

# Uso de descritores de textura extraídos de GLCM para o reconhecimento de padrões em diferentes domínios de aplicação

Jefferson Gustavo Martins<sup>1,2</sup>

*Universidade Tecnológica Federal do Paraná  
Toledo, Brazil*

Yandre Maldonado e Gomes da Costa<sup>1,3</sup>

*Universidade Estadual de Maringá  
Maringá, Brazil*

Diego Bertolini Gonçalves<sup>1,4</sup>

*Universidade Federal do Paraná  
Curitiba, Brazil*

Luiz Eduardo Soares Oliveira<sup>1,5</sup>

*Universidade Federal do Paraná  
Curitiba, Brazil*

---

## Abstract

This paper presents the application of a statistical approach for texture representation in classification problems situated in three different domains. The applied approach is the well-known gray level co-occurrence matrix, which stores the probability of two gray scale values are involved by a certain spatial relationship. In order to represent the image texture, statistical measures are extracted from this matrix. The final results presented here show that the use of these descriptors is feasible to achieve good classification rates in different application domains. The best rates were achieved by employing the Support Vector Machine classifier. These rates were: 97.43% ( $\sigma = 0.59$ ) for forest species classification; 94.49% ( $\sigma = 0.66$ ) for autorship verification; and 60.11% ( $\sigma = 9.06$ ) for music genre recognition.

*Keywords:* gray level co-occurrence matrix, pattern recognition, texture.

---

# 1 Introdução

Reconhecimento de padrões vem sendo empregado com sucesso em tarefas de classificação voltadas para os mais diferentes domínios de aplicação. O modelo clássico para a construção de um sistema de reconhecimento de padrões prevê as etapas de aquisição, pré-processamento, segmentação, extração de características e classificação. Na maioria das vezes, o sinal que representa os modelos a serem classificados se apresentam, após adquiridos, no formato de imagens digitais. Este formato tem se mostrado uma solução natural para o problema de buscar uma representação comum para as mais variadas situações.

Uma vez representados no formato de imagem digital, procura-se extrair características desta imagem a partir das quais se possa descrever da melhor forma os conteúdos presentes na mesma. É vasta a literatura acerca de técnicas para representar conteúdos presentes em imagens digitais relacionados aos principais atributos visuais encontrados nelas, tais como: cor, forma, estrutura e textura entre outros. Dentre estes atributos, a textura é particularmente promissora, dado o seu potencial de discriminação do conteúdo das imagens.

A textura é um importante atributo visual presente nas imagens do mundo real. Esta corresponde a um padrão visual, geralmente relacionado à distribuição de pixels em uma região e características do objeto da imagem, como tamanho, brilho e cor. Este atributo geralmente contém informações bastante significativas acerca do conteúdo da imagem e é amplamente explorado em aplicações de visão computacional.

Algumas texturas apresentam uma regularidade no que diz respeito a repetição de padrões que aparecem na mesma, enquanto outras não. Assim como cor e forma, a textura é facilmente percebida pelo olho humano e contribui com a identificação de objetos em uma dada cena. Porém, este atributo não possui uma definição formal e, conforme Jain e Farrokhnia [10], a diversidade de texturas naturais e artificiais torna impossível estabelecer uma definição universal para a mesma. Neste sentido, a figura 1(c) exemplifica uma textura regular, enquanto as demais presentes na figura 1 são texturas irregulares.

Considerando esta diversidade, este trabalho apresenta a aplicação de Matriz de Co-ocorrências em níveis de cinza para representar conteúdo de textura em tarefas de classificação voltadas a diferentes domínios de aplicação. Assim, sua principal contribuição consiste em demonstrar o potencial destes descritores para a extração de características em: classificação de espécies florestais, a partir de imagens obtidas de amostras de madeiras; verificação de autoria

---

<sup>1</sup> Os autores agradecem a CAPES e a Fundação Araucária pelo financiamento da pesquisa.

<sup>2</sup> Email: [martins@utfpr.edu.br](mailto:martins@utfpr.edu.br)

<sup>3</sup> Email: [yandre@din.uem.br](mailto:yandre@din.uem.br)

<sup>4</sup> Email: [diegob@inf.ufpr.br](mailto:diegob@inf.ufpr.br)

<sup>5</sup> Email: [lesoliveira@inf.ufpr.br](mailto:lesoliveira@inf.ufpr.br)

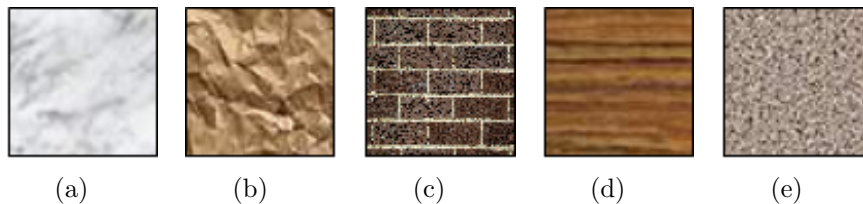


Figura 1. Amostras de textura: (a) mármore, (b) papel, (c) parede, (d) madeira, (e) areia.

de textos, a partir de amostras de manuscritos colhidos de diferentes escritores; e classificação de gêneros musicais, a partir de imagens de espectrogramas extraídas do sinal de áudio de músicas latinoamericanas.

A automação de processos de classificação é, em geral, muito oportuna nestes domínios, pois além de eliminar a necessidade de mão de obra humana, evitam problemas relacionados à subjetividade da percepção humana. Além disso, os resultados aqui apresentados mostram que estes descritores produzem resultados comparáveis aos alcançados por outros sistemas de classificação recentemente apresentados na literatura.

