16$



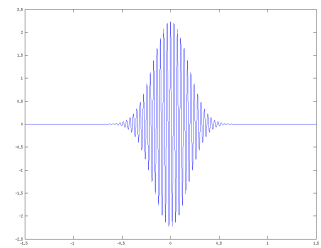
(i) $j=4$ e $Q=32$



(j) $j=5$ e $Q=8$

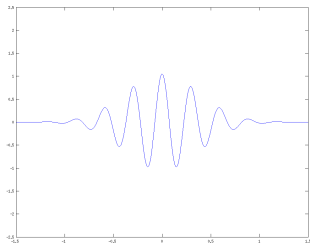


(k) $j=5$ e $Q=16$

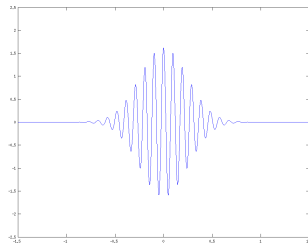


(l) $j=5$ e $Q=32$

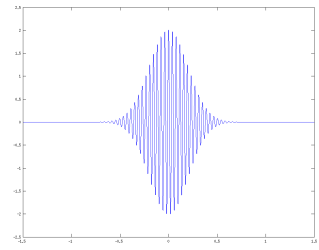
Figura 3.3: ψ_j



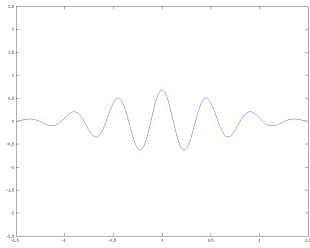
(a) $j=10$ e $Q=8$



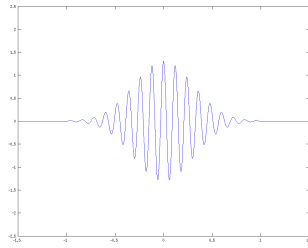
(b) $j=10$ e $Q=16$



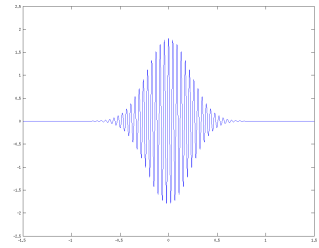
(c) $j=10$ e $Q=32$



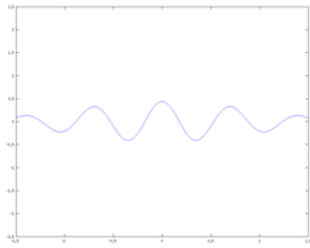
(d) $j=15$ e $Q=8$



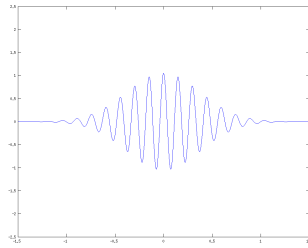
(e) $j=15$ e $Q=16$



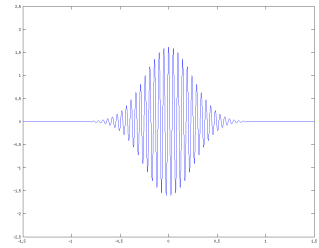
(f) $j=15$ e $Q=32$



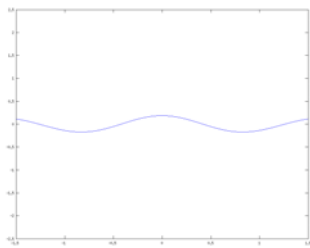
(g) $j=20$ e $Q=8$



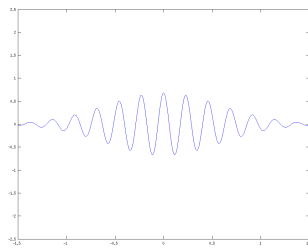
(h) $j=20$ e $Q=16$



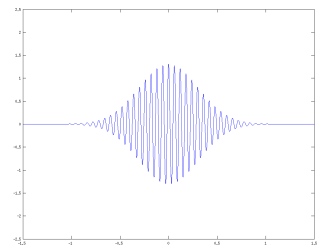
(i) $j=20$ e $Q=32$



(j) $j=30$ e $Q=8$



(k) $j=30$ e $Q=16$



(l) $j=30$ e $Q=32$

Figura 3.4: ψ_j

$$1 - \epsilon < \|\hat{\phi}(w|^{2+})\frac{1}{2} \sum_{j < J+P} \|\psi_j(w)\|^2 + \|\psi_j(-w)\|^2 \leq 1$$

Sabendo que $\|x\|^2 = \int_{\mathbb{R}} \|x(t)\|^2 dt$, assim como $\|W_{Jx}\|^2 = \|x * \phi_J\|^2 + \sum_{j < J+P} \|x * \psi_j\|^2$ temos que:

$$(1 - \epsilon) \|x\|^2 < \|\hat{\phi}(w|^{2+})\frac{1}{2} \sum_{j < J+P} \|\psi_j(w)\|^2 + \|\psi_j(-w)\|^2\|x\|^2 = \|W_{Jx}\|^2 \leq \|x\|^2$$

Logo, se $\epsilon > 0$ a transformada wavelets é uma contração, e se $\epsilon = 0$ ela é unitária.

Um exemplo de wavelets que satisfaz as propriedades listadas anteriormente são os filtros de Gabor $\psi(t) = \theta(t)e^{i2\pi Qt}$, onde θ é uma gaussiana. Verificou-se que para $\epsilon = 0.02$ a Condição de Admissibilidade é satisfeita, com $Q = 16$, $P = 23$ e $f = 1000Hz$ (neste caso $J = 80$). A Figura 3.5 mostra o módulo da Transformada de Fourier dos filtros de Gabor para os parâmetros supracitados. A Figura 3.5(b) consiste em um zoom nas baixas frequências. Observe que, existem baixas frequência que não são cobertas por esse *Filter Bank*. A Figura 3.5(c) mostra esse mesmo Filter Bank normalizado (resultado apresentado em [1]).

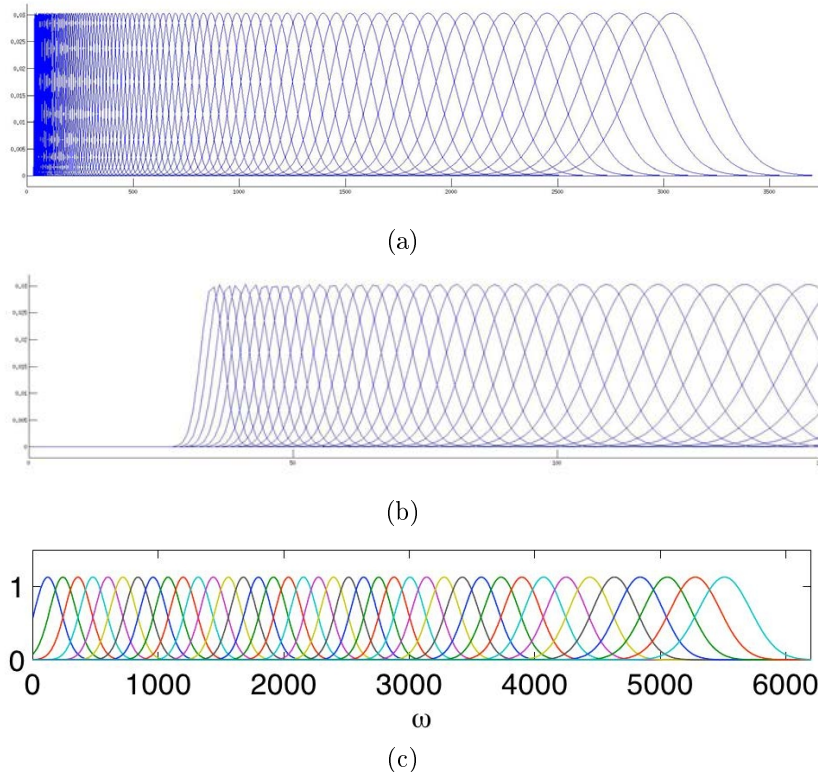


Figura 3.5: Função de Transferência do Filtro de Gabor: $Q = 16$, $f = 1024$, $J = 80$, $P = 23$

3.2 Scattering Wavelets

Como vimos na seção anterior os MFSCs $M_T x(t, j)$ referem-se a uma média do quadrado da energia do sinal x , na escala a^j (coberta pela wavelet ψ_j), em um intervalo de tamanho T tomado na vizinhança de t . Para o cálculo do MFCC será aplicado um logaritmo a esses coeficientes, dessa forma, o termo quadrático na integral será linearizado. Por essa razão, podemos utilizar uma outra medida $|x * \psi_j| * \phi_J(t)$, baseada apenas na amplitude do sinal filtrado, em vez da amplitude ao quadrado.

A medida $|x * \psi_{j_1}| * \phi_J(t)$ perde informações ao fazer uma média temporal, numa janela de tempo de largura $T = a^J$ (quanto maior o valor T maior a perda) no intervalo de frequência coberto pela ψ_j .

Vale observar que $|x * \psi_{j_1}| * \phi_J(t)$ é a componente de baixa frequência da Transformada de Wavelet, previamente apresentada, aplicada ao sinal $|x * \psi_{j_1}|$.

$$W_{J|x*\psi_{j_1}|}(t) = \begin{pmatrix} |x * \psi_{j_1}| * \phi_J(t) \\ |x * \psi_{j_1}| * \psi_{j_2}(t) \end{pmatrix}_{j_2 < J+P}$$

Por essa razão, se essa Transformada de Wavelet for invertível a informação perdida pela filtragem por ϕ_J pode ser recuperada pelos coeficientes $|x * \psi_{j_1}| * \psi_{j_2}(t)$. Eles medem a amplitude do sinal $|x * \psi_{j_1}|$ no intervalo de frequência coberto por ψ_{j_2} . Mais adiante nesse relatório iremos comentar sobre uma condição de invertibilidade dessa classe de Transformada de Wavelets.

Podemos aplicar um filtro passa baixa nos coeficientes de scattering para obter a *co-ocorrência* das informações na escala a^{j_1} e a^{j_2} :

$$||x * \psi_{j_1}| * \psi_{j_2}| * \phi_J(t) \tag{3.1}$$

Esses coeficientes são chamados de coeficientes de scattering (*scattering coefficients*) porque calculam a interferência das wavelets ψ_{j_1} e ψ_{j_2} no sinal x . Eles medem a amplitude das variações do tempo de $x * \psi_{j_1}$, no intervalo de frequência coberto pela wavelet ψ_{j_2} .

Novamente, a média realizada por ϕ_J em (3.1) causa perda de informação de altas frequências que pode ser recuperada se $W_{J||x*\psi_{j_1}|*\psi_{j_2}|}(t)$ for invertível. Podemos iterar esse procedimento criando uma cascata de filter banks e operadores de módulos. Essa estrutura define a Transformação de Scattering.

Para definir formalmente a Transformação de Scattering vamos primeiro definir o operador de módulo. Seja U_J o operador que calcula o módulo dos coeficientes da wavelet complexa, enquanto mantém a fase de $x * \phi_J$:

$$U_J x(t) = \begin{pmatrix} x * \phi_J(t) \\ |x * \psi_j(t)| \end{pmatrix}_{j < J+P}$$

A Transformação de Scattering consiste em um encadeamento do operador de módulo U_J . Primeiro aplica-se no sinal o filtro passa baixa: $x * \phi_J$, obtendo os coeficientes de ordem 0. Na camada seguinte, pega-se módulo do sinal filtrado pela wavelet ($|x * \psi_{j_1}|$) e filtra-o analogamente: $|x * \psi_{j_1}| * \phi_J$, obtendo os coeficientes de ordem 1. Em seguida, pega-se o módulo do módulo do sinal filtrado pela wavelet e filtra-o com as demais wavelets: $||x * \psi_{j_1}| * \psi_{j_2}|$. Além disso, aplica-se o filtro passa baixa: $||x * \psi_{j_1}| * \psi_{j_2}| * \phi_J$, obtendo os coeficientes de ordem 2. E assim sucessivamente. Generalizando esses passos, temos que a transformação scattering de ordem m é dada por:

$$S_{Jx}(t) = \begin{pmatrix} x * \phi_J(t) \\ |x * \psi_{j_1}| * \phi_J(t) \\ ||x * \psi_{j_1}| * \psi_{j_2}| * \phi_J(t) \\ \vdots \\ ||\dots|x * \psi_{j_1}|\dots| * \psi_{j_m}| * \phi_J(t) \end{pmatrix}_{j < J+P}$$

Essa transformação de scattering é uma cascata de filter banks modulados e retificações não lineares. Esse processo é análogo ao feito nas *convolution networks*. A Figura 3.6 mostra um exemplo do processo de construção da transformação de scattering.

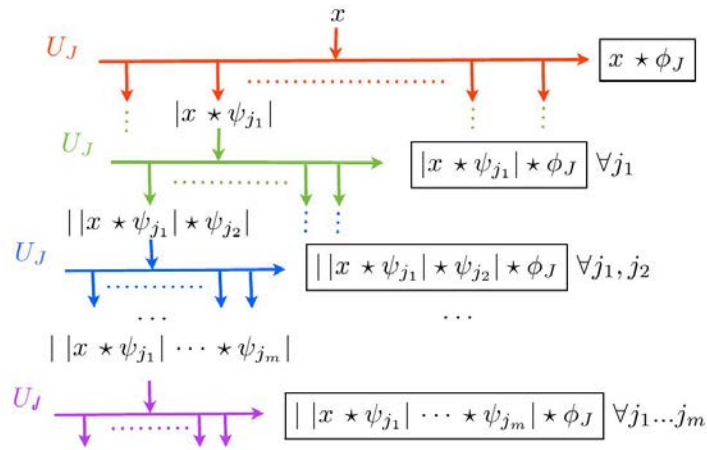


Figure 3. A scattering operator is a cascade of wavelet modulus operators U_J . It outputs convolutions with ϕ_J shown in boxes.

