

Álgebra de Mapas: Agregações e Vizinhanças

João Pedro Cerveira Cordeiro^{1,2}
Guaraci José Erthal²

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{jpedro, gaia}@dpi.inpe.br

Abstract. This paper illustrates the use of new functionalities available in LEGAL, the map algebra language module of the GIS Spring, resulting from a conceptual revision in which concepts such as local, zone and neighborhood were unified into the region concept, with implications in all classes of operations. Besides incorporating neighborhood operations functionalities, the language can now explore differences in resolution among data in more explicit and formal ways, thus allowing operations such as generalization and aggregation, typical in multi-scale and multi-resolution data analysis, to be naturally incorporated to map algebra jargon. We also illustrate the discussion with some applications aimed to noise removal from classified image and target detection taking deforestation patterns as target from images.

Palavras-chave: map algebra, geographic information systems, image processing, spatial analysis, environment modeling

1. Introdução

Uma álgebra essencialmente define uma estruturação sobre tipos de dados, que podem ser desde números inteiros, reais ou valores lógicos como “TRUE” e “FALSE”, até grandezas das mais variadas naturezas e dimensionalidades, como os mapas. Para isso deve ser definida claramente, uma maneira de representar objetos e operações sob a forma de expressões de uma linguagem que reflita claramente as propriedades dessa álgebra.

A Linguagem LEGAL (Cordeiro et al., 1996), disponível como um módulo do GIS Spring (Camara et al., 1994; 1996), resulta da implementação de uma estrutura algébrica sobre tipos de dados categorizados como imagens, grades numéricas regulares e mapas em geral, tipicamente representados sob a forma de matrizes. Outros objetos geográficos, representados em tabelas de bancos de dados e associados a representações vetoriais de polígonos, linhas e pontos, também podem explorar alguns conceitos da álgebra de mapas. Sobre esse modelo de dados, foram definidas operações algébricas com base não apenas na clássica álgebra de mapas (Tomlin, 1990), hoje disponível na maioria dos softwares GIS, mas também em outros formalismos como a álgebra de imagens (Ritter, 1990) e a geoálgebra (Takeyama, 1997) que vêm a contribuir principalmente no contexto em que se busca estender sua aplicabilidade além dos aspectos estáticos da modelagem cartográfica, para os aspectos dinâmicos cada vez mais exigidos em modelagem ambiental.

Operações em álgebra de mapas são classificadas em “locais”, “zonais” e de “vizinhanças”, segundo o modo como são caracterizados cada local de uma área de trabalho, ora em função de valores associados ao mesmo local através de mapas, ora por uma síntese de valores associados a conjuntos específicos de locais. Em suas primeiras versões, a linguagem LEGAL contempla apenas as classes de operações locais e zonais da taxonomia introduzida por Tomlin, embora seus recursos sintáticos, além de permitirem a descrição de várias operações de vizinhança como se fossem locais, também prevêm alguma compatibilidade entre dados de resoluções diferentes, a partir do uso de técnicas de interpolação.

A partir da versão 5.0 do Spring, a linguagem LEGAL incorpora uma revisão conceitual com implicações em todas as classes de operações, que parte da unificação dos conceitos de local, zona e vizinhança, em um único conceito – regiões –, adotado para se referir a quaisquer subconjuntos de locais de uma área de trabalho. Desse modo, além de incorporar novas funcionalidades da classe de operações de vizinhanças, a linguagem permite ainda

explorar as diferenças de resolução entre dados de uma maneira mais explícita e formal, permitindo que operações como “generalização” e “agregação”, muito usadas na integração de dados de sensores remotos e GIS sejam naturalmente incorporadas à álgebra de mapas

Cada operação em LEGAL essencialmente se divide em três fases: a primeira consiste de selecionar conjuntos de locais que exercem influencia sobre cada local específico de uma área de trabalho; na segunda, os valores associados a esses locais através de mapas, são recuperados e, eventualmente, processados mais ainda; finalmente um valor que sintetize adequadamente a influencia desses conjuntos de locais e valores, é obtido e usado para caracterizar a cada local específico da área de trabalho resultando em um novo mapa em uma base de dados. No caso de operações ditas “locais”, essas três etapas se colapsam em uma única, se não houver resoluções diferentes entre os dados envolvidos. A importância das operações locais vem ainda do fato de que são a básicas também no processo de seleção dos locais e valores de influencia, bem como em eventuais processamentos sobre eles.