A organização deste trabalho encontra-se da seguinte forma: na seção 2 a Matriz de Co-ocorrência é descrita; na seção 3 são descritos os resultados experimentais obtidos nos três diferentes domínios de aplicação supracitados; na seção 4 são apresentadas as considerações finais e trabalhos futuros.

## 2 Matriz de Co-ocorrências

Matriz de co-ocorrência de níveis de cinza (Gray Level Co-occurrence Matrix - GLCM) é um método estatístico para descrição de texturas. A GLCM armazena a probabilidade de que dois valores de intensidade de cinza estejam envolvidos por uma determinada relação espacial. A partir desta matriz de probabilidades, diferentes medidas estatísticas são extraídas a fim de caracterizar a textura presente na imagem. Este método foi proposto em [9] e ainda é empregado com sucesso em diferentes domínios de aplicação.

A partir da matriz de co-ocorrência, pode-se extrair várias medidas relacionadas a características de uma textura. Nos experimentos aqui relatados, para representar textura foram extraídas as medidas de contraste, energia, entropia, correlação, homogeneidade, momento de terceira ordem e probabilidade máxima. Este conjunto corresponde apenas à uma parte das 14 características originalmente apresentadas por Haralick *et al.*. Entretanto, alguns trabalhos posteriores demonstraram que muitas das características propostas originalmente eram correlacionadas, e seu uso simultâneo não traz benefícios na maioria dos casos. Para cada uma das aplicações aqui abordadas foi empregado um subconjunto de características diferentes, maiores detalhes são apresentados na subseções 3.1, 3.2 e 3.3.

A GLCM armazena a probabilidade de que dois valores de intensidades de cinza estejam envolvidos por uma determinada relação espacial. Assim, uma GLCM corresponde a uma matriz de ordem  $N_g \times N_g$ . Parâmetros como a distância  $d$  entre os pixels e o ângulo  $\theta$  caracterizado pela orientação da reta que passa pelos pixels são importantes na caracterização da relação espacial. Sendo  $p(i, j)$  a probabilidade de ocorrência das intensidades de cinza  $i$  e  $j$  observando a distância  $d=1$  e com um dado ângulo  $\theta$ , as características utilizadas neste trabalho são encontradas pelas seguintes equações:

$$\begin{aligned}
 (1) \quad \text{Contraste} &= \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i, j) \\
 (2) \quad \text{Energia} &= \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p(i, j))^2 \\
 (3) \quad \text{Entropia} &= - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log p(i, j) \\
 (4) \quad \text{Homogeneidade} &= \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i, j)}{1 + (i - j)^2} \\
 (5) \quad \text{Momento de terceira ordem} &= \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) (i - j)^3 \\
 (6) \quad \text{Probabilidade máxima} &= \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \max p(i, j) \\
 (7) \quad \text{Correlação} &= \frac{p(i, j) - \mu_x \mu_y}{\sigma_x^2 \sigma_y^2}
 \end{aligned}$$

onde  $\mu_x = \sum_{i=1}^{N_g} i \times p_x(i)$ ,  $p_x(i) = \sum_{j=1}^{N_g} p(i, j)$ ,  $\sigma_x^2 = \sum_{i=1}^{N_g} (i - \mu_x)^2 p_x(i)$ ,  $\mu_y = \sum_{j=1}^{N_g} j \times p_y(j)$ ,  $p_y(j) = \sum_{i=1}^{N_g} p(i, j)$  e  $\sigma_y^2 = \sum_{j=1}^{N_g} (j - \mu_y)^2 p_y(j)$ .

A próxima seção mostra detalhes acerca de quais e como estas características foram empregadas nas tarefas de classificação.

### 3 Experimentos realizados

O classificador empregado nestes experimentos foi o *Support Vector Machine* (SVM)[17]. A normalização dos dados foi feita de forma que os valores ficassem compreendidos em uma escala de variação entre -1 e 1. Os parâmetros  $C$  e  $\gamma$  do kernel Gaussiano foram otimizados utilizando uma busca gulosa. Para uma

melhor análise dos resultados, gerou-se três diferentes conjuntos e empregou-se validação cruzada *3-fold* a fim de se avaliar a variação entre os conjuntos. Assim, as taxas de acertos e os desvios padrão apresentados a seguir referem-se à média destes três experimentos, nos quais duas partições foram combinadas e utilizadas para treinamento enquanto a terceira foi utilizada para teste.

Nos experimentos descritos na subseção 3.1 também foram utilizados LDA (*Linear Discriminant Analysis*) e kNN (*k-Nearest Neighbors*) para realizar a classificação.

### 3.1 Reconhecimento de espécies florestais

Um modo de simplificar a tarefa de classificação de espécies florestais consiste na identificação de suas famílias ou classes taxonômicas. Neste contexto, espécies florestais podem ser categorizadas em Angiospermas e Gimnospermas. Angiospermas são as plantas que apresentam formas mais sofisticadas e se adaptaram para sobreviver em um conjunto mais diversificado de ambientes e condições climáticas. Este grupo é constituído por plantas ornamentais que apresentam flores e por todos os vegetais que produzem frutos comestíveis. Além disso, os exemplares pertencentes a este grupo também compreendem espécies com valor comercial, são fonte de madeira para todo o mundo e são vulgarmente denominadas *hardwood*. Gimnospermas, de outro modo, compreendem um conjunto de exemplares considerados primitivos em razão de sua anatomia e o menor número de características, dificultando inclusive a diferenciação entre suas espécies. Também denominados *softwood* as espécies que compõem este grupo possuem menor valor comercial devido às propriedades da madeira e a suas possibilidades de aplicação. Neste grupo, a produção de madeira provém de um subconjunto denominado coníferas, dentre as quais tem-se cedros, sequóias, pinheiros e pinos. Além da produção de madeira, algumas espécies do grupo *softwoods* são empregadas para a fabricação de papel e compensados em geral, além de também serem fontes de matéria-prima para a produção de essências e resinas.

