



Ligia Marcela Tarazona Alvarado

**Comparação de Métodos de Classificação Multitemporal
em Cascata**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre pelo Programa de Pós-
Graduação em Engenharia Elétrica do Departamento
de Engenharia Elétrica, PUC-Rio.

Orientador: Prof. Raul Queiroz Feitosa
Co-orientador: Prof. Gilson A. O. P. da Costa

Rio de Janeiro

Maio 2014



Ligia Marcela Tarazona Alvarado

Comparação de Métodos de Classificação Multitemporal em Cascata

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada

Prof. Raul Queiroz Feitosa

Orientador

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

Prof. Gilson Alexandre Oswalt Pedro da Costa

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

Prof. Karla Figueiredo

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

Prof. Guilherme Lúcio Mota Abelha

Universidade Estadual do Rio de Janeiro (UERJ)

Prof. José Eugenio Leal

Coordenador do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 16 de maio de 2014

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Ligia Marcela Tarazona Alvarado

Graduou-se em Engenharia de Controle-Eletrônico e Instrumentação pela Universidad Distrital Francisco Jose de Caldas (2008) em Bogotá, Colômbia. Trabalhou de 2009 a 2011 em diferentes projetos de engenharia na SEAL Eletronics em Bucaramanga, Colômbia. Juntou-se ao Laboratório de Visão Computacional (LVC/PUC-Rio), na qual desenvolveu pesquisas nas áreas de visão computacional, principalmente em aplicações voltadas para sensoriamento remoto.

Ficha Catalográfica

Alvarado, Ligia Marcela Tarazona

Comparação de métodos de classificação multitemporal em cascata / Ligia Marcela Tarazona Alvarado ; orientador: Raul Queiroz Feitosa, co-orientador: Gilson A. O. P. da Costa– 2014.

89 f. : il. (color.) ; 30 cm

Dissertação (mestrado)—Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2014.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Análise multitemporal. 3. Sensoriamento remoto. 4. SVM. 5. HMM. 6. FMC. I. Feitosa, Raul Queiroz. II. Costa, Gilson A. O. P. da. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Na sua paciência e compreensão, tu preferiste sacrificar teu tempo para me ajudar.
Por tua bondade e sacrifício, inspiraste-me a ser melhor, agora posso dizer que
esta tese tem um monte de ti, obrigada para sempre junto a mim.
Bonito.

Agradecimentos

À Deus, em primeiro lugar, desde do meu primeiro dia no Rio de Janeiro me protegeu estendeu Sua bondade.

À minha amada mãe, minha companheira e a amiga, quem ao longo da minha vida entregou tudo por mim. Sou grata por sua constante atenção para ajudar me a ser um ótimo ser humano e por entregar-me todas as ferramentas para que eu pudesse realizar meus objetivos.

Ao meu orientador e amigo, o Prof. Raul Queiroz Feitosa, pela forma que me orientou na realização deste trabalho. As suas contribuições foram parte fundamental bem como a cordialidade que sempre me estendeu. Sou grata por ambas. A sua participação contribuiu para o meu desenvolvimento profissional e pessoal.

Ao meu co-orientador Prof. Gilson A. O. P. da Costa pela disponibilidade na orientação e pela paixão transmitida a mim pela ciência de sensoriamento remoto.

À CAPES e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Ao meu namorado, amigo, confidente e apoio Diego Pereira Dias - o que seria de mim sem você? Muito obrigada, meu amor por ser meu bastão ao longo deste caminho de alegrias, mas também de dificuldades. Muito obrigada por cada uma das tuas ajudas, por cada um dos sacrifícios em tempo e atividades para me apoiar. Você foi fundamental neste processo.

Ao meu Tio Francisco Tarazona, pela confiança depositada em meu potencial e pelo incentivo em aventurar-me a um país desconhecido.

À minha irmã Juliana Tarazona Alvarado, que sempre me escuta, com quem sempre posso contar.

Aos membros da minha banca: a Profa. Dr. Karla Figueiredo por compartilhar seus conhecimentos em horas de aula e quando foi necessário esclarecer minhas dúvidas, e ao Prof. Dr. Guilherme Lúcio Mota Abelha pelas contribuições e sugestões nesta dissertação.

À minha avó Carmen Villamizar, a quem, mesmo eu não conseguindo me

despedir dela antes ir a sua viagem sem retorno, sou muito agradecida. Tu me deste os primeiros alentos para eu vir para o Brasil.

Às minhas irmãs, que o Rio me deu de presente, Jeniffer Muñoz Trujillo e Judy Marcela Ochoa. Obrigada por compartilharem estes anos comigo. Mais do que companheiras de apartamento, somos uma família.

Aos meus colegas da PUC–Rio, Pedro, Victor Hugo, Victor Andrés, Patrick, Rodrigo, Marcelo, Dario, José, Paolo, Pierre de quem sempre recebi colaboração e fizeram do ambiente de trabalho um lugar especial.

Aos meus amigos de sempre, que, apesar da distância continuam fazendo parte da minha vida.

Aos professores do departamento de Engenharia Elétrica Prof. Marcos, Prof. Marley, Prof Ricardo que compartilharam comigo seus conhecimentos.

Aos funcionários do departamento de Engenharia Elétrica pela a ajuda de todos os dias, em particular Alcina e Mara.

À cidade Maravilhosa do Rio de Janeiro e a todos seus cariocas, por sua hospitalidade, suas lindas paisagens, pelas oportunidades e por me presentear com gente de grande coração.

Resumo

Alvarado, Ligia Marcela Tarazona; Feitosa, Raul Queiroz; Costa, Gilson A. O. P. da. **Comparação de Métodos de Classificação Multitemporal em Cascata**. Rio de Janeiro, 2014. 89p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Esta dissertação faz uma comparação de três métodos de classificação em cascata de imagens multitemporais. Os classificadores se baseiam nas seguintes técnicas: (1) Máquina de Suporte Vetorial (SVM), (2) Modelos Ocultos de Markov (HMM) e (3) Cadeias de Markov Nebulosas(FMC). Para verificar a robustez dos modelos de classificação, introduziram-se nos dados de entrada *outliers*, avaliando-se assim, a robustez dos classificadores. Adicionalmente, avaliou-se o desempenho dos métodos quando a proporção de ocorrências de cada transição de classe no conjunto de treinamento difere da proporção no conjunto de teste. Determinou-se também qual o benefício do uso de conhecimento a priori sobre as transições possíveis. A análise experimental foi realizada sobre dois conjuntos de imagens de diferentes características, um par de imagens IKONOS do Rio de Janeiro, Brasil e um par de imagens LANDSAT7 de Alcinópolis, Mato Grosso do Sul. O estudo revelou que acurácia global das três abordagens tem um comportamento similar nos diferentes experimentos. Mostrou também que todas as três abordagens multitemporais apresentam desempenho superior aos seus homólogos monotemporais.

Palavras-chave

Análise multitemporal; Sensoriamento Remoto; SVM; HMM; FMC.

Abstract

Alvarado, Ligia Marcela Tarazona; Feitosa, Raul Queiroz (Advisor); Costa, Gilson A. O. P. da (Co-advisor). **A Comparison of Cascade Multitemporal Image Classification Methods**. Rio de Janeiro, 2014. 89p. MSc. Dissertation - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This dissertation compares three cascade multitemporal image classification methods based on: (1) Support Vector Machines (SVM), (2) Hidden Markov Models (HMM) and (3) Fuzzy Markov Chains (FMC). The robustness of the classification models is verified, by introducing outliers in the data set. Additionally, performance of each method is evaluated when the number of occurrences of each class transition is different in the training and in the testing set. The gain of exploiting a prior knowledge regarding the admissible transitions in each target site is also investigated. The experimental analysis is conducted over two data sets with different characteristics; specifically a pair of IKONOS images of Rio de Janeiro and a pair of LANDSAT7 images of Alcinópolis, Mato Grosso do Sul. This study has concluded that the overall accuracy of the three approaches are similar through all experiments. The superiority of the multitemporal approaches over the monotemporal counterparts was confirmed.

Keywords

Multitemporal Analysis; Remote Sensing; SVM; HMM; FMC.

Sumário

1 Introdução	19
1.1. Objetivos da dissertação	21
1.1.1. Objetivos específicos	22
1.2. Estrutura da dissertação	22
2 Trabalhos Relacionados	23
2.1. Classificação de imagens de sensoriamento remoto	23
2.2. Classificação multitemporal	24
2.3. Classificação multitemporal em cascata	25
2.4. Comparação de classificadores aplicados em SR	26
2.4.1. Máquinas de suporte vetorial aplicado ao SR	27
2.4.2. Classificadores baseados em Modelos de Markov Ocultos	31
2.4.3. Classificação baseada em Cadeias de Markov Nebulosas	32
3 Fundamentos teóricos dos classificadores multitemporais	34
3.1. Introdução	34
3.2. Máquina de suporte vetorial	34
3.2.1. Classificação multi-classe	35
3.3. Modelos de Markov ocultos	36
3.3.1. Cadeias de Markov	36
3.4. Modelos de Markov ocultos	38
3.5. Baseado em cadeias de Markov nebulosas	40
4 Modelos dos classificadores em cascata	42
4.1. Introdução	42
4.2. Classificação monotemporal	42
4.3. Abordagens de classificação multitemporal em cascata	43
4.3.1. Classificação multitemporal em única data	44
4.3.2. Classificação multitemporal em duas datas	44
4.4. Particularização dos modelos de classificação	44

4.4.1. Método baseado em SVM	45
4.4.2. Parâmetros de seleção	46
4.4.3. Métodos baseados em HMM	48
4.4.4. Método baseado em FCM	51
5 Análise e resultados	54
5.1. Introdução	54
5.2. Descrição do conjunto de dados	54
5.2.1. Área de teste: Alcinópolis	54
5.2.2. Área de teste: Rio de Janeiro	57
5.3. Resultados	59
5.3.1. Experimento 1	60
Objetivo e procedimento	60
Resultados e Discussão	61
5.3.2. Experimento 2	63
Objetivo e procedimento	63
Resultados e Discussões	63
5.3.3. Experimento 3	64
Objetivo e procedimento	64
Resultados e Discussões	65
5.3.4. Experimento 4	68
Objetivo e procedimento	68
Resultados e Discussões	69
5.3.5. Experimento 5	73
Objetivo e procedimento	73
Resultados e Discussão	74
6 Conclusão e trabalhos futuros	77
Referência Bibliográfica	79

Lista de Figuras

Figura 1: Processo de classificação aplicando SVM com duas classes se- paradas por um hiperplano (adaptado de (Mountrakis, et al., 2011))	27
Figura 2: Cadeia de Markov com 4 estados (ω_1 , a ω_2), onde as setas indicam as transições possíveis entre os estados (adaptado de (Mota, et al., 2007), (Costa & Feitosa, 2014) e (Leite, 2008)).	37
Figura 3: Exemplo de um HMM com estados observáveis, onde ($\omega_i \rightarrow$ estados, $v_k \rightarrow$ símbolos observáveis, $\tau_{ij} \rightarrow$ probabilidade de transição de estado, $b_{jk} \rightarrow$ probabilidade de emissão de símbolo) (adaptado de (Costa e Feitosa, 2014), (Leite, 2008) e (Ra-biner, 1989))	39
Figura 4: Classificador monotemporal	42
Figura 5: Classificador multitemporal em cascata.	43
Figura 6: Rotina de treinamento SVM (adaptado de (Chang & Lin, 2011)).	46
Figura 7: Classificador monotemporal baseado em SVM (adaptado de (Feitosa, et al., 2013).	47
Figura 8: Esquemas de classificação multitemporal com fusão de dados para uma única data (considerem-se as linhas solidas e as linhas pontilhadas) (adaptado de (Feitosa, et al., 2013)).	47
Figura 9: Classificador monotemporal baseado em ML.	49
Figura 10: Classificador multitemporal HMM baseado em fusão de dados para uma única data.	50
Figura 11: Classificador multitemporal baseado em HMM	51
Figura 12: Classificador monotemporal fuzzy (adaptado de (Feitosa, et al., 2009))	51
Figura 13: Classificador multitemporal FMC, para um modelo <i>forward</i> (adaptado de (Costa & Feitosa, 2014)).	52
Figura 14: Área de teste na sub-bacia do rio Taquari (extraído de (Costa, et al., 2009))	55

Figura 15: Subconjunto da imagem do site de teste de Alcinópolis, MS (extraído de (Mota, 2004))	55
Figura 16: Subconjunto da imagem da área de teste do Rio de Janeiro 2008 (extraído de (Costa & Feitosa, 2014)).	58
Figura 17: Acurácia Global (GAc) para experimento 1, imagens do Rio de Janeiro e Alcinópolis, ressaltando o desempenho geral dos classificadores dos esquemas propostos.	62
Figura 18: Acurácia Global GAc para experimento 2, imagens do Rio de Janeiro e Alcinópolis, ressaltando o desempenho geral dos classificadores dos esquemas propostos quando a verdade terrestre é conhecida.	64
Figura 19: Acurácia Global (GAc) para experimento 3, ressaltando o desempenho dos classificadores dos multitemporais sob a presença variada de <i>outliers</i> .	66
Figura 20: Acurácia Global (GAc) para experimento 3, ressaltando o desempenho geral dos classificadores dos multitemporais sob a presença variada de <i>outliers</i> .	67
Figura 21: Acurácia Global (GAc) para experimento 4, Imagem do Rio de Janeiro ressaltando o desempenho dos classificadores dos multitemporais com conhecimento prévio sobre as transição de classe e na ausência de <i>outliers</i> .	70
Figura 22: Acurácia Global (GAc) para experimento 4, Imagem Alcinópolis ressaltando o desempenho dos classificadores dos multitemporais com conhecimento prévio sobre as transições de classe e na ausência de <i>outliers</i> .	71
Figura 23: Acurácia Global (GAc) para experimento 4, Imagem do Rio de Janeiro ressaltando o desempenho dos classificadores dos multitemporais com conhecimento prévio sobre as transições de classe e na presença de <i>outliers</i> .	72
Figura 24: Acurácia Global (GAc) para experimento 4, Imagem Alcinópolis ressaltando o desempenho dos classificadores dos multitemporais com conhecimento prévio sobre as transições de classe e na presença de <i>outliers</i> .	73

Figura 25: Acurácia Global (GAc) para experimento 5, imagem do Rio de Janeiro, ressaltando o desempenho geral dos classificadores quando as probabilidades de transição de classe estão mal estimadas e na ausência de *outliers*.

75

Figura 26: Acurácia Global (GAc) para experimento 5, imagem Alcinópolis, ressaltando o desempenho geral dos classificadores quando as probabilidades de transição de classe estão mal estimadas e na ausência de *outliers*.

76

Lista de Tabelas

Tabela 1: Classes consideradas nos experimentos e a suas descrições (extraída de (Costa, 2009),Tabela 2, pg 62)	56
Tabela 2: Simulação da dinâmica de classes 100 (Alta dinâmica)	57
Tabela 3: Simulação da dinâmica de classes 1000 (Baixa dinâmica)	57
Tabela 4: Classes consideradas no local de teste Rio de Janeiro (extraída de (Costa & Feitosa, 2014))	58
Tabela 5: Número de transições entre 2008 e 2009 para o local de teste urbano.	59
Tabela 6: Simulação da dinâmica de classes 100 (Alta dinâmica)	59
Tabela 7: Simulação da dinâmica de classes 1000 (Baixa dinâmica)	59

Lista de Acrônimos

<i>AIDA</i>	<i>Automatic Image Data Analyser.</i>
<i>CV</i>	Validação Cruzada, do inglês, <i>Cross-Validation</i>
<i>FMC</i>	Cadeias de Markov Nebulosa do inglês, <i>Fuzzy Markov Chain</i>
<i>GA</i>	Algoritmos Genéticos do inglês, <i>Genetic Algorithm</i>
<i>GAc</i>	Acurácia Global do inglês, <i>Global Accuracy</i>
<i>GEOBIA</i>	Geographic Object Based Image Analysis
<i>GIS</i>	Sistema de Informação Geográfica, do inglês, <i>Hidden Markov Models</i>
<i>INPE</i>	Instituto Nacional de Pesquisa Espaciais
<i>HMM</i>	Modelos Ocultos de Markov, do inglês, <i>Hidden Markov Models</i>
<i>LIBSVM</i>	<i>Library for Support Vector Machines</i> Desenvolvida por Chih-Chung Chang and Chih-Jen Lin (2011)
<i>LULC</i>	Uso e Coberto do Solo do inglês, <i>land-use / land-cover</i>
<i>LC</i>	Cobertura do Solo, do inglês, <i>land-use</i>
<i>ML</i>	Máxima verossimilhança do inglês, <i>maximum likelihood</i>
<i>OBIA</i>	<i>Object-based Image Analysis</i>
<i>RBF</i>	Função de Base Radial, do inglês, <i>Radial Basis Function</i>
<i>SR</i>	Sensoriamento Remoto
<i>SVM</i>	Máquina de Suporte Vetorial do inglês, <i>Support Vector Machine</i>

Lista de Notação e Símbolo

x	vetor normal ao hiperplano de separação da SVM
C	(SVM) parâmetro de regulação (penalidade do erro)
C	(FMC) classificador monotemporal
f	função decisão (estimação de pares de classe)
H	função que seleciona à classe de maior probabilidade
$K(.,.)$	(SVM) função kernel do problema de otimização primal
k	(SVM) número total de classe
I	imagem
${}^B M$	(FMC) classificador multitemporal, modelo backward
${}^F M$	(FMC) classificador multitemporal, modelo forward
p	Estimativa da probabilidade de cada objeto
Q	Matriz positiva semi-definida
v	Estados não-observáveis de ω
x_i	Vetores de treinamento
y	Classes
w	Vetor de alta dimensionalidade da SVM
tA	Instante de tempo (normalmente, imagem alvo)
tB	Instante de tempo (imagem adicional)

Alfabeto Grego

α	(SVM) multiplicador de Lagrange
α	(HMM/FMC) vetor nebuloso em um instante de tempo
β	(HMM/FMC) vetor nebuloso em outro instante de tempo
γ	(SVM) parâmetro de ajuste
ϕ	(SVM) função kernel do problema de otimização dual
v	(SVM) vetor suporte
ρ	(SVM) ponderação para o problema de determinação do vetor suporte v
τ	Probabilidades de transição de estados

Sub-índices

α	(SVM) multiplicador de Lagrange
α	(HMM/FMC) vetor nebuloso em um instante de tempo

Operadores ou funções usuais

e^x	Função exponencial de base e
$\ \cdot \ $	Norma (de um vetor ou uma matriz)
$\ \cdot \ _{\infty}$	Norma infinita (de um vetor ou uma matriz)
max	Máximo (de um vetor ou de uma matriz)
min	Mínimo (de um vetor ou de uma matriz)
<i>sgn</i>	Função sinal
\circ	Multiplicação de matrizes
\perp	t-norma
T	s-norma

*“Tenho a impressão de ter sido uma criança brincando
à beira-mar, divertindo-me em descobrir
uma pedrinha mais lisa ou uma concha mais bonita
que as outras, enquanto o imenso oceano
da verdade continua misterioso diante de meus
olhos.”*

Isaac Newton

1

Introdução

O monitoramento de atividades tais como agricultura, extração marítima, extração vegetal, é essencial para o desenvolvimento sustentável (Gladstone, et al., 2012) . O Sensoriamento Remoto (SR) tem se mostrado indispensável para o acompanhamento das alterações na superfície da Terra provocadas pela dinâmica natural ou por influência humana.

O desenvolvimento da disciplina científica de SR teve início a no século XIX com o aprimoramento da teoria da luz e utilização de câmaras fotográficas a bordo de balões. Não muito depois, ficou evidente a necessidade de identificar e analisar os fenômenos naturais, bem como os efeitos das atividades humanas sobre a superfície da Terra. O SR, como ciência, desenvolveu técnicas de aquisição, ferramentas de observação da Terra e finalmente, métodos para descrever ou quantificar padrões da superfície da Terra a fim de contribuir para a compreensão dos processos subjacentes (Blaschke & Strobl, 2001)

O volume sem precedentes de dados produzidos por sensores aéreos e especialmente orbitais requer métodos automáticos de análise que transformem os dados de SR em informação útil. Dados de SR são hoje coletados a partir de inúmeras plataformas orbitais. O Instituto Nacional de Pesquisas Espaciais (INPE) mantém um banco de imagens dos seguintes satélites: CBERS-2 (China-Brasil), CBERS-2B (China-Brasil), LANDSAT1, LANDSAT2, LANDSAT3, LANDSAT4, LANDSAT5, LANDSAT7 (EUA), RESOURCESAT-1 (Índia), TERRA (EUA) e AQUA (EUA) (Veras, 2013).

Além disso, a disponibilidade de dados de características distintas, em diferentes resoluções espaciais, incluindo dados ópticos, como SAR e LIDAR, permitem a criação de um conjunto diversificado de aplicações como o gerenciamento e monitoramento ambiental, mapeamento de cobertura do solo, planejamento urbano, inventariação agrícola, gestão, etc. (Leite, 2008). Neste estudo, um dos conjuntos de imagens foi adquirido pelo satélite LANDSAT7.