Na próxima seção deste artigo se discute um pouco mais os conceitos de local e agregação de locais junto com uma aplicação simples voltada para a eliminação de ruídos em uma imagem classificada. Na seção 3 os conceitos de região e interação entre regiões e outros tipos de dados são discutidos; e como essa interação seguida de um processo de sumarização estatística, permite descrever indiferentemente operações tanto de vizinhança, quanto zonais. Como, afinal, a maior motivação nesse trabalho refere-se à introdução de operações ditas “de vizinhança” em LEGAL, um método para detecção de padrões de desmatamento florestal a partir de imagens de satélite é esboçado, afim de ilustrar alguns recursos sintáticos para descrição de regiões vizinhas. Na seção 6, alguns aperfeiçoamentos sobre os exemplos discutidos nas seções anteriores são incorporados a fim de ilustrar um pouco da integração entre as novas funcionalidades num contexto mais amplo da linguagem.

2. Operações locais e generalizações

O conceito de local, embora por um lado modele a noção de menor unidade representável de um mapa, por outro também pode ser visto como a agregação de locais representados a escalas cartográficas maiores, quando a resolução é mantida constante, ou a resoluções mais finas, quando a escala cartográfica é mantida constante. Cada local de uma área de trabalho, quando visto sob uma certa resolução espacial, pode ser caracterizado por um valor, que representa a síntese do conjunto de valores associados às celas (locais, pixels) que contém ou que interceptam seu interior a uma resolução mais fina. Valores selecionados dessa maneira são como amostras para estatísticas básicas tais como maioria, máximo, média e mediana. Expressar a compatibilidade entre dados a diferentes resoluções e escalas implica em discriminar, de modo explícito, o processo de sumarização que será utilizado em operações, permitindo assim avaliar a perda de informação resultante. A Figura 1 ilustra o uso de um critério baseado na maioria (ou moda) para inferir a classe predominante em celas a resoluções variadas.

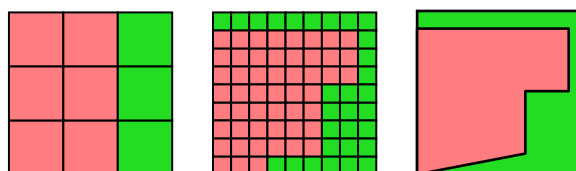


Figura 1. Dados de diferentes resoluções, na mesma escala cartográfica.

A análise estatística da dependência à escala, entre variáveis e processos ambientais permite avaliar escalas apropriadas para inventário, monitoramento e modelagem (Walsh et al., 1997); quanto à quantidade de detalhes observáveis, à área afetada ou ao tamanho de objetos. Isso muitas vezes depende da análise de dados criados a partir de dados originais de

sensores, por algum tipo de operação de generalização (De Cola, 1997). Os chamados efeitos de escala são evidentes nesse processo em que padrões espaciais podem ser revelados ou desaparecer, e relações entre variáveis podem ser afetadas (Cão et al., 1997). Na área de ecologia da paisagem, por exemplo, o uso de critérios de agregação baseados em maioria tem sido comum na avaliação do crescimento e variação de padrões espaciais de classes dominantes e o desaparecimento de menores, a partir de imagens classificadas (He et al, 1997).