A base de dados empregada neste experimento é composta por 600 imagens de 15 diferentes espécies florestais adquiridas no Laboratório de Anatomia da Madeira da Universidade Federal do Paraná (UFPR). Estas espécies estão relacionadas na Tabela 1, sendo 5 espécies pertencentes ao grupo Gimnospermas e 10 ao grupo Angiospermas, cada uma com 40 imagens e totalizando 200 e 400 imagens, respectivamente.

Os procedimentos adotados no processo de coleta das imagens compreenderam 7 passos. O primeiro consistiu na extração de amostras caracterizadas por pequenos blocos com aproximadamente  $2cm^3$  de peças maiores de cada uma das espécies. Em seguida realizou-se o cozimento dos blocos de madeira, por tempo variado de acordo com a espécie florestal, para seu amolecimento.

Tabela 1  
Espécies que compõem a Base de Dados.

| Gimnospermas                  | Angiospermas                 |
|-------------------------------|------------------------------|
| <i>Araucaria Angustifolia</i> | <i>Dinizia excelsa</i>       |
| <i>Pinus elliottii</i>        | <i>Cedrela sp</i>            |
| <i>Pinus maximinoi</i>        | <i>Dalbergia nigra</i>       |
| <i>Pinus taeda</i>            | <i>Eucalyptus grandis</i>    |
| <i>Podocarpus lambertii</i>   | <i>Grevillea robusta</i>     |
|                               | <i>Tabebuia sp</i>           |
|                               | <i>Hymenaea sp</i>           |
|                               | <i>Nectandra sp</i>          |
|                               | <i>Porcelia macrocarpa</i>   |
|                               | <i>Swietenia macrophylla</i> |

Após o cozimento, realizou-se os cortes histológicos (pequenas amostras) de madeira, com espessura de aproximadamente 25 micras<sup>6</sup>, de forma paralela ao eixo transversal (eixo X na Figura 2(a)) e com o emprego de um micrótomo de deslizamento. Por fim, no sétimo passo realizou-se a coleta das imagens com o auxílio de um microscópio Olympus modelo CX40 produzindo imagens com aproximação ótica de 100 vezes e resolução de 1024×768 pixels.

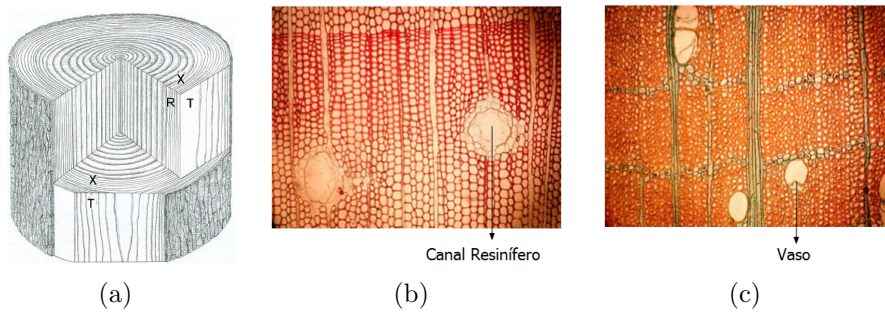


Figura 2. (a) Planos de corte (b) Amostra de Gimnosperma e (c) Amostra de Angiosperma.

### 3.1.1 Caraterísticas

Conforme já mencionado, a maioria dos trabalhos que utilizam GLCM descritos literatura consideram diferentes subconjuntos destes e afirmam haver cor-

<sup>6</sup> Uma micra equivale à milionésima parte de um metro ou  $1 \times 10^{-6}$  metro.

relações ou redundâncias entre as informações provenientes de alguns deles. Assim, neste experimento optou-se por considerar o seguinte subconjunto com seis características: Energia, Contraste, Entropia, Homogeneidade, Probabilidade Máxima e Momento de Terceira Ordem. Estas características foram extraídas de diferentes combinações quanto a distâncias e ângulos. Os melhores resultados foram obtidos com  $d = 1$  e  $\theta = [0, 45, 90, 135]$ , sendo que os seis descritores mencionados foram calculados para os quatro sentidos e concatenados, gerando um vetor de características com 24 componentes.

### 3.1.2 Resultados

No experimento, as 24 características extraídas das GLCM foram aplicadas aos 3 classificadores. Os resultados são apresentados na Tabela 2 e demonstram que todos os classificadores alcançaram resultados similares com aproximadamente 2% de vantagem para SVM. Diante do fato de que estes são os primeiros experimentos realizados com a referida base, não há referências que permitam a avaliação do desempenho por meio de comparações de experimentos.

Tabela 2  
Taxa de reconhecimento por classificador.

| Classificador                     | LDA   | kNN( $k = 4$ ) | SVM   |
|-----------------------------------|-------|----------------|-------|
| <b>Taxa de reconhecimento (%)</b> | 95.06 | 95.55          | 97.43 |
| $\sigma$                          | 0.18  | 0.87           | 0.59  |
| <b>Distância</b>                  | 4     | 5              | 4     |

### 3.2 Verificação de autoria

O formato digital vem sendo cada vez mais empregado para representar e armazenar dados. Contudo, sabe-se que textos manuscritos são documentos que não se apresentam originalmente neste formato e existem em grande quantidade. Ao se observar um conjunto de textos escritos por um mesmo autor, percebe-se muitas características implícitas e únicas que podem ser utilizadas de maneira a distinguir um indivíduo de outro. A partir disso, a ciência forense concentra seus estudos buscando, através de peritos, analisar a possibilidade da existência de fraudes e falsificações de documentos.