Se por um lado a crescente disponibilidade de dados de SR fornece informação útil para lidar com a atual taxa de atividade humana e mudanças ambientais, por outro, apresenta aos gestores de políticas e tomadores de decisão uma enorme quantidade de dados para explorar e analisar. No entanto, a maioria das técnicas de análise de imagens de SR ainda requer intensa intervenção humana. A maioria dos mapas de cobertura do solo, por exemplo, são criados por meio de interpretação visual. Em contra partida, técnicas de interpretação automática das imagens de SR estão sendo aperfeiçoadas (de Almeida Crusco, 2006), (Coppin & Bauer, 1996), (Gladstone, et al., 2012), (Bontemps, et al., 2008), (Bruzzone & Cossu, 2002), (Gong, 1993). Por conseguinte, a classificação multitemporal faz parte de este grupo de técnicas que procuram automatizar a interpretação de imagens de SR.

Atualmente, as técnicas de classificação multitemporal de imagens de SR seguem diferentes abordagens: pré-classificação, pós-classificação e classificação em cascata.

As técnicas de pré-classificação apenas indicam as áreas em que ocorreram mudanças, e eventualmente informações sobre a intensidade e a direção da mudança. Métodos de pré-classificação (Coppin & Bauer, 1996) baseados na subtração de imagens (Ridd & Liu, 1998), em regressão (Ridd & Liu, 1998), na razão entre bandas (Gupta & Prakash, 1998), na análise de componentes principais (Gong, 1993), são alguns das técnicas demais frequentemente utilizadas em métodos de pré-classificação.

A pós- classificação considera classificações independentes de cada imagem que compõe uma série temporal, cujos resultados são posteriormente comparados (Weismiller, et al., 1977). Uma vantagem da pós- classificação é a minimização dos problemas causados pelas variações das condições atmosféricas nos sensores, além de que para os fenômenos da vegetação a reflectância não precisa ser necessariamente ajustada para uma direta comparação entre as datas (Fuller, et al., 2003).

Uma alternativa mais eficiente é a classificação em cascata (Swain, 1978), que usa toda informação contida na sequência de imagens, explorando a correlação entre elas. A maioria das abordagens em cascata encontradas na literatura trata a classificação multitemporal como um problema de fusão de dados, tendo evoluído da pesquisa sobre a classificação de dados de múltiplos

sensores (Aach & Kaup, 1995), (Bruzzone, et al., 1999), (Bruzzone & Cossu, 2002), (Khazenie & Crawford, 1990). Um dos primeiros relatos sobre a classificação multitemporal é encontrado em (Jeon & Landgrebe, 1992) Neste trabalho os autores propuseram um classificador contextual que considera as dependências interpixel por classe, tanto espaciais quanto temporais, sendo as últimas modeladas através de probabilidades de transição de classes. Em (Bruzzone & Cossu, 2002) um método multitemporal em cascata é proposto para a atualização do mapa da cobertura terrestre. Classificadores da Máxima Verossimilhança (ML) e redes neurais de função de base radial são usadas no método proposto. O problema principal é a determinação de como explorar os dados de uma imagem mais antiga para o treinamento dos classificadores, quando a verdade terrestre da imagem a ser classificada não está disponível.

Estudos anteriores (Costa, et al., 2009) conduzidos no LVC-DEE/PUC-Rio demonstram que ganhos expressivos de acurácia podem ser alcançados quando se explora a correlação temporal de imagens. Em outros estudos, (Mota, et al., 2007) propor um modelo que utiliza Cadeias de Markov Nebulosas (FMC – do inglês *Fuzzy Markov Chain*). Posteriormente em (Feitosa, et al., 2009) se apresenta uma extensão do FMC. Em Leite e co-autores (Leite, 2008) propõe-se uma metodologia baseada em modelos de Markov ocultos (HMM – do inglês *Hidden Markov Models*) para a classificação de culturas agrícolas, e em (Feitosa, et al., 2013) comparam-se cinco esquemas de classificação em cascata baseados em FMC.

Esta dissertação está inserida neste contexto e apresenta uma comparação entre métodos de classificação em cascata de imagens de SR baseados em: (1) FMC, (2) HMM e (3) Máquina de Suporte Vetorial (SVM).

1.1.

Objetivos da dissertação

O objetivo central desta dissertação é comparar alternativas de classificação multitemporal em cascata com base em modelos SVM, HMM e FMC seguindo o paradigma *Geographic Object Based Image Analysis* (GEOBIA).

1.1.1.

Objetivos específicos

Os objetivos secundários de este trabalho de dissertação procuram:

1. Avaliar os métodos estudados sobre dois conjuntos de imagens de SR multitemporais com resoluções espaciais e características distintas.
2. Investigar a influência da dinâmica entre as classes na acurácia nos esquemas de classificação em cascata multitemporal estudados.
3. Avaliar como a “qualidade” das classificações monotemporais afeta o desempenho da classificação multitemporal.
4. Avaliar a robustez de cada método de classificação contra *outliers*.
5. Estimar o benefício do uso do conhecimento a priori sobre as transições de classe possíveis.

1.2.

Estrutura da dissertação

A presente dissertação se encontra dividida em 5 capítulos além deste, organizados da seguinte maneira. O próximo capítulo apresenta alguns trabalhos científicos produzidos anteriormente, utilizando metodologias de classificação baseadas em análise multitemporal. O Capítulo 3 faz uma introdução sobre os fundamentos teóricos dos classificadores multitemporais baseados nas três técnicas a comparar, de forma facilitar o entendimento do trabalho desenvolvido nesta dissertação.

No capítulo seguinte são apresentados os modelos das classificações em cascata investigados neste trabalho.

No Capítulo 5 é apresentada a análise experimental. Finalmente, são apresentadas as conclusões obtidas ao longo deste trabalho e sugestões para continuidade desta pesquisa.

2

Trabalhos Relacionados

Neste capítulo apresentam-se alguns dos trabalhos anteriores sobre análise multitemporal de imagens de sensoriamento remoto, relacionados com as técnicas usadas neste trabalho. Além disso, mostram-se aplicações que utilizam individualmente os métodos de Máquina de Suporte Vetorial (SVM), Modelos Ocultos de Markov (HMM) e Cadeias de Markov Nebulosas (FMC).

2.1.

Classificação de imagens de sensoriamento remoto

Sensoriamento Remoto (SR) fornece meios para o monitoramento de áreas extensas em termos de florestas, recursos hídricos, ecossistemas, áreas produtivas e espaços urbanos, entre outras. Apesar do aumento da quantidade de novos sensores orbitais nos últimos anos e dos recentes avanços na tecnologia de software de apoio ao processamento de imagens de SR, a interpretação das imagens é ainda realizada de forma predominantemente manual. O conhecimento do especialista humano sobre a região analisada é de importância e influência no resultado da interpretação (Leite, 2008). Essa tarefa requer grande esforço, experiência e perícia do intérprete.

Atualmente a maior parte dos algoritmos de interpretação de imagens de SR, produz resultados fragmentados, incompletos e até inconsistentes. Uma etapa posterior manual é necessária para corrigir estes resultados. Neste processo de pós-edição, informações adicionais de Sistemas de Informação Geográfica (GIS) são utilizadas pelo especialista para corrigir os resultados. Devido à crescente demanda faz-se necessário o desenvolvimento de uma tecnologia capaz de aumentar o grau de automação da interpretação de imagens e de sua acurácia a fim de melhorar a produtividade e rapidez do processo.

A interpretação automática de imagens consiste de três etapas: (1) criação de uma descrição simbólica das imagens, (2) identificação automática do

conteúdo (3) agregação da extensão geográfica relevante para uma aplicação particular. Tais descrições simbólicas podem ser utilizadas para diferentes fins, como entradas de processos de apoio à decisão, na busca de imagens com um conteúdo específico. Por tanto, cada classificador de imagens baseado em conhecimento incorpora alguma forma de representar o conhecimento. Conforme (Mota, et al., 2007), as estruturas de representação do conhecimento e a lógica do processo de interpretação são os principais aspectos que diferenciam os sistemas baseados em conhecimento das outras abordagens automáticas de interpretação. O conhecimento multitemporal, nesse contexto, está associado evolução de uma classe de objetos em diferentes instantes de tempo.

2.2.

Classificação multitemporal

Um dos primeiros e relevantes sistemas para análise multitemporal baseada em conhecimento de imagens de SR é descrito em (Liedtke, et al., 2001). Desenvolveu-se uma plataforma que veio a chamar-se do *Automatic Image Data Analyser* (AIDA), na qual o conhecimento temporal é representado por relações temporais.

Outras abordagens para a análise multitemporal de imagens de Sensoriamento Remoto (SR) têm sido propostas desde então, muitas delas preocupadas apenas com a detecção de mudanças, ao invés de proporcionar uma classificação dos objetos que estão presentes na imagem em termos do Uso e Cobertura do Solo (LULC).

A detecção de mudanças geralmente inclui: (1) pré-processamento de imagem, incluindo correções geométricas e atmosféricas, (2) seleção de uma técnica aplicável à análise de detecção de mudanças e (3) medição da acurácia na avaliação dos resultados.

Em (Lu, et al., 2004), uma revisão é feita sobre os métodos aplicáveis a detecção de mudanças, destacando-se que análise de componentes principais (PCA) e a pós-classificação são os métodos mais usados em aplicações de SR. Uma limitante daquela proposta é que a classificação é pixel-a-pixel.

Alternativamente, em (Büscher & Buck, 2007) apresenta-se uma proposta para detecção de mudanças baseada em *Object-based Image Analysis* (OBIA), no

contexto do projeto (*DeCover*) (Büscher & Buck, 2007) . O método compara segmentos de duas imagens da mesma região tomadas em diferentes instantes de tempo. Cria-se objetos de mudança dentro dos limites dos objetos LULC tal como estão definidos em uma base de dados GIS. Indicadores de mudança são associados a cada objeto. A classificação de cada objeto leva em conta a probabilidade de ter ocorrido a mudança levando em consideração os atributos do objeto de imagem num instante de tempo anterior. Requer-se a definição ou estimativa de uma matriz de probabilidade de transição de classes.

Doutro modo a classificação em cascata mostra-se uma abordagem eficaz, pois utiliza toda a informação contida na sequência de imagens, tentando explorar a correlação temporal entre imagens (Swain, 1978). Esta característica foi explorada em diversas linhas de pesquisa por diferentes técnicas incluindo: métodos Bayesianos (Serpico & Melgani, 2000), redes neurais (Bruzzone, et al., 1999) e abordagens de multi-classificadores (Bruzzone, et al., 2004). Em (Mota, et al., 2007) um método nebuloso multitemporal é proposto para aplicações de classificação de LULC.

2.3.

Classificação multitemporal em cascata

A abordagem de classificação em cascata usa toda a informação contida em uma sequência de imagens multitemporais (Swain, 1978) e tenta explorar a correlação contida em um conjunto multitemporal de dados de SR. A maioria das abordagens em cascata encontradas na literatura trata o problema de classificação multitemporal com um problema de fusão de dados. A evolução da técnica teve contribuição das pesquisas anteriores baseadas na classificação multisensor (Khazenie & Crawford, 1990), (Aach & Kaup, 1995), (Jeon & Landgrebe, 1992), (Solberg, et al., 1996), (Bruzzone & Cossu, 2002), (Bruzzone, et al., 2004).

Como mencionado anteriormente, uma das primeiras publicações sobre um método de classificação em cascata Jeon baseado em fusão de decisões foi encontrado em (Jeon & Landgrebe, 1992) que propõem uma classificação contextual. Neste classificador foi considerado tanto as dependências de classes inter-pixel espacial e temporal. Outra abordagem que incorpora o método de fusão de decisões, considerando tanto os aspectos multissensoriais e temporais de

imagens multi-data, foi apresentada em (Jeon & Landgrebe, 1999). Este método de classificação multitemporal baseia-se no resultado de classificadores monotemporais. A seleção de cada um dos objetos de imagem é dada pela classe associada à maior probabilidade produzida por classificadores monotemporais. O método apresentou um aumento de 10% na Acurácia Global (GAc) global, em comparação com um classificador monotemporal de Máxima Verossimilhança (ML).

Em (Melgani, et al., 2003) é proposto um método para agregar contextos: espectrais, espaciais e temporais para classificação multitemporal de imagens multi-sensor. Múltiplas redes neurais são utilizadas para a fusão de dados de múltiplos sensores relativos à uma mesma data que são combinados iterativamente para produzir a classificação final.

Um método para a classificação de culturas agrícolas em uma série de imagens SAR através de uma abordagem estatística é proposto em (S.M. Tavakkoli Sabour, 2008). A média e o desvio padrão dos valores dos pixels cobrindo a extensão de um conjunto de amostras de campo são usados como assinaturas de cada classe de culturas para cada imagem. Em (Bruzzone & Cossu, 2002) um método multitemporal em cascata é proposto para atualização de mapas de LULC. Classificadores de ML e redes neurais de função de base radial compõem o sistema de classificação proposto. A maioria das abordagens multitemporais em cascata assumem a independência condicional das classes no domínio do tempo (Bruzzone, et al., 1999). Outros trabalhos apresentam métodos que não assumem independência temporal e visam capturar as correlações inter-fonte (espectral, espacial e temporal) através de redes neurais (Melgani, et al., 2001).

2.4.

Comparação de classificadores aplicados em SR

Pode-se dizer que a classificação de dados de SR pode tornar-se difícil porque a maioria dos sistemas de aprendizagem supervisionada exigem uma grande quantidade de amostras de treinamento.

As seções seguintes fazem uma breve revisão do tema.

2.4.1.

Máquinas de suporte vetorial aplicado ao SR

As Máquinas de Suporte Vetorial (SVMs) consistem em uma técnica de aprendizado de máquina, fundamentada nos princípios da Minimização do Risco Estrutural¹. Na sua formulação original (Vapnik, 1979), esta busca minimizar o risco empírico do conjunto de treinamento. A metodologia consiste em determinar um hiperplano que melhor separe o conjunto de dados de treinamento em um número predefinido discreto de classes. O hiperplano de separação é definido como a fronteira de decisão que minimiza erros de classificação, obtidos no passo de treinamento. O processo de aprendizagem é obtido iterativamente na busca de uma fronteira de decisão ótima que separa os padrões de treinamento (em um espaço potencialmente de maior dimensão do que o espaço de atributos originais), como ilustrado na Figura 1 (Zhu & Blumberg, 2002). SVMs são em sua formulação original, classificadores binários que atribuem a uma determinada amostra de teste uma de duas classes possíveis. No caso de classificação de imagens de SR as amostras podem ser pixels ou segmentos da imagem.

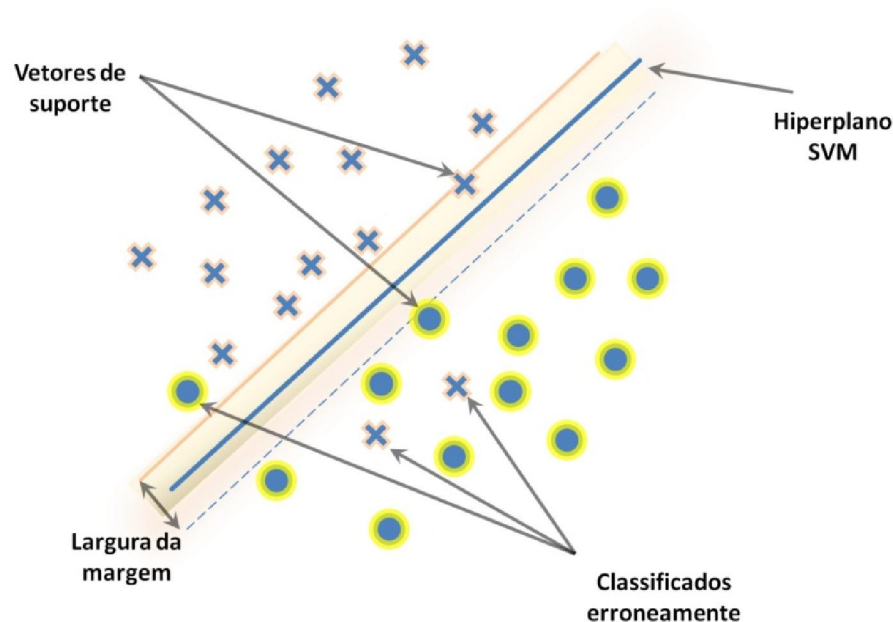


Figura 1: Processo de classificação aplicando SVM com duas classes separadas por um hiperplano (adaptado de (Mountrakis, et al., 2011))

¹ Nomenclatura orinda do inglês, *Structural Risk Management*.

A Figura 1 exemplifica um cenário simples de um problema de classificação separável em duas classes, em um espaço de entrada de duas dimensões. A generalização das SVMs tem uma característica interessante: muitas vezes nem todas as amostras de treinamento disponíveis são utilizadas na descrição e especificação do hiperplano separador. O subconjunto de amostras que ficam próximos à margem do hiperplano, também conhecidos como vetores suportes são os únicos que definem o hiperplano de margem máxima².

Uma SVM linear assume que os dados são linearmente separáveis no espaço de entrada. Na maioria dos casos práticos, contudo, as amostras de dados de diferentes classes não se localizam em regiões cujos limites são segmentos de reta. Isso limita a aplicação dos SVM lineares. O método de margem suave foi proposto por (Cortes & Vapnik, 1995) como solução alternativa para lidar com esse problema. Uma abordagem mais flexível admite fronteiras de decisão não lineares utilizando-se os chamados *kernels* que mapeiam o espaço de atributos original num outro espaço de maior dimensão em que as classes têm maior chance de serem linearmente separáveis.

Uma função deve satisfazer o Teorema de Mercer para se qualificar como um *kernel* válido para SVM (Scholkopf & Smola, 2001), (Andreola & Haertel, 2009). A influência da escolha da função *kernel* foi estudada em (Scholkopf & Smola, 2001), (Andreola & Haertel, 2009), (Burgess, 1998), (Negri, et al., 2011). Alguns dos *kernels* mais usados são: linear, polinomial, função de base radial (RBF), sigmoide.

Além disso, problemas típicos das aplicações de SR geralmente envolvem várias classes. O problema multiclasse é contornado com o ajuste do classificador SVM binário usando métodos como: um-contra-um, um-contra-todos, um-contra-outros e grafo dirigido acíclico (Knerr, et al., 1990).

Em resumo, as SVMs são particularmente atraentes para aplicações de SR devido a sua capacidade de lidar com pequenos conjuntos de dados de treinamento com sucesso.

O resultado da classificação usando SVM muitas vezes é mais exato do que os métodos tradicionais (Mantero, et al., 2005).

² A margem máxima, assume-se que as amostras de treinamento das diferentes classes são linearmente separáveis, consiste na distância maximizada entre os conjuntos das amostras de treinamento. A função de decisão que maximiza esta separação é dita ótima (Andreola & Haertel, 2009) e (Meloni, 2009).

A formulação do processo de classificação de imagens multitemporais utilizando um classificador SVM requer a solução de alguns desafios. A escolha dos *kernels* foi estudada genericamente visando a aplicabilidade em SR em vários trabalhos, como (Andreola & Haertel, 2009), (Scholkopf & Smola, 2001), (Negri, et al., 2011). Há várias evidências empíricas que indicam que a seleção dos *kernels* pode produzir resultados de classificação substancialmente diferentes (Zhu & Blumberg, 2002).

Em (Gualtieri & Crompt, 1998) foi utilizado um classificador SVM para a classificação de vegetação em imagens de SR e o seu desempenho foi avaliado. Posteriormente, imagens hiperespectrais AVIRIS³ foram usadas e os resultados sugeriram a superioridade dos classificadores SVMs que foram comparados com outros classificadores para o mesmo conjunto de dados (Zhang, et al., 2009). Aplicada a outro conjunto de imagens ASTER, a técnica de SVM foi também utilizada para o mapeamento de áreas urbanas (Zhu & Blumberg, 2002). O resultado mostrou que classificadores baseados em SVMs são pouco sensíveis ao tamanho do conjunto de treinamento. As SVMs obtiveram igualmente sucesso na classificação de quantidades limitadas de amostras de treinamento e teste (Foody & Mathur, 2004). Em (Mantero, et al., 2005) foram utilizadas apenas um quarto das amostras de treinamento originais adquiridas sobre imagens do satélite SPOT HRV sem prejuízo importante para a acurácia da classificação.

Uma outra abordagem incorpora a teoria de aprendizagem por inferência transdutiva ao classificador SVM na interpretação de imagens de SR (Bruzzone, et al., 2006). Neste trabalho o hiperplano de separação foi definido através de um processo de integração de amostras não rotuladas em um conjunto com as amostras de treinamento. Todavia os resultados esperados não foram alcançados, uma vez que o método não se mostrou robusto contra a presença de *outliers* (Castillo, et al., 2008).

Em (Zhang, et al., 2009) foi abordada a questão da seleção das amostras de treinamento. Inicialmente, modificaram-se as SVMs de forma a torná-las recursivas, pois o intuito era reduzir a dimensionalidade do conjunto de amostras de treinamento. Os resultados de classificação foram ligeiramente melhores. No

³ Do inglês, *Airborne Visible / Infrared Imaging Spectrometer*.

entanto este método apresentou uma carga computacional maior do que a de métodos alternativos baseados em SVM.