Figura 3.6:

3.3 Decomposição e Reconstrução

Diferentemente das convolution networks, a transformação de scattering itera o processo sobre a saída no nível anterior, fazendo com que a energia dos coeficientes decaia para zero a medida que a ordem cresce. Em experimentos apresentados em [1] verificou-se que para amostras de som, tomadas na base de dados GTZAN, os coeficientes de ordem 2 capturaram mais do que 98% da energia do sinal, conforme a Tabela 1.

T	$m = 0$	$m = 1$	$m = 2$	$m = 3$
23 ms	23.7%	98.9%	99.6%	99.6%
93 ms	1.9%	97.7%	99.4%	99.4%
370 ms	1.2%	92.7%	99.3%	99.4%
1.5 s	1.0%	82.0%	98.9%	99.3%
5.9 s	0.99%	73.0%	98.1%	99.1%
22 s	0.97%	67.5%	96.5%	99.0%

Table 1. Averaged ratio $\|S_J x\|/\|x\|$ on the GTZAN dataset, as a function of the maximum scattering order m and of $T = a^J$.

Nesta seção discutiremos como podemos obter uma representação para o sinal através dos coeficientes scattering e como é possível reconstruir o sinal a partir dessa representação.

Como visto na Tabela 1, tomando os coeficientes de ordem menor do que ou igual a dois é possível capturar grande parte da energia do sinal (no caso de decomposição de amostras de um sinal de áudio apresentado na Tabela 1, esses coeficientes capturaram 98% do sinal). Isso significa que tais coeficientes não são suficientes para obter uma representação exata do sinal, mas apenas uma boa aproximação.

Como o suporte do filtro ϕ_J tem largura a^J podemos tomar amostras para $t = na^J$, com $n \in \mathbb{N}$. Analogamente, para calcular a convolução entre x e ψ_j basta tomar amostras $t = na^j$. Se $j_2 < j_1 + \log_a \frac{Q}{2}$ então $||x * \psi_1| * \psi_2| * \phi_J \approx 0$, ou seja, para calcular os coeficientes de segunda ordem, basta calcular as correlações entre ψ_1 e ψ_2 tais que $j_1 + \log_a \frac{Q}{2} \leq j_2 \leq J + P$.

A Figura 3.7 mostra os coeficientes de scattering de ordem um ($q = 1$) e dois ($q = 2$) de um sinal de áudio. Note que nos coeficientes de ordem dois, muitos valem quase zero (azul é zero e vermelho é o valor máximo). Isso está de acordo com a relação exposta no parágrafo anterior.

Um frame de duração $T = a^J$, contendo N amostras do sinal produz $Q \log_2 \frac{N}{Q}$ coefi-

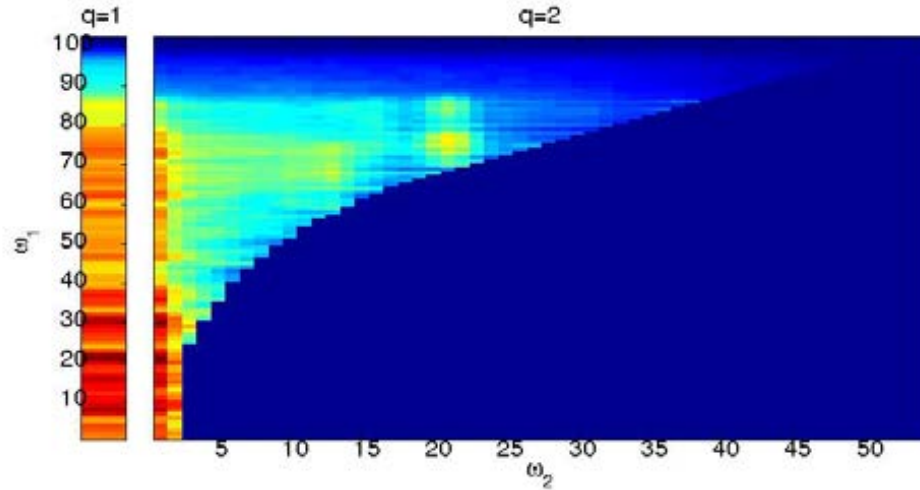


Figura 3.7: Decaimento dos coeficientes de scattering.

coeficientes de primeira ordem e $\frac{Q^2}{2} \log_2^2 \frac{N}{Q^2}$ coeficientes de segunda ordem. Se usar FFT no cálculo desses coeficientes são necessárias $O(N \log(\frac{N}{Q}))$ operações.

Se a Transformação de Scattering for calculada através de iterações do operador U_J então o processo de reconstrução consiste em obter uma inversa para esse operador. A Transformação Wavelet W_J é invertível (com uma inversa estável), mas U_J não é capaz de armazenar a fase dos coeficientes da wavelet (complexa). Dessa forma, obter a inversa de U_J significa recuperar a fase complexa da parte que aplica o módulo na convolução do sinal com a wavelet ($|x * \psi_j|$).

Waldspurger e Mallat [12] mostraram que utilizando uma Wavelet apropriada o operador U_J é invertível e possui inversa contínua (o que é um resultado de estabilidade fraca). Nesse trabalho os autores apresentaram um algoritmo para calcular a inversa de U_J . Porém esse algoritmo não é exato pois depende de um processo de otimização não convexa (portanto pode convergir para um mínimo local). Além disso, como dito no início desta seção, apesar de U_J ser invertível não é possível reconstruir o sinal a partir dos coeficientes de $S_J x$ de ordem m finito. Porém, como dito, para $m = 2$ é possível captar grande parte da energia do sinal, e portanto realizar uma boa aproximação. Mais adiante discutiremos melhor as condições de invertibilidade da Transformada de Wavelet.

3.4 Cosine Log-Scattering - CLS

Os coeficientes MFCC são calculados aplicando a Transformada dos Cossenos no logaritmo do MFSC. Analogamente, calculamos os Coeficientes CLS (Cosine Log-Scatering) aplicando a DCT ao logaritmo dos coeficientes de scattering.

Como dito na seção anterior, se um frame contem N amostras, então teremos $Q \log_2 \frac{N}{Q}$ coeficientes de primeira ordem e $\frac{Q^2}{2} \log_2^2 \frac{N}{Q^2}$ coeficientes de segunda ordem. Porém, a DCT decorrelaciona os coeficientes *log-scattering* e concentra a energia em poucos desses coeficientes. Desta forma, para a representação final do sinal, podemos armazenar apenas alguns coeficientes CLS, o que viabiliza implementações computacionais.

A Figura 3.4 mostra um exemplo aonde a energia dos coeficientes CLS estão agrupados em poucos termos. Esse exemplo compara a variância dos coeficientes das amostras retiradas na base GTZAN (base de dados com amostras de áudio de diferentes gêneros). Esses coeficientes foram representados em três bases: log-scattering (linha sólida), base PCA dos log-scattering (linha tracejada) e os CLS (linha pontilhada). Observe que a decorrelação obtida pelo CLS é semelhante a obtida usando a base PCA.

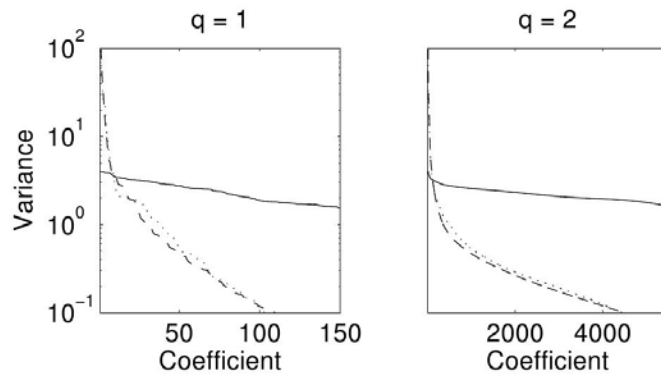


Figure 4. Variances, in decreasing order, of log-scattering coefficients in different bases for $q = 1$ and $q = 2$ computed on GTZAN for $T = 1.5$ s. Solid curve: Variance of log-scattering coefficients. Dashed curve: Variance of a PCA basis computed on log-scattering coefficients. Dotted curve: Variance of cosine log-scattering coefficients.

ATENÇÃO: Observe que, para obtermos discriminabilidade do sinal, necessária para o processo de classificação, a transformada de scattering aumentou a dimensão da representação do sinal. Porém esse aumento tem um custo computacional (de armazenamento e processamento da representação do sinal). Por outro lado, a DCT nos permite reduzir a dimensionalidade dessa representação sem perder a discriminabilidade. Dessa forma, os coeficientes CLS são uma representação mais compacta do sinal, mas que ainda apresentam a discriminabilidade desejada, conseguida pela transformação de scattering.

3.5 Classificação

Existem muitas técnicas para classificação de sinais de áudio. Muitos delas baseiam-se nos coeficientes MFCC ou em informações obtidas por Delta-MFCC. Os coeficientes Delta-MFCC são obtidos tomando a diferença entre coeficientes MFCC de dois frames consecutivo. Assim como MFCC e Delta-MFCC, pode-se utilizar os coeficientes CLS como vetor de características utilizados para a classificação. Algoritmos mais sofisticados utilizam técnicas como GMM, HMM, AdaBoost, etc. Nesta seção serão discutidos aspectos relacionados ao processo de classificação desses sinais a partir de características dos coeficientes de scattering utilizando classificadores específicos.

O processo de classificação pode ser feito utilizando SVM, com filtro Gaussiano com kernel $k(x_1, x_2) = e^{-\gamma\|x_1-x_2\|^2}$, ou através de algum classificador que reduz os dados em subespaços afins (como por exemplo PCA).

Devido ao grande número de amostras de treinamento que devem ser avaliadas quando toma-se pequenas janelas no tempo (como por exemplo 0.023s ou 0.19s), o custo de treinar uma SVM é muito alto. Neste caso, PCA é uma alternativa mais viável. Para janelas maiores (por exemplo 1.5s), e portanto menos amostras, os resultados usando SVM são favoráveis. A Tabela 2 apresenta os resultados obtidos usando MFCC, Delta-MFCC e CLS (de ordem 1 e 2). O melhor resultado obtido é usando SVM, para frames de tamanhos 1.5s. O Estado da Arte em classificação de música possui método cuja taxa de erro (nessa mesma base de dados) é de 7.6%.

T/classifier	0.023 s/PCA	0.19 s/PCA	1.5 s/SVM
MFCC	46	36	28
Delta-MFCC	37	33	26
CLS, $m = 1$	46	36	28
CLS, $m = 2$	34	23	18

Table 2. Error rates (in percent) on GTZAN using five-fold cross-validation for different window sizes (T) and features.

Observe que a taxa de erro obtida nos experimentos desse trabalho é significativamente alta. Isso deve-se ao fato que os coeficientes de scattering representam muito bem sinais estacionários. Sinais sonoros, embora tenham certas estruturas estacionárias, tem grandes variações em baixa frequência que podem prejudicar a discriminabilidade entre as classes. A seguir, abordaremos sinais mais semelhantes a processos estacionários. Nestes casos, os resultados obtidos são bem melhores.

Capítulo 4

Classificação de Sinais bidimensionais com Operadores de Scattering

No capítulo anterior tratamos de sinais unidimensionais. Nesse, analisaremos os sinais bidimensionais, como por exemplo imagens monocromática (imagens de dígitos e texturas). Nesse caso, precisamos estender as ferramentas apresentadas anteriormente para esse contexto. Este capítulo apresenta parte da teoria apresentada por Bruna e Mallat nos artigos *Classification with Scattering Operators* [3] e *Invariant Scattering Convolution Network* [4].

4.1 Transformação de Scattering

Como citado no Capítulo 2, dois exemplos de filtros que podem ser usados na construção da Transformada de Scattering são o Filtro de Gabor ($\psi(x) = e^{i\xi x_1} e^{-\frac{|x|^2}{2}}$ [3]) e a Wavelet de Morlet ($\psi(x) = C_1(e^{i\langle x, \xi \rangle} - C_2)e^{-\frac{|x|^2}{2\sigma^2}}$, onde C_2 é escolhida de modo que $\int \psi(u)du = 0$ [4]). A Figura 4.1 mostra um exemplo da Wavelet de Morlet construída com $\sigma = 0.85$ e $\xi = 3\pi/4$.

Seja G um grupo finito de rotações em \mathbb{R}^2 , $r \in G$ e $j \in \mathbb{Z}$, um filtro wavelet multiescala e direcional é definido por $\psi_{2^j r}(u) = 2^{2j} \psi(2^j r^{-1}u)$. Se a $\hat{\psi}$ está centralizada em η então $\hat{\psi}_{2^j r}(w) = \hat{\psi}(2^{-j} r^{-1}w)$ está centralizada em $2^j r \eta$, com largura de banda proporcional a 2^j (semelhantemente como visto no caso unidimensional). Para simplificar a notação, denotaremos $\lambda = 2^j r$.

A obtenção dos coeficientes de scattering é semelhante à apresentada no Capítulo 3, acrescentando a variação dos ângulos. A construção do Operador de Scattering também é análoga.