O conceito de agregação também pode ser usado em situações menos formais como ilustrado na Figura 2 em que um mapa obtido por um processo de classificação apresenta informação supostamente ruidosa, caracterizada pela ocorrência inesperada, de pequenas regiões da classe “pinus” (cor rosa), no interior de regiões maiores da classe “floresta” (cor verde). Um critério para eliminação pode ser definido em que, a cada local na resolução mais fina do mapa original, um novo valor é associado que depende da comparação entre o valor original e o valor generalizado de resolução mais grossa que o contém. O impacto visual do mapa resultante sugere que alguns ruídos foram eliminados como mostra a Figura 2(c). Entretanto também muita informação pode ter sido eliminada junto com os ruídos, de modo que algum método de reconstrução deve ainda ser aplicado, como veremos no final da seção 3, depois de tratarmos do tema “operações de vizinhança”.

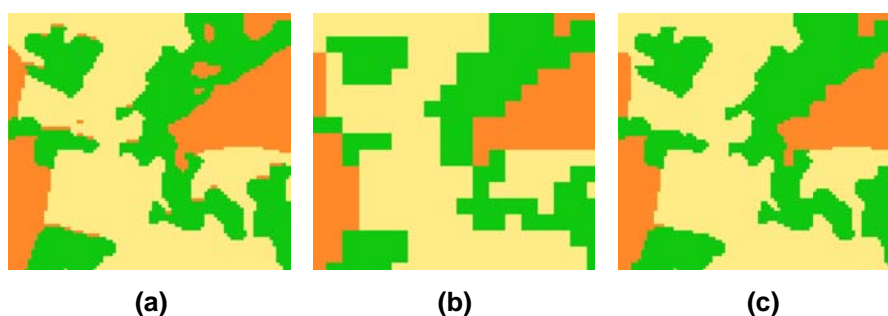


Figura 2. Extração de ruídos sobre o mapa original (a) com base em um mapa generalizado por maioria (b), resultando em um mapa sem ruídos (c)

Exemplo 1.

```
maior = Novo(Nome = "Maioria", ResX=300, ResY=300);
final = Novo(Nome = "Resultado", ResX=30, ResY=30);

maior = Maioria(inicial);

final = inicial != "pinus" ? inicial
       : maior != "pinus" ? maior
       : inicial;
```

A idéia de selecionar valores regionais para sumarizar a influencia de locais sobre locais a diferentes resoluções é essencialmente a mesma usada quando se explora propriedades de zoneamentos ou a influencia exercida por conjuntos de locais vizinhos sobre cada local de uma área de trabalho, como ocorre em muitas operações em processamento de imagem e modelagem numérica de terreno. Antes de discorrer mais sobre tais operações “não locais”, vamos falar um pouco sobre o conceito de região, uma abstração que pretende agrupar os conceitos de zonas, vizinhanças e locais.

3. Regiões

Uma região é um qualquer conjunto de locais de uma área de trabalho, podendo, portanto, ser identificada a uma regra de pertinência ao conjunto que define. Uma tal regra pode ser dada em termos de relações como ordem e igualdade entre valores associados a locais; ou relações espaciais tais como proximidade, adjacência, contigüidade e conexidade como ilustra o exemplo 2.

Exemplo 2.

```
veg = vegetação == "forest" ;
dec = declividade <= 30 ;
reg = vegetação == "forest" && declividade <= 30 ;
```

Regiões podem ser representadas por variáveis como qualquer outro tipo de dados em LEGAL, embora não possam ser diretamente realizadas sob a forma de um mapa ou qualquer representação física, elas funcionam como sinônimos para as expressões que as definem, cuja avaliação resulta em um valor '1' (true) para indicar a pertinência de um local a uma região, ou '0' (false) caso contrário. A interação entre esses valores com valores de qualquer natureza irá resultar na seleção ou não desses valores para fins de sumarização, como ilustrado na Figura 3.

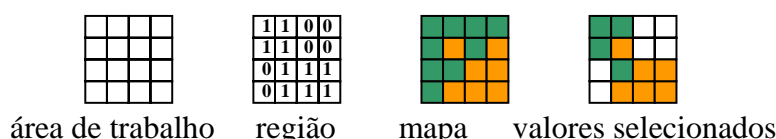


Figura 3. Interação entre regiões e mapas para seleção de valores

Na verdade as operações em modelagem espacial se fundamentam em coleções, ou famílias de regiões que recobrem uma área de trabalho. Quando cada local pode ser funcionalmente associado a uma única região, se estabelecem os fundamentos para a caracterização das classes de operações locais, zonais e de vizinhança de uma álgebra de mapas.