Em (Ghoggali, et al., 2009) a integração de Algoritmos Genéticos (GA) e SVM para classificação de imagens SR foi avaliada. As imagens eram provenientes dos satélites IKONOS e Landsat-TM. Os resultados experimentais revelaram que a fusão das técnicas GA e SVM aumentou a acurácia. Novamente, as SVM mostram a sua capacidade de trabalhar com um pequeno tamanho de amostras de treinamento. Contudo, a carga computacional foi significativa, principalmente devido à convergência lenta do GA.

Um modelo de classificação SVM semi-supervisionado também foi proposto em (Ghoggali, et al., 2009). Como consequência, esta abordagem os resultados envolveram um esforço computacional menor e as amostras de treinamento não confiáveis foram descartados. Finalmente, foram comparados os resultados com algoritmos de classificação amplamente utilizados, tais como SVM progressivo sem-supervisionado, Máxima Verossimilhança (ML) e o algoritmo *k-Nearest Neighbors*. O algoritmo baseado em SVM superou os outros modelos de classificação em termos de robustez e eficácia, especialmente quando foram utilizadas amostras de treinamento não totalmente confiáveis.

Em (Gomez-Chova, et al., 2010) foi proposto um método para aumentar a confiabilidade e acurácia de classificação combinando pixels (amostras) rotulados e não rotulados usando o agrupamento (*clustering*) e a média do mapa de *kernel*. A classificação foi aplicada a imagens de média resolução MERIS. Verificou-se que o método foi particularmente bem sucedido quando a seleção da amostra viés, ou seja mesmo quando as amostras de treinamento e teste seguiram diferentes distribuições.

Recentemente (Prasad, et al., 2011) propôs uma técnica eficiente para a classificação de imagens multiespectrais para Uso e Cobertura do Solo (LULC) utilizando SVM. A técnica utiliza lógica neulosa para segmentação da imagem, posteriormente utiliza SVM no processo de seleção das amostras de treinamento e, finalmente na etapa de classificação os centroides dos clusters são submetidos à SVM treinada para obter classificação. Na técnica de classificação proposta o desempenho foi superior ao de um o algoritmo de agrupamento tradicional.

Finalmente, (Mountrakis, et al., 2011) faz uma revisão extensa da aplicação de SVM em SR e indica que a técnica é um ótimo método para ser utilizado

quando o número de amostras de treinamento é reduzido. Neste trabalho, foi desenvolvido o classificador de segmentos, ou objetos de imagem, conforme *Geographic Object Based Image Analysis* (GEOBIA) (Hay & Castilla, 2008). Estes objetos estão associados a um vetor de atributos, que podem incluir espectrais, textuais, morfológicas e topológicas. Sendo a extensão do trabalho proposto em (Feitosa, et al., 2013).

2.4.2.

Classificadores baseados em Modelos de Markov Ocultos

Os Modelos Ocultos de Markov (HMM) surgiram em meados da década de 60, com o intuito de resolver problemas de processamento de voz (Baum & Petrie, 1966). Posteriormente, trabalhos como de (Aarnio, 1999) e (Juang & Rabiner, 1991), seguindo a mesma linha de pesquisa utilizaram HMM para auxiliar no reconhecimento de voz. Em (Boys, et al., 2000) aplicaram HMM em biologia, neste caso eles estavam interessados em detectar segmentos homogêneos em uma sequência de DNA. Concorrentemente, (Conroy, et al., 2001) aplicou-se HMM para identificar cromossomos. O interesse pela técnica de HMM estendeu-se para outras áreas, em (Otsuka & Ohya, 1997) HMM foram usados para fazer o reconhecimento de expressões faciais e para identificar caracteres em documentos (Huang, et al., 2008), inclusive aplicações de processamento de imagens (Otsuka & Ohya, 1997), (Huang, et al., 2008) SR (Viovy & Saint, 1994), (Aurdal, et al., 2005), (Leite, 2008).

Em (Viovy & Saint, 1994) foi proposta a utilização de HMM como ferramenta para a classificação de diferentes tipos de vegetação. O objetivo foi identificar estágios fenológicos numa região africana. A técnica mostrou que com um modelo relativamente simples, informações importantes sobre a dinâmica da vegetação (início e duração do ciclo) em grande escala podem ser extraídas automaticamente.

Um trabalho análogo (Aurdal, et al., 2005) utiliza HMMs para classificação de tipos de vegetação na Noruega. A eficácia da técnica foi comparada com a técnica ML. A HMM obteve um resultado que superou a taxa de acerto em comparação a ML. Seguindo uma das linhas de pesquisa do Laboratório de Visão

Computacional da PUC-Rio, foi proposta a aplicação de HMM para classificar culturas agrícolas (Leite, 2008). Adicionalmente, foram exploradas sequências temporais de imagens de sensores orbitais, visando reconhecer diferentes tipos de culturas agrícolas através de modelos que consideram a fenomenologia associada às culturas investigadas. O desempenho do método de classificação multitemporal proposto foi comparado com classificações monotemporais, baseadas em ML e os resultados mostraram a superioridade notável do método baseado em HMM.

2.4.3.

Classificação baseada em Cadeias de Markov Nebulosas

Apesar do crescente interesse por sistemas de classificação baseados em conhecimento estruturado através de regras nebulosas, são relativamente poucos os trabalhos encontrados na literatura sobre métodos de classificação em cascata baseados em regras nebulosas, em comparação com as demais abordagens. Em (Mota, et al., 2007) métodos multitemporais nebulosos foram propostos para aplicações de classificação de LULC. O método estima as possibilidades de transições entre classes por GA com base em um conjunto de treinamento. A partir de uma classificação *crisp* da data anterior da região de interesse e dos valores das possibilidades de transição de classe, uma classificação nebulosa é obtida e, em seguida, combinada com a saída da classificação (monotemporal) da mesma região em uma imagem de data posterior. O método é, porém, restrito a aplicações em que a classe correta dos objetos a serem classificados é previamente conhecida em um instante de tempo anterior.

Posteriormente (Costa, 2009) introduziu uma extensão do método proposto em (Mota, et al., 2007) que oferece a opção de utilizar atributos dos objetos no passado, ao invés da informação sobre suas classes verdadeiras. Em (Feitosa, et al., 2011) foram avaliados alguns projetos alternativos do Cadeias de Markov Nebulosas (FMC). O objetivo foi discutir o impacto dos esquemas alternativos e como afetam a acurácia de classificação. Este trabalho é uma ótima orientação para a concepção e aplicação de métodos de classificação baseados em FMC para aplicações de SR. Uma extensão deste trabalho foi apresentada em (Costa & Feitosa, 2014), como uma generalização conceitual do método baseado em FMC.

O método está centrado na hipótese da inversibilidade do modelo classificação com respeito ao tempo, conduzindo a um modelo que pode classificar objetos de imagem em dois pontos no tempo simultaneamente. O resultado é aplicado para a classificação multitemporal de dois locais diferentes rurais e urbano. Todavia, não foi encontrado relatos de uma comparação com SVM e HMM.

3

Fundamentos teóricos dos classificadores multitemporais

3.1.

Introdução

Neste capítulo são apresentadas as formulações matemáticas associadas aos classificadores usados neste trabalho de modo a facilitar a compreensão do presente texto.

Abordam-se os modelos teóricos dos classificadores: Máquina de Suporte Vetorial (SVM), Modelos Ocultos de Markov (HMM) e Cadeias de Markov Nebulosas (FMC).

Descrições mais detalhadas são encontradas em (Vapnik, 1979), (Rabiner, 1989), (Leite, 2008), (Costa & Feitosa, 2014) e (Costa & Feitosa, 2014).

3.2.

Máquina de suporte vetorial

A formulação matemática subjacente às máquinas de suporte vetorial não é passível de ser descrita em poucas páginas, de modo que esta seção se limita a uma apresentação informal do tema.

Em sua concepção original SVM é voltada para problemas em que há duas classes apenas. Em essência, admite-se que os padrões x_i no espaço original de atributos podem ser mapeados por uma função $\phi(x_i)$ em um outro espaço de dimensão mais elevada no qual as classes do problema são linearmente separáveis com razoável acurácia.

O treinamento de uma SVM busca no espaço de dimensão elevada por uma margem em torno de uma reta que separa a maioria dos padrões de uma e outra classe (Boser, et al., 1992), (Cortes & Vapnik, 1995).

Na prática, raras são as aplicações em que existe uma tal faixa que separe perfeitamente as classes do problema. No caso geral haverá padrões de uma classe que recaem no outro lado da faixa, o lado correspondente à classe complementar.

A escolha da reta que separa as classes é guiada por um critério que privilegia soluções que, por um lado, maximizem a margem, e ao mesmo tempo reduzem o número de padrões de treinamento erroneamente classificados.

Assim, um classificador baseado em SVM envolve 2 parâmetros principais:

- O parâmetro de regularização C que evita o *overfitting*; o parâmetro C determina um compromisso entre a importância de maximizar a margem e de minimizar o erro de classificação sobre os padrões de treinamento.
- A função $K(x_i, x_j) \triangleq \phi(x_i)^T \phi(x_j)$ que engloba o mapeamento de dois padrões do espaço de atributos original para o outro espaço de maior dimensão e o produto escalar entre eles neste espaço multidimensional.

A função $K(x_i, x_j)$ é chamada de *kernel*. Inúmeras funções qualificam como *kernel*, como se verá mais adiante. A escolha da função *kernel* é feita a priori pelo projetista do classificador com base em sua experiência e em características particulares da aplicação. Em geral, envolve parâmetros, cujos valores devem também ser estimados via treinamento

3.2.1.

Classificação multi-classe

Na literatura existem vários métodos de classificação multi-classes (Knerr, et al., 1990). Em (Hsu & Lin, 2002) foi realizada a análise de qual seria a melhor técnica para uma classificação multi-classe.

No presente trabalho utilizou-se a proposta de Fu e co-autores (Fu & Lin, 2004) para problemas de classificação que envolvem mais de duas classes. Nesta abordagem a SVM atribui uma probabilidade de que cada padrão pertence a cada classe do problema.

Esta variante do classificador SVM permite uma analogia mais próxima dos demais esquemas de classificação investigados neste trabalho, o que justificou sua escolha.

3.3.

Modelos de Markov ocultos

Nesta seção alguns conceitos básicos sobre HMM são apresentados. Discussões mais completas estão disponíveis em (Rabiner, 1989) e (Leite, 2008).

Trata-se de um processo estocástico definido como uma família de variáveis aleatórias $X(t)$, sendo t geralmente a variável tempo. $X(t)$ representa uma característica mensurável de interesse no tempo t . Os processos estocásticos são usados para descrever o comportamento ao longo de um período de tempo de um sistema onde a incerteza é significativa. Em termos mais formais, a variável aleatória $X(t)$ representa o estado do sistema no instante t .

Um processo estocástico é dito **markoviano**⁴, se a probabilidade do processo estar em um dado estado em um instante de tempo for condicionada a uma sequência finita de estados anteriores. Usando a notação q_t para designar a variável aleatória $X(t)$, as probabilidades condicionais $P(q_{t+1} = \omega_j | q_t = \omega_i)$ são denominadas probabilidades de transição e representam, a probabilidade do estado q_{t+1} ser ω_j no instante $t + 1$ dado que o estado q_t é ω_i no instante t .

3.3.1.

Cadeias de Markov

Um processo markoviano é dito uma **cadeia de Markov** quando as variáveis aleatórias $X(t)$ estão definidas em um espaço de estados discreto. Consideram-se sistemas que podem ser descritos como estando em qualquer instante de tempo em um estado de um conjunto de N estados distintos $(\omega_1, \omega_2, \dots, \omega_N)$ como ilustrado na Figura 2 (onde $N = 4$ para simplificar).

⁴ Homenagem a *Andreyevich Markov*, matemático russo que estudou os processos estocásticos.

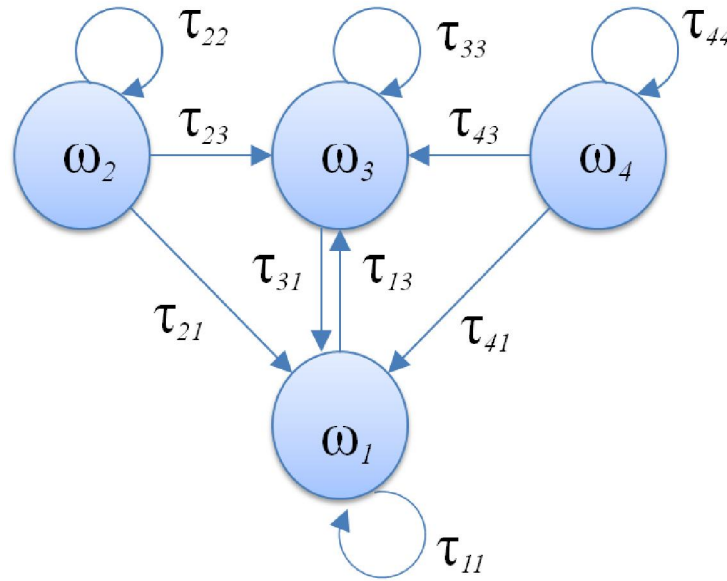


Figura 2: Cadeia de Markov com 4 estados ($\omega_1, \omega_2, \omega_3, \omega_4$), onde as setas indicam as transições possíveis entre os estados (adaptado de (Mota, et al., 2007), (Costa & Feitosa, 2014) e (Leite, 2008)).

Para instantes de tempo igualmente espaçados, o sistema passa por uma mudança de estado (podendo eventualmente retornar ao mesmo estado) de acordo com um conjunto de probabilidades de transição. Cada instante de tempo, onde ocorrem as mudanças de estado, é descrito *por* $t = 1, 2, \dots$, e o estado no tempo t é denotado como o símbolo q_t , considerando um espaço de tempo discreto. Uma descrição probabilística completa do sistema requer a especificação do estado presente (no tempo t) e também dos estados anteriores. Para o caso especial de uma cadeia de Markov de primeira ordem, essa descrição pode ser truncada apenas no estado presente e no seu antecessor:

$$P(q_t = \omega_j | q_{t-1} = \omega_i, q_{t-2} = \omega_k \dots) = P(q_t = \omega_j | q_{t-1} = \omega_i) \quad (1)$$

Considerando processos em que o lado direito da Eq. (1) é independente do tempo, tem-se um conjunto de probabilidades de transição de estado τ_{ij} :

$$\tau_{ij} = P(q_t = \omega_j | q_{t-1} = \omega_i), \quad 1 \leq i, j \leq N, \quad (2)$$

As probabilidades de transição de estado têm as seguintes propriedades:

$$\begin{aligned} \tau_{ij} &\geq 0 \\ \sum_{i=1}^N \tau_{ij} &= 1 \end{aligned} \quad (3)$$

Quando as probabilidades de transição não se alteram ao longo do tempo, diz-se que o processo é estacionário. O processo estocástico mencionado acima poderia ser chamado de modelo de Markov observável uma vez que a saída do processo é um conjunto de estados em cada instante de tempo, onde cada estado corresponde a um evento físico observável. Essa característica torna esses modelos muito restritos para solução de diversos problemas de interesse. Para estes casos aplica-se uma extensão chamada de Modelos Ocultos de Markov (HMM) onde o estado não é diretamente observável, mas uma observação é uma função probabilística do estado. Um HMM é um processo duplamente estocástico composto por um processo latente que não é observável (*hidden*), mas que se manifesta através de outro processo estocástico o qual produz uma sequência de símbolos observáveis (Bunkle & Caelli, 2001).

3.4.

Modelos de Markov ocultos

Em um HMM, as observações (v_i) são símbolos emitidos por estados não observáveis (ω_i) de acordo com determinadas funções probabilísticas, sendo que cada sequência de estados é uma cadeia de Markov de primeira ordem. A Figura 3 ilustra um HMM, sendo N o número de estados do modelo (o conjunto dos estados individuais é representado por $\omega = \omega_1, \dots, \omega_N$ e o estado no tempo t pelo símbolo q_t) e M o número de símbolos distintos observáveis (o conjunto dos símbolos individuais é representado por $V = v_1, \dots, v_M$).

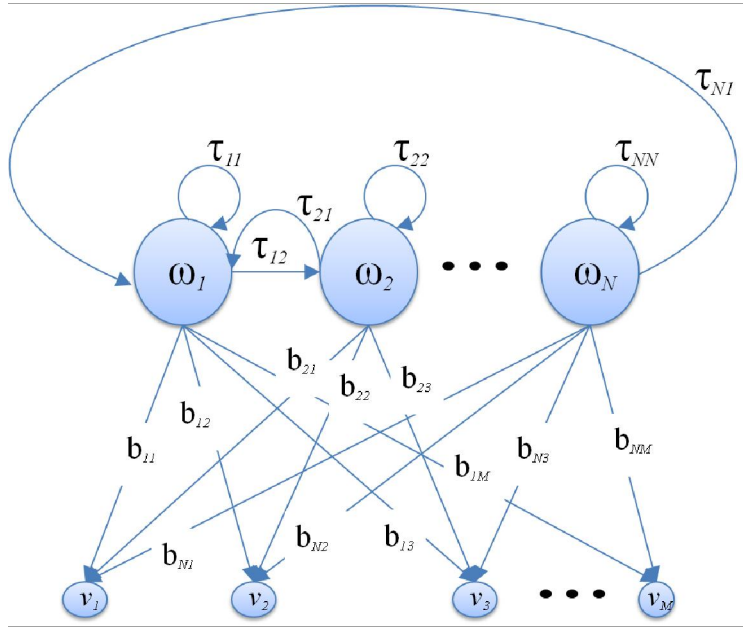


Figura 3: Exemplo de um HMM com estados observáveis, onde $(\omega_i \rightarrow$ estados, $v_k \rightarrow$ símbolos observáveis, $\tau_{ij} \rightarrow$ probabilidade de transição de estado, $b_{jk} \rightarrow$ probabilidade de emissão de símbolo) (adaptado de (Costa e Feitosa, 2014), (Leite, 2008) e (Ra-biner, 1989))

Um HMM básico é constituído de três conjuntos de parâmetros principais:

1. as probabilidades de emissão de símbolo b_{jk} : a probabilidade do símbolo v_k ser emitido pelo estado ω_j , i.e.

$$b_{jk} = P(v_k \text{ em } t | q_t = \omega_j) \\ 1 \leq j \leq N \quad e \quad 1 \leq k \leq M \quad (4)$$

2. as probabilidades de transição de estado τ_{ij} : a probabilidade de estar no estado ω_j no instante de tempo subsequente dado que o estado atual é ω_i , i.e.

$$\tau_{jk} = P(q_{t+1} = \omega_j | q_t = \omega_i) \quad 1 \leq ij \leq N \quad (5)$$

3. a probabilidade a priori π_i do sistema estar em um dado estado ω_i no instante inicial de tempo (não mostrado na Figura 3), i.e.

$$\pi_i = P(q_1 = \omega_i) \quad 1 \leq i \leq N \quad (6)$$

Pode-se observar na Figura 3 que o sistema pode passar de um estado ω_i para outro ω_j , desde que $\tau_{ij} > 0$. Apesar dos estados serem escondidos, em

muitas aplicações práticas, estes possuem significado físico relevante para o problema em questão. Esta revisão foi baseada em (Rabiner, 1989) e (Leite, 2008).

3.5.

Baseado em cadeias de Markov nebulosas

Esta seção apresenta as Cadeias de Markov Nebulosas (FMC). Muitos dos conceitos foram discutidos nas seções anteriores. Seja, $\Omega = \omega_1, \omega_2, \dots, \omega_n$ um conjunto de n classes a matriz. A matriz $T = \{t_{ij}\}, N \times N$ denominada de matriz de transição, representa uma relação nebulosa binária definida pelo produto cartesiano $\Omega \times \Omega$. O símbolo τ_{ij} é a possibilidade de um objeto pertencer a classe $\omega_i \in \Omega$ no tempo t e a classe $\omega_j \in \Omega$ em $t + 1$ com $0 \leq \tau_{ij} \leq 1$, para $i, j = 1, \dots, n$. Esta relação nebulosa pode ser representada graficamente através de um diagrama de transição de classes na Figura 2 Um grafo ponderado e direcionado cujos nós correspondem às classes e ligações às possíveis transições de classes entre t e $t + 1$. Cada ligação é rotulada com a possibilidade de transição de classe τ_{ij} . Por simplicidade as ligações $\tau_{ij} = 0$ não são desenhadas.