Assim, muitos estudos nesta área concentram-se na busca por métodos para análise e autenticação de manuscritos geralmente subdivididos em: identificação e verificação de autoria. Identificação refere-se ao processo de recuperação de amostras a partir de uma base de dados de manuscritos, enquanto verificação refere-se o processo de verificação basicamente classifica uma amostra de manuscrito como sendo, pertencente ou não pertencente à um determi-



nado autor [3][5][1]. O método aqui apresentado tem o objetivo de verificar autoria através de seu conteúdo textural.

### 3.2.1 Descrição do Problema

Os principais fatores que fazem com que sistemas de verificação e identificação de autoria não sejam triviais são: as fortes variações de características intrapessoal e a possíveis similaridades existentes entre falsificações e amostras genuínas de manuscritos (interpessoal). Outro problema é que a escrita de uma pessoa pode sofrer alterações ao longo dos anos, isso devido a uma série de fatores físicos e psicológicos intrínsecos a cada um.

Na amostra de manuscritos assim como em assinaturas, podemos observar que a variação intrapessoal e a similaridade entre uma falsificação e um exemplo genuíno pode ser alta. A figura 3 demonstra tais particularidades.

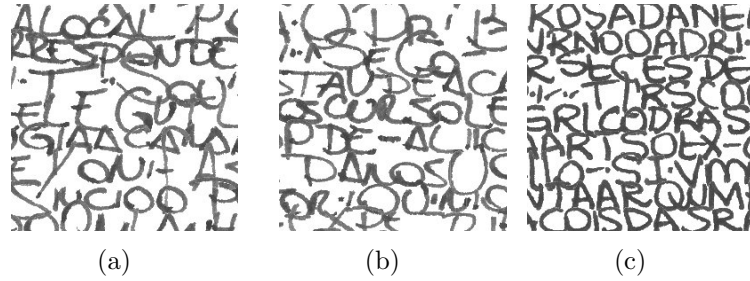


Figura 3. Exemplos de amostras de cartas fragmentadas compactadas: 3(a) e 3(b) variação intrapessoal. 3(b) e 3(c) similaridade entre amostras de diferentes escritores.

### 3.2.2 Base PUC-PR

Proposta por Baranoski [2], a base de cartas original da PUC-PR (Pontifícia Universidade Católica do Paraná), obtida entre 2002 e 2005 através da coleta com voluntários, foi elaborada devido a falta de uma base latina com escrita em língua portuguesa. Cada voluntário redigiu uma carta, cujo conteúdo foi elaborado de forma a maximizar o uso do conjunto de letras do alfabeto da língua portuguesa (minúsculas e maiúsculas). A carta utiliza um total de 124 palavras da língua portuguesa e envolve também algumas particularidades da língua escrita, como símbolos de acentuação e mínimos gráficos. Segundo o autor, todas as regras quanto a transcrição foram obedecidas, como uso de caneta esferográfica azul ou preta, não hifenização em palavras (caso falte espaço) e a escrita do texto sem auxílio de linhas guia.

Os exemplares físicos (manuscritos) foram digitalizados e as imagens foram obtidas com resolução de 300 *dpi*, 256 tons de cinza e foram armazenadas no formato *bitmap*. Ao todo, os manuscritos foram produzidos por 315 autores,



sendo 3 amostras por autor, totalizando assim 945 imagens. A Figura 4(a), apresenta a imagem original de um manuscrito da base PUC-PR.

A partir da base de cartas da base original, foi elaborada a base compactada, uma base sem espaços em branco entre linhas e entre palavras. De posse da base compactada, foi realizada uma fragmentação das imagens. De cada imagem da base de cartas compactada foram produzidos 9 fragmentos de tamanho  $256 \times 256$ , resultando um total de 27 fragmentos por autor 4(b).

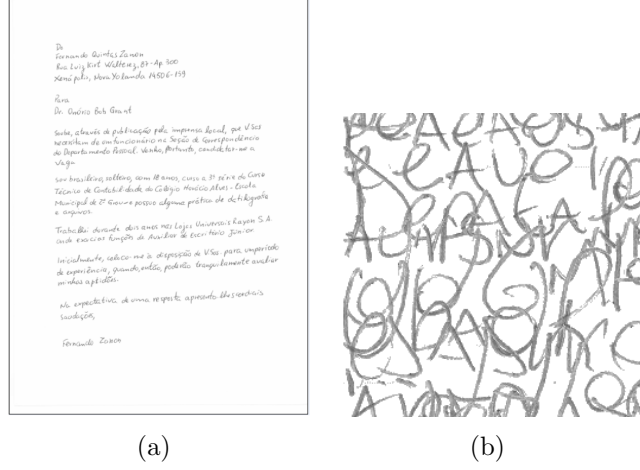


Figura 4. Exemplos de Cartas Compactadas: 4(a) Imagem de amostra da base PUC-PR, 4(b) Fragmento de tamanho  $256 \times 256$ .

### 3.2.3 Independência de escritor e dissimilaridade

Atualmente, duas abordagens são bastante conhecidas e discutidas para aplicações de identificação e verificação de autoria, a dependente do escritor e a independente do escritor [4], [16]. Em sistemas de verificação e identificação com independência de escritor, o interesse está em classificar uma amostra de texto em termos de sua autenticidade, o que reduz o problema a duas classes.