3.1. Zonas

Quando um conjunto de regiões particiona uma área de trabalho, dizemos que cada região é uma zona. Neste caso cada local está associado a uma única região, embora diferentes locais podem pertencer à mesma região, como ilustra a Figura 4.

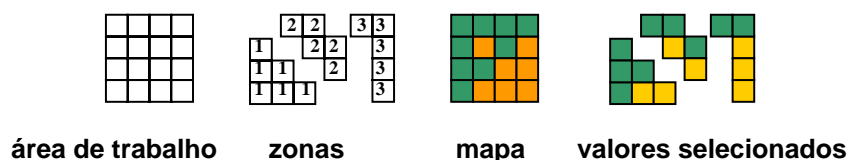


Figura 4. Cobertura de uma área de trabalho por zonas

Uma coleção de zonas pode ainda ser dada através de outras formas sintáticas que facilitam expressar regiões representadas por mapas vetoriais através de polígonos, linhas e pontos. O exemplo 3 ilustra algumas dessas situações.

Exemplo 3.

```
est = estado NoMapa estados ;
tem = temas.All ;
```

3.2. Vizinhanças

Quando cada local é associado a uma região distinta, uma família de regiões é definida, que não constitui uma partição da área de trabalho. Os locais de cada uma dessas regiões são usualmente chamados de vizinhos do local de referencia, ou foco. Os critérios para especificação de tais famílias de regiões são tipicamente espaciais, baseados em noções de proximidade e distância. Alguns modelos de vizinhança baseados na adjacência como critério de proximidade são clássicos, particularmente nas áreas de processamento de imagens, morfologia matemática e modelagem numérica de terreno, nas quais são referidos por máscaras, janelas móveis, elementos estruturantes etc. A Figura 5 ilustra alguns desses arranjos estruturais de celas ao lado de expressões em LEGAL que os definem através de listas de pares de ordenados [linha, coluna] que indicam o deslocamento, em termos do número de linhas e colunas, com relação ao foco, associado ao par [0,0], de cada vizinhança.

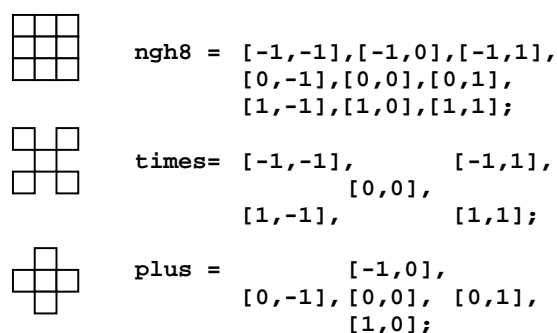


Figura 5. Formas gráficas de especificação de vizinhanças.

A interação de uma família de vizinhanças com um mapa, resulta na seleção de conjuntos de valores cuja influencia sobre cada local da área de trabalho se deseja avaliar, como ilustra a Figura 6.

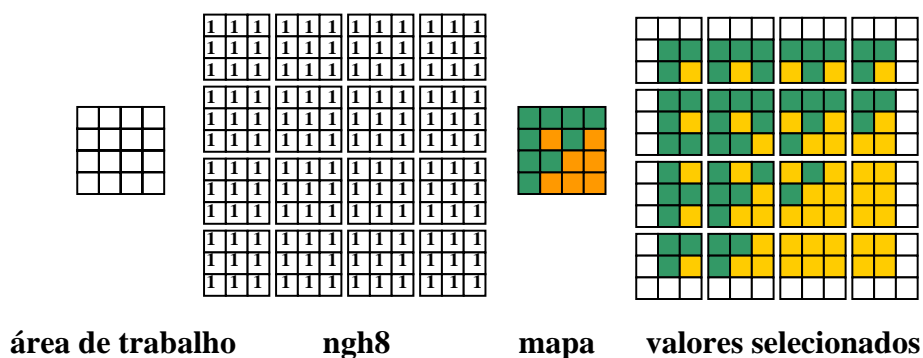


Figura 6. Seleção de valores por vizinhanças à partir de dados de mapas.