Define-se o vetor ${}^t\alpha = [{}^t\alpha_1, \dots, {}^t\alpha_n]$ para $0 \leq {}^t\alpha_i \leq 1$ como a classificação nebulosa de um objeto no tempo t . Assume-se ainda que ${}^t\alpha_i$ é uma função dos valores dos atributos do objeto de imagem no tempo t . Baseado no vetor de rótulos nebulosos ${}^t\alpha$ e na matriz de transição T , o FMC calcula os valores de pertinência às classes, representados pelo vetor ${}^{t+1}\beta = [{}^{t+1}\beta_1, \dots, {}^{t+1}\beta_n]$ para o mesmo objeto uma unidade de tempo mais tarde, através da aplicação da seguinte formulação:

$${}^{t+1}\beta_j = \bigoplus_{i=1, \dots, n} \left\{ \top({}^t\alpha_i, \tau_{ij}) \right\} \quad (7)$$

onde $i, j = 1, \dots, n$. Os símbolos \bigoplus e \top respectivamente representam a t-norma e a s-norma. A lei de transição introduzida em Eq. (7) pode ser expressa em forma mais compacta pela Eq. (8).

$${}^{t+1}\beta = {}^t\alpha \circ T \quad (8)$$

O símbolo " \circ " denota um tipo especial de multiplicação de matrizes, análogo à multiplicação de matrizes convencionais, onde o produto é substituído por o operador *t-norma* e a somatória é substituído por o operador *s-norma*. Os símbolos $^{t+1}\beta_j$ e $^{t+1}\alpha_j$ denotam diferentes pertinências, embora ambos se refiram ao mesmo objeto na data $t + 1$. Porém, enquanto, $^{t+1}\alpha$ foi calculado com base unicamente em valores das características de tempo $t + 1$, $^{t+1}\beta$ é o resultado da aplicação da lei de transição FMC sobre os valores das pertinências em t calculados a partir dos valores das propriedades do objeto no tempo t . Para uma descrição geral do conceito FMC, o leitor deve referir-se a (Avrachenkov & Sanchez, 2002). Todavia, em (Costa & Feitosa, 2014) explora-se a generalização do problema de classificação de imagens multitemporais em cascata.

4

Modelos dos classificadores em cascata

4.1.

Introdução

Neste capítulo apresentam-se os três esquemas de classificação em cascata analisados nesta dissertação.

Os termos monotemporal e multitemporal foram utilizados para designar classificadores, cujas entradas referem-se, respectivamente, a uma data única ou a múltiplas datas.

4.2.

Classificação monotemporal

A Figura 4 ilustra o modelo de classificação monotemporal geral, em que ${}^{tA}\mathbf{I}$, denota uma imagem adquirida em um instante de tempo tA . Admite-se há K classes de LULC, denotadas aqui como $\{\omega_0, \dots, \omega_{K-1}\}$. Neste caso, os padrões a classificar são segmentos descritos por um vetor de atributos. O vetor de atributos pode incluir características espectrais, textuais, morfométricas. Daqui em diante o vetor de atributos de um determinado segmento no instante de tempo tA é representado por ${}^{tA}\mathbf{X} = [{}^{tA}X_0, \dots, {}^{tA}X_{n-1}]$.



Figura 4: Classificador monotemporal

O problema de classificação monotemporal pode ser apresentado da seguinte maneira: dado ${}^{tA}\mathbf{X}$ deseja-se encontrar ${}^{tA}\omega$. Define-se ${}^{tA}\alpha_i$ como sendo a probabilidade de um segmento pertencer à classe ω_i em tA para $i = 0, \dots, k - 1$.

4.3.

Abordagens de classificação multitemporal em cascata

Diferentemente do caso monotemporal, consideram-se duas imagens adquiridas em dois instantes distintos. A Figura 5 ilustra o classificador multitemporal em cascata. Considera-se ${}^{tB}\mathbf{I}$ e ${}^{tA}\mathbf{I}$ duas imagens co-registradas, cobrindo a mesma área geográfica, adquiridas respectivamente nas datas tA e tB . Neste trabalho, adotam-se os termos "imagem alvo" e a "imagem complementar" como associados às imagens adquiridas nos tempos tA e tB respectivamente.



Figura 5: Classificador multitemporal em cascata.

Em outras palavras, ${}^{tA}\mathbf{I}$ é a imagem que vai ser classificada, enquanto ${}^{tB}\mathbf{I}$ representa uma fonte de informações adicionais para a classificação de ${}^{tA}\mathbf{I}$.

Consequentemente, utilizando o conceito de vetor de atributos para uma imagem capturada em um instante de tempo, ${}^{tA}\mathbf{X} = [{}^{tA}X_0, \dots, {}^{tA}X_{n-1}]$ e ${}^{tB}\mathbf{X} = [{}^{tB}X_0, \dots, {}^{tB}X_{n-1}]$ representam vetores de atributos das datas alvo e complementar.

4.3.1.

Classificação multitemporal em única data

Os vetores de atributos ${}^{tA}\mathbf{X} = [{}^{tA}X_0, \dots, {}^{tA}X_{n-1}]$ e ${}^{tB}\mathbf{X} = [{}^{tB}X_0, \dots, {}^{tB}X_{n-1}]$, formam as entradas do classificador.

Neste caso, a "**imagem alvo**" é a imagem a ser classificada e a "**imagem complementar**" fornece dados adicionais para a classificação da imagem alvo. Em suma, a abordagem de classificação multitemporal de única data com fusão de dados, utiliza as informações dos vetores de atributos calculados a partir de ambas as imagens.

O classificador gera um vetor de probabilidades, cujos elementos indicam a probabilidade de um segmento de entrada pertencer a uma das classes de LULC na data da imagem alvo. Após o cálculo das probabilidades uma função, (e.g. **H**), seleciona a classe correspondente à maior probabilidade.

4.3.2.

Classificação multitemporal em duas datas

Neste caso o esquema multitemporal classifica um segmento em duas datas, simultaneamente. O classificador calcula vetores de probabilidade ${}^{tAtB}\alpha$, cujos elementos são as probabilidades de que um segmentos pertença a um determinado par de classes em tA e tB .

Uma função **H** seleciona o par de classes com maior probabilidade. É importante ressaltar que se tem um conhecimento prévio das possíveis transições.

4.4.

Particularização dos modelos de classificação

Nesta seção são apresentados as diferentes abordagens de classificação multitemporal estudadas nesta dissertação:

- Máquina de Suporte Vetorial (SVM)
- Modelos Ocultos de Markov (HMM)
- Cadeias de Markov Nebulosas (FMC)

4.4.1.

Método baseado em SVM

O modelo usado nesta análise foi implementado com a (LIBSVM). A Figura 8 ilustra o procedimento de classificação.

Para o treinamento das SVMs e posterior classificação foram usadas as funções *svmtrain* e posteriormente *svmpredict* (Chang & Lin, 2011) da LIBSVM (Chang & Lin, 2011).

Antes do treinamento e classificação os dados dos segmentos são normalizados, conforme o código abaixo (ver código 1).

Código 1: Exemplo de normalização dos dados (objetos) para LibSVM

```
1 Normaliza segmentos
2 Lmin=min([TREINA; AMOSTRA]);
3 TREINA=TREINA-repmat(Lmin,[size(TREINA,1) 1]);
4 AMOSTRA=AMOSTRA-repmat(Lmin,[size(AMOSTRA,1) 1]);
5 Lmax=max([TREINA; AMOSTRA]);
6 TREINA=2*(TREINA./repmat(Lmax,[size(TREINA,1) 1]))-1;
7 AMOSTRA=2*(AMOSTRA./repmat(Lmax,[size(AMOSTRA,1) 1]))-1;
```

Para estimar os parâmetros C e γ definiu-se a grade de busca, conforme o código abaixo (ver código 2).

Código 2: Definição do Grid-search

```
1 Grid-search / Grade para estimativa de valores ótimos
2 C e Gama
3 log2cRange = [-1:1:3];
4 log2gRange = [-4:1:2];
```

Após varrer o espaço de valores para os parâmetros C e γ , os candidatos que apresentam melhor acurácia sobre os dados de treinamento são selecionados.

De posse destes parâmetros é possível gerar o modelo que irá calcular as estimativas de probabilidade.

No Código 3 mostra-se um exemplo da função *svmpredict*, que utiliza o modelo gerado para estimar a classe a que pertence cada segmento.

Código 3: Geração de modelo para predição com SVM

```

1 % Treinamento do SVM e geração de modelo
2 MODELO =svmtrain(CLASSES, DADOS_TREINAMENTOS, ...
3               Melhores_Parametros_CeGama);
4 % Classifica dados de treinamento
5 [~,~,prob_treinamento] = svmpredict(CLASSES, ...
6               DADOS_TESTE, MODELO);
7 % Fim da estimativa

```

A Figura 6 ilustra a rotina de treinamento da LIBSVM.

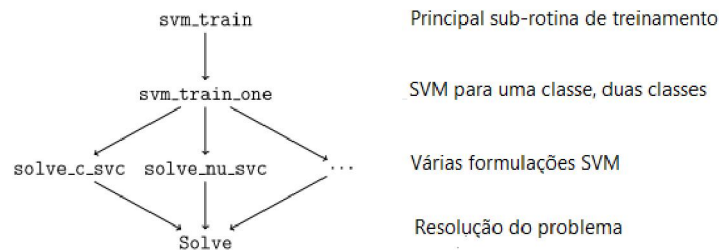


Figura 6: Rotina de treinamento SVM (adaptado de **(Chang & Lin, 2011)**).

4.4.2.

Parâmetros de seleção

LIBSVM fornece uma ferramenta simples para selecionar parâmetros. Para cada parametrização, LIBSVM calcula a acurácia da validação cruzada (do inglês, *cross-validation*) CV.

Finalmente, os parâmetros que proporcionam maior acurácia são devolvidos como resposta. A ferramenta de seleção de parâmetro utiliza uma função de base radial (RBF) $\phi(\mathbf{x}, \mathbf{c}) = \phi(\|\mathbf{x} - \mathbf{c}\|)$ definida como:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (9)$$

Os parâmetros (C, γ) são estimados da seguinte forma. O usuário pode definir um possível intervalo para C e γ para que seja gerada uma grade de valores. Todos os pontos da grade são avaliados para se descobrir, qual é o par (C, γ) que alcança a melhor acurácia. Os valores assim obtidos para (C, γ) foram

usados para o treinamento das SVM. Finalmente, gera-se assim um modelo de classificação.

Para classificação multi-classe, o problema de encontrar o melhor par (C, γ) é tratado como um problema "um-contra-outro".

Classificação monotemporal baseada em SVM

Uma SVM produz as estimativas de probabilidade ${}^tA\alpha$ com base nas entradas tAX conforme representado na Figura 7.

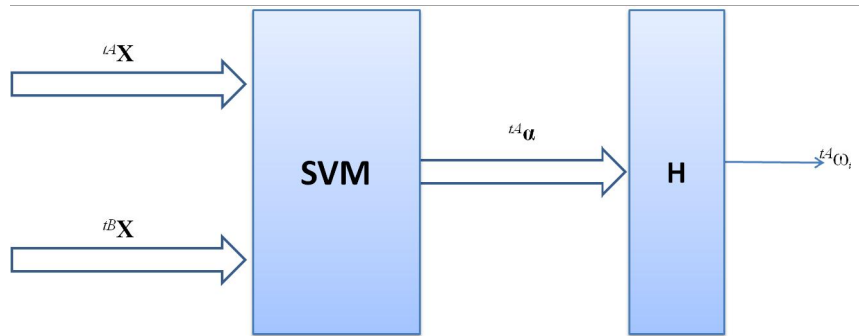


Figura 7: Classificador monotemporal baseado em SVM (adaptado de (Feitosa, et al., 2013)).

A função H seleciona a classe com maior probabilidade.

Classificação multitemporal baseada em SVM e fusão de dados para uma única data

A Figura 8 apresenta o procedimento de classificação baseado em fusão de dados para uma única data.

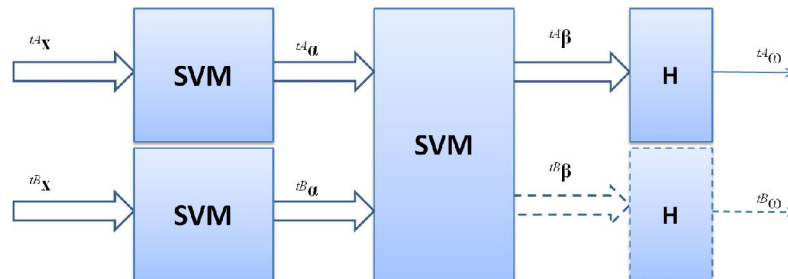


Figura 8: Esquemas de classificação multitemporal com fusão de dados para uma única data (considerem-se as linhas solidas e as linhas pontilhadas) (adaptado de (Feitosa, et al., 2013)).

Um classificador SVM produz para cada segmento da imagem ${}^{tA}\mathbf{I}$ um vector das estimativas de probabilidade ${}^{tA}\boldsymbol{\alpha} = [{}^{tA}\alpha_0, \dots, {}^{tA}\alpha_{k-1}]$ com base numa entrada formada pelo empilhamento de ${}^{tA}\mathbf{X}$ e ${}^{tB}\mathbf{X}$. A função \mathbf{H} selecciona a classe que apresenta a maior probabilidade.

Este conceito é ilustrado na Figura 8, este método de classificação considera somente as linhas sólidas.

Classificação multitemporal baseada em SVM e fusão de dados para duas datas

O método que SVM para duas datas funciona como o método anterior. A diferença está na definição das classes. Neste caso cada rótulo é composto por um par de classes, por exemplo (ω_1, ω_2) , (ω_1, ω_3) , etc., em que o primeiro rótulo se refere à classe em tA e o segundo à classe em tB .

Para efeito de avaliação de desempenho considera-se apenas o rótulo do par correspondente à data alvo.

Então, se o problema envolve, por exemplo, k classes, haveria em teoria até k^2 classes a serem classificadas. Na prática, contudo, o número de classes será menor, ou mesmo bem menor do que k^2 , sendo que só foram considerados os pares de classes que ocorrem no conjunto de treinamento. Na Figura 8 este esquema é representado pelas linhas sólidas e pelas linhas pontilhadas.

4.4.3.

Métodos baseados em HMM

O modelo probabilístico multitemporal utilizado nesta análise baseia-se em (Leite, et al., 2011).

Seja $\Omega = \{\omega_1, \dots, \omega_n\}$ um conjunto de n classes de LULC. ${}^{tA}\omega$ e ${}^{tB}\omega \in \Omega$ são os rótulos correspondentes às classificações de um segmento nos instantes de tempo tA e tB . Similarmente, ${}^{tA}\mathbf{X}$ e ${}^{tB}\mathbf{X}$ são vetores de atributos de uma mesmo segmento extraídos das imagens ${}^{tA}\mathbf{I}$ e ${}^{tB}\mathbf{I}$.

O classificador HMM determina a classe mais provável ${}^{tA}\Omega_p \in \Omega$ para o segmento na imagem alvo, segundo o seguinte critério:

$${}^{tA}\Omega_p = \arg \max_{{}^{tA}\omega} \left[p({}^{tA}\mathbf{X} | {}^{tA}\omega) \sum_{{}^{tB}\omega} ({}^{tB}\omega) p({}^{tB}\mathbf{X} | {}^{tB}\omega) p({}^{tA}\omega | {}^{tB}\omega) \right], \quad (10)$$

onde $p({}^{tA}\mathbf{X} | {}^{tA}\omega)$ e $p({}^{tB}\mathbf{X} | {}^{tB}\omega)$ representam a verossimilhança dos atributos observados para as classes ${}^{tA}\omega$ e ${}^{tB}\omega$ respectivamente, $P({}^{tB}\omega)$ denota a probabilidade a priori da classe ${}^{tB}\omega$ para o instante de tempo tB e $p({}^{tA}\omega | {}^{tB}\omega)$ a probabilidade de transição da classe ${}^{tB}\omega$ em tB para ${}^{tA}\omega$ em tA , para ${}^{tB}\omega, {}^{tA}\omega \in \Omega$.

Classificação monotemporal baseada em máxima verossimilhança

A Figura 9 ilustra a classificação monotemporal baseada em um classificador de máxima verossimilhança (ML). A verossimilhança um objeto de imagem em relação à classe ω_i no instante tA pode ser definida como:

$${}^{tA}\alpha_i = P({}^{tA}\omega_i) p({}^{tA}\mathbf{X} | {}^{tA}\omega_i) \quad (11)$$

Neste trabalho admite-se em todos os casos que os atributos de cada classe em todas as épocas estão distribuídos como uma Gaussiana multidimensional.

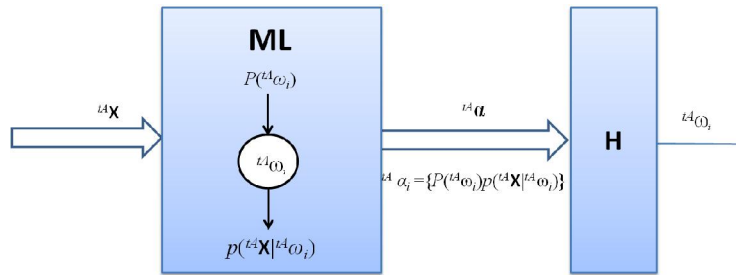


Figura 9: Classificador monotemporal baseado em ML.

A função **H** seleciona a classe para a qual a verossimilhança é maior.

Classificação multitemporal baseada em HMM e fusão de dados para uma única data

O classificador multitemporal com fusão de dados para uma única data tem como entrada os vetores de atributos ${}^{tA}\mathbf{X}$ e ${}^{tB}\mathbf{X}$, e como saída o vetor de

probabilidades ${}^{tA}\alpha$, , cujos elementos ${}^{tA}\alpha_i$, referem-se à probabilidade de um objeto pertencer à classe ω_i em tA :

$${}^{tA}\alpha_i = p({}^{tA}\mathbf{X}|{}^{tB}\omega_j) \sum_{j=0}^{k-1} P({}^{tB}\omega_j) p({}^{tB}\mathbf{X}|{}^{tB}\omega_j) P({}^{tA}\omega_i|{}^{tB}\omega_j) \quad (12)$$

onde $P({}^{tA}\omega_i|{}^{tB}\omega_j)$ é a probabilidade de transição da classe ω_i em tA para ω_j em tB . Observe-se o esquema de classificação na Figura 10.

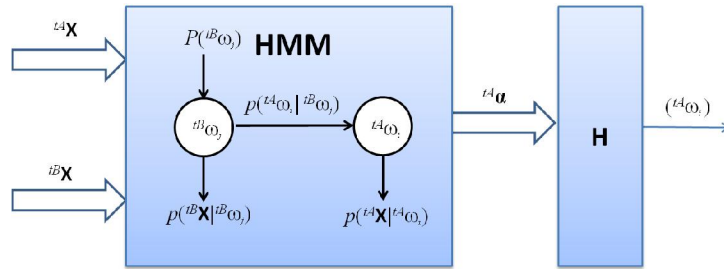


Figura 10: Classificador multitemporal HMM baseado em fusão de dados para uma única data.

A função **H** seleciona a classe para a qual a verossimilhança é a maior.

Classificação multitemporal baseada em HMM e fusão de dados para duas datas

O classificador HMM para duas datas baseado em fusão de dados calcula a probabilidade do objeto pertencer à classe ω_i de acordo com a Eq.(12). A classe a que o objeto pertence em tB é selecionada como a classe que maximiza o termo à direita na equação.

$${}^{tA}\alpha_i = p({}^{tA}\mathbf{X}|{}^{tB}\omega_i) \max_j [P({}^{tB}\omega_j) p({}^{tB}\mathbf{X}|{}^{tB}\omega_j) P({}^{tA}\omega_i|{}^{tB}\omega_j)] \quad (13)$$

A Figura 11 apresenta o esquema de classificação. A função **H** seleciona as classes em tA e tB para as quais o valor da Eq. (12) é máximo.

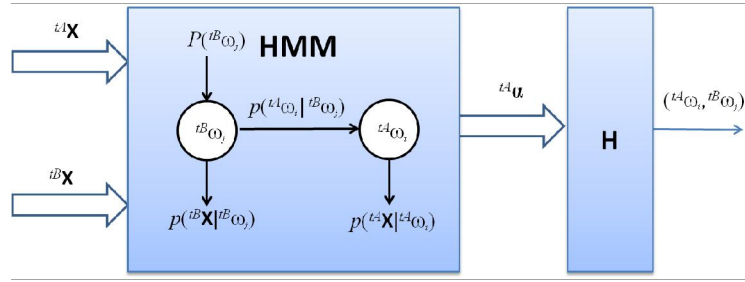


Figura 11: Classificador multitemporal baseado em HMM

Maiores detalhes sobre o procedimento para estimativa dos parâmetros do modelo, podem ser encontrados em (Leite, et al., 2011).

4.4.4.

Método baseado em FCM

Classificador Monotemporal Fuzzy

O classificador monotemporal utilizado foi originalmente proposto por (Mota, et al., 2007).

O classificador chamado de monotemporal, baseia-se exclusivamente em descrições espectrais e espaciais de objetos ${}^{tA}\mathbf{X}$ no instante de tempo tA . Este produz um vetor de rótulos nebulosos de n dimensões denotados por ${}^{tA}\boldsymbol{\alpha} = [{}^{tA}\alpha_1, \dots, {}^{tA}\alpha_n]^T$, onde ${}^{tA}\alpha_i$ representa a pertinência do objeto da imagem associada pelo classificador monotemporal à classe ω_i para $i = 1, 2, \dots, n$, sendo que para pelo menos um i , ${}^{tA}\alpha_i \neq 0$.

A figura 12 ilustra o esquema de classificação. O classificador genericamente, é representado pela função ${}^{tA}C$.