Pekalska *et al.* [13], representam as amostras envolvidas em um problema de classificação por meio das diferenças que possuem em relação a um conjunto de amostras de referência  $R$ . Cada amostra  $s_i$  pode ser representada por um vetor de dissimilaridade dado pela Equação 8, na qual tem-se a diferença (em módulo) entre dois vetores e  $m$  representa o número total de características utilizadas para descrever as referências  $r_j \in R$ ,  $j = 1, 2, \dots, n$ . Pela regra do vizinho mais próximo, uma nova amostra  $s_i$  é classificada considerando  $D(s_i, R)$  e associada à classe  $j$  para a qual  $d(s_i, r_j) = \min_{r \in R} D(s_i, R)$ .

$$(8) \quad D(s_i, r_j) = [|s_{i,1} - r_{j,1}|, |s_{i,2} - r_{j,2}|, \dots, |s_{i,n} - r_{j,m}|]$$

Um conceito relacionado é o de vetor de dissimilaridade (figura 5). A idéia consiste em extrair vetores de características das amostras questionadas ( $s_i$ )

e amostras de referência ( $s_j$ ), calculando o vetor de dissimilaridade de características. Assim, amostras provenientes do mesmo autor (genuínas) possuem vetor de dissimilaridade com valores próximos de 0. Em caso de falsificação, estes mesmos valores devem divergir de 0.

Neste caso, utiliza-se um conjunto de referências composto por  $n$  exemplos de amostras genuínas  $s_{j,i}$ ,  $i = 1, 2, 3, \dots, n$ . Comparando então cada  $s_j$  com um exemplo de amostra questionada  $s_i$ . Seja  $V_i$  o conjunto das características extraídas das amostras de cartas de referência ( $s_j$ ) e  $Q$  o conjunto das características extraídas das amostras questionadas ( $s_i$ ), então, o vetor de dissimilaridade de características  $Z_i = |V_i - Q|$  é computado para alimentar os classificadores  $C_i$  que proporcionam uma decisão parcial. A decisão final  $D$  é dada através de esquemas que combinam as saídas dos classificadores a fim de obter um consenso na decisão.

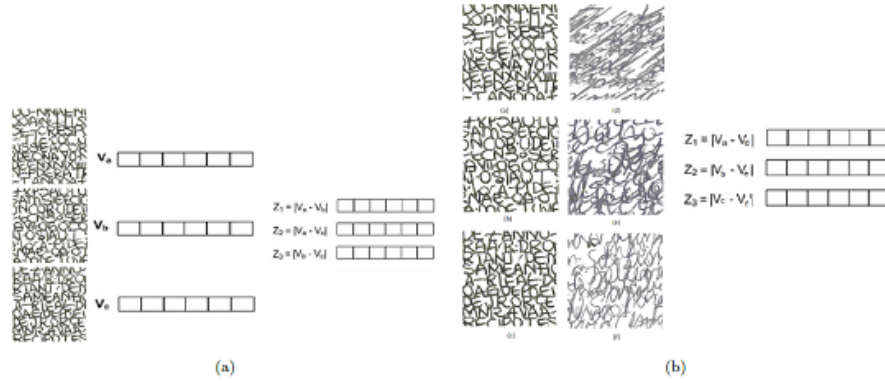


Figura 5. Esquema para dissimilaridade.

### 3.2.4 Resultados

Foram utilizados 115 autores no conjunto de teste e até 200 no de treinamento. Para verificar o impacto da variação do número de autores nos resultados, o conjunto de treino assumiu quatro diferentes tamanhos: 25, 50, 100 e 200.

A influência da quantidade de níveis de cinza também foi investigada. Diante da variação deste parâmetro, os testes indicaram que o uso de imagens monocromáticas produziu os melhores resultados (Tabela 3). Em cada experimento o vetor de características empregou a medida de energia obtida com MC, com variação da distância de 1 a 5, para os ângulos  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ , num total de 20 características. No processo de dissimilaridade, utilizou-se cinco referências para o conjunto de teste e cinco para o conjunto de treinamento. Avaliou-se ainda o impacto da variação do número de autores no processo de treinamento.

Um segundo experimento verificou o impacto de se utilizar diferentes

Tabela 3  
Uso do descritor energia com distâncias variando de 1 a 5.

|        | Número de Autores no Treinamento do Modelo |          |              |          |       |          |       |          |
|--------|--|----------|--------------|----------|-------|----------|-------|----------|
|        | 25   |          | 50           |          | 100   |          | 200   |          |
| Regra  | Média                                      | $\sigma$ | Média        | $\sigma$ | Média | $\sigma$ | Média | $\sigma$ |
| Voto   | 84,92                                      | 1,25     | 85,07        | 3,29     | 83,48 | 1,74     | 88,55 | 1,09     |
| Máximo | 90,29                                      | 2,05     | 91,01        | 2,65     | 92,17 | 0,87     | 90,58 | 1,75     |
| Soma   | 91,88                                      | 1,53     | <b>94,06</b> | 2,00     | 92,89 | 0,90     | 93,04 | 0,75     |

números de referências para treinamento. Foram utilizados 3, 5, 7 e 9 referências e os melhores resultados foram obtidos para a regra da soma (Tabela 4).