A linguagem LEGAL oferece ainda algumas formas simplificadas para referir-se a famílias de vizinhanças típicas como “3x3”, “5x5” entre outras, definidas pelo número de linhas e colunas que as caracterizam, como no exemplo 4.

Exemplo 4.

$\text{ngh8} = 3 \times 3 ;$
 $\text{ngh24} = 5 \times 5 ;$

Muitas vezes a noção de proximidade depende de medidas de distância dadas em termos do número de celas contíguas ou unidades métricas usuais definidas sobre locais de uma área de trabalho. Vizinhanças desse tipo em LEGAL são especificadas através de operações de comparação entre o valor medido por uma certa função “Distancia” entre locais e focos, contra um valor que corresponde ao raio de uma região circular, no caso de distancia Euclidiana, ou quadrangular no caso de distância medida pelo número mínimo de celas entre locais, como ilustra a Figura 7.

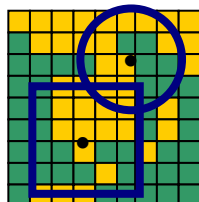


Figura 7. Vizinhanças dadas pela distância a um local de referência, ou foco.

Assim, com essas formas sintáticas, é possível definir vizinhanças mais adequadas para diversas disciplinas de modelagem ambiental, tais como autômatos celulares e agentes.

3.3. Sumarização

Após a seleção de valores por zonas, vizinhanças, ou mesmo por celas a diferentes resoluções, segue-se a última etapa de uma operação envolvendo regiões, a sumarização de um valor que represente cada região envolvida. A Figura 8 a seguir ilustra a sumarização por um critério baseado em “maioria”, referente às situações ilustradas nas Figuras 4 e 6.

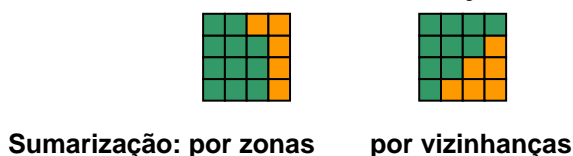


Figura 8. Sumarizando valores com base em estatísticas simples, tais como maioria.

Usando-se o símbolo ‘*’ para representar o operador de interação entre regiões e mapas, é possível finalmente, descrever operações regionais como ilustra o exemplo 5.

Exemplo 5.

```
res1 = Maioria (mapa * zonas);
res2 = Maioria (mapa * ngh8);
```

Observe que as regras sintáticas usadas são idênticas, não há distinção sob o ponto de vista de regiões entre as duas situações descritas, embora a primeira possa ser classificada como uma operação zonal e a segunda como uma operação de vizinhança.

Voltando agora para o exemplo de aplicação usado na seção 2 para ilustrar operações de generalização, a fim de recompor parte da informação que foi eliminada junto com os ruídos pode-se agora usar um critério baseado em operações de vizinhança, para extrair o valor predominante (maioria) e o valor menos freqüente (minoridade) em uma vizinhança 3x3, como ilustra o trecho de programa em LEGAL no exemplo 6.

Exemplo 6.

```
final = inicial;
n = 0;
While (n <= 2)
{
```

```

final = final == "pinus" ? Minoria(final * ngh8) == "pinus" ?
Maioria(final * ngh8) : null : null ;
n = n+1;
} ;