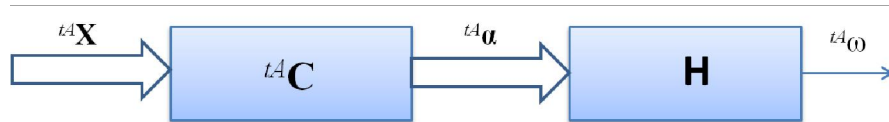


Figura 12: Classificador monotemporal fuzzy (adaptado de (Feitosa, et al., 2009))

O classificador monotemporal ${}^{tA}C$ produz um vetor nebuloso ${}^{tA}\alpha = [{}^{tA}\alpha_1, \dots, {}^{tA}\alpha_n]$, onde ${}^{tA}\alpha_i$ representa o grau de pertinência do objeto à classe ω_i , Eq. (14).

$${}^{tA}\alpha = {}^{tA}C({}^{tA}\mathbf{X}) \quad (14)$$

Analogamente, o classificador nebuloso pode ser usado para classificar o objeto em tB conforme a Eq.(15)

$${}^{tB}\alpha = {}^{tB}C({}^{tB}\mathbf{X}) \quad (15)$$

O passo final é a defuzzificação, a função **H** seleciona a classe para a qual foi calculado o maior grau de pertinência.

Classificação multitemporal FMC

O método multitemporal consiste essencialmente em combinar os resultados de dois classificadores monotemporais de duas datas distintas, produzindo um resultado final único.

Mostra-se na Figura 13 como os classificadores monotemporais participam no esquema multitemporal, eles calculam graus de pertinência difusos, os quais são combinados, produzindo outra classificação (melhorada) *fuzzy*, a qual posteriormente é defuzzificada.

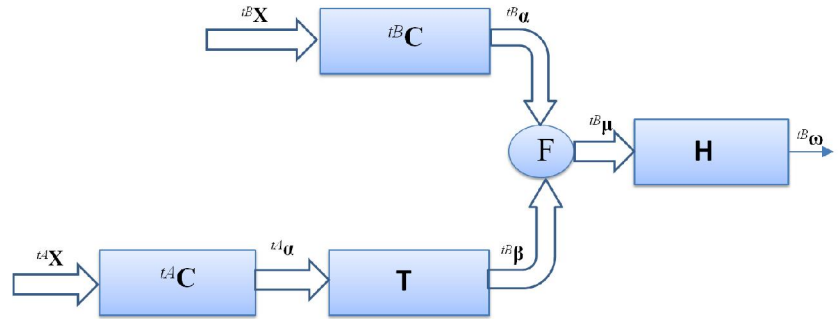


Figura 13: Classificador multitemporal FMC, para um modelo *forward* (adaptado de (Costa & Feitosa, 2014)).

A partir da classificação monotemporal nas duas datas, pode-se estimar a classificação multitemporal no instante de tempo desejado, (i.e., tB) usando a lei de transição expressa na Eq(16).

$${}^{tA}\boldsymbol{\beta} = {}^{tA}\boldsymbol{\alpha} \circ \mathbf{T}, \quad (16)$$

que, combinada com a Eq. (15) toma a seguinte forma:

$${}^{tB}\boldsymbol{\beta} = {}^{tA}CC({}^{tA}\mathbf{X}) \circ \mathbf{T}, \quad (17)$$

onde \mathbf{T} representa uma relação nebulosa.

Os vetores de rótulos nebulosos ${}^{tB}\boldsymbol{\alpha}$ e ${}^{tB}\boldsymbol{\beta}$ são combinados no próximo passo por uma função de agregação \mathbf{F} para formar o vetor de pertinências multitemporal fuzzy ${}^{tB}\boldsymbol{\mu} = [{}^{tB}\mu_1, \dots, {}^{tB}\mu_n]$, como apresenta a Eq (18).

$${}^{tB}\boldsymbol{\mu} = \mathbf{F}({}^{tB}\boldsymbol{\alpha}, {}^{tB}\boldsymbol{\beta}) = \mathbf{F}[{}^{tB}C({}^{tB}\mathbf{X}), {}^{tA}C({}^{tA}\mathbf{X}) \circ \mathbf{T}] \quad (18)$$

O passo final é a defuzzificação, executada por uma função \mathbf{H} que seleciona a partir do vetor nebuloso ${}^{tB}\boldsymbol{\mu}$ a classe com maior pertinência em tB .

Juntando tudo, na Eq. (19) o classificador multitemporal é descrito \mathbf{M} graficamente na Figura 13.

$${}^{tB}\boldsymbol{\Omega}_p = \mathbf{M}({}^{tB}\mathbf{X}, {}^{tA}\mathbf{X}) = \mathbf{H}\{\mathbf{F}[{}^{tB}C({}^{tB}\mathbf{X}), {}^{tA}C({}^{tA}\mathbf{X}) \circ \mathbf{T}]\} \quad (19)$$

Detalhes sobre o procedimento para estimativa dos parâmetros do modelo, podem ser encontrados em (Leite et al., 2011).

5

Análise e resultados

5.1.

Introdução

Os experimentos descritos neste capítulo foram projetados para avaliar e comparar os métodos de classificação descritos no capítulo anterior. Os conjuntos de dados usados nos experimentos descritos foram usados em trabalhos anteriores, como (Mota, et al., 2007), (Feitosa, et al., 2013) e (Costa & Feitosa, 2014). Posteriormente, uma descrição dos conjuntos de dados é apresentada. Foram testados os modelos de classificação monotemporais e multitemporais para uma única data e para duas datas aplicando os métodos SVM, HMM e FMC. Os resultados da aplicação dos métodos em várias experiências diferentes são apresentados e discutidos, ressaltando suas particularidades.

5.2.

Descrição do conjunto de dados

Foram usados dois conjuntos de dados com características diferentes, para avaliar o desempenho dos métodos propostos, em áreas rurais e áreas urbanas.

5.2.1.

Área de teste: Alcinópolis

A área rural de teste está situada no município de Alcinópolis, no Estado de Mato Grosso do sul, Brasil, (vide Figuras 14 e 15). Os rios da região estão localizados na sub-bacia do rio Taquari, parte da bacia do rio Paraguai, e cabeceiras do Pantanal, um dos ecossistemas mais importantes da América do Sul.

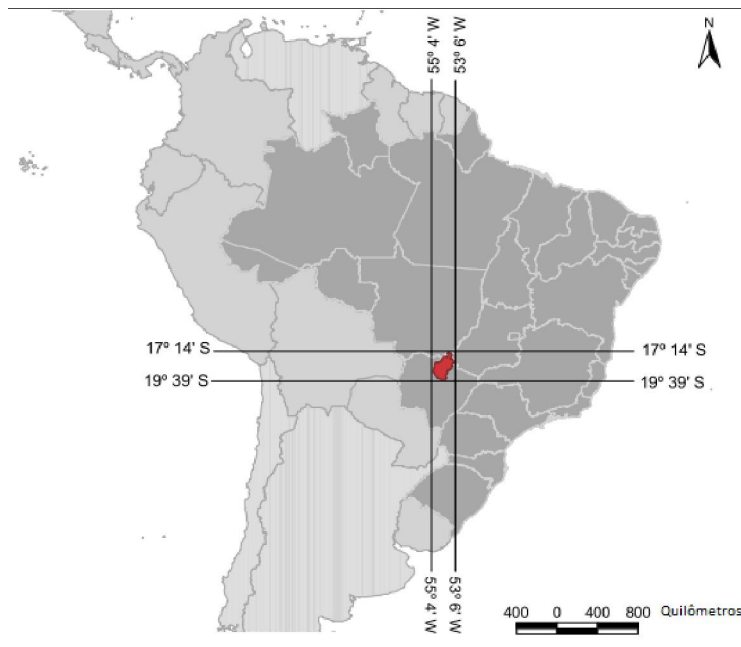


Figura 14: Área de teste na sub-bacia do rio Taquari (extraído de **(Costa, et al., 2009)**)

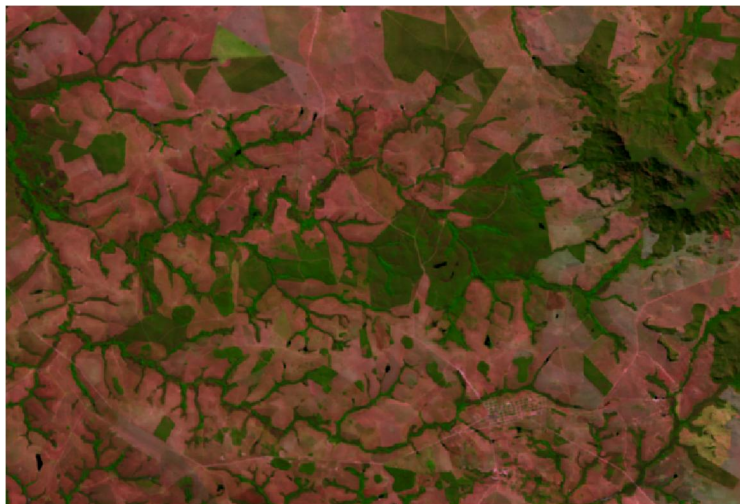


Figura 15: Subconjunto da imagem do site de teste de Alcinópolis, MS (extraído de (Mota, 2004))

Duas imagens LANDSAT corrigidas, capturadas em 1999 e 2001 foram selecionadas para os experimentos realizados. A área de teste tem uma superfície de 459km (Costa & Feitosa, 2014).

Procedimento de segmentação

As imagens usadas neste experimento possuem as bandas três, quatro e cinco do satélite. A segmentação seguiu o procedimento descrito em (Costa, 2009). O resultado da segmentação gerou 18385 objetos.

Validação dos dados

Um foto-interprete classificou um total de 442 segmentos em cada um dos anos. Como dados de entrada, foram empregados dados de uma videografia realizada pela Embrapa Solos em conjunto com o Instituto Conservação Internacional do Brasil em Outubro de 2001. A descrição completa dos resultados de referência desta área de teste pode-se encontrar em (Mota, 2004). A Tabela 1 contém a descrição das classes consideradas nos experimentos.

Tabela 1: Classes consideradas nos experimentos e a suas descrições (extraída de (Costa, 2009), Tabela 2, pg 62)

Rótulo	Classe	Descrição
ω_1	Solo descoberto	Um solo que se degradou pela erosão ou que está sendo preparado para o cultivo
ω_2	Florestas Ripárias	Densas florestas ao longo de rios e córregos
ω_3	Pasto	Pastagem cultivada para alimentação de gado bovino
ω_4	Corpos d'água	Água e pantanais
ω_5	Savana Densa	Formação densa de pequenas árvores (8-12m) de altura

Diferentemente das referências encontradas em (Costa, 2009) e (Mota, 2004) simularam-se nestes experimentos condições de dinâmica diferentes. Entende-se aqui por dinâmica a proporção dos segmentos que mudam de classe de uma data para outra. Cabe notar que a classe da maioria dos segmentos não se altera. O expediente utilizado para simular dinâmicas distintas consistiu em limitar em nossos experimentos o número de mostras que representam cada transição a 100 ou a 1000 para treinamento e teste. Assim, o limite 100 corresponde a alta dinâmica enquanto o limite de 1000 a baixa dinâmica. As Tabelas 2 e 3 mostram o número de amostras para cada tipo de transição.

Tabela 2: Simulação da dinâmica de classes 100 (Alta dinâmica)

Classes 1999	Classes 2001				
	Solo Descoberto	Ribeirinho	Pasto	Água	Savana
Solo Descoberto	38		38		
Ribeirinho		56			
Pasto	8		200		
Água				26	
Savana					126

Tabela 3: Simulação da dinâmica de classes 1000 (Baixa dinâmica)

Classes 1999	Classes 2001				
	Solo Descoberto	Ribeirinho	Pasto	Água	Savana
Solo Descoberto	38		38		
Ribeirinho		56			
Pasto	8		448		
Água				26	
Savana					126

Importa desde já realçar que os resultados apresentados mais adiante são resultados de vários experimentos em que se fizeram escolhas aleatórias das amostras de cada transição segundo os números apresentados nas Tabelas 2 e 3.

5.2.2.

Área de teste: Rio de Janeiro

O segundo local de teste foi uma região urbana na cidade do Rio de Janeiro. A uma área de 14.4 km, situada na parte do Parque Nacional Tijuca. Utilizaram-se dados colhidos pelo projeto PIMAR (PIMAR, 2010).

Duas imagens IKONOS II *Pan-sharpened* e ortorretificadas sem correções radiométricas foram usadas nos experimentos.

As imagens foram adquiridas em março 2008 (fim do verão) e junho de 2009 (final do outono), e apresentam condições de iluminação bem diferentes devido às épocas distintas do ano. A Figura 16 mostra um subconjunto da região monitorada.

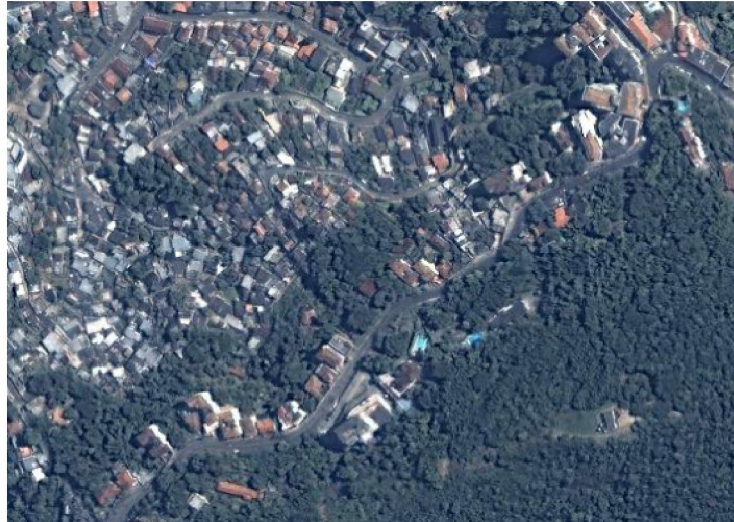


Figura 16: Subconjunto da imagem da área de teste do Rio de Janeiro 2008 (extraído de (Costa & Feitosa, 2014)).

As imagens ortorretificadas foram segmentadas usando o algoritmo de segmentação *multi-resolution* proposto em (Baatz & Schäpe, 2000), utilizando o *Definiens Software Developer 7R*. Todos os segmentos gerados para as duas imagens foram visualmente classificados por especialistas. A Tabela 4 descreve as LC em estudo.

Tabela 4: Classes consideradas no local de teste Rio de Janeiro (extraída de (Costa & Feitosa, 2014))

Rótulo	Classe	Descrição
ω_1	Rocha	Formação de pedra (granito) Compos de grama formados naturalmente criados por formações antrópicas
ω_2	Campo	
ω_3	Urbano	Área construída (edifícios, estrada, etc.)
ω_4	Árvores	Individual o grupo de árvores

Os segmentos que possuíam sombras foram classificados também, porém não foram considerados nos experimentos.

Após a classificação, todos os segmentos de cada classe em 2008 foram fundidos com os segmentos adjacentes da mesma classe, gerando segmentos de área maior. Então, somente os segmentos de 2009, contidos completamente dentro dos segmentos 2008 (gerados pelo o processo de fusão) foram selecionados. Para todos os segmentos selecionados, os valores de atributos foram calculados para

cada uma das duas imagens. Os atributos calculados para cada segmento foram: os valores médios das quatro bandas espectrais e a entropia (para todas as bandas em todas as direções).

A Tabela 5 apresenta o número de segmentos designados para cada classe de cada ano e as transições de classe observadas 2008-2009.

Tabela 5: Número de transições entre 2008 e 2009 para o local de teste urbano.

Classes 2008	Classes 2009				
	Rocha	Campo	Zona Urbana	Árvores	Total
Rocha	188	10	0	1	199
Campo	11	421	66	153	651
Zona Urbana	-	9	5594	220	5814
Árvores	1	194	390	33947	34532
Total	200	634	6050	34321	41205

A dinâmica temporal também foi alterada neste conjunto de dados, como apresentado nas tabelas 6 e 7.

Tabela 6: Simulação da dinâmica de classes 100 (Alta dinâmica)

Classes 2008	Classes 2009			
	Rocha	Campo	Zona Urbana	Árvores
Rocha	188	10	0	
Campo	10	200	64	152
Zona Urbana	-	8	200	200
Árvores	1	194	200	200

Tabela 7: Simulação da dinâmica de classes 1000 (Baixa dinâmica)

Classes 2008	Classes 2009			
	Rocha	Campo	Zona Urbana	Árvores
Rocha	188	10	0	
Campo	10	420	64	152
Zona Urbana	-	8	2000	220
Árvores	1	194	390	2000

5.3.

Resultados

Os objetivos dos experimentos descritos deste ponto em diante podem ser resumidos da seguinte forma:

- Experimento 1: Avaliar o desempenho geral dos modelos de classificação;
- Experimento 2: Avaliar como a "qualidade" dos dados trazidos da outra data afeta o desempenho da classificação multitemporal;
- Experimento 3: Verificar a robustez dos modelos de classificação contra *outliers* ;
- Experimento 4: Avaliar o benefício do uso de conhecimento a priori sobre as transições de classe possíveis;
- Experimento 5: Verificar a robustez quando o número de ocorrências de transição de classe é diferente no conjunto de treinamento e teste.

Para cada transição de classe a metade das amostras foram selecionadas aleatoriamente para formar o conjunto de treinamento e a outra metade das amostras foi selecionada como conjunto de teste. Os experimentos foram executados 20 vezes cada vez com uma seleção diferente dos conjuntos de treinamento e de teste. Os resultados apresentados representam a média de 20 iterações. Para HMM $p({}^{tB}\mathbf{X}|{}^{tB}\omega)$. Além disso, em certos experimentos, informação sobre a verdade terrestre da data complementar é considerada conhecida. Neste caso, para HMM $p({}^{tA}\omega)p({}^{tA}\mathbf{X}|{}^{tA}\omega)$ foi definido como 1 (um) se ${}^{tA}\omega$ é a classe verdadeira e 0 (zero) como contrário. Similarmente, para FMC ${}^{tA}\alpha_i$ recebe 1 quando i é o índice da classe correta, e os outros ${}^{tA}\alpha_j$ recebem 0.

5.3.1.

Experimento 1

Objetivo e procedimento

No primeiro experimento foi avaliado o desempenho geral dos classificadores. Este experimento teve a finalidade verificar a sensibilidade de cada abordagem à dinâmica das áreas alvo (vide as seções 5.2.1 e 5.2.2).

Neste experimento os dados foram distribuídos de modo que cada transição de classe ocorra com a mesma frequência relativa tanto no treinamento quanto no teste.

Resultados e Discussão

As figuras a seguir devem ser interpretadas da seguinte forma: as colunas vermelhas e azuis, representam respectivamente as situações de dinâmica baixa e alta respectivamente. Os resultados são apresentados em dois grupos, cada grupo relativo a uma data de aquisição da imagem da área alvo.

A legenda indicada abaixo de cada par de colunas azul/vermelha corresponde à seguinte convenção:

- SVM_M: classificador monotemporal baseado em SVM;
- SVM_DOU: classificador multitemporal baseado em SVM de par de datas;
- ML_M: classificador monotemporal baseado em ML;
- HMM_SIN: classificador multitemporal baseado em HMM de data única;
- HMM_DOU: classificador multitemporal baseado em HMM de par de datas;
- FMC_M: classificador monotemporal usado no esquema FMC;
- FMC_MUL: classificador multitemporal baseado em FMC.

A Figura 17 mostra os resultados de acurácia global para as duas áreas de teste para ambas as dinâmicas consideradas.

Chama atenção que as situações de maior dinâmica (barras azuis) apresentaram resultados muito inferiores aos alcançados para baixa dinâmica. Há que se considerar que em todos os experimentos metade das amostras disponíveis foram utilizadas para treinamento e metade para teste.

Assim, houve mais padrões de treinamento nos experimentos que simularam menor dinâmica (barras vermelhas), o que pode ter conduzido a um melhor treinamento e, por consequência, a um melhor desempenho nos experimentos de menor dinâmica. Reforça esta conclusão o fato que os classificadores monotemporais, nos quais a dinâmica das classes em pouco ou nada afetam a acurácia, apresentaram desempenho muito distintos. Não é possível, portanto, extrair dos resultados conclusões de validade geral quanto ao desempenho relativo dos classificadores analisados em função da dinâmica presente nas imagens.