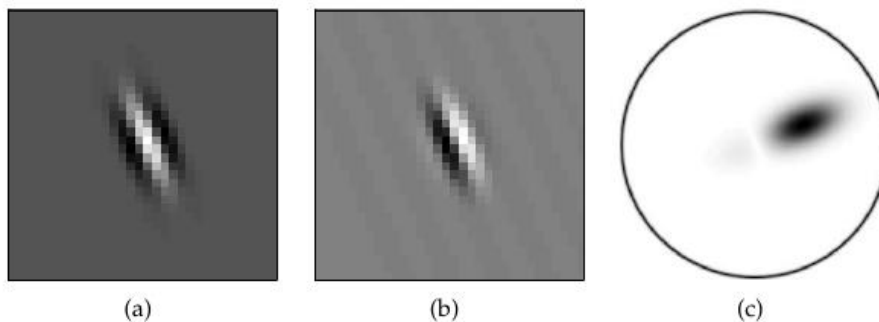


Figura 4.1: Wavelet de Morlet: (a) parte real de ψ ; (b) parte imaginária de ψ ; e (c) é o módulo de $\hat{\psi}$ ($|\hat{\psi}|$)

Novamente, queremos que a Transformada de Scattering seja invariante à pequenas deformações elásticas. Para alcançarmos esse objetivo partiremos de alguns fatos descritos por Bruna et. al. [2] e Waldspurger et. al. [12].

- A transformada de wavelets comuta com operadores de translação, portanto não é invariante por translação.
- Seja R um operador, linear ou não, que comuta com o operador de translação, então o operador $T(u) = \int Rx(u)du$ é invariante por translação.

Portanto, seja R um operador, tal que $Rx = x * \psi_\lambda$, então $\forall x$ temos que $\int Rx(u)du = \int x * \psi_\lambda(u)du = 0$, pois $\int \psi_\lambda(u)du = 0$. Neste caso, o operador R é trivialmente invariante por translação.

Agora tomemos um operador linear M que comute com um operador de deformações (ie, $\forall \tau(u)$ tem-se que $ML_\tau = L_\tau M$), $\forall x, y \in L^2(\mathbb{R}^2)$ temos que $\|Mx\| = \|x\|$, e $\|Mx - My\| \leq \|x - y\|$. Neste caso, temos que $Mx = e^{i\alpha}|x|$. Tomando $\alpha = 0$, temos que $Mx = |x|$ satisfazendo às propriedades acima listadas. Neste caso, podemos definir um operador $Rx = M(x * \psi_\lambda) = |x * \psi_\lambda| = \|x * \psi_\lambda\|_1 = \int |x * \psi_\lambda(u)|du$ que é invariante por translações.

Os coeficientes $\{\|x * \psi_\lambda\|_1\}$ são uma representação pobre do sinal, capaz de medir a esparsidade dos coeficientes da wavelet. Utilizando wavelets adequadas é possível reconstruir o sinal x a partir desses coeficientes [12, 13]. Mas em geral, essa representação é limitada, pois a integral remove todas as componentes de frequência diferentes de zero. Uma forma de recuperar esses dados é calcular também os coeficientes de segunda ordem $\{\| |x * \psi_{\lambda_1}| * \psi_{\lambda_2} \|_1\}$, $\forall \lambda_1, \lambda_2$ ($\| |x * \psi_{\lambda_1}| * \psi_{\lambda_2} \|_1 = \int |x * \psi_{\lambda_1}| * \psi_{\lambda_2}(u)du$). Esse processo é análogo ao descrito no caso unidimensional.

Seja $U(\lambda)x = |x * \psi_\lambda|$ e um caminho $p = (\lambda_1, \dots, \lambda_m)$, podemos definir $U[p]x = U(\lambda_m) \dots U(\lambda_2)U(\lambda_1)x = \| \dots \|x * \psi_{\lambda_1}| * \psi_{\lambda_2} | \dots * \psi_{\lambda_m} |$, onde $U[\emptyset] = x$. A Transformação

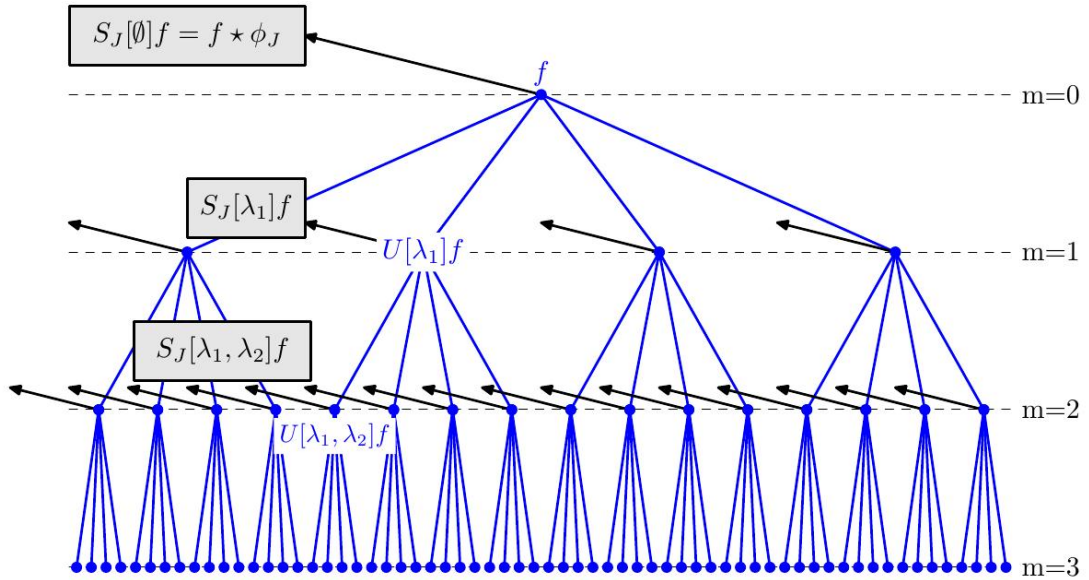


Figura 4.2: Representação em multi-camadas da Transformação de Scattering de ordem $m = 3$ aplicada em um sinal f .

de Scattering ao longo do caminho p é definida por $\bar{S}x(p) = \mu_p^{-1} \int U[p]x(u)du$, onde $\mu_p = \int U[p]\delta(u)du$.

Atenção: Para usarmos os coeficientes scattering no processo de classificação é mais interessante que tenhamos descritores que sejam localizados e invariantes à translações menores do que uma certa escala, digamos 2^J , mantendo a variabilidade para escalas maiores do que 2^J . Isso faz com que o descritor seja invariante aos detalhes, mas discriminante à variações de baixa frequência.

Para obter a propriedade de localização, basta aplicar uma janela no domínio do espaço que esteja devidamente escalada. Ou seja, definiremos $\phi_{2^J}(u) = 2^{-2J}\phi(2^{-J}u)$ e $S_J[p]x(u) = U[p]x * \phi_{2^J}(u) = \int U[p]x(v)\phi_{2^J}(u-v)dv$. Portanto, $S_J[p]x(u) = |\dots|x * \psi_{\lambda_1}| * \psi_{\lambda_2}|\dots| * \psi_{\lambda_m}| * \phi_{2^J}(u)$

Se p é um caminho de tamanho m então $S_J[p]x(u)$ é chamado de coeficiente de scattering de ordem m em uma escala 2^J . Esse coeficiente é calculado com uma convolution network [7]. A representação por scattering é um conjunto de coeficientes de coocorrências nas diferentes ordens (diferentes caminhos possíveis).

A Transformação de Scattering invariante à deformações em escalas menores do que 2^J pode ser definida como recursões do operador de módulo U_J (análogo ao definido no caso unidimensional). Os coeficientes da wavelet calculam frequências $2^j \geq 2^{-J}$ e as baixas frequências são filtradas por uma gaussiana escalada $\phi_{2^J}(x) = 2^J\phi(2^{-J}x)$, onde ϕ é uma gaussiana.

Seja G^+ um conjunto de rotações, com ângulos r entre zero e π e $\Lambda_J = \{\lambda = 2^j r; r \in$

$G^+, j \geq -J$. Definiremos o operador de módulo como $U_J x(u) = \{x * \phi_{2^j}, |x * \psi_\lambda|\}_{\lambda \in \Lambda_j}$.

Denotaremos $U[\Lambda_j^m] = \{U[p]x\}_{p \in \Lambda_j^m}$ e $S[\Lambda_j^m] = \{S[p]x\}_{p \in \Lambda_j^m}$. Sendo assim, o coeficiente $S_j[p]x(u)$ pode ser calculado primeiramente tomando $U_J x = \{S_j[\emptyset]x, U[\Lambda_j^1]\}$ e, para cada nível $m \leq m_{max}$, aplicando o operador de módulo U_J a $U[\Lambda_j^m]$ iterativamente. Esse processo iterativo pode ser escrito como:

$$U_J U[\Lambda_j^m] x = \{U_J U[p]x\}_{p \in \Lambda_j^m} = \{S[\Lambda_j^m], U[\Lambda_j^{m+1}]\}$$

A Figura 4.2 mostra um exemplo de construção de coeficientes de scattering até ordem $m = 3$ aplicado em um sinal f .

Atenção: 2^j é uma escala espacial (espaço de escala), mas $\lambda = 2^j r$ é uma variável no domínio da frequência, onde j é escolhido tal que $2^j \geq 2^{-j}$. Logo, a transformada de scattering combina coeficientes referentes às altas frequências (escala baixa), que são calculados de modo a satisfazer invariância à pequenas deformações; e os coeficientes referentes às baixas frequências (escalas altas), armazenados para garantir variabilidade entre objetos de diferentes classes.

4.2 Transformação de Scattering e Convolution Network

A transformação de scattering é uma generalização das convolution networks. Como dito, o cálculo de um coeficiente de scattering, em uma determinada ordem, é feito usando uma convolution network.

As convolutions networks são estáveis a pequenas deformações, e por isso tem sido usadas em algumas tarefas de classificação de imagens [7]. Esse operador apresenta uma arquitetura de uma rede neural. Porém, há várias propriedades desse operador que são difícil de entender devido ao encadeamento de transformações não lineares. Podemos incluir entre essas dúvidas: quantas camadas devem ser usadas? como otimizar os filtros e as funções de pooling? quantos neurônios serão usados em cada camada? Essas perguntas, geralmente, são respondidas com argumentos empíricos.

Por outro lado, substituindo o processo de aprendizagem para definir os filtros por wavelets predefinidas, Mallat et. al. [9, 1, 3, 4, 10] conseguiram construir o operador de scattering, com uma arquitetura semelhante a das convolution networks, com o qual podiam responder matematicamente às perguntas do parágrafo anterior. Esse operador é Lipschitz contínuo, e portanto invariante a movimentos rígidos e a lineariza pequenas

deformações, preservando a energia do sinal. Além disso, a energia do sinal se concentra em poucas camadas, de baixa ordem.

Uma diferença entre a transformação de scattering e a maioria das convolutions networks é que esta contém coeficientes em todas as camadas $m \leq m_{max}$, e não apenas na última camada.

Outra diferença é que os filtros não são obtidos a partir de um processo de aprendizagem que depende dos dados. No caso da transformação de scattering, os filtros são wavelets especiais predefinidas. A escolha das wavelets deve-se ao fato dessas serem estáveis à pequenas deformações e proverem uma representação esparsa do sinal.

O operador de módulo U_J é semelhante a função *pooling* das convolution networks. Ele recombina a parte real e a imaginária, e agrega coeficientes que possibilitam certas invariâncias. As altas frequências perdidas no processo de filtragem são recuperadas nas camadas seguintes, o que motiva a estrutura em multi-camadas da transformação de scattering.

4.3 Reconstrução a partir dos coeficientes de Scattering

Apesar de conseguirmos recuperar as altas frequências perdidas no processo de filtragem com a estrutura de multi-camadas, o processo de filtragem perde a fase do sinal. Porém, a fase ainda pode ser recuperada graças à redundância da transformação de wavelets. Waldspurger et. al. [12, 13] mostraram que o operador de módulo $U_J x$ é invertível, com inversa contínua. Portanto a fase complexa de cada $x * \psi_\lambda$, e portanto de x , pode ser recuperada.

Embora U_J tenha inversa, a transformação de scattering não é exatamente invertível, devido à instabilidades. A transformação de scattering é calculada com uma ordem m_{max} finita. Logo, existe uma energia nos coeficientes $m \geq m_{max}$ que é desconsiderada, introduzindo um pequeno erro que é propagado em todas as camadas quando calcula-se U_J^{-1} . Porém, vale destacar que, conforme será melhor discutido mais adiante, a energia dos coeficientes decai rapidamente pra zero a medida que a ordem cresce. Portanto, quanto maior o m_{max} menor será o erro da reconstrução. Por outro lado, cada nova camada introduz uma grande quantidade de coeficientes (maior do que a soma das quantidades de coeficientes das camadas anteriores), que cresce a medida que a ordem cresce.

Atenção: *Para classes de objetos invariantes à grandes escalas a representação por coeficientes de scattering deve ter uma ordem mais elevada, ou seja, é necessário adicionar mais camadas de coocorrência para evitar a perda de informações essenciais.*

4.4 Propriedades da Transformada de Scattering

A convolution network é uma estrutura altamente não linear, o que dificulta a entender as relações entre os coeficientes e as propriedades do sinal. Como a transformada de scattering consiste em iterações do operador de módulo, que consiste em uma variação da transformada de wavelet, as propriedades dos coeficientes de scattering dependem da transformada de wavelets.

As propriedades do operador de módulo $U_J x(u) = \{x * \phi_{2^j}(u), |x * \psi_\lambda(u)|\}_{\lambda \in \Lambda_J}$ dependem das propriedades da transformada de wavelets $W_J x(u) = \{x * \phi_{2^j}(u), x * \psi_\lambda(u)\}_{\lambda \in \Lambda_J}$.