```

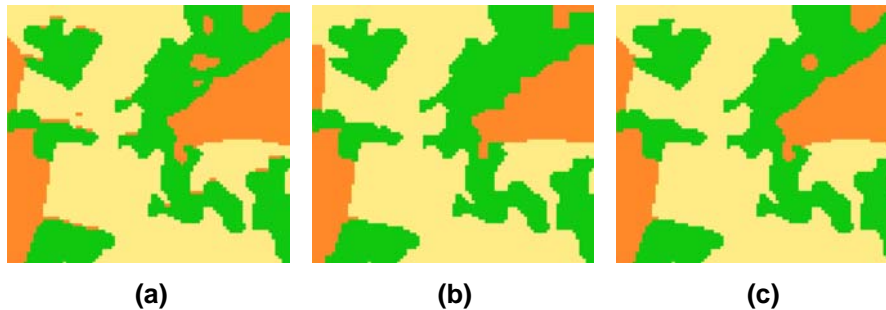


Figura 9. Vizinhanças distintas, contíguas (a) ou desconexas (b)

A Figura 9 permite confrontar o mapa original em (a), com o mapa após a redução de ruídos descrita na seção 2 em (b), e o mapa restaurado em três iterações (c). Este exemplo ilustra ainda o uso do comando iterativo “While” que permite controlar o total de repetições de um mesmo procedimento, o que no caso tem um efeito de recuperar mais e mais locais cujos valores foram alterados erroneamente na extração de ruídos.

4. Aplicações

Alguns critérios estatísticos de detecção de alvos são baseados no contraste entre parâmetros como média e mediana, das distribuições que caracterizam fundo e alvo. Em (Gam e Wang, 2006), num estudo sobre a detecção de alvos em imagens de radar, alguns métodos são avaliados com base em diferentes distribuições. A fim de ilustrar a expressividade da LEGAL, escolhemos implementar um desses métodos, em que se assume uma condição de normalidade tanto para a distribuição de valores associados aos locais de uma região de teste para detecção do alvo, quanto para a distribuição de valores sobre a região que irá representar o fundo nesse processo. Quando se associa um tal par de regiões a cada local de uma área de trabalho, pode-se então avaliar a diferença entre as médias nas áreas de teste, de alvo (μ_0) e fundo (μ_1), que quando ponderadas pelo valor inverso do desvio padrão (σ_1) da região de fundo, resulta num eficiente método de detecção de alvos, cuja expressão matemática pode ser dada pela equação:

$$d = (\mu_0 - \mu_1) / \sigma_1 \quad (1)$$

A partir de uma imagem que cobre uma área de trabalho sobre a região amazônica são inicialmente obtidas imagens representando valores médios nas vizinhanças 3x3, 5x5 e 7x7, e outras imagens representando médias sobre vizinhanças definidas pelo conjunto diferença entre as vizinhanças 9x9 e cada uma das anteriores. A Equação 1 é aplicada para cada local da área de trabalho, resultando em imagens que destacam regiões que coincidem parcialmente com as regiões mais claras, na imagem original. que podem tanto corresponder a acidentes geográficos naturais quanto dar indícios de ações de desmatamento, muitas vezes ilegal e de difícil detecção visual.

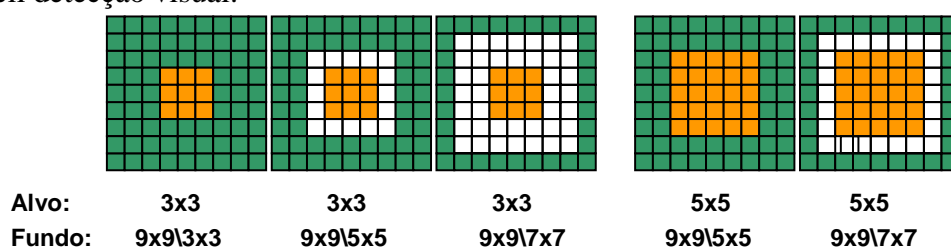


Figura 10. Pares de vizinhanças usados para representar alvos (cor laranja) e fundo (cor verde). Locais em branco não participam

A Figura 11 ilustra as diversas configurações possíveis para o desenho de detectores com base nas vizinhanças ilustradas na Figura 10. Como em uma tabela cujos elementos são imagens, a escolha de um par de imagens representando as médias de alvo e fundo permite localizar o resultado do detector correspondente, na intercessão da linha e coluna a que pertencem.