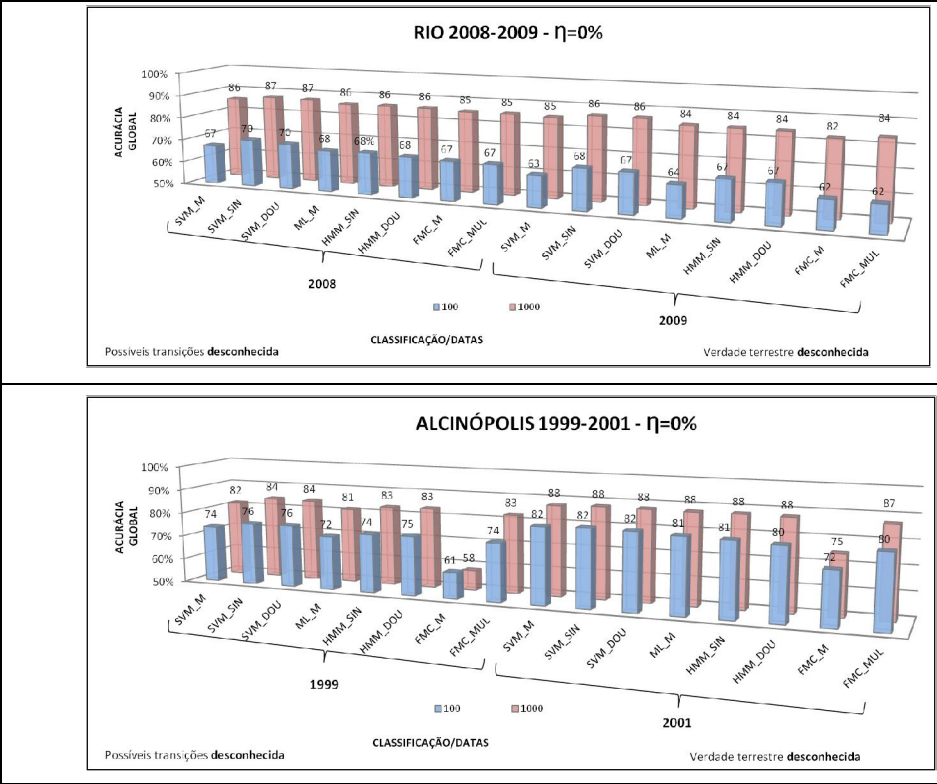


Figura 17: Acurácia Global (GAc) para experimento 1, imagens do Rio de Janeiro e Alcinópolis, ressaltando o desempenho geral dos classificadores dos esquemas propostos.

O resultado pode ser resumido na Tabela 5.8.

Tabela 8: Comparação de desempenho entre classificadores aplicados ao esquema monotemporal

Classificadores	Desempenho percentual das amostras de	
	Baixa Dinamica (100)	Alta Dinamica (1000)
SVM	86 %	67 %
HMM	86 %	68 %
FCM	85 %	67 %

A Figura 17 revela ainda que, de um modo geral, o desempenho dos esquemas multitemporais são superiores aos correspondentes monotemporais. A diferença é, contudo, pequena em praticamente todos os casos. Uma exceção se observa em relação aos classificadores nebulosos para Alcinópolis. Nota-se um ganho substancial da versão multitemporal (FMC_MUL) sobre a monotemporal (FMC_M) para Alcinópolis.

Chama atenção o baixo desempenho da versão monotemporal *fuzzy* quando comparado ao dos demais classificadores monotemporais. Ainda assim, a estratégia multitemporal *fuzzy* demonstrou-se capaz de alcançar acurácia similar à dos demais classificadores multitemporais.

5.3.2.

Experimento 2

Objetivo e procedimento

Neste experimento foi avaliado como a "qualidade" dos dados trazidos da outra data afeta o desempenho da classificação multitemporal da data alvo. Qualidade neste contexto refere-se à exatidão que se pode alcançar ao classificar os dados da "outra data".

O classificador ideal é o que tem 100% de acerto, ou que conhece a verdade de campo da imagem da "outra data" tA . Para simular esta situação, cada padrão a ser classificado na "outra data" foi expresso por um vetor, cuja dimensão é igual ao número de classes do problema. Tal vetor possui o valor "1" no elemento correspondente à classe verdadeira e o valor "0" nas demais posições.

Resultados e Discussões

A Figura 18 mostram os resultados.

Comparando-os com os resultados da Figura 17, por exemplo, fica claro que a qualidade da imagem trazida da outra data pode ter impacto substancial sobre o desempenho dos classificadores multitemporais. Tal observação confirma conclusão já publicada, por exemplo, em (Feitosa, et al., 2011). Baseada em estudos anteriores pode inclusive ocorrer que a abordagem multitemporal seja deletéria, no caso da qualidade da "outra imagem" ser inferior à da imagem que está sendo classificada.

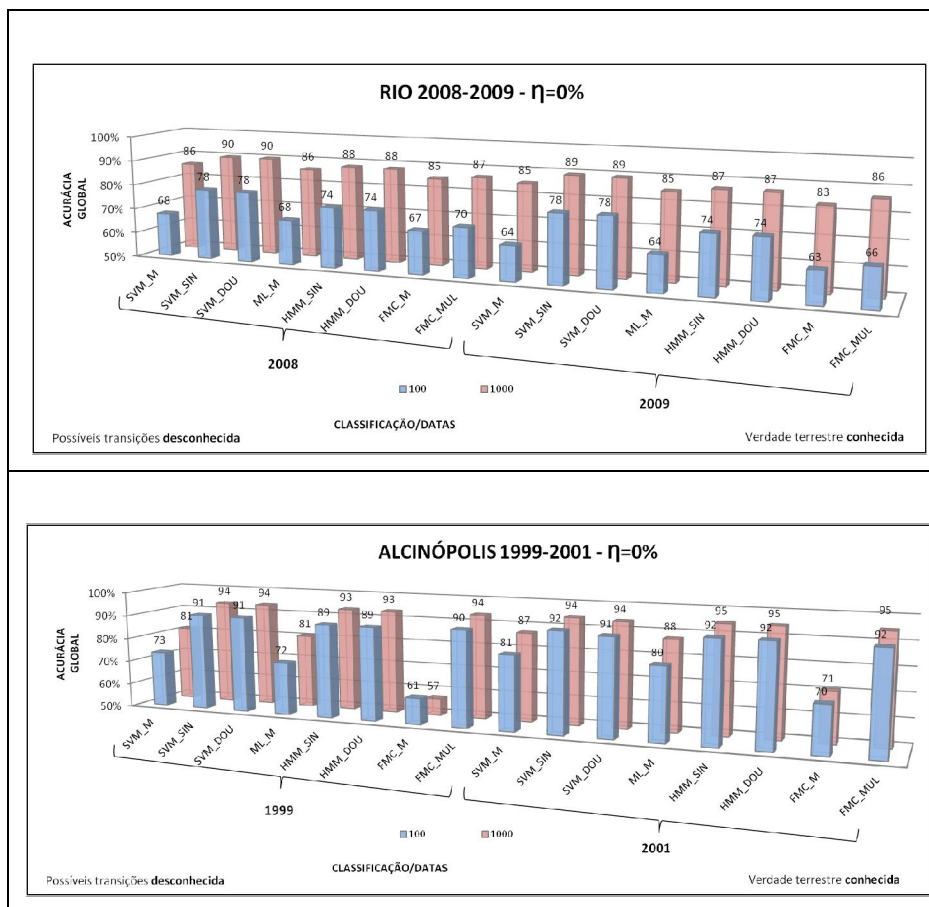


Figura 18: Acurácia Global GAc para experimento 2, imagens do Rio de Janeiro e Alcinópolis, ressaltando o desempenho geral dos classificadores dos esquemas propostos quando a verdade terrestre é conhecida.

Novamente neste experimento nenhuma das abordagens multitemporais se sobressaiu substancialmente em relação às demais.

5.3.3.

Experimento 3

Objetivo e procedimento

Um dos aspectos mais importantes na classificação de imagens de SR é a capacidade dos esquemas de classificação se acomodar à presença de *outliers*.

Por este motivo neste experimento foi avaliada a sensibilidade de cada método de classificação quanto a erros que podem ocorrer quando, em situações

práticas, o analista comete erros ao atribuir rótulos de classes aos padrões de treinamento.

A presença de *outliers* foi simulada neste experimento da seguinte forma: $\eta\%$ dos rótulos das amostras que representam cada transição presente na base de dados foi alterada aleatoriamente, onde η assumiu os valores 20% e 50%.

Nos experimentos 1 e 2 o valor de η foi zero.

Resultados e Discussões

As Figuras 19 e 20 mostram claramente o impacto negativo de *outliers* em todos os casos considerados, muito especialmente no caso de Alcinópolis.

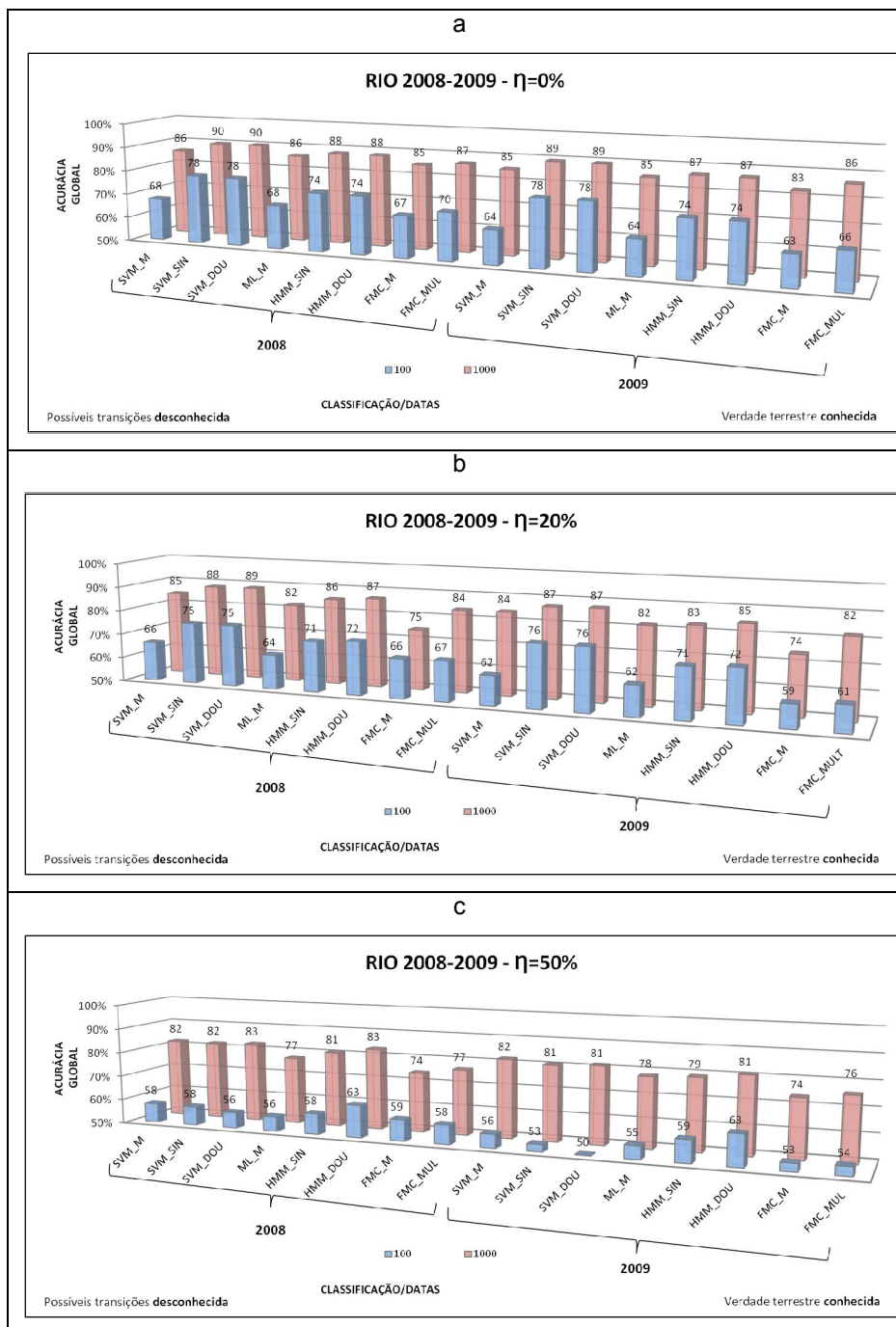


Figura 19: Acurácia Global (GAc) para experimento 3, ressaltando o desempenho dos classificadores dos multitemporais sob a presença variada de *outliers*.

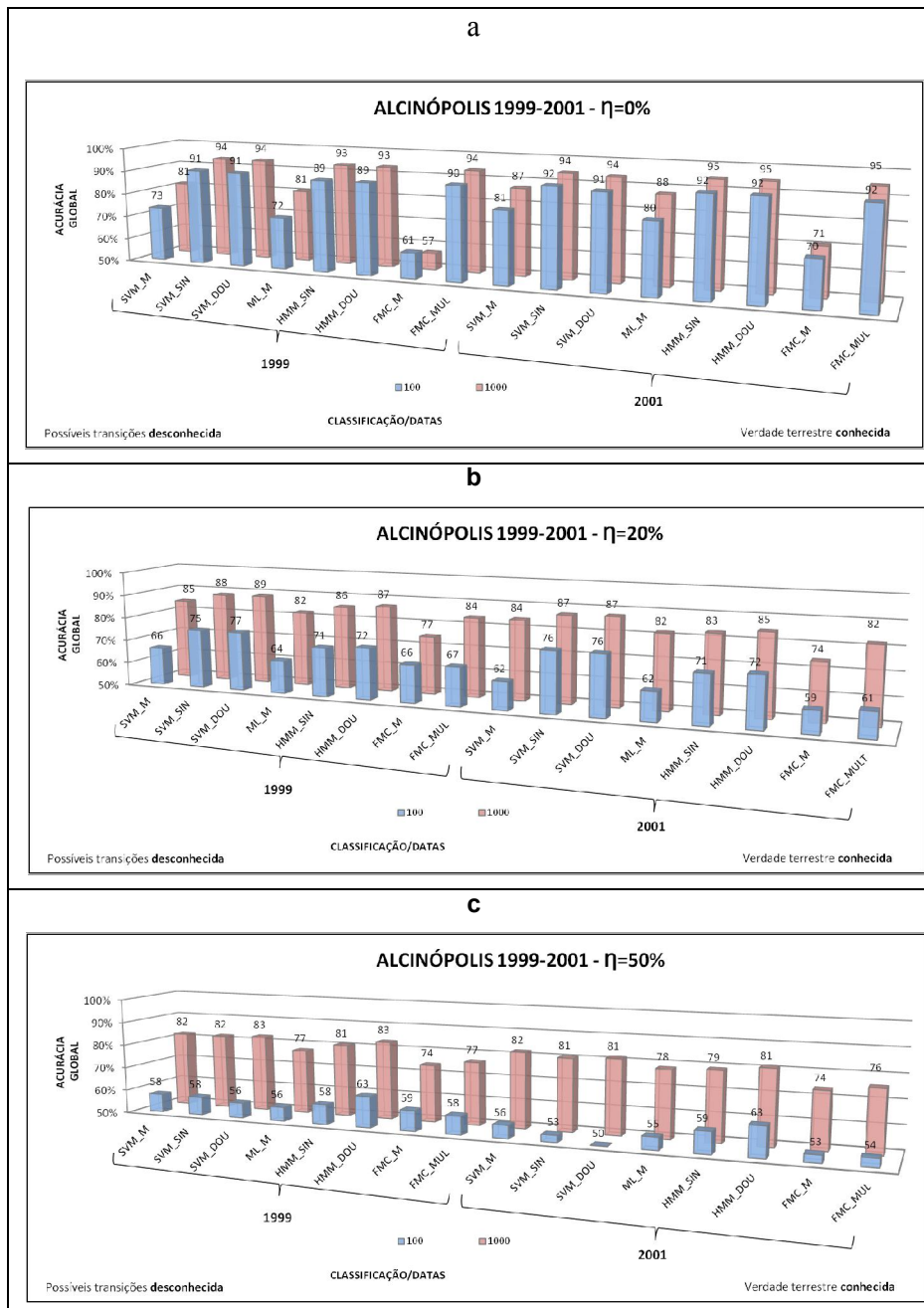


Figura 20: Acurácia Global (GAc) para experimento 3, ressaltando o desempenho geral dos classificadores dos multitemporais sob a presença variada de *outliers*.

Os resultados sugerem que a degradação de desempenho decorrente é mais expressiva nas variantes multitemporais do que nas monotemporais.

De fato, *outliers* produzem dois impactos danosos para a classificação.

Em primeiro lugar, falseiam a relação entre a classe e os atributos dos segmentos usados para treinamento. Em segundo lugar, alteram a frequência com

que ocorrem cada tipo de transição e, por último, podem criar no conjunto de treinamento transições que não ocorrem na área teste.

Os dois últimos impactos afetam apenas as abordagens multitemporais que de uma forma ou de outra procuram modelar a partir dos dados de treinamento a dinâmica de transição de classes. Os métodos monotemporais que ignoram a dinâmica temporais não sofrem com estes impactos. Isso explica a mais acentuada sensibilidade dos métodos multitemporais à presença de *outliers*.

Adicionalmente, pode-se observar nas Figuras 19c e 20c que o método multitemporal baseado em FMC é significativamente mais sensível à presença de *outliers* do que os métodos HMM e SVM.

5.3.4.

Experimento 4

Objetivo e procedimento

Uma das causas apontadas para o mau desempenho dos classificadores multitemporais diante de *outliers* é resultante, por exemplo, das transições no treinamento que de fato não ocorrem na área alvo.

Este problema pode ser evitado explorando-se conhecimento prévio sobre as transições de classes admissíveis.

Em termos práticos, o que se fez neste experimento foi submeter o conjunto de treinamento a um teste prévio que expurga qualquer amostra que represente uma transição de classe inadmissível na área alvo.

Esta medida impacta o processo de classificação de modo diferente em cada um dos métodos multitemporais analisados, como se explica a seguir.

Nos métodos multitemporais baseados em HMM e FMC, a filtragem prévia do conjunto de treinamento quanto às transições possíveis, leva a que a probabilidade/possibilidade associada às transições inadmissíveis sejam estimadas no treinamento como "zero".

Esta condição é suficiente para garantir que a classificação produzida por estas abordagens seja consistente com o conhecimento prévio a respeito das

transições de classe passíveis de ocorrer na área alvo no intervalo de tempo analisado.

Os rótulos considerados pelo classificador SVM para duas datas SVM_DOU são, na verdade, formados pelos pares de classes que ocorrem no conjunto de treinamento associados a um mesmo segmento.

Resta, portanto, evidente que o teste prévio do conjunto de treinamento inviabiliza uma resposta inconsistente com a dinâmica admissível na área alvo.

O efeito do teste prévio do conjunto de treinamento sobre a operação dos classificadores SVM voltados à data alvo apenas SVM_SIN não é facilmente explicável teoricamente.

Espera-se, contudo, que esta medida seja benéfica ao seu desempenho como se discute mais adiante.

Resultados e Discussões

As Figuras 21 e 22 correspondem à situação em que não há *outliers* respectivamente para as áreas do Rio de Janeiro e de Alcinópolis.

Nota-se neste caso os resultados com ou sem exploração de conhecimento prévio sejam praticamente os mesmos.

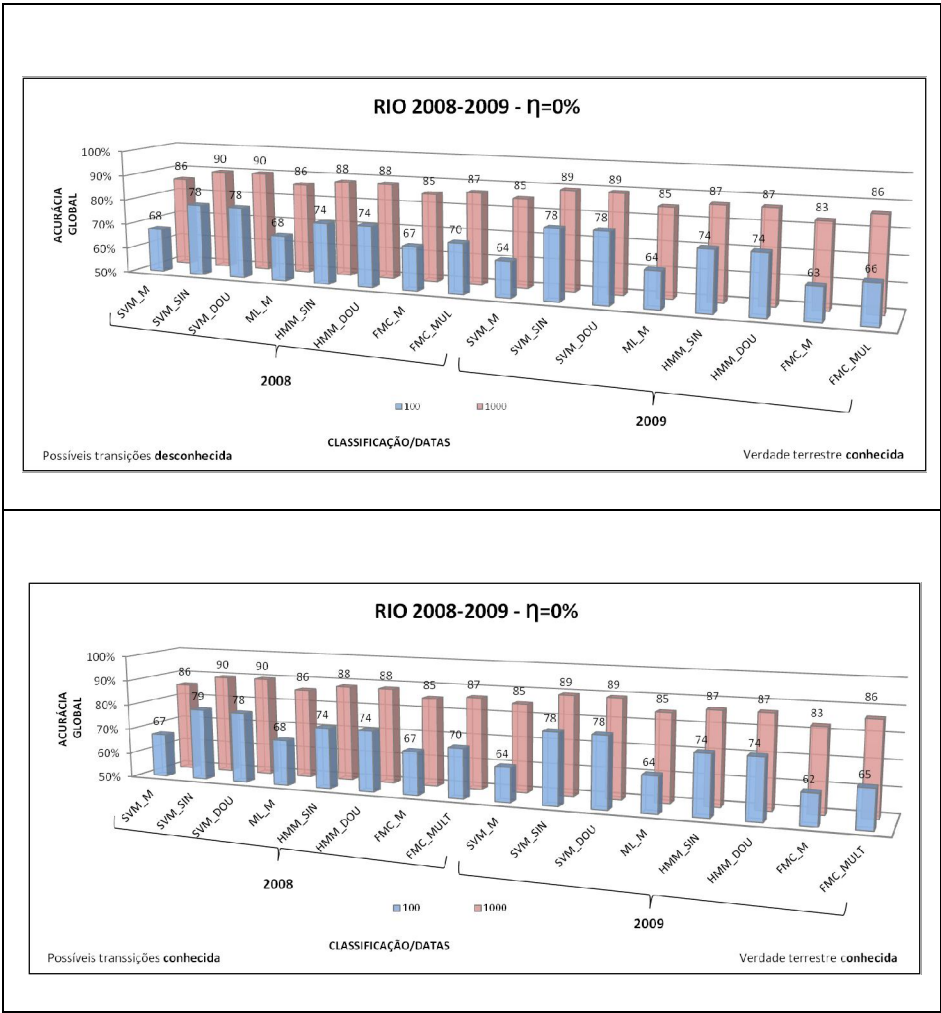


Figura 21: Acurácia Global (GAc) para experimento 4, Imagem do Rio de Janeiro ressaltando o desempenho dos classificadores dos multitemporais com conhecimento prévio sobre as transição de classe e na ausência de *outliers*.