No caso de x ser um sinal real, a *condição de admissibilidade* da wavelet consiste em:

$$\exists \epsilon > 0 \text{ tq } \forall w \in \mathbb{R}^2 \text{ tem-se que: } 1 - \epsilon \leq |\hat{\phi}(w)|^2 + \frac{1}{2} \sum_{j=1}^{\infty} \sum_{r \in G} |\hat{\psi}(2^{-j}rw)|^2 \leq 1$$

Se a wavelet ψ satisfaz a condição de admissibilidade, então podemos provar que $(1 - \epsilon)\|x\|^2 \leq \|W_J x\|^2 \leq \|x\|^2$, onde $\|W_J x\| = \|x * \phi_J\|^2 + \sum_{\lambda \in \Lambda_J} \|x * \psi_\lambda\|^2$.

Se $\epsilon < 1$ então a transformada de wavelet é uma contração e é invertível, com inversa estável. Se $\epsilon = 0$ então W_J é unitária. Um exemplo de wavelet unitaria é a spline cúbica de Battle-Lemarie [10]. A wavelet de Morlet satisfaz a condição de admissibilidade, para $\epsilon = 0.25$, com $\phi(u) = C e^{-\frac{|u|^2}{(2\sigma_0)^2}}$, onde $\sigma_0 = 0.07$ e C é tal que $\int \phi(u) du = 1$.

Atenção: *A condição de admissibilidade caracteriza quais wavelets podem ser usadas na construção da transformada de scattering.*

Se a transformada de wavelet é uma contração, então o operador de módulo também será. Com efeito $\|U_J x - U_J y\| \leq \|W_J x - W_J y\| \leq \|x - y\|$, onde a primeira desigualdade deve-se a relação $\||a| - |b|\| \leq |a - b|$ e a segunda é consequência de W_J ser uma contração. É possível verificar que se W_J é unitária então U_J também é. Como a transformada de scattering consiste em iterações de operador de módulo, temos que se U_J é uma contração então S_J também será.

Se W_J for unitária, e portanto U_J e S_J também são, pode-se provar que $\|S_J x\|^2 = \sum_{m=0}^{\infty} \|S_J[\Lambda_J^m]x\|^2 = \sum_{m=0}^{\infty} \sum_{p \in \Lambda_J^m} \|S_J[p]x\|^2 = \|x\|^2$. Neste caso, a série $\sum_{m=0}^{\infty} \|S_J[\Lambda_J^m]x\|^2$ é convergente, e portanto $\lim_{m_{max} \rightarrow \infty} \sum_{m \geq m_{max}} \|S_J[\Lambda_J^m]x\|^2 = 0$.

Atenção: *Esta propriedade significa que a energia dos coeficientes decai para zero a medida que a ordem da camada cresce. Esse resultado é essencial para a viabilidade computacional de se usar os coeficientes de scattering como representação do sinal, ie, podemos*

utilizar uma quantidade finita de camadas para representar o sinal (embora na verdade queremos um resultado um pouco mais forte, queremos utilizar poucas camadas).

A Tabela abaixo mostra um exemplo de experimento realizado com uma textura da base de dados Caltech-101, usando uma wavelet unitária (spline cúbica) para definir o Operador de Scattering. Essa tabela apresenta a porcentagem de energia dos coeficientes de uma determinada ordem m em função da escala J .

J	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m \leq 3$
1	95.1	4.86	-	-	-	99.96
2	87.56	11.97	0.35	-	-	99.89
3	76.29	21.92	1.54	0.02	-	99.78
4	61.52	33.87	4.05	0.16	0	99.61
5	44.6	45.26	8.9	0.61	0.01	99.37
6	26.15	57.02	14.4	1.54	0.07	99.1
7	0	73.37	21.98	3.56	0.25	98.91

Uma condição essencial para o processo de classificação é que a representação seja estável à pequenas deformações. Podemos satisfazer essa condição com a propriedade da representação ser Lipschitz-contínua.

Para sinais de suporte compacto Mallat [10] provou que:

$$\|S_J(L_\tau x) - S_J x\| \leq C m_{max} \|x\| (2^{-J} \|\tau\|_\infty + \|\nabla \tau\|_\infty)$$

Se $2^J \geq \frac{\|\tau\|_\infty}{\|\nabla \tau\|_\infty}$ então temos que a transformada de scattering é Lipschitz contínua para deformações. Ou seja, $\|S_J(L_\tau x) - S_J x\| \leq C m_{max} \|x\| \|\nabla \tau\|_\infty$.

Atenção: A condição de a transformada de scattering ser Lipschitz contínua para deformações é essencial para usar os coeficientes de scattering no processo de classificação.

4.5 Viabilidade Computacional

Nossa motivação inicial para representar um sinal com coeficientes de scattering é obter discriminabilidade entre as classes, preservando certas invariâncias. Para obter a discriminabilidade foi necessário analisar as diferenças entre as estruturas estacionárias do sinal. Isso é feito através dos cálculos das coocorrências de certas frequências. Como consequência, a dimensão da representação do sinal aumenta consideravelmente. Porém, algumas propriedades supracitadas já ajudam a diminuir a dimensão dessa representação. Nesta seção, citaremos alguns fatos que permitem reduzir a dimensão da representação.

A primeira observação essencial é que podemos usar poucas camadas no operador de scattering. Cada nova camada acrescenta uma quantidade de coeficientes que cresce substancialmente, a medida que a ordem cresce.

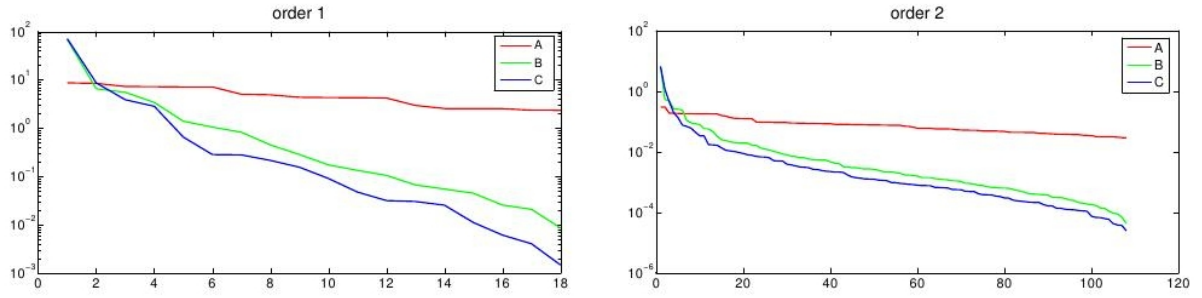


Figura 4.3: Variância dos coeficientes de scattering de ordem 1 e dois: A linha A representa os coeficientes de scattering, a linha B é a DCT dos coeficientes de scattering e a C é a variância desses coeficientes na base de Karhunen-Loeve.

Em cada camada, os coeficientes de scattering estão altamente correlacionado. Para cada m , se representarmos os coeficientes de scattering com a base de Karhunen-Loeve, que diagonaliza a matriz de covariância, verificaremos de fato essa correlação, pois a variância se concentra em poucos coeficientes. Bruna e Mallat [4] mostraram que a DCT aproxima a base de Karhunen-Loeve, e portanto descorrelaciona os coeficientes de scattering. A Figura 4.3 mostra que aplicando a DCT nos coeficientes de scattering estes ficam quase tão descorrelacionados como se aplicássemos a base de Karhunen-Loeve.

Para o cálculo dos coeficientes da camada m tomaremos os caminhos $p = (2^{-j_1}r_1, \dots, 2^{-j_m}r_m)$ tal que $0 < j_k \leq j_{k+1} \leq J$. Se calcularmos os coeficientes para K rotações teremos $K^m \binom{J}{m}$ caminhos de profundidade m . Seja N o número de pixels de uma imagem x , então $S[p]x(u) = U[p]x * \phi_{2^J}$ pode ser uniformemente amostrada a intervalos de $\alpha 2^J$, com $\alpha = 1$, ou $\alpha = 0.5$ (para evitar aliasing). Cada $S[p]x$ é uma imagem com $\alpha^{-2} 2^{-2J} N$ coeficientes. A quantidade de coeficientes depende exclusivamente da escala J e decresce exponencialmente a medida que essa escala aumenta. A quantidade de caminhos p varia de acordo com a profundidade máxima m_{max} . Portanto o total de coeficientes de scattering é $P = N \alpha^{-2} 2^{-2J} \sum_{m=0}^{m_{max}} K^m \binom{J}{m}$. O Experimento 5, descrito em 5.5 exemplifica esses cálculos.

Seja P_{\downarrow}^m o conjunto dos caminhos com frequências decrescentes e comprimento $m \leq m_{max}$. Além disso, para $\lambda = 2^{-j}r$, definamos $p + \lambda = (2^{-j_1}r_1, \dots, 2^{-j_m}r_m, 2^{-j}r)$ (nesse caso $U[p + \lambda] = |U[p]x(u) * \psi_{\lambda}(u)|$). Fixando $\alpha = 1$, ou $\alpha = 0.5$, o Algoritmo 1 mostra como calcular os coeficientes de scattering.

A camada m tem $K^m \binom{J}{m}$ sinais $U[p]x$. Eles são amostrados a intervalos $\alpha 2^{j_m}$. Nessa camada, tem-se um total de $\alpha^{-2} (\frac{K}{3})^m N$ amostras. Os $K^m \binom{J}{m}$ sinais $S[p]x$ podem ser subamostrados a taxa de 2^J , reduzindo a quantidade de coeficientes. Portanto, usando FFT para calcular as convoluções, o número de operações para calcular cada camada é de $O((\frac{K}{3})^m N \log N)$ (assumindo $K > 3$). Portanto, a complexidade do algoritmo é de $O((\frac{K}{3})^{m_{max}} N \log N)$.

Algorithm 1 Fast Scattering Transform

```
for  $m = 1$  to  $\bar{m}$  do
  for all  $p \in \mathcal{P}_{\downarrow}^{m-1}$  do
    Output  $S[p]x(\alpha 2^J n) = U[p]x \star \phi_{2^J}(\alpha 2^J n)$ 
  end for
  for all  $p + \lambda_m \in \mathcal{P}_{\downarrow}^m$  with  $\lambda_m = 2^{-j_m} r_m$  do
    Compute
      
$$U[p + \lambda_m]x(\alpha 2^{j_m} n) = |U[p]x \star \psi_{\lambda_m}(\alpha 2^{j_m} n)|$$

  end for
end for
for all  $p \in \mathcal{P}_{\downarrow}^{\bar{m}}$  do
  Output  $S[p]x(\alpha 2^J n) = U[p]x \star \phi_{2^J}(\alpha 2^J n)$ 
end for
```

4.6 Classificação

Uma dificuldade no processo de classificação de sinais refere-se a incapacidade de identificar similaridades entre sinais de uma determinada classe usando distância euclideana na representação do sinal. Sinais de uma mesma classe apresentam variações relacionadas com movimentos rígidos, pequenos escalamentos e pequenas deformações. Essas variações devem ser eliminadas pela representação, aproximando assim, sinais de uma mesma classe.

Um tipo de variação do sinal que é muito importante refere-se às deformações não rígidas. Porém, não podemos considerar qualquer tipo de deformações, pois existem transformações capazes de deformar um sinal de uma classe em um sinal de outra classe. Desse modo, estamos interessados em uma representação invariante à pequenas deformações. Essa representação deve linearizar as pequenas deformações para não afetar a discriminabilidade de classificadores lineares.

Atenção: Neste caso, entendemos por linearização o fato de a representação ser Lipschitz contínua para pequenas deformações.

Na introdução desse relatório abordamos que o processo de classificação depende das escolhas de descritores e de classificadores. Nessa seção, abordaremos o processo de clas-

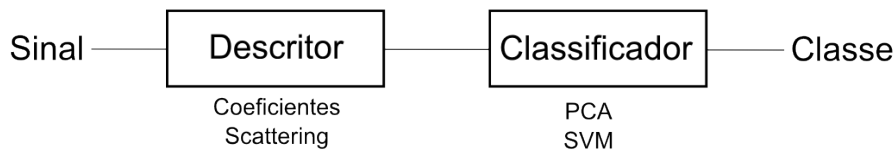


Figura 4.4: Pipeline do processo de classificação.

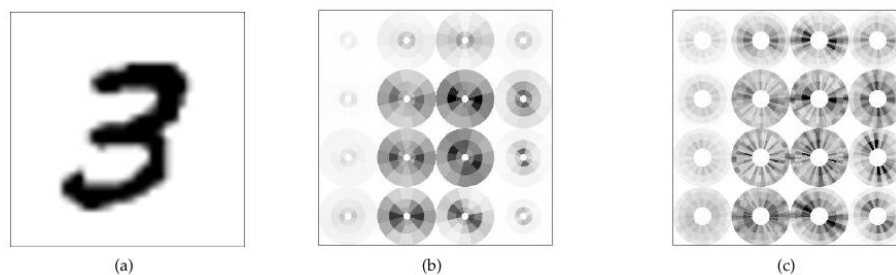


Figura 4.5: (a) Exemplo, (b) Coeficientes de primeira ordem; (c) coeficientes de segunda ordem.

sificação de sinais usando a representação por coeficientes de scattering e utilizando como classificadores modelos generativos como PCA, e modelos discriminantes como SVM.

Os coeficientes de scattering apresentam algumas propriedades interessantes para o processo de classificação. A primeira é que eles são invariantes à translações. A segunda é que a transformação de scattering lineariza pequenas deformações. Além disso, a decomposição realizada permite capturar padrões estacionários que se repetem entre amostras de uma classe. Ao longo desse relatório comparamos essa representação com outros descritores conhecidos, tais como SIFT e DAISY. Propriedades como essas citadas, são a justificativa de preferirmos usar a descrição por scattering.

Como, anteriormente citado, podemos aplicar a DCT nos coeficientes de scattering para reduzir a dimensão da representação. Como a DCT é um operador linear ortogonal, aplicá-lo aos coeficientes de scattering não modifica o resultado de nenhum classificador linear. Verificou-se empiricamente que mantendo apenas 50% desses coeficientes, a perda no processo de classificação é insignificante.