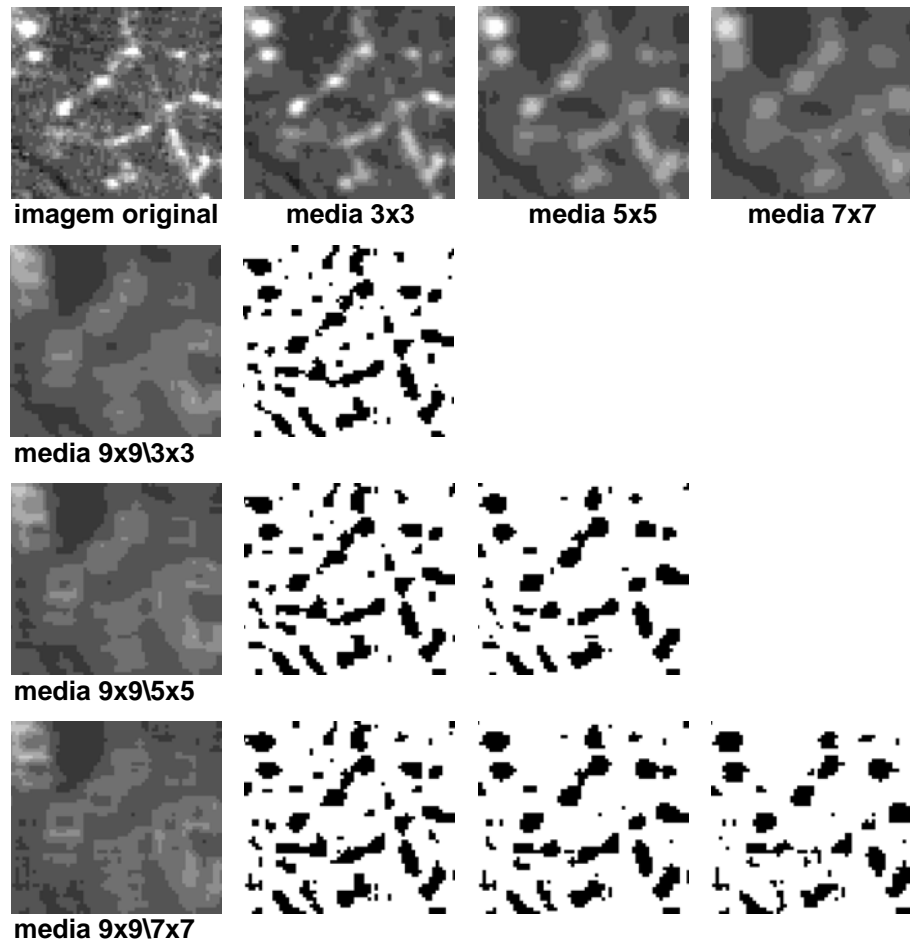


Figura 11. Detecção de alvos baseada no contraste entre médias obtidas sob diferentes formas de especificação de vizinhanças.

Para descrever em LEGAL as famílias de vizinhanças envolvidas nesse processo de detecção, vamos utilizar a noção de distancia dada pelo número mínimo de celas contíguas, discutido na Seção 3.2. A Tabela 1 indica a correspondência entre as vizinhanças da Figura 10 e suas descrições dessa forma.

Tabela 1. Equivalência entre formas de especificar vizinhanças

3x3	Distance() < 1
9x9\3x3	1 < Distance() < 4
9x9\5x5	2 < Distance() < 4
9x9\7x7	3 < Distance() < 4

O Exemplo 7 ilustra a implementação do método discutido nos parágrafos anteriores, para um dos resultados apresentados na Figura 11, envolvendo a imagem média 3x3 para caracterizar o alvo, e a imagem média 9x9\3x3 para caracterizar o fundo.