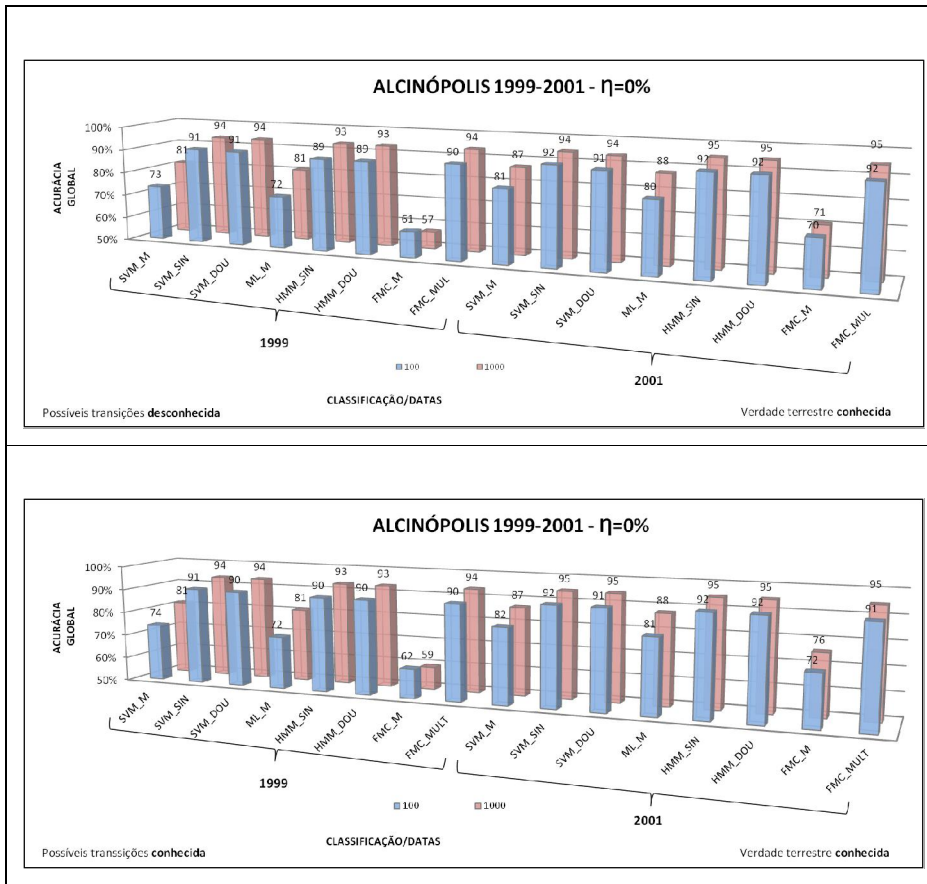


Figura 22: Acurácia Global (GAc) para experimento 4, Imagem Alcinópolis ressaltando o desempenho dos classificadores dos multitemporais com conhecimento prévio sobre as transições de classe e na ausência de *outliers*.

As Figuras 23 e 24 referem-se à situação em que 50% de *outliers* são inseridos.

Nota-se para ambos os lugares que a exploração do conhecimento prévio reduz substancialmente o prejuízo provocado pelos *outliers*.

Nota-se ainda que o impacto é maior sobre os classificadores SVM e FMC, o que sugere que estas abordagens multitemporais são mais sensíveis à presença de *outliers* no treinamento.

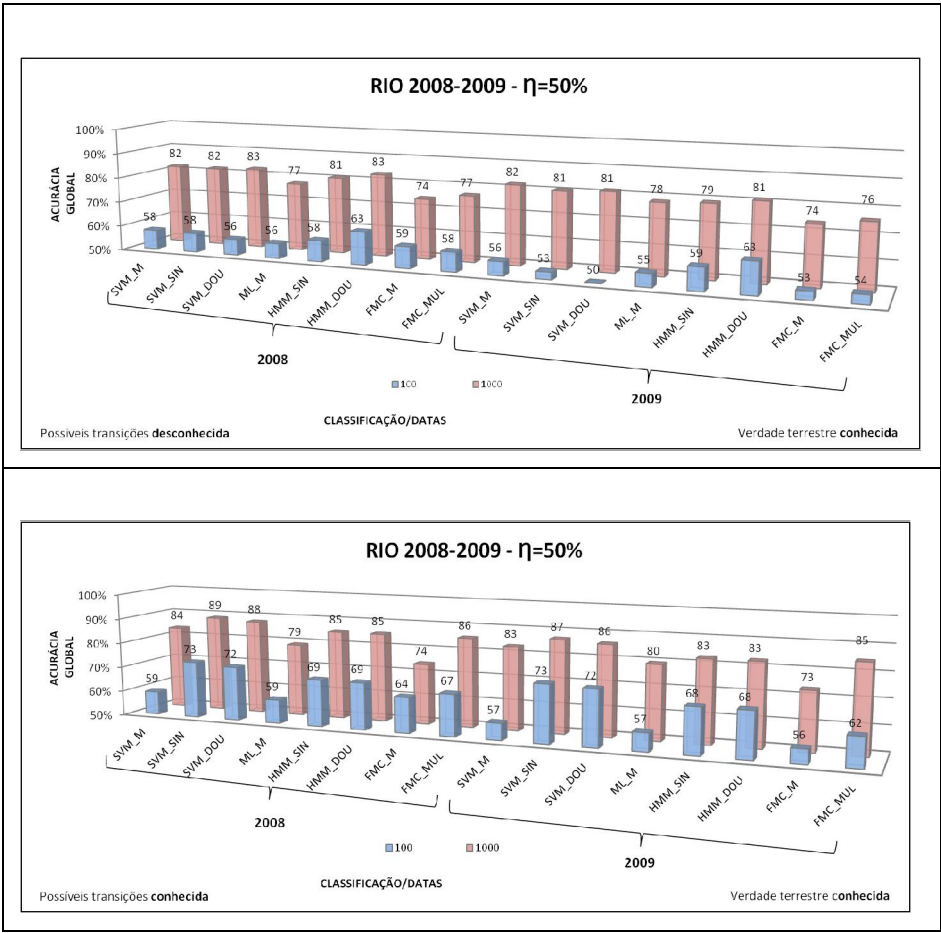


Figura 23: Acurácia Global (GAc) para experimento 4, Imagem do Rio de Janeiro ressaltando o desempenho dos classificadores dos multitemporais com conhecimento prévio sobre as transições de classe e na presença de *outliers*.

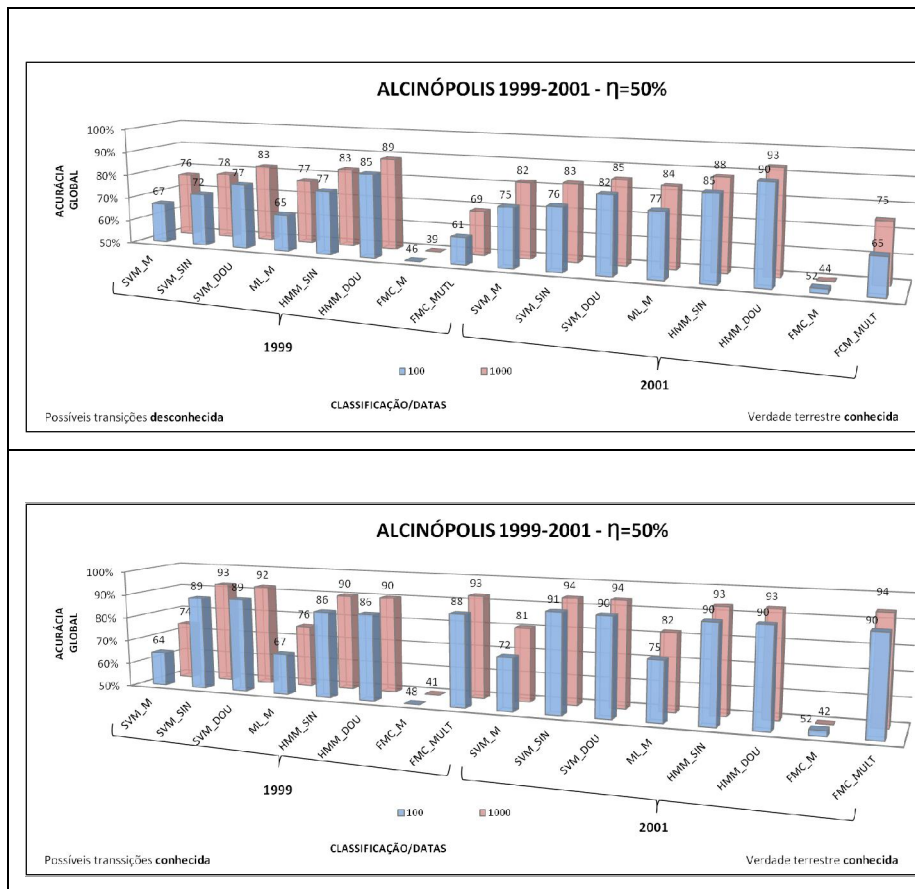


Figura 24: Acurácia Global (Gac) para experimento 4, Imagem Alcinópolis ressaltando o desempenho dos classificadores dos multitemporais com conhecimento prévio sobre as transições de classe e na presença de *outliers*.

5.3.5.

Experimento 5

Objetivo e procedimento

Finalmente, foi realizado um experimento para determinar a robustez de cada modelo quando o número de ocorrências de cada transição de classe no conjunto de treinamento difere do número das mesmas transições de classe no conjunto de teste.

Os dados foram distribuídos neste experimento de modo que todas as transições de classe ocorrem com a mesma frequência no conjunto de treinamento.

Note-se que, relativamente aos experimentos anteriores, este procedimento implica numa redução do número total de padrões de treinamento. Se NP_{min}

denota o número de amostras da transição com menos padrões de treinamento, escolheram-se de cada uma das demais transições aleatoriamente NP_{min} amostras.

Desta forma, a frequência relativa com que ocorrem transições entre pares de classe é diferente no conjunto de treinamento e no de teste.

Resultados e Discussão

As Figuras 25a e 25b ilustram para a área teste Rio de Janeiro respectivamente a situação em que não se tem e em que se tem o mesmo número de amostras de treinamento para cada transição.

De um modo geral, ao se impor que cada transição seja representada pelo mesmo número de mostras de treinamento reduz-se o número total de padrões de treinamento o que impacta negativamente sobre a acurácia em todos os casos considerados.

Destarte, não surpreende que os valores sejam inferiores. O que importa neste caso é analisar o quanto o desempenho de cada classificador multitemporal foi afetado.

A Figure 25 mostra que o desempenho dos classificadores nebulosos e especialmente dos baseados em SVM se degradaram mais do que os baseados em HMM para a situação de teste.

Os métodos baseados em HMM apresentaram-se pouco sensíveis à estimativa das probabilidades condicionais de uma classe numa data dada a classe na outra data - a probabilidade de transição.

Esta observação corrobora a conclusão levantada no experimento 4 de que os métodos baseados em HMM foram menos sensíveis à presença de *outliers* nos conjuntos de treinamento.

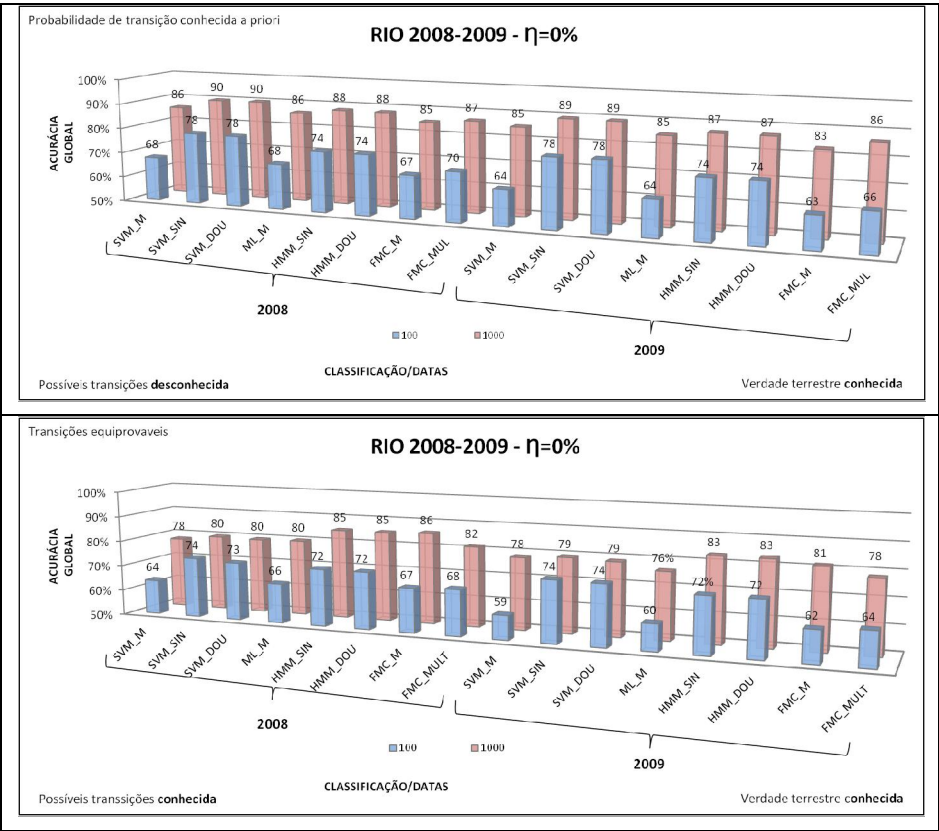


Figura 25: Acurácia Global (GAc) para experimento 5, imagem do Rio de Janeiro, ressaltando o desempenho geral dos classificadores quando as probabilidades de transição de classe estão mal estimadas e na ausência de *outliers*.

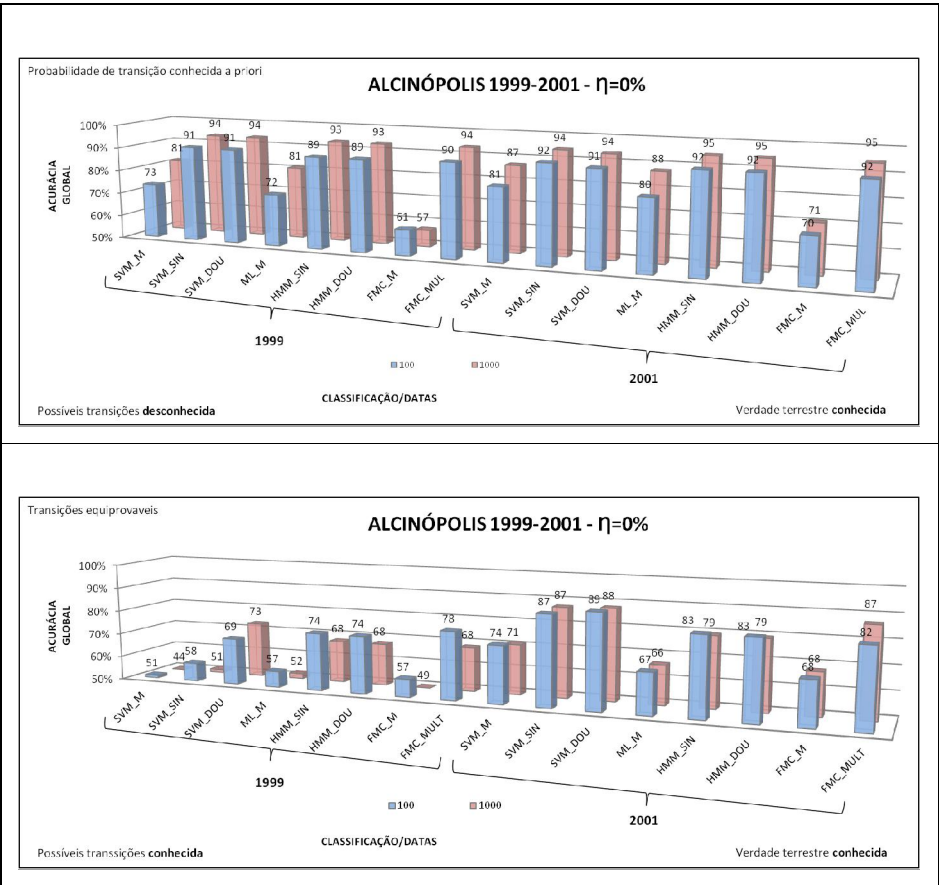


Figura 26: Acurácia Global (GAc) para experimento 5, imagem Alcinópolis, ressaltando o desempenho geral dos classificadores quando as probabilidades de transição de classe estão mal estimadas e na ausência de *outliers*.

6

Conclusão e trabalhos futuros

Neste trabalho foi comparado diferentes abordagens para classificação multitemporal em cascata, usando como base de dados um par de imagens IKONOS do Rio de Janeiro e em um par de imagens LANDSAT 7 de Alcinópolis (MS). Consideraram-se modelos baseados em SVM, FMC e baseados em HMM.

Simularam-se diversas condições de operação, envolvendo, diferentes dinâmicas das classes nas áreas de teste, a qualidade das imagens, e a presença de *outliers* nos conjuntos de treinamento.

Em condições ideais de operação, todos os métodos apresentaram desempenho similares.

Os resultados confirmaram conclusões de trabalhos anteriores de que a qualidade das imagens trazidas de outras datas tem impacto significativo sobre o desempenho dos métodos multitemporais.

Os métodos baseados em HMM apresentaram entre os métodos considerados maior robustez quanto à presença de *outliers* nos conjuntos de treinamento.

De outra forma, os métodos multitemporais nebulosos FMC e baseados em SVM são mais sensíveis a erros presentes no conjunto de treinamento, e requerem mais cuidado e maior experiência do analista humano ao reunir as amostras de treinamento.

Conclui-se ademais que, de uma maneira geral, o conhecimento a priori sobre as transições classes possíveis de ocorrer no sítio alvo pode tornar todos os métodos multitemporais estudados mais robustos contra a presença de *outliers*.

Como trabalhos futuros visualizam-se duas linhas de atuação.

A primeira envolve a incorporação de outras estratégias de classificação, como por exemplo, *Conditional Random Fields*. A segunda envolve testes com sequências temporais mais longas de datas.

Como comentário final, ressalte-se que métodos multitemporais envolvem um maior volume de dados a analisar com impactos sobre os tempos de

processamento. Caberia ainda uma comparação do ponto de vista dos tempos de processamento. Os experimentos aqui descritos foram implementados em MATLAB, que não se presta adequadamente a comparações deste tipo. Considerações sobre implementações paralelas destes métodos são temas de interesse para pesquisas futuras.

Referência Bibliográfica

Aach, T. & Kaup, A., 1995. Bayesian algorithms for adaptive change detection in image sequences using Markov random fields. *Signal Processing: Image Communication*, 7(2), pp. 147-160.

Aarnio, T., 1999. *Speech Recognition with Hidden Markov Models in Visual Communication*, Finlândia: s.n.

Andreola, R. & Haertel, V., 2009. *Support Vector Machines na Classificação de Imagens Hiperespectrais*. Natal, RN, Brasil, s.n., pp. 6757-6764.

Arbeláez, P., Maire, M., Fowlkes, C. C. & Malik, J., 2011. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(33), pp. 898-916.

Aurdal, L. et al., 2005. *Use of hidden Markov models and phenology for multitemporal satellite image classification: applications to mountain vegetation classification*. s.l., s.n., pp. 220-224.

Avrachenkov, K. & Sanchez, E., 2002. Fuzzy Markov chains and decision-making. *Fuzzy Optimization and Decision Making*, Volume v1, pp. 143-159.

Ayma, V. A., 2013. Avaliação de Métodos de Otimização Aplicados no Modelamento da Estrutura dos Vasos Sanguíneos. *Dissertação de Mestrado - Pontifícia Universidad Católica do Rio de Janeiro - PUC-Rio.*, p. 77.

Baatz, M. & Schäpe, A., 2000. Multiresolution Segmentation - an optimization approach for high quality multi-scale image segmentation. *Strobl/Blaschke/Griesebner*, pp. 12-23.

Baatz, M. & Schäpe, A., 2000. Multiresolution Segmentation -- An Optimization Approach for High Quality Multi-scale Image Segmentation.. Em: Strobl/Blaschke/Griesebner, ed. *Angewandte Geographische Informationsverarbeitung XII*. Heidelberg, Deutschland: Wichmann-Verlag, pp. 12-23.

Bandyopadhyay, S., 2005. Satellite image classification using genetically guided fuzzy clustering with spatial information. 26(3), pp. 579-593.