Seja X_k o conjunto de amostras da classe k , onde cada amostra tem P coeficientes. Ao utilizar PCA como classificador, escolhendo d , calculamos inicialmente as componentes principais V_k de cada classe k (os d autovetores da matriz de covariância, associados aos d maiores autovalores). Podemos aproximar a região que contém as amostras da classe k por $A_k = E\{SX_k\} + V_k$, onde $E\{SX_k\}$ é a média das representações de scattering das amostras da classe k . Uma propriedade da PCA é que poucas componentes têm tamanho significativo. Logo, podemos aproximar a região que contém essas amostras por um subespaço de dimensão $d \ll P$.

Seja $P_{A_k}Sx$ a projeção da representação do sinal x no subespaço afim referente a classe k . Dizemos que x pertence a classe k quando a $\|Sx - P_{A_k}Sx\| \leq \|Sx - P_{A_{k'}}Sx\|, \forall k' \neq k$.

É necessário se determinar dois parâmetros. O primeiro é a escala de invariância 2^J . O segundo é a dimensão dos subespaços afins que representam cada classe. Nos exemplos mostrados por Bruna e Mallat [3, 4] utilizou-se cross-validation para determiná-los, de modo a minimizar o erro de classificação em subconjuntos das amostras de treinamento.

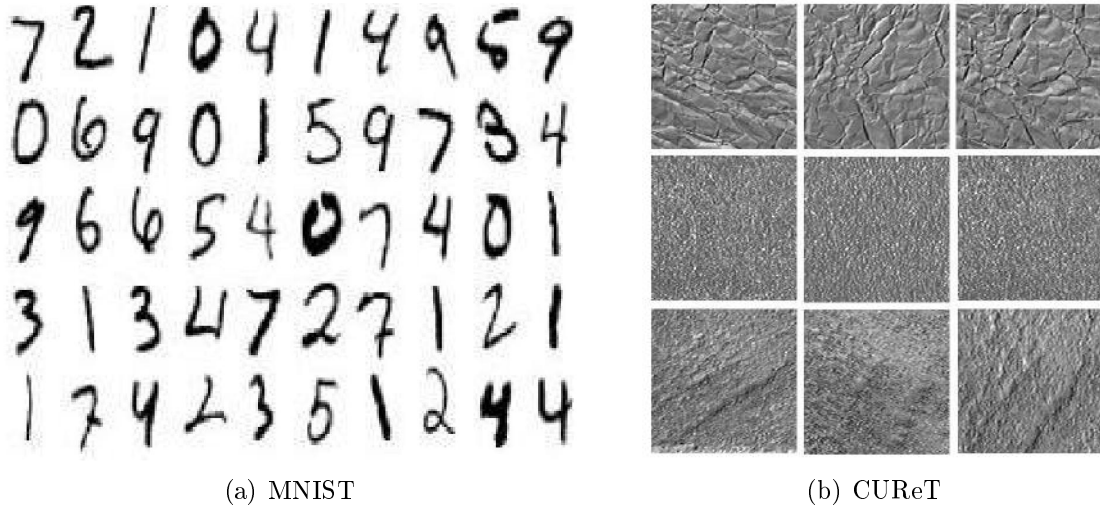


Figura 4.6: Amostras das bases de dados MNIST e CURET.

A Figura 4.5 mostra um exemplo de uma imagem com tamanho 32×32 aonde a escala ótima encontrada é $2^J = 8$. Nesse exemplo, ao aumentar $m_{max} = 1$ para $m_{max} = 2$ a o erro de classificação reduz aproximadamente 30%. Porém, se aumentar $m_{max} = 3$, o ganho é insignificante. Para $m_{max} = 2$, o valor de d obtido por *cross-validation* para o caso de 300, 5000 e 40000 amostras de dígitos é, respectivamente, 5, 100 e 140.

Os testes apresentados por Bruna e Mallat [3, 4] combinam coeficientes de scattering com PCA ou SVM na classificação de sinais, como por exemplo, dígitos e texturas. Os dígitos referem-se a base de dados MNIST [6]. Já as texturas, foram retiradas da base de dados CURET (Columbia-Utrecht Reflectance and Texture Database) [5]. A Figura 4.6 mostra exemplos de amostras dessas duas bases de dados.

Para o problema de classificação de dígitos, Bruna e Mallat [3, 4] fizeram testes comparando PCA e SVM com outros classificadores. No caso de SVM, utilizou-se kernels RBF, com variância ajustada aplicando *cross-validation* sob um subconjunto dos dados de treinamento. Todos os coeficientes foram renormalizados para o intervalo $[-1, 1]$. No caso de PCA, adotou-se um fator de normalização $\gamma(p) = \max_{x_i} (\sum_u |S[p]x_i(u)|)^{\frac{1}{2}}$.

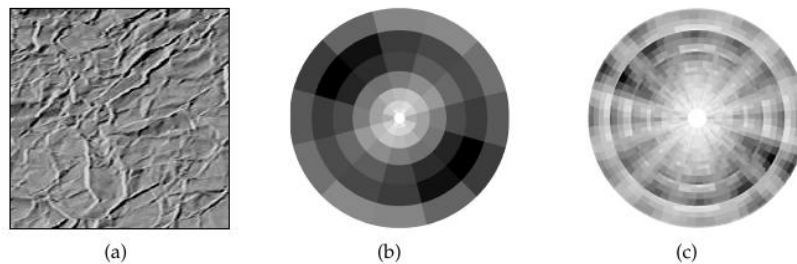


Figura 4.7: (a) Exemplar, (b) Coeficientes de primeira ordem e (c) de segunda ordem.

A Tabela abaixo mostra uma comparação entre o processo de classificação usando diferentes descritores e diferentes classificadores. O resultado apresentado é a taxa de erro em função do número de samples utilizado no treinamento. Nesse caso, exceto em um caso, os melhores resultados foram obtidos usando a representação por coeficientes de scattering. Além disso, como esperava-se, na comparação entre PCA e SVM, o primeiro apresentou melhores resultados quando o classificador foi treinado com poucas amostras, enquanto o segundo foi o melhor com mais amostras.

Training size	x		Wind. Four.		Scat. $m_{\max} = 1$		Scat. $m_{\max} = 2$		Conv. Net.
	PCA	SVM	PCA	SVM	PCA	SVM	PCA	SVM	
300	14.5	15.4	7.35	7.4	5.7	8	4.7	5.6	7.18
1000	7.2	8.2	3.74	3.74	2.35	4	2.3	2.6	3.21
2000	5.8	6.5	2.99	2.9	1.7	2.6	1.3	1.8	2.53
5000	4.9	4	2.34	2.2	1.6	1.6	1.03	1.4	1.52
10000	4.55	3.11	2.24	1.65	1.5	1.23	0.88	1	0.85
20000	4.25	2.2	1.92	1.15	1.4	0.96	0.79	0.58	0.76
40000	4.1	1.7	1.85	0.9	1.36	0.75	0.74	0.53	0.65
60000	4.3	1.4	1.80	0.8	1.34	0.62	0.7	0.43	0.53

O outro caso de teste apresentado por Bruna e Mallat [3, 4] refere-se a classificação de texturas. O uso de PCA junto com a representação por scattering remove grande parte das variabilidade de $SX - E\{SX\}$ em cada classe. Essa variabilidade deve-se a resíduo estocástico (que decai a medida que a escala aumenta), variação de iluminação, rotações, escalas, ou deformações perspectivas.

O processo de classificação de texturas foi realizado sobre a base de dados CURET [5]. Cada exemplar contém $N = 200 \times 200$ pixels. Cada uma das 61 classes de texturas, apresenta exemplares do mesmo material mas com variações de posição e iluminação, o que agrega variabilidade devido a especularidade, sombras e variações de normais.

Nesse experimento utilizou-se 46 exemplares de cada classe para teste de 10 classes distintas. Utilizando validação cruzada obteve-se como melhor escala $2^J = 200$, ie, o tamanho do exemplar. Além disso, também por validação cruzada, escolheu-se a dimensão do espaço como $d = 16$. A Figura 4.7 mostra os coeficientes de scattering de um exemplar dessa base de dados. A Tabela abaixo mostra uma comparação no processo de classificação usando diferentes descritores e classificadores. Observe que o melhor resultado foi obtido usando a representação por scattering e PCA.

Training size	X PCA	MRF 36	Textons 15	BIF 11	Histo. 6	Four. Spectr. PCA	Scat. $\bar{m} = 1$ PCA	Scat. $\bar{m} = 2$ PCA
46	17	2	1.5	1.4	1	1	0.5	0.2

Capítulo 5

Experimentos

5.1 Experimento 1: Wavelets de Gabor

Nesse experimento, foi implementada a Wavelet de Gabor: $\psi(t) = \theta(t)e^{i2\pi Qt}$, onde θ é uma gaussiana, $Q = 16$ e $P = 23$, o que satisfaz a desigualdade da condição de admissibilidade para $\epsilon = 0.02$. A Figura 5.1 apresenta a Wavelet de Gabor e a Figura 5.2 mostra sua função de transferência.

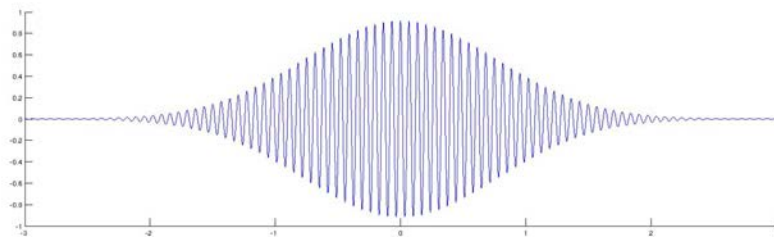


Figura 5.1: Wavelet de Gabor

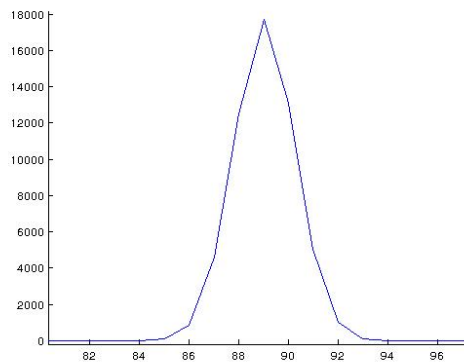


Figura 5.2: Função de Transferência da Wavelet de Gabor

5.2 Experimento 2: Filter Bank

Nesse experimento, tentamos reproduzir o esquema relacionado ao filter bank, apresentado na Figura 5.3. As Figuras 5.4 e 5.5, mostram a transformada de Fourier das ψ_j , com $j = 2, \dots, 40$. O gráfico refere-se a $\frac{2\pi|FFT(\psi_j(k))|}{n}$, com $k = 1, \dots, 3200$.

Há uma diferença de escalas entre a Figura 5.3 e nas Figuras 5.4 e 5.5. Essa escala não ficou clara no artigo, e por isso não foi exatamente reproduzida.

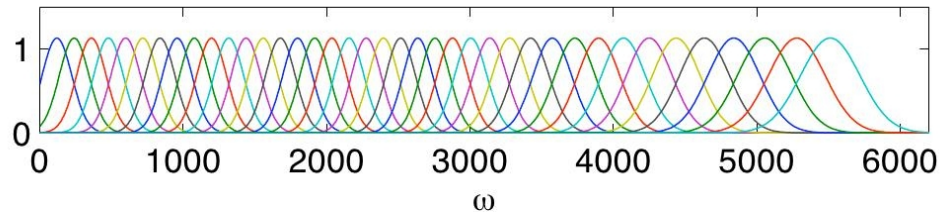


Figura 5.3: Wavelet Filter bank of Gabor Filters at sampling frequency 11025 Hz [1].

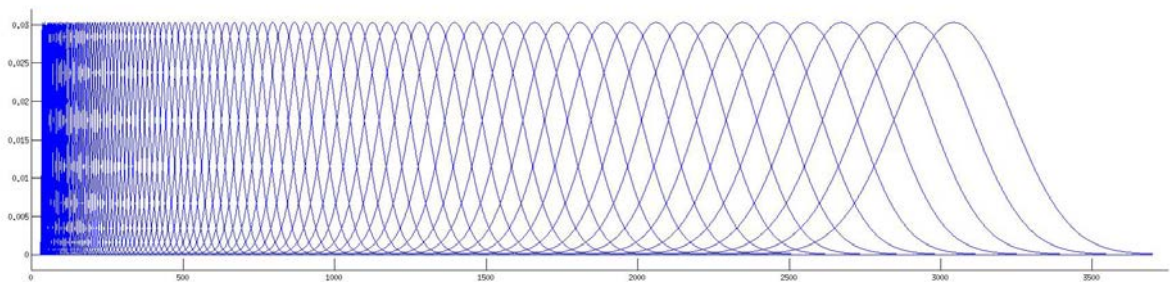


Figura 5.4: Transformada de Fourier das wavelets ψ_j .

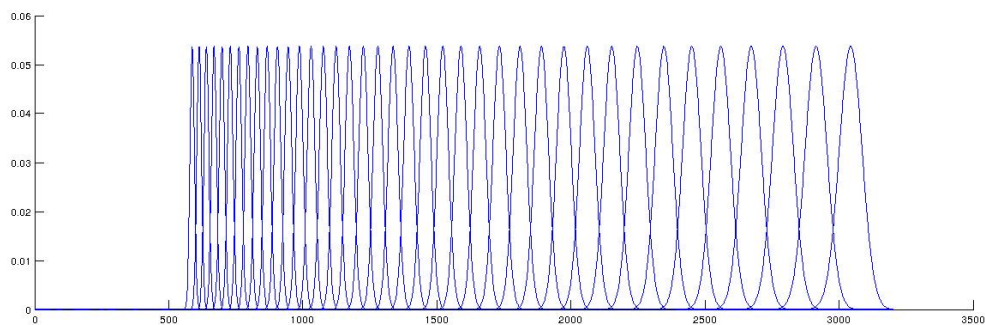


Figura 5.5: Baixas frequências do espectro mostrado na Figura 5.4

5.3 Experimento 3: Transformação Scattering

Dado o sinal apresentado na Figura 5.6 (modelo contido no pacote handel do MatLab) calculamos a transformação scattering com os parâmetros: $q = 1$, $q = 2$, $Q = 16$, $J = 80$ e $T = Q2^{\frac{J}{Q+1}}$.