Exemplo 7.

```
med3 = Media((Distance() < 1) * img);

med9\3 = Media((1 < Distance() < 4) * img);
dev9\3 = img > med9\3 ?
    sqrt (Media((1<(Distance())< 4) * (img - med9_3)^2)) : null ;

det9\3 = med3 > med9_3 ? (img01- img02) / stdev : null ;
```

Observe o uso de operações de atribuição condicional na determinação do desvio padrão (variável “dev9\3”) e do detector propriamente dito (variável “det9\3”). O uso desse recurso sintático se justifica pelo fato de que os alvos que buscamos detectar têm valores digitais mais altos que os que ocorrem das regiões de fundo. Essa informação permite evitar a geração desnecessária de muitas situações de alarme-falso no processo de detecção.

5. Conclusões e trabalho futuro

Acreditamos que a principal contribuição deste trabalho consiste em mostrar que a partir de um maior nível de abstração na especificação de operações de uma álgebra de mapas, é possível obter-se um ganho significativo na expressividade de problemas que demandam de critérios mais flexíveis e variados para sua representação, como vizinhanças de dimensões e formas arbitrárias. Um outro aspecto é o de permitir que operações como agregação, frequentemente tratada como uma operação de preparação de dados, possa também contar com esse ferramental algébrico.

Ao descrever operações não locais através da distinção entre duas fases – seleção e sumarização –, estamos também criando um contexto sintático e semântico no qual essencialmente, qualquer operação possa no futuro ser concebida como a interação entre critérios locais de fácil implementação e outros conjuntos de ferramentas, por exemplo, estatísticas que possam ser incorporadas ao contexto de uma linguagem algébricas sem, contudo, estarem atreladas a funções específicas, como sugere a proposta inicial de Tomlin (1990). Operações como MaioriaZonal, MaioriaFocal e MaioriaLocal, por exemplo não precisam apresentar versões diferentes dependendo da natureza local, zonal ou focal em que se inserem, desde que, ocorra a separação entre selecionar valores e efetivamente aplicar uma estatística.

O projeto completo da linguagem envolve também regiões definidas pela distância a locais específicos ou mesmo a conjuntos específicos de locais de uma área de trabalho, tornando possível a modelagem de problemas de propagação e difusão, eventualmente ponderados por custos e atritos atribuídos ao longo de uma área de trabalho. Outra classe de operações interessante é voltada para a análise de visibilidade, para isso a noção de direção também deve ser incorporada à linguagem. É claro que essa unificação em torno do conceito de região discutido aqui é ainda um pouco mais teórica do que prática, e existe muito a fazer no sentido de conformar uma linguagem que possa explorar ao máximo suas propriedades.

6. Referências

- Barbosa, C.C.F.; Camara, G.; Medeiros, J.S.; Crepani, E.; Novo, E.M.L.M. ; Cordeiro, J.P. “Operadores Zonais em Álgebra de Mapas e Sua Aplicação a Zoneamento Ecológico-Econômico”. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 9, 1998, Salvador. **Anais...** São José dos Campos, INPE, 1998 (em CD-ROM).
- Camara, G., Souza, R.C., Freitas, U.M., Garido, J.C. SPRING: Integrating Remote Sensing and GIS with Object-Oriented Data Modeling **Computers and Graphics**, 15, 6, 1994.

Cao, C., Lam, N.S. Understanding the Scale and Resolution Effects in Remote Sensing and GIS. In Quattrochi, D.A., Goodchild, M.F. (editors) *Scale in Remote Sensing and GIS*, chapter2, p.57-72, 1st edition, CRC, Boca Raton, FL, USA, january, 1997.

Cordeiro, J. P.; Amaral, S.; Freitas, U.M.; Câmara, G. Álgebra de geo-campos e suas aplicações. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 8, 1996, Salvador. **Anais...** São José dos Campos: INPE, 1996. Artigos, p. 1641-1648.

Cordeiro J.P., Câmara G., Freitas U.M., Almeida F. Yet Another Map Algebra **GeoInformatica**, doi: 10.1007/s10707-008-0045-4, apr, 2008.

De Cola, L