O único trabalho publicado com a mesma base e passível de comparação é o de Hanusiak *et al.* [14]. Este utiliza cinco amostras como referências, 100 autores para treinamento e 115 autores para testes, avaliando diferentes descritores (Entropia, Homogeneidade, Dissimilaridade, Variância Inversa e Energia). As taxas obtidas por Hanusiak ficaram entre 94% e 95,4%, sendo que a diferença entre estes e os aqui apresentados também podem ser decorrentes do fato das amostras utilizadas no processo terem sido sorteadas em ambos os trabalhos, além de algumas outras pequenas diferenças em suas metodologias.

Tabela 4  
Avaliação do Número de Referências no Processo de Treinamento.

|                       | Número de Autores no Treinamento do Modelo |          |              |          |       |          |              |          |
|-----------------------|--|----------|--------------|----------|-------|----------|--------------|----------|
|                       | 25   |          | 50           |          | 100   |          | 200          |          |
| Número de Referências | Média                                      | $\sigma$ | Média        | $\sigma$ | Média | $\sigma$ | Média        | $\sigma$ |
| 3                     | 91,88                                      | 1,81     | 93,18        | 0,25     | 91,45 | 3,89     | 94,34        | 1,89     |
| 5                     | 91,88                                      | 1,53     | <b>94,06</b> | 2,09     | 92,89 | 0,90     | 93,04        | 0,75     |
| 7                     | 92,46                                      | 2,95     | 93,33        | 1,81     | 92,46 | 0,66     | <b>94,49</b> | 0,66     |
| 9                     | 91,59                                      | 0,90     | 93,47        | 1,14     | 92,60 | 0,43     | 93,76        | 0,24     |

### 3.3 Reconhecimento de gêneros musicais

Esta subseção descreve alguns resultados obtidos em uma pesquisa voltada à classificação automática de gêneros musicais através da análise de texturas presentes nas imagens de espectrogramas gerados a partir do sinal de áudio

das músicas. Outros resultados, relacionados inclusive a combinação do esquema de classificação aqui apresentado com outros esquemas encontram-se em [7]. Os espectrogramas podem representar dados referentes a um sinal de áudio no domínio de tempo e frequência, e pode ser um instrumento bastante útil para discernir detalhes importantes acerca do mesmo [8]. A extração de características de textura das imagens dos espectrogramas produz um formato de descritor do conteúdo do sinal da música diferente dos já tradicionalmente utilizados.

Para realizar os experimentos, tomou-se como fonte de dados a *Latin Music Database*, uma base particularmente desafiadora, composta por músicas latinas de 10 diferentes gêneros musicais (Axé, Bachata, Bolero, Forró, Gaúcha, Merengue, Pagode, Salsa, Sertaneja, Tango) [15]. A LMD é baseada na percepção de especialistas humanos e foi desenvolvida com propósitos também voltados à classificação automática de gêneros musicais. No total, foram tomadas 900 músicas desta base, divididas em 3 conjuntos, cada um com 300 músicas, doravante denominados *dataset1*, *dataset2* e *dataset3*. Em cada um destes conjuntos foram incluídas 30 músicas de cada gênero e foi empregado o “*artist filter*” na distribuição dos títulos entre os conjuntos. Proposto em [12], o “*artist filter*” determina que os títulos musicais pertencentes à um interprete sejam todos atribuídos à um mesmo conjunto quando da divisão da base de músicas. Devido ao uso do “*artist filter*” não foi possível utilizar todas as músicas da LMD no experimento, já que existe uma diferença significativa entre as quantidades de músicas de cada intérprete presentes na base.

Adicionalmente, adotou-se a estratégia descrita em [6] na qual os autores utilizam três diferentes segmentos dispersos ao longo do sinal do áudio. Para isto, foram tomados segmentos do início, meio e final de cada música. A fim de evitar que efeitos como “*fade in*”, “*fade out*” e vibração de platéia em músicas gravadas ao vivo tornassem trechos da amostra pouco discriminantes, utilizou-se como amostra do início da música o segmento compreendido entre os segundos 11 e 20 da música, e como amostra do final da música o segmento compreendido entre os segundos  $n-20$  e  $n-11$ , sendo  $n$  a duração da música em segundos. O segmento central foi extraído do intervalo compreendido entre os segundos  $m-5$  e  $m+5$ , sendo  $m$  o segundo que se encontra exatamente no meio do sinal da música. A figura 6 ilustra esta estratégia.

Foram gerados espectrogramas a partir de arquivos de áudio com a concatenação dos três segmentos para cada música utilizando-se o software SoX 14.3.0 (*Sound eXchange*), disponível em <http://sox.sourceforge.net>. As imagens dos espectrogramas foram geradas com valores *default* de largura (800 pixels) e altura (550 pixels).

Depois de extraídas as imagens dos espectrogramas das músicas, elas foram convertidas para níveis de cinza para melhor se adequarem aos processos subseqüentes. A figura 7 permite visualizar uma imagem de espectrograma.

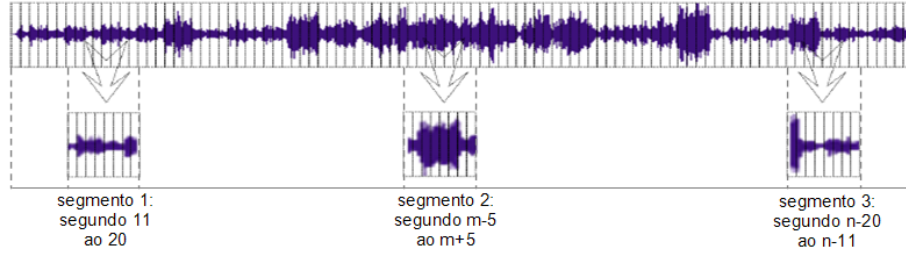


Figura 6. Extração de segmentos do sinal.