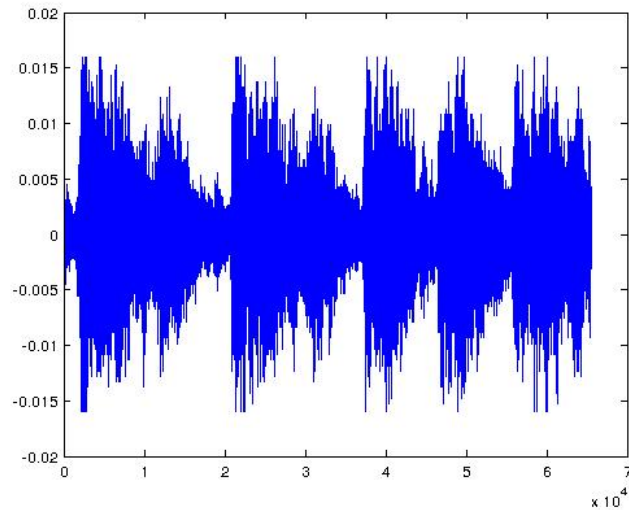


Figura 5.6: Sinal.

A Figura 5.7 mostra os coeficientes scattering de primeira ($q=1$) e segunda ordem ($q=2$).

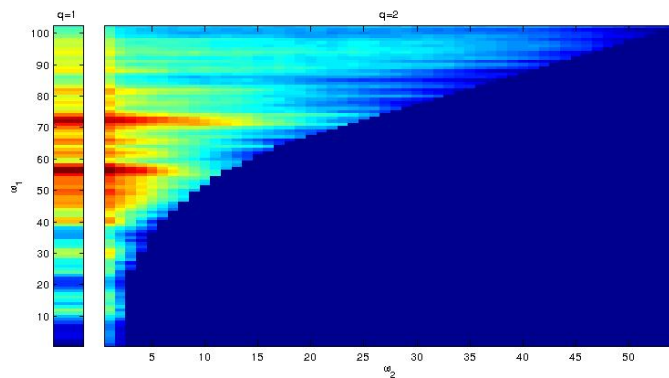


Figura 5.7: Coeficientes Scattering.

5.4 Experimento 4: Transformação Scattering - áudios

Nesse experimento, calculamos os coeficientes de scattering para um conjunto de amostras de áudio de diferentes gêneros.

Vale observar, nesse experimento, as semelhanças entre os coeficientes de amostras da mesma classe, e as diferenças entre amostras de classes distintas.