Baum, L. E. & Petrie, T., 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*, 37(6), pp. 1554-1563.

Blaschke, T., 2010. Object based image analysis for remote sensing. 65(1), pp. 2-16.

Blaschke, T. & Strobl, J., 2001. What is wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS Zeitschrift für Geoinformationssysteme*, Volume 14(6), pp. 12-17.

Bontemps, S. et al., 2008. An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution. *Remote Sensing of Environment*, 112(6), pp. 3181-3191.

Boser, B. E., Guyon, I. M. & Vapnik, V. N., 1992. *A Training Algorithm for Optimal Margin Classifiers*. s.l., ACM Press, pp. 144-152.

Boys, R. J., Henderson, D. A. & Wilkinson, D. J., 2000. Detecting homogeneous segments in DNA sequences by using hidden Markov models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(2), pp. 269-285.

Bradski, G. R. & Kaehler, A., 2008. *Learning OpenCV - Computer Vision with the OpenCV Library: software that sees*. s.l.:O'Reilly.

Bruzzone, L., Chi, M. & Marconcini, M., 2006. A Novel Transductive SVM for Semisupervised Classification of Remote-Sensing Images. *Geoscience and Remote Sensing, IEEE Transactions on*, Nov, 44(11), pp. 3363-3373.

Bruzzone, L. & Cossu, R., 2002. A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps. *Geoscience and Remote Sensing, IEEE Transactions on*, Sep, 40(9), pp. 1984-1996.

Bruzzone, L., Cossu, R. & Vernazza, G., 2004. Detection of land-cover transitions by combining multiclass classifiers. *Pattern Recognition Letters*, 25(13), pp. 1491-1500.

Bruzzone, L., Prieto, D. & Serpico, S., 1999. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, May, 37(3), pp. 1350-1359.

Bunkle, H. & Caelli, T., 2001. Hidden Markov Models -- Applications in Computer Vision. *World Scientific*.

Burges, C. J. C., 1998. A tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, Volume 2, pp. 121-167.

Büscher, O. & Buck, O., 2007. *DeCover - Geoinformation Services to Update and Supplement Land Cover Data for German Decision Makers*. s.l., s.n.

Carleer, A. P. & Debeir, O., 2005. Assessment of very high spatial resolution satellite image segmentations. 71(11), pp. 1285-1294.

Castillo, C., Chollett, I. & Klein, E., 2008. Enhanced duckweed detection using bootstrapped SVM classification on medium resolution RGB MODIS imagery. *International Journal of Remote Sensing*, 29(19), pp. 5595-5604.

Chang, C.-C. & Lin, C.-J., 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, Volume 2, pp. 27:1--27:27.

Cheng, H. D., Jiang, X. H., Sun, Y. & Wang, J. L., 2001. Color image segmentation: Advances & prospects. 34(12), pp. 2259-2281.

Comaniciu, D. & Meer, P., 2002. Mean Shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 24, pp. 603-619.

Conroy, J. et al., 2001. *Hidden Markov models for chromosome identification*. s.l., s.n., pp. 473-477.

Coppin, P. R. & Bauer, M. E., 1996. Digital change detection in forest ecosystems with remote sensing imagery. *Remote Sensing Reviews*, 13(3-4), pp. 207-234.

Cormen, T., Leiserson, C. & Rivest, R., 1990. *Introduction to Algorithms*. s.l.:The MIT Press.

Cortes, C. & Vapnik, V., 1995. Support-Vector Networks. *Machine Learning*, 20(3), pp. 273-297.

Costa, G. A. d. et al., 2009. *Um Método de Classificação Multitemporal em Cascata de Imagens de Sensoriamento Remoto*. Natal, RN, Brasil, s.n., pp. 1291-1298.

Costa, G. A. O. P., 2009. *A Knowledge-Based Approach for Automatic Interpretation of Multidate Remote Sensing Data*, Rio de Janeiro, RJ, Brasil: s.n.

Costa, G. A. O. P. & Feitosa, R. Q., 2014. A generalized fuzzy Markov chain-based model for classification of remote-sensing multitemporal images. *International Journal of Remote Sensing*, 35(1), pp. 341-364.

Csurka, G. & Perronnin, F., 2010. An Efficient Approach to Semantic Segmentation. *Int. Journal on Computer Vision*, April, 95(2), pp. 198-212.

de Almeida Crusco, N., 2006. *Sensoriamento Remoto Para Análise Multitemporal da Dinâmica de Áreas Agrícolas*, São José Dos Campos, SP, Brasil: s.n.

Dey, V., Zhang, Y. & Zhong, M., 2010. A Review on Image Segmentation techniques with Remote Sensing perspective. 5-7 July, XXXVIII(Part 7A).

Domke, J., 2013. Learning Graphical Model Parameters with Approximate Marginal Inference. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(10), pp. 2454-2467.

Estrada, F. J. & Jepson, A. D., 2005. Quantitative evaluation of a novel image segmentation algorithm. Volume 2, pp. 1132-1139.

Everingham, M., Muller, H. & Thomas, B., 2002. Evaluating image segmentation algorithms using monotonic hulls in fitness/cost space. pp. 363-372.

Feitosa, R., Alvarado, L. & Costa, G. A., 2013. *A comparison of SVM-based cascade multitemporal classifiers*. s.l., s.n., pp. 3455-3458.

Feitosa, R. Q. et al., 2009. Cascade multitemporal classification based on fuzzy Markov chains. *Journal of Photogrammetry and Remote Sensing (ISPRS)*, 64(2), pp. 159-170.

Feitosa, R. Q., Costa, G. A. O. P., Cazes, T. B. & Feijo, B., 2006. A genetic approach for the automatic adaptation of segmentation parameters. *1st International Conference on Object Based Image Analysis*, May.

Feitosa, R. Q., da Costa, G. A. O. P., Mota, G. L. A. & ao, B. F., 2011. Modeling alternatives for fuzzy Markov chain-based classification of multitemporal remote sensing data. *Pattern Recognition Letters*, 32(7), pp. 927-940.

Felzenszwalb, P. F. & Huttenlocher, D. P., 2004. Efficient graph-based image segmentation. 59(2), pp. 167-181.

Foody, G. M. & Mathur, A., 2004. Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93(1--2), pp. 107-117.

Fu, K. S. & Mui, J. K., 1981. A survey on image segmentation. January, 13(1), pp. 3-16.

Fuller, R., Smith, G. & Devereux, B., 2003. The characterisation and measurement of land cover change through remote sensing: problems in operational applications?. *International Journal of Applied Earth Observation and Geoinformation*, 4(3), pp. 243-253.

Fu, T.-F. & Lin, C.-J., 2004. Probability Estimates for Multi-class Classification by Pairwise Coupling. *Journal of Machine Learning Research*, Volume 5, pp. 976-1005.

Ghoggali, N., Melgani, F. & Bazi, Y., 2009. A Multiobjective Genetic {SVM} Approach for Classification Problems With Limited Training Samples. *Geoscience and Remote Sensing, IEEE Transactions on*, June, 47(6), pp. 1707-1718.

Gladstone, C. S., Gardner, A. & Holland, D., 2012. A SEMI-AUTOMATIC METHOD FOR DETECTING CHANGES TO ORDNANCE SURVEY TOPOGRAPHIC DATA IN RURAL ENVIRONMENTS. *Proceedings of the 4th GEOBIA*, pp. 396-401.

Gomez-Chova, L., Camps-Valls, G., Bruzzone, L. & Calpe-Maravilla, J., 2010. Mean Map Kernel Methods for Semisupervised Cloud Classification. *Geoscience and Remote Sensing, IEEE Transactions on*, Jan, 48(1), pp. 207-220.

Gondra, I. & Xu, T., 2010. A multiple instance learning based framework for semantic image segmentation. 48(2), pp. 339-365.

Gong, P., 1993. Change Detection using principal component analysis and fuzzy theory. *Canadian Journal of Remote Sensing*, Volume 19, pp. 22-29.

Gonzales, R. & Woods, R. E., 2008. *Digital Image Processing*. 3rd ed. s.l.:Prentice Hall.

Gualtieri, J. A. & Crompton, R. F., 1998. *Support Vector Machines for Hyperspectral Remote Sensing Classification*. s.l.:s.n.

Gupta, R. P. & Prakash, A., 1998. {Land-use mapping and change detection in a coal mining area - a case study in the Jharia coalfield, India}. *International Journal of Remote Sensing*, #feb#, Volume 19, pp. 391-410.

Hansen, F. R. & Elliott, H., 1982. Image Segmentation using simple Markov random field models. 20(2), pp. 101-132.

Happ, P., Feitosa, R., Bentes, C. & Farias, R., 2013. A Region-Growing Segmentation Algorithm for GPUs. *IEEE Geoscience and Remote Sensing Letters*, November, 10(6), pp. 1612-1616.

Hay, G. & Castilla, G., 2008. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. Em: T. Blaschke, S. Lang & G. Hay, eds. *Object-Based Image Analysis*. s.l.:Springer Berlin Heidelberg, pp. 75-89.

Hoover, A. et al., 1996. An Experimental Comparison of Range Image Segmentation Algorithms. 18(7), pp. 673-689.

Hsu, C.-W. & Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, Mar, 13(2), pp. 415-425.

Huang, S., Ahmadi, M. & Sid-Ahmed, M., 2008. A hidden Markov model-based character extraction method. *Pattern Recognition*, 41(9), pp. 2890-2900.

Jeon, B. & Landgrebe, D., 1992. Classification with spatio-temporal interpixel class dependency contexts. *IEEE Trans. Geosci. Remote Sens*, Volume 30, pp. 663-672..

Jeon, B. & Landgrebe, D., 1999. Decision fusion approach for multitemporal classification. *Geoscience and Remote Sensing, IEEE Transactions on*, May, 37(3), pp. 1227-1233.

Jeon, B. & Landgrebe, D. A., 1992. Classification with spatio-temporal interpixel class dependency contexts. 30(4), pp. 663-672.

Jiang, X., Marti, C., Irniger, C. & Bunke, H., 2006. Distance Measures for Image Segmentation Evaluation. *EURASIP Journal on Applied Signal Processing*, Volume 1, pp. 1-10.

Juang, B. H. & Rabiner, L. R., 1991. Hidden {Markov Models} for Speech Recognition. *Technometrics*, 33(3), pp. 251-272.

Khazenie, N. & Crawford, M., 1990. Spatial-temporal Autocorrelated Model For Contextual Classification. *Geoscience and Remote Sensing, IEEE Transactions on*, Jul, 28(4), pp. 529-539.

Knerr, S., Personnaz, L. & Dreyfus, G., 1990. Single-layer learning revisited: a stepwise procedure for building and training a neural network. Em: F. F. Soulie & J. Héroult, eds. *Neurocomputing*. s.l.:Springer Berlin Heidelberg, pp. 41-50.

Kohli, P., Ladický, L. & Torr, P. H. S., 2008. Robust higher order potentials for enforcing label consistency. 23-28 June, pp. 1-8.

Kolmogorov, V., 2006. Convergent tree-reweighted message passing for energy minimization. 28(10), pp. 1568-1583.

Kuma, S. & Hebert, M., 2003. Discriminative random fields: A discriminative framework for contextual interaction in classification. Oct., Volume 2, pp. 1150-1157.

Leite, P. B. C., 2008. *Identificação de Tipos de Culturas Agrícolas a partir de Sequências de Imagens Multitemporais Utilizando Modelos de Markov Ocultos*, s.l.: s.n.

Leite, P. B. C. et al., 2011. Hidden {Markov} Models for crop recognition in remote sensing image sequences. *Pattern Recognition Letters*, Jan, 32(1), pp. 19-26.

Liedtke, C.-E., Buckner, J., Pahl, M. & Stahlhut, O., 2001. *Knowledge Based System for the Interpretation of Complex Scenes*. s.l.:s.n.

Li, F., Peng, J. & Zheng, X., 2004. Object-based and Semantic Image Segmentation using MRF. Volume 3, pp. 833-840.

Lucchese, L. & Mitra, S. K., 2001. Color image segmentation: a state-of-the-art survey. March, 67(2), pp. 207-221.

Lucieer, A., 2004. Uncertainties in Segmentation and their Visualisation. *PhD Thesis Utrecht University*, p. 174.

Lu, D., Mausel, P., Brondizio, E. & Moran, E., 2004. Change detection techniques. *International Journal of Remote Sensing*, 25(12), pp. 2365-2401.

Mantero, P., Moser, G. & Serpico, S., 2005. Partially Supervised classification of remote sensing images through SVM-based probability density estimation. *Geoscience and Remote Sensing, IEEE Transactions on*, March, 43(3), pp. 559-570.

Marpu, P. R., Neubert, M., Herold, H. & Niemeyer, I., 2010. Enhanced evaluation of image segmentation results.

Martin, D., 2003. An Empirical Approach to Grouping and Segmentation. *PhD Thesis, EECS Department, University of California, Berkeley*, August.

Mathews, J. H. & Fink, K. D., 2004. *Numerical Methods Using Matlab*. 4th ed. United States: Prentice Hall.

Maxwell, T. & Zhang, Y., 2005. A fuzzy logic approach to optimization of segmentation of object-oriented classification. Volume 5909, pp. 1-11.

Melgani, F., Serpico, S. B. & Vernazza, G., 2003. Fusion of Multitemporal Contextual Information by Neural Networks for Multisensor Remote Sensing Image Classification. *Integr. Comput.-Aided Eng.*, #jan#, 10(1), pp. 81-90.

Melgani, F., Serpico, S. & Vernazza, G., 2001. *Fusion of multitemporal contextual information by neural networks for multisensor image classification*. s.l., s.n., pp. 2952-2954 vol.7.

Meloni, R. B. d. S., 2009. *Classificação de Imagens de Sensoriamento Remoto usando SVM*, Rio de Janeiro. RJ: s.n.

Mota, G. L. A., 2004. *Interpretação baseada em conhecimento aplicada a imagens multitemporais de satélite de baixa resolução*, s.l.: s.n.

Mota, G. L. et al., 2007. Multitemporal fuzzy classification model based on class transition possibilities. *Journal of Photogrammetry and Remote Sensing (ISPRS)*, 62(3), pp. 186-200.

Mountrakis, G., Im, J. & Ogole, C., 2011. Support vector machines in remote sensing: A review. *Journal of Photogrammetry and Remote Sensing (ISPRS)*, 66(3), pp. 247-259.

Musci, M., 2013. Uma Metodologia para Análise de Imagens usando Segmentações Específicas por Classe. *Tese de Doutorado - Pontifícia Universidad Católica do Rio de Janeiro - PUC-Rio.*, p. 169.

Negri, R. G., Dutra, L. V. & ao Siqueira Sant'Anna, S. J., 2011. *Desenvolvimento de função kernel para aplicação de SVM na classificação baseada em regiões*. s.l., s.n.

Nelder, J. A. & Mead, R., 1965. A Simplex Method for Function Minimization. 7(4), pp. 308-313.

Neubert, M., Herold, H. & Meinel, G., 2006. Evaluation of remote sensing image segmentation quality - further results and concepts. 4-5 July, Volume XXXVI-4/C42, p. 6.

Neubert, M., Herold, H. & Meinel, G., 2008. Assessment of Remote Sensing Image Segmentation Quality. 6-7 August, Volume XXXVIII-4/C1, p. 5.

Neubert, M. & Meinel, G., 2003. Evaluation of Segmentation programs for high resolution remote sensing applications. *Proc. Joint ISPRS/EARSel Workshop "High Resolution Mapping from Space 2003"*, p. 8.

Otsuka, T. & Ohya, J., 1997. *Recognizing multiple persons' facial expressions using HMM based on automatic extraction of significant frames from image sequences*. s.l., s.n., pp. 546-549 vol.2.

Pal, N. R. & Pal, S. K., 1993. A review on image segmentation techniques. 26(9), pp. 1277-1294.

Palus, H. & Kotyczka, T., 2001. Evaluation of Colour Image Segmentation Results. Volume 7, p. 4.

Petersen, M. E., De Ridder, D. & Handels, H., 2002. Image processing with neural networks - A review. 35(10), pp. 2279-2301.

PIMAR, P., 2010. *Programa Integrado de Monitoria Remota de Fragmentos Florestais e de Crescimento Urbano no Rio de Janeiro*. s.l.:s.n.

Pont-Tuset, J. & Marques, F., 2013. Measures and Meta-Measures for the Supervised Evaluation of Image Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition- CVPR*, 23-28 June, pp. 2131-2138.

Prasad, S. V. S., Savitri, D. T. S. & Krishna, D. I. V. M., 2011. Classification of Multispectral Satellite Images Using Clustering With SVM Classifier. *International Journal of Computer Applications*, December, 35(5), pp. 32-44.

Rabiner, L. R., 1989. *A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*. s.l., s.n., pp. 257-286.

Rand, W., 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, Volume 336, pp. 846-850.

Ridd, M. K. & Liu, J., 1998. A Comparison of Four Algorithms for Change Detection in an Urban Environment. *Remote Sensing of Environment*, 63(2), pp. 95-100.

Rother, C., Kolmogorov, V. & Blake, A., 2004. GrabCut: interactive foreground extraction using iterated graph cuts. 23(3), pp. 309-314.

Russ, J. C., 1998. *The Image Processing Handbook*. 3rd ed. s.l.:Raleigh, NC, USA: Materials Science and Engineering Department North Carolina State University.

S.M. Tavakkoli Sabour, P. L. U. S., 2008. MONITORING AGRICULTURAL ACTIVITIES USING MULTI-TEMPORAL ASAR ENVISAT DATA. *IntArchPhRS*, Volume XXXVII B7-2, pp. 735-742.

Scholkopf, B. & Smola, A. J., 2001. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press.

Serpico, S. & Melgani, F., 2000. *A fuzzy spatio-temporal contextual classifier for remote sensing images*. s.l., s.n., pp. 2438-2440 vol.6.

Shi, J. & Malik, J., 2000. Normalized Cuts and Image Segmentation. 22(8), pp. 888-905.

Solberg, A. H. S., Taxt, T. & Jain, A., 1996. A Markov random field model for classification of multisource satellite imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, Jan, 34(1), pp. 100-113.

Strasters, K. & Gerbrands, J., 1991. Three-dimensional segmentation using a split, merge and group approach. Volume 12, pp. 307-325.

Swain, P., 1978. Bayesian Classification in a Time-Varying Environment. *Systems, Man and Cybernetics, IEEE Transactions on*, 8(12), pp. 879-883.

Van Rijsbergen, C., 1979. *Information Retrieval*. 2nd ed. s.l.:Dept. of Computer Science, Univ. of Glasgow.

Vantaram, S. R. & Saber, E., 2012. Survey of contemporary trends in color image segmentation. *Journal of Electronic Imaging*, 21(4), pp. 040901-1-040901-28.

Vapnik, V., 1979. *Estimation of Dependences Based on Empirical Data*. New York: Springer Verlag.

Viovy, N. & Saint, G., 1994. Hidden Markov models applied to vegetation dynamics analysis using satellite remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on*, Jul, 32(4), pp. 906-917.

Visa, A., Valkealahti, K. & Simula, O., 1991. Cloud detection based on texture segmentation by neural network methods. Volume 2, pp. 1001-1006.

Wang, J. et al., 2005. Interactive video cutout. 24(3), pp. 585-594.

Weismiller, R. A. et al., 1977. Evaluation of Change Detection Techniques for Monitoring Coastal Zone Environments. *American Society for Photogrammetry and Remote Sensing*, 12, Volume 43, pp. 1533-1539.

Wu, Z. & Leahy, R., 1993. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. 15(11), pp. 1101-1113.

Yang, L., Albregtsen, F., Lonnestad, T. & Grottum, P., 1995. A Supervised Approach to the evaluation of image segmentation methods. Volume 970, pp. 759-765.

Yang, W., Zhang, X., Chen, L. & Sun, H., 2010. Semantic Segmentation of Polarimetric SAR Imagery using Conditional Random Fields. pp. 1593-1596.

Zhang, H., Fritts, J. A. & Goldman, S. A., 2004. An entropy-based objective segmentation evaluation method for image segmentation. pp. 38-49.

Zhang, H., Fritts, J. A. & Goldman, S. A., 2005. A Co-Evaluation Framework for Improving Segmentation Evaluation. pp. 420-430.

Zhang, R., Ma, J., Chen, X. & Tong, Q., 2009. *Feature selection for hyperspectral data based on modified recursive support vector machines*. s.l., s.n., pp. II-847-II-850.

Zhang, Y., 1996. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8), pp. 1335-1346.

Zhang, Y., 2001. A review of recent evaluation methods for image segmentation. *Int. Symp. on Signal Processing and its Applications (ISSPA)*, pp. 148-151.

Zhang, Y. & Gerbrands, J., 1994. Objective and quantitative segmentation evaluation and comparison. *Signal Processing*, 39(1-2), pp. 43-54.

Zhong, C., Zhongmin, Z., DongMei, Y. & Renxi, C., 2005. Multi-scale segmentation of the high resolution remote sensing image. Volume 5, pp. 3682-3684.

Zhu, G. & Blumberg, D. G., 2002. Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel. *Remote Sensing of Environment*, 80(2), pp. 233-240.