Por ela pode-se perceber a textura como atributo visual preponderante.

Depois de segmentadas as imagens dos espectrogramas, a fim de se remover legendas e outras áreas que não a de interesse, foi estabelecida uma divisão das imagens em zonas para que se pudesse preservar parcialmente informações espaciais relacionadas às características a serem extraídas. Com a divisão em zonas (Figura 7) foi possível discriminar as características presentes em diferentes faixas de frequência representadas no espectrograma.

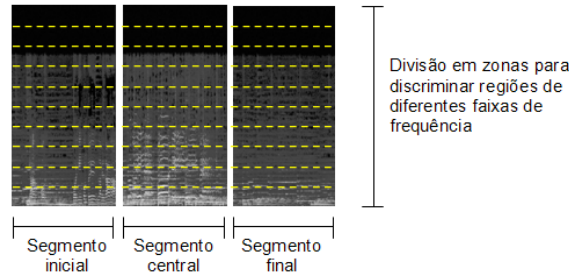


Figura 7. Esquema de divisão das imagens dos espectrogramas em zonas.

Para cada zona da imagem do espectrograma foram extraídas 7 características (Equações 1 a 7) para as orientações  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$  com  $d=1$ . Assim, para cada zona foi formado um vetor com um total de 28 características. A próxima subseção descreve o emprego destes vetores no processo de classificação.

### 3.3.1 Esquema geral de classificação utilizado

Foi realizada a validação cruzada entre os três *datasets*, com 300 músicas cada, previamente descritos. Os conjuntos de teste e treinamento utilizados nos experimentos foram formados pelos vetores de características extraídos de cada zona da imagem do espectrograma e a divisão de conjuntos utilizada foi a mesma empregada em [11], a fim de que os resultados pudessem ser comparados. Assim, considerando que de cada música foram extraídos 3 segmentos, de cada segmento de música foi extraído um espectrograma e que, a partir de cada espectrograma foram criadas 10 zonas, o conjunto de teste foi formado por 9.000 vetores enquanto o conjunto de treinamento foi formado por 18.000 vetores em cada execução do classificador.

Depois de executado o classificador SVM para cada um dos cruzamentos previstos na validação cruzada, obteve-se como resultado uma classe, neste caso gênero musical, atribuída ao vetor extraído de cada zona. A classificação final de cada música foi encontrada através da votação entre as classes atribuídas para as 30 zonas da mesma. Assim, o gênero para o qual foi atribuído o maior número de zonas foi escolhido como gênero ao qual a música pertence. A figura 8 mostra o esquema de votação.

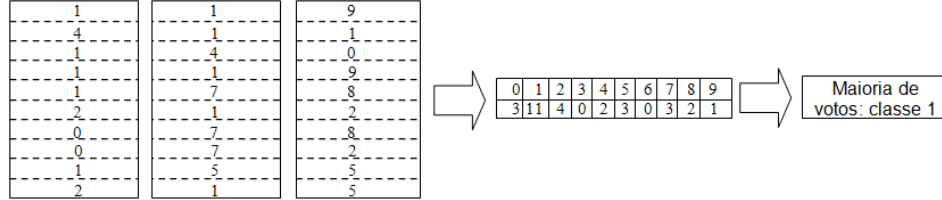


Figura 8. Votação para a decisão final.

Este esquema de votação é frequentemente empregado em trabalhos similares, que envolvem o reconhecimento de gêneros musicais. Em geral, ele oferece bons resultados, já que é bastante comum que partes de uma música sejam parecidas com um gênero ao qual ela não pertence. O esquema de votação ajuda a evitar algumas confusões que poderiam ser causadas por este motivo, já que toma a decisão com base na maioria das zonas criadas.

### 3.3.2 Resultados

A LMD tem como característica o fato de reunir muitos gêneros com significativa similaridade entre si no que diz respeito a instrumentalização, estrutura rítmica e conteúdo harmônico. Isto acontece porque muitos gêneros presentes na base são originários de um mesmo país ou de países com grandes semelhanças no que diz respeito a aspectos culturais. Este fato faz com que a tentativa de discriminar tais gêneros automaticamente seja particularmente desafiadora. As próximas subseções descrevem os resultados obtidos.

A tabela 5 mostra o percentual médio de acerto obtido por gênero com a estratégia de classificação aqui apresentada, seus respectivos desvios padrão e a média geral.

De forma geral, os resultados obtidos nos experimentos mostram taxas de acerto comparáveis às do primeiro trabalho realizado com a LMD [15]. É importante observar que naquele trabalho não foi empregado o “*artist filter*” durante a divisão dos conjuntos para posterior classificação. Além disso, os experimentos aqui descritos apresentaram uma taxa de acerto média ligeiramente superior quando comparados com outros trabalhos recentemente realizados sobre a mesma base de dados, como [11]. Entretanto, houve uma redução significativa do desvio padrão relacionado a taxa média de acerto entre os diferentes gêneros envolvidos no estudo, conforme mostra a tabela 5.