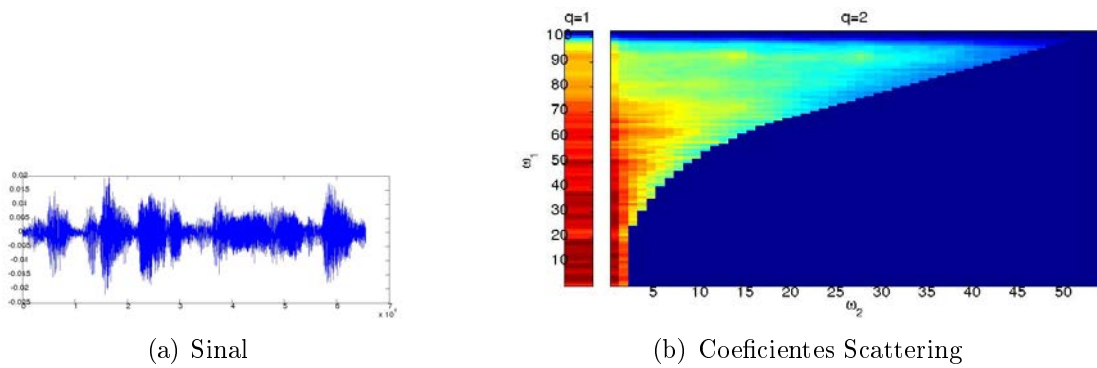


Figura 5.8: Blues - Amostra 1

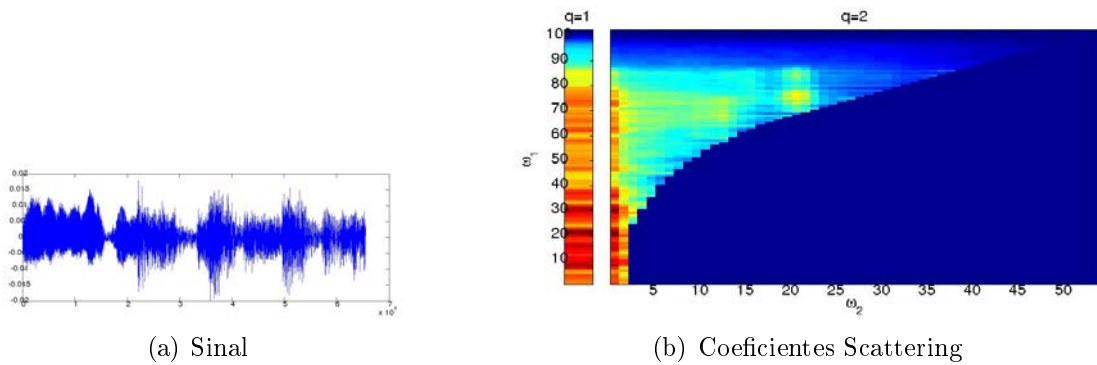


Figura 5.9: Blues - Amostra 2

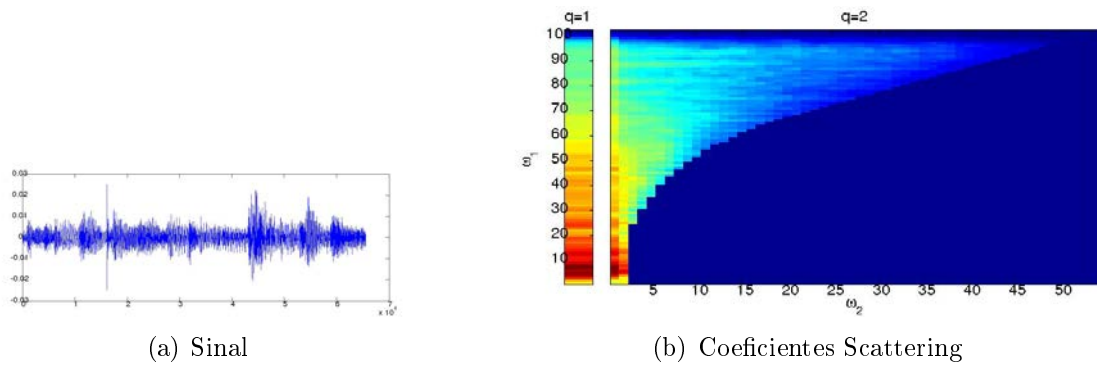


Figura 5.10: Blues - Amostra 3

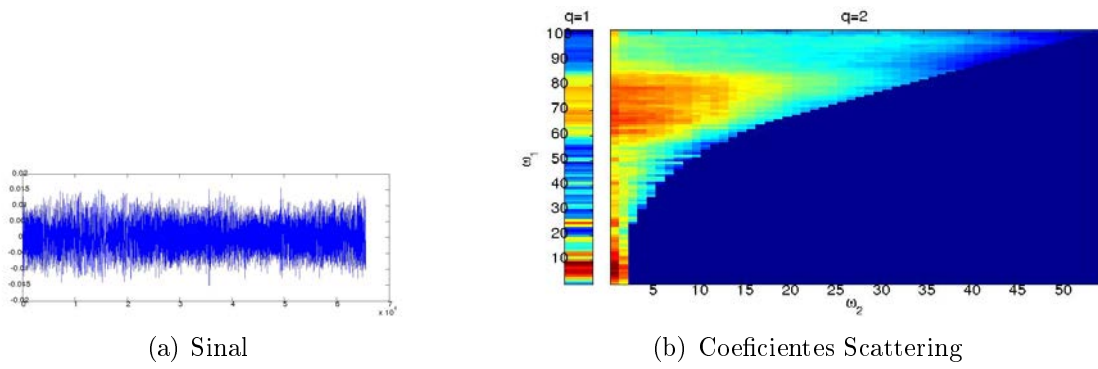


Figura 5.11: Metal - Amostra 1

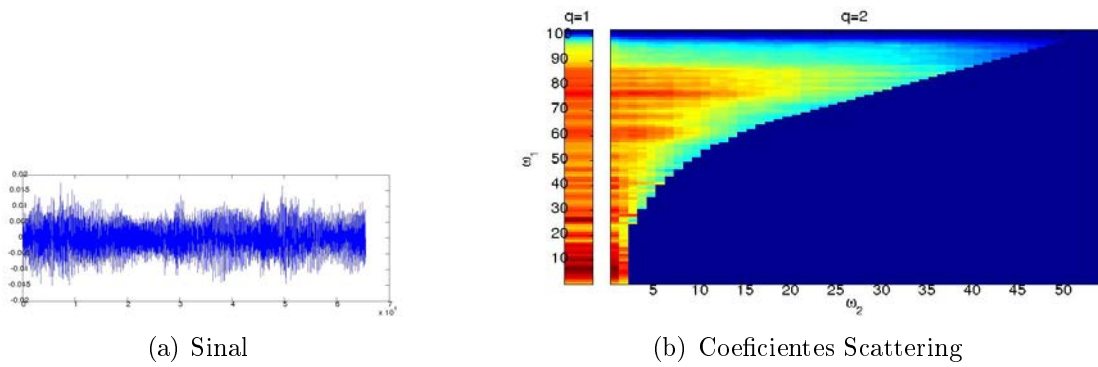


Figura 5.12: Metal - Amostra 2

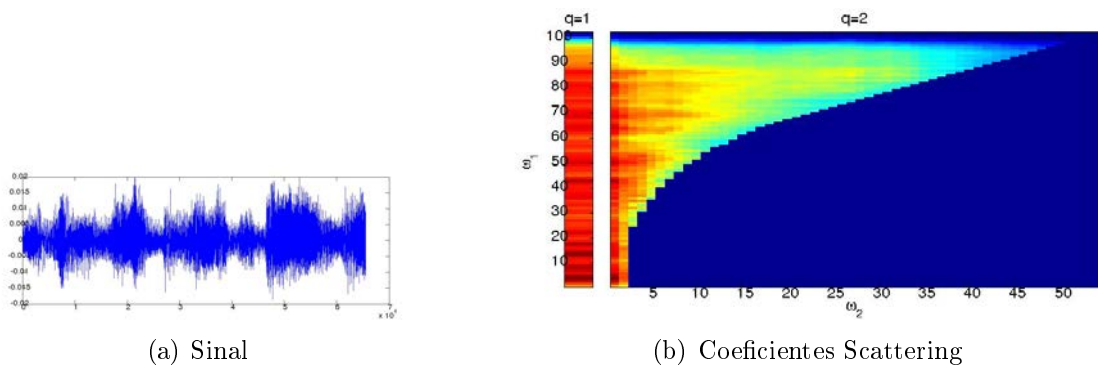


Figura 5.13: Metal - Amostra 3

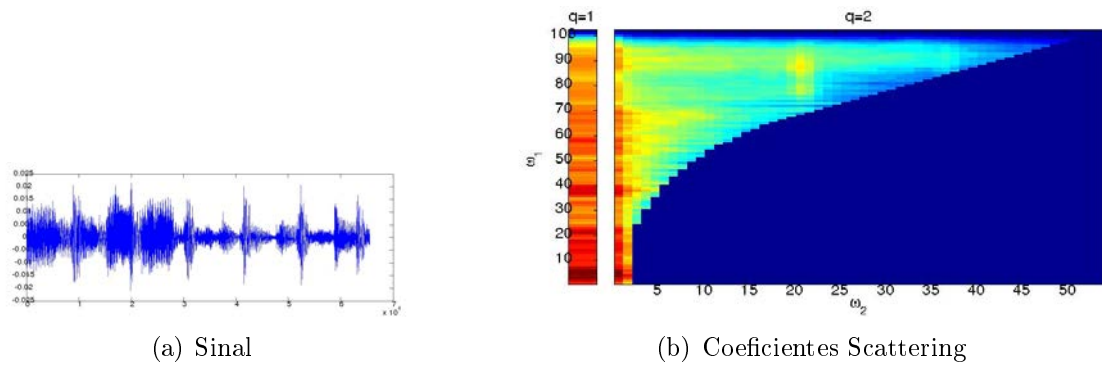


Figura 5.14: Reggae - Amostra 1

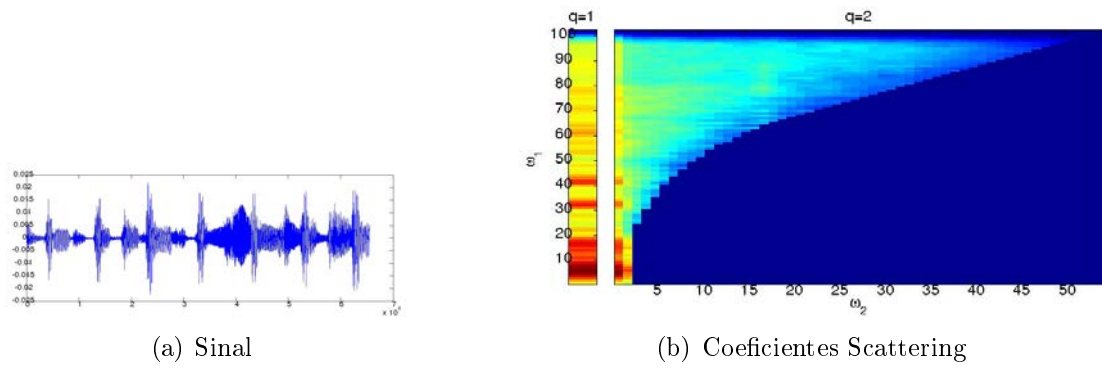


Figura 5.15: Reggae - Amostra 2

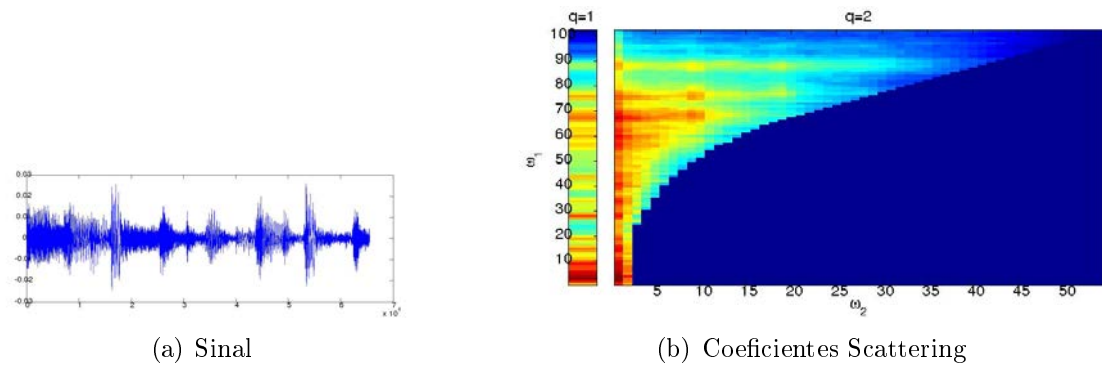


Figura 5.16: Reggae - Amostra 3

5.5 Experimento 5:

Coeficientes de Scattering - exemplares de terreno

Nesse experimento, tomamos uma amostra de terreno x cuja resolução é $N = 64 \times 64$ pixels, conforme mostramos na Figura 5.17.



Figura 5.17: Amostra

A escala adotada foi $J = 5$, ou seja, para cada amostra tomamos uma vizinhança de largura $2^J \times 2^J = 32 \times 32$. Calculamos os coeficientes de scattering com profundidade máxima $m_{max} = 3$, obtendo um total de $P = 11249$ caminhos. Amostrando o exemplar com $\alpha = 1$, para cada caminho p os coeficientes de scattering $S[p]x$ refere-se a uma imagem cuja resolução é $\alpha^{-2}2^{-2J}N = 2^{-2 \times 5}64 \times 64 = 4$. A Figura 5.18 mostra alguns exemplos de $S[p]x$.

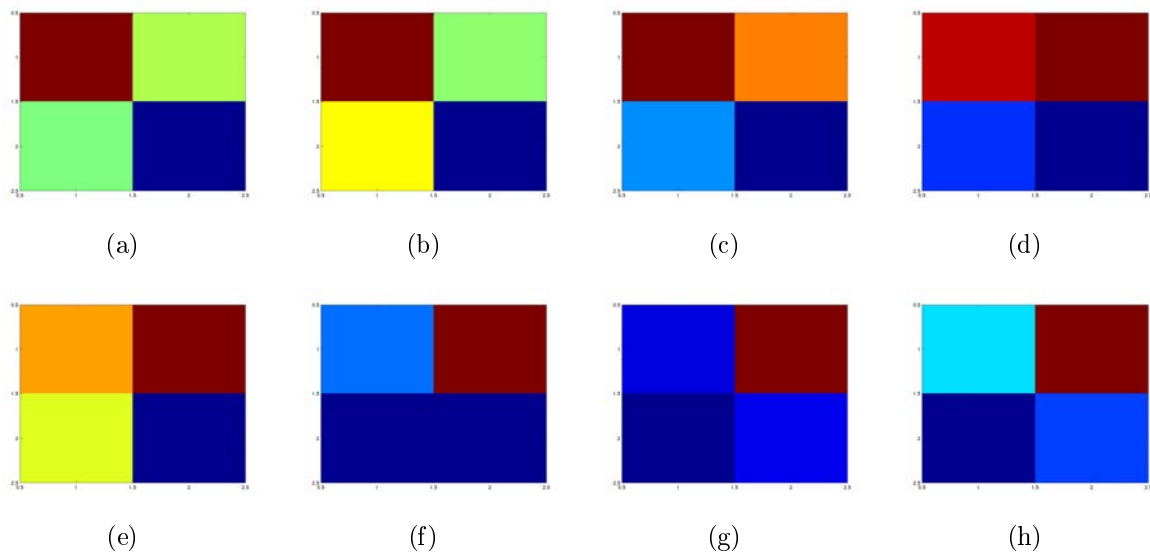
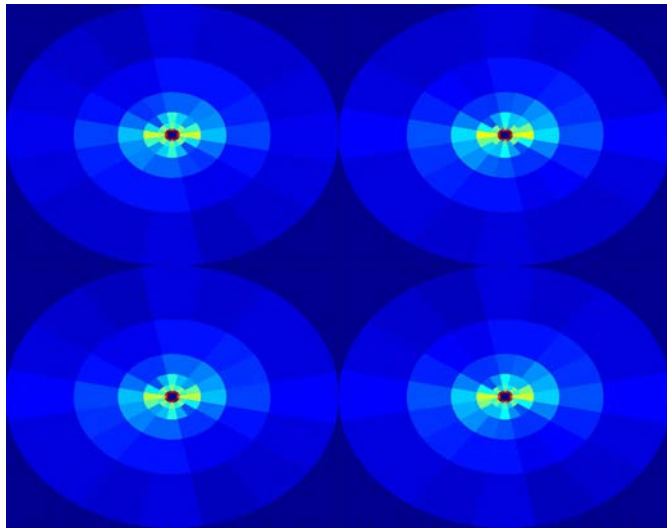
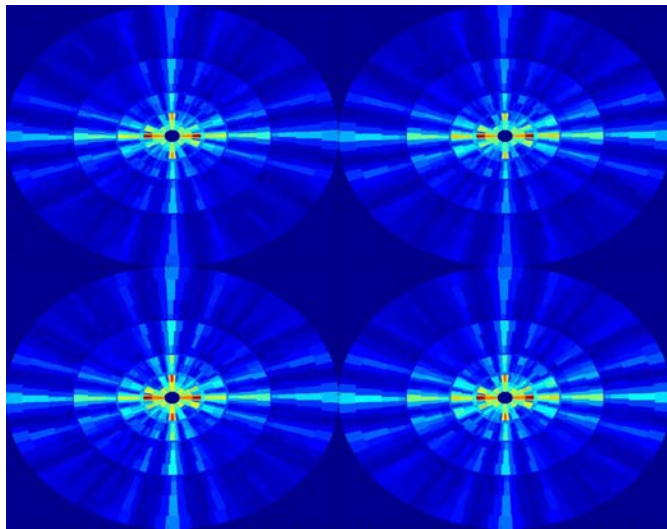


Figura 5.18: $S[p]x$

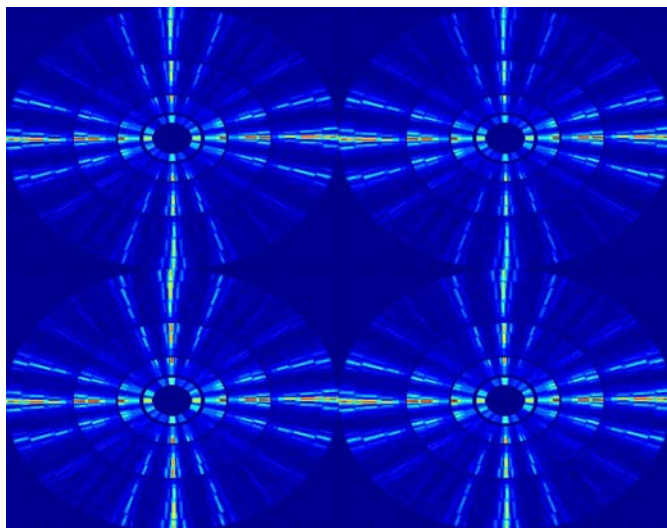
Usando a escala $J = 5$, para cada $m \leq m_{max}$ temos 4 blocos de coeficientes de scattering, como mostra a Figura 5.19.



(a) $m = 1$



(b) $m = 2$



(c) $m = 3$

Figura 5.19:

5.6 Experimento 6:

Transformação Scattering - exemplares de terreno

Nesse experimento, calculamos os coeficientes de scattering para um conjunto de amostras topográficas de diferentes tipos de terrenos.

Abaixo apresentaremos exemplares de diferentes categorias de terrenos e seus respectivos coeficientes de scattering de ordem $m = 1$ e $m = 2$. Todos os coeficientes foram calculados para escala $J = 4$. Os exemplares tem resolução 256×256 pixels. Nessa escala, tomou-se 16×16 vizinhanças de $2^J \times 2^J = 16 \times 16$ pixels.

Atenção: *Os coeficientes de scattering exibidos referem-se a uma janela de tamanho 16×16 da imagem. Como os exemplares têm resolução 256×256 , para cada exemplar mostraremos apenas um dos 256 conjuntos de coeficientes ($\frac{256 \times 256}{16 \times 16} = 16 \times 16 = 256$).*

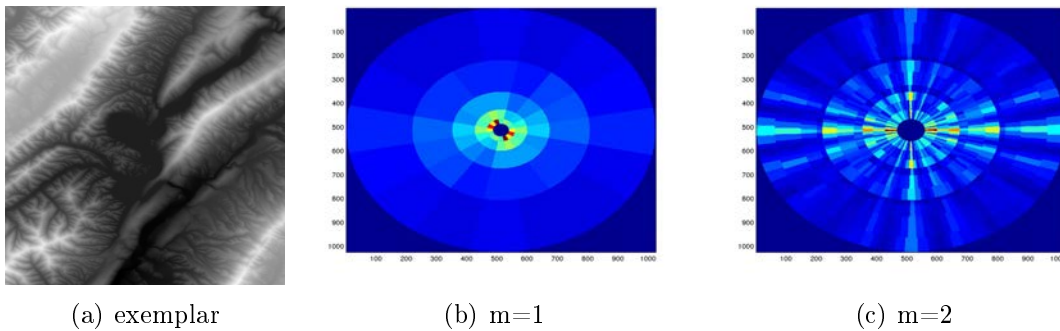


Figura 5.20: Chest - Amostra 1;

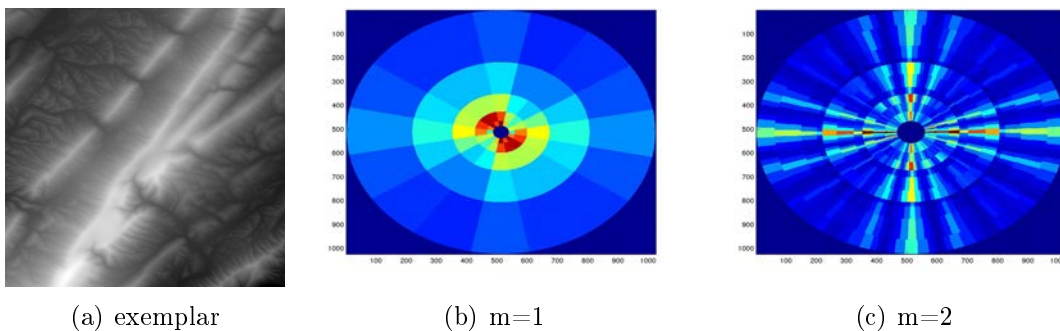
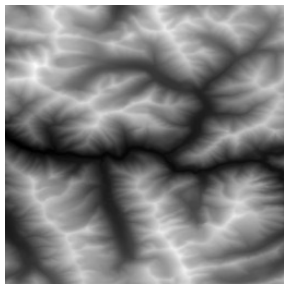
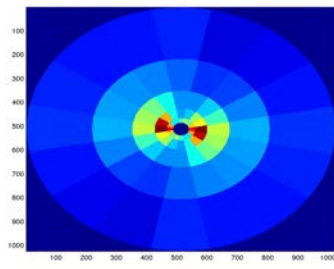


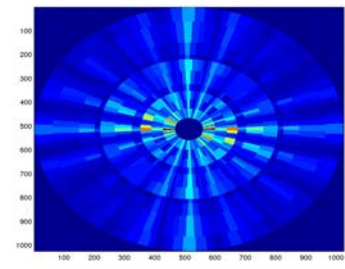
Figura 5.21: Chest - Amostra 2;



(a) exemplar

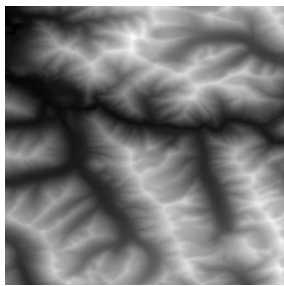


(b) $m=1$

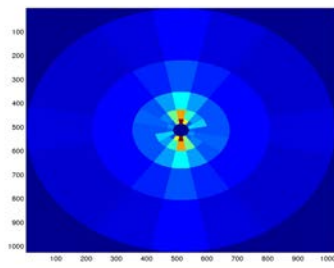


(c) $m=2$

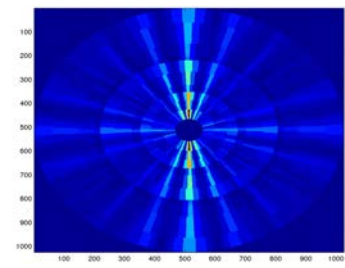
Figura 5.22: Flat - Amostra 1;



(a) exemplar

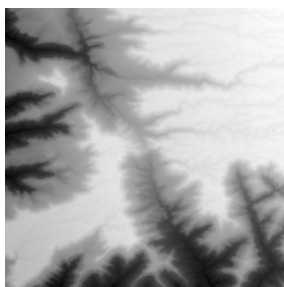


(b) $m=1$

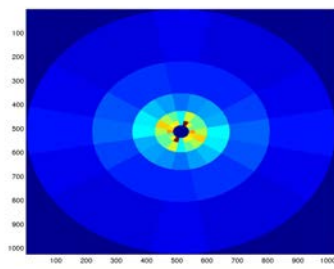


(c) $m=2$

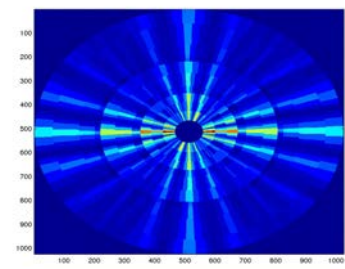
Figura 5.23: Flat - Amostra 2;



(a) exemplar

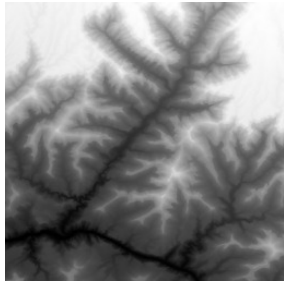


(b) $m=1$

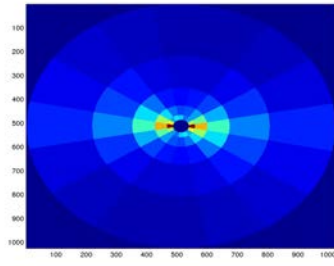


(c) $m=2$

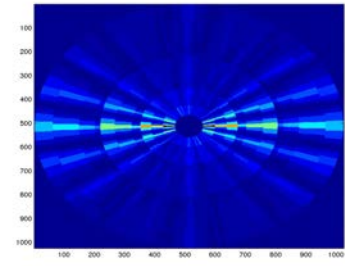
Figura 5.24: Grand - Amostra 1;



(a) exemplar

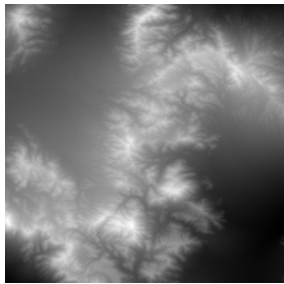


(b) $m=1$

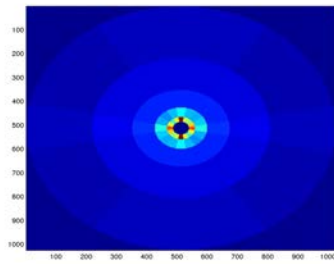


(c) $m=2$

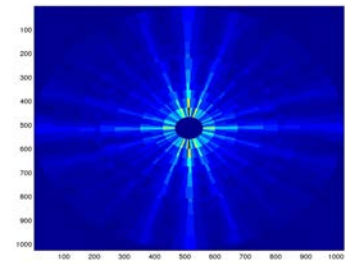
Figura 5.25: Grand - Amostra 1;



(a) exemplar

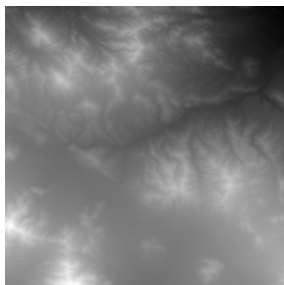


(b) $m=1$

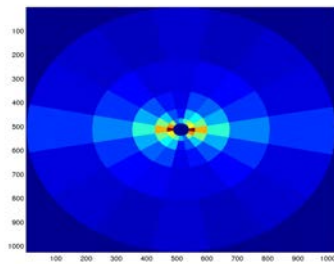


(c) $m=2$

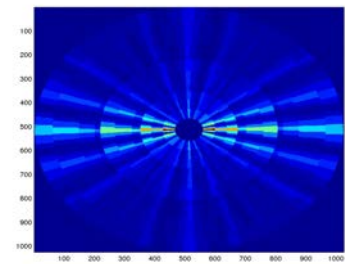
Figura 5.26: Nest - Amostra 1;



(a) exemplar



(b) $m=1$



(c) $m=2$

Figura 5.27: Nest - Amostra 1;

5.7 Experimento 7:

Transformação Scattering - Variando a escala dos coeficientes de scattering em exemplares de terreno

Esse experimento consiste no cálculo de coeficientes de scattering do exemplar de terreno apresentado na Figura 5.28 em diferentes escalas, com oito rotações.

Atenção: Para $J = 1$ não temos co-ocorrência de frequências, pois o filter bank é composto por apenas uma wavelet. Logo não há coeficientes de ordem maior do que 1. Analogamente, para $J = 2$, só há coeficientes de ordem até $m = 2$. Para os demais casos apresentamos os coeficientes com $m_{max} = 3$. Note que os coeficientes de ordem 3 possuem energia bem baixa. Os de ordem maior, tem energia quase nula, por isso optamos por não exibi-los.

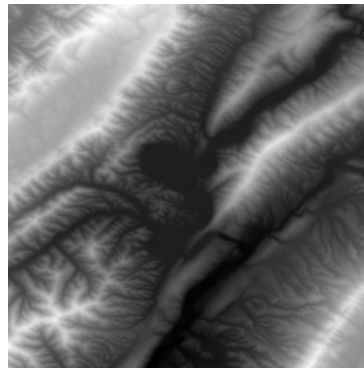
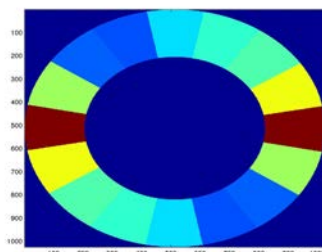
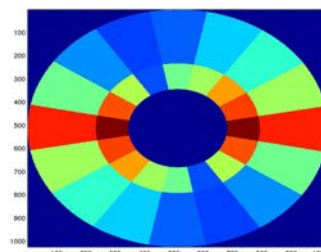


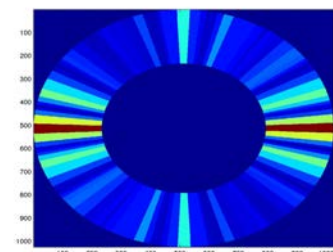
Figura 5.28: Exemplar escolhido - Chest



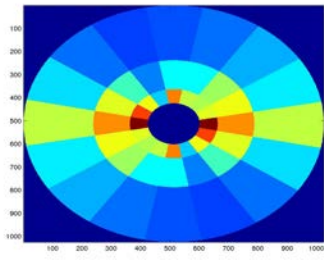
(a) $J = 1$ e $m = 1$



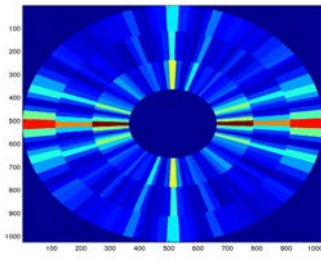
(b) $J = 2$ e $m = 1$



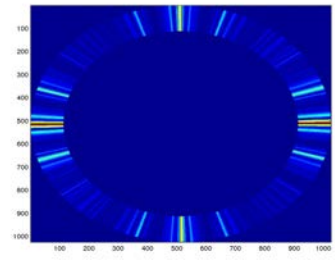
(c) $J = 2$ e $m = 2$



(d) $m=1$

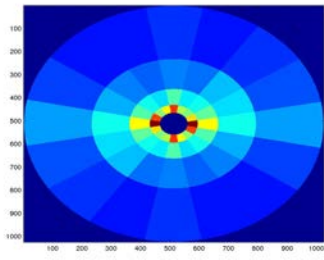


(e) $m=2$

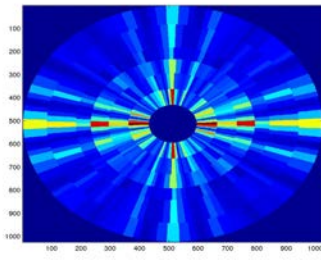


(f) $m=3$

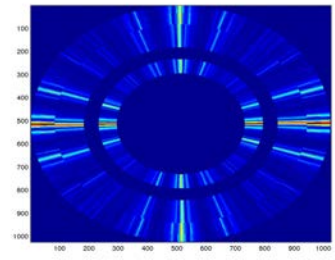
Figura 5.29: $J = 3$



(a) $m=1$

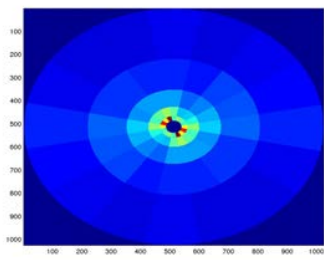


(b) $m=2$

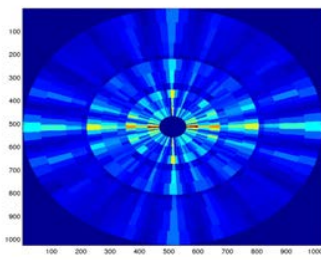


(c) $m=3$

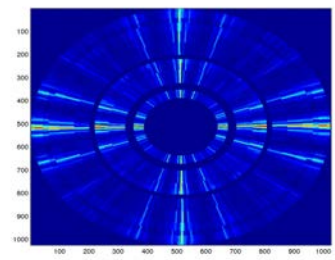
Figura 5.30: $J = 4$



(a) $m=1$

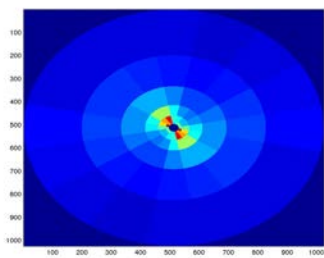


(b) $m=2$

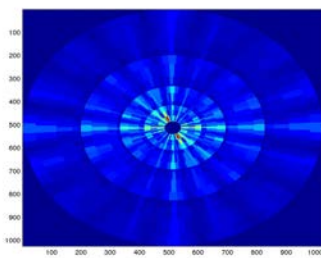


(c) $m=3$

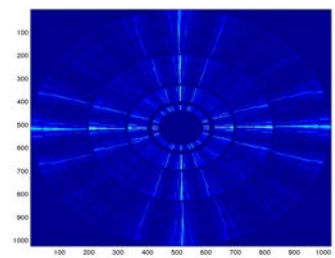
Figura 5.31: $J = 5$



(a) $m=1$



(b) $m=2$



(c) $m=3$

Figura 5.32: $J = 6$

Capítulo 6

Resultados e Conclusões

Ao longo desse relatório discutimos o problema de buscar bons descritores para um sinal de modo que estes maximizem a taxa de acerto no processo de classificação. Uma conclusão obtida é que a combinação da representação por coeficientes de scattering com classificadores como PCA e SVM é muito boa para sinais com comportamentos estacionários.

Ao longo desse relatório, apresentamos alguns testes realizados com essa combinação. Esses testes referiram-se a classificação de áudio (exemplares de trechos de músicas de diferentes estilos), classificação de dígitos e de texturas. Desses experimentos, o que apresentou melhores resultados foi o relacionado com texturas.

Por que a representação por scattering funcionou tão bem no caso de texturas? Pois estas podem ser pensadas como processos estacionários, com estruturas que se repetem ao longo do espaço. Isso está muito relacionado com as invariâncias dos coeficientes de scattering. Esses são invariantes a translações, rotações e pequenas deformações.

Nesse relatório, também apresentamos alguns experimentos relacionados com o problema de representação e classificação de exemplares de terrenos. Um dos objetivos futuros é descobrir como adequar a teoria de scattering para representar os terrenos, de modo que possamos realizar classificação desses exemplares. Além disso, outra etapa que desejamos fazer com a representação por scattering é sintetizar exemplares de uma determinada classe.

Referências Bibliográficas

- [1] Joakin Anden and Stephane Mallat. Multiscale scattering for audio classification. ISMIR, 2011.
- [2] Joan Bruna. Operators commuting with diffeomorphisms. Technical Report of CMAP - Ecole Polytechnique, 2012.
- [3] Joan Bruna and Stephane Mallat. Classification with scattering operators. IEEE CVPR, 2011.
- [4] Joan Bruna and Stephane Mallat. Invariant scattering convolution network. IEEE Transactions on PAMI, 2012.
- [5] Kristin J. Dana, Bram Van Ginneken, Shree K. Nayar, and Jan J. Koenderink. Curet: Columbia-utrecht reflectance and texture database. <http://www.cs.columbia.edu/CAVE/software/curet/>.
- [6] Yann LeCun and Corinna Cortes. The mnist database of handwritten digits. <http://pr.cs.cornell.edu/sceneunderstanding/data/data.php>.
- [7] Yann LeCun, K. Kavukvuoglu, and C. Farabet. Convolution networks and applications in vision. ISCAS, 2010.
- [8] David G Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004.
- [9] Stephane Mallat. Recursive interferometric representation. EUSIPCO, 2010.
- [10] Stephane Mallat. Group invariant scattering. Communications in Pure and Applied Mathematics., 2012.
- [11] E. Tola. Daisy: An efficient dense descriptor applied to wide-baseline stereo. PAMI, 2010.
- [12] I. Waldspurger and Stephane Mallat. Wavelet and scattering phase retrieval. Technical Report of CMAP - Ecole Polytechnique.

- [13] Irene Waldspurger, Alexandre D'Aspremont, and Stephane Mallat. Phase recovery, maxcut and complex semidefinite programming. IEEE CVPR, 2011.