| Gênero       | Taxa de acerto | $\sigma$    | Taxa de acerto obtida em [11] | $\sigma$     |
|--------------|----------------|-------------|-------------------------------|--------------|
| axé          | 73,33%         | 8,82        | 61,11%                        | 13,47        |
| bachata      | 82,22%         | 15,03       | 91,11%                        | 6,94         |
| bolero       | 64,44%         | 8,39        | 72,22%                        | 9,62         |
| forró        | 65,56%         | 8,39        | 17,78%                        | 17,11        |
| gaúcha       | 35,56%         | 5,09        | 44,44%                        | 8,39         |
| merengue     | 80,00%         | 6,67        | 78,89%                        | 11,71        |
| pagode       | 46,67%         | 17,64       | 61,11%                        | 8,39         |
| salsa        | 42,22%         | 6,94        | 40,00%                        | 12,02        |
| sertaneja    | 17,78%         | 6,94        | 41,11%                        | 36,57        |
| tango        | 93,33%         | 6,67        | 88,89%                        | 9,62         |
| <b>geral</b> | <b>60,11%</b>  | <b>9,06</b> | <b>59,67%</b>                 | <b>13,38</b> |

Tabela 5  
Desempenhos obtidos neste trabalho e em Lopes *et al.*

## 4 Considerações finais

Este trabalho destacou a potencialidade de características extraídas de MC na descrição de padrões presentes em imagens digitais. A partir de descritores formados por estas características foram obtidos resultados bastante satisfatórios em domínios de aplicação bastante diversos entre si.

Os resultados aqui apresentados demonstram a viabilidade do emprego destes descritores. Em reconhecimento de espécies florestais, não se encontra na literatura trabalho equivalente, visando a classificação de grupos de espécies florestais. Entretanto, a taxa de 97,43% ( $\sigma = 0,59$ ) é particularmente animadora, e sugere que a investigação seja apropriada. Em verificação de autoria, a taxa de acerto de 94,49% ( $\sigma = 0,66$ ) é ligeiramente superior a outros resultados obtidos com a mesma base. Em reconhecimento de gêneros musicais, a taxa de acerto de 60,11% ( $\sigma = 9,06$ ) foi ligeiramente superior à alcançada em outros trabalhos recentemente apresentados na literatura, conforme mostrado na tabela 5. Resultados descritos nesta tabela também mostram que há complementaridade entre os dois classificadores, fato que permite vislumbrar trabalhos futuros relacionados à combinação entre os mesmos. Adicionalmente, o caráter desafiador da base de músicas utilizada e o fato de ter sido empregado o “*Artist filter*” tornam os resultados bastante expressivos.

Como trabalhos futuros, pretende-se utilizar outros formatos de descritores de textura, permitindo uma melhor avaliação dos resultados apresentados.



Técnicas das abordagens estrutural e espectral são potenciais candidatas para a realização destes experimentos, e os desempenhos obtidos com as mesmas devem ser confrontados com os resultados aqui descritos.

## Referências

- [1] Schlapbach, A. *Writer Identification And Verification (Dissertations In Artificial Intelligence: Infix)*. Ios Press,Us, 2008.
- [2] Baranoski, F. L. Verificação da autoria em documentos manuscritos usando svm. Master's thesis, Pontifícia Universidade Católica do Paraná, 2005.
- [3] Bensefia, A., Paquet, T., and Heutte, L. A writer identification and verification system. *Pattern Recognition Letters*, 26(13):2080 – 2092, 2005.
- [4] Gonçalves, D. B., Oliveira, L. E. S., Justino, E., and Sabourin, R. Ensemble of classifiers for off-line signature verification. In *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, pages 283 –288, 2008.
- [5] Bulacu, M., and Schomaker, L. Text-independent writer identification and verification using textural and allographic features. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):701 –717, 2007.
- [6] Costa, C.H.L., Valle Jr, J.D., and Koerich, A.L., *Automatic classification of audio data*, IEEE Transactions on systems, man and cybernetics, **3** (2003), 562–567.
- [7] Costa, Y. M. G., Oliveira, L. E. S., Koerich, A. L., and Gouyon, F., *Classificação de gêneros musicais por textura no espaço de frequência*, XXXVIII Seminário Integrado de Software e Hardware (2011).
- [8] French, M., and Handy, R., *Spectrograms: turning signals into pictures*, Journal of engineering technology, **24** (2007), 32–35.
- [9] Haralick, R. M., Shanmugam, K., and Dinstein, I. H., *Textural features for image classification*, IEEE Transactions on systems, man and cybernetics, **3** (1973), 610–621.
- [10] Jain, A. K., and F. Farrokhnia, *Unsupervised texture segmentation using Gabor filters*, Pattern Recognition, **24** (1991), 1167–1186.
- [11] Lopes, M., Gouyon, F., Koerich, A. L., and Oliveira, L. E. S., *Selection of Training Instances for Music Genre Classification*, ICPR 2010 - 20th International Conference on Pattern Recognition, (2010).
- [12] Pampalk, E., Flexer, A., and Widmer, G., *Improvements of audio-based music similarity and genre classification*, proc. ISMIR, **5** (2005).
- [13] Pekalska, E. and Robert P., and Duin, W. Dissimilarity representations allow for building good classifiers. *Pattern Recogn. Lett.*, 23(8):943–956, 2002.
- [14] Hanusiak, R. and Oliveira, L. and Justino, E. and Sabourin, R. Writer verification using texture-based features. *International Journal on Document Analysis and Recognition* (2011), 1–14.
- [15] Silla Jr, C. N., Koerich, A. L., and Kaestner, C. A. A., *The latin music database*, Proceedings of the 9th International Conference on Music Information Retrieval (2008), 451–456.
- [16] Sargur, N., Srihari, A. X., and Meenakshi, K. K. Learning strategies and classification methods for off-line signature verification. In *IWFHR '04: Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition*, pages 161–166, Washington, DC, USA, 2004. IEEE Computer Society.
- [17] Vapnik, V. N., *Estimation of Dependences Based on Empirical Data*, NY: Springer-Verlag, (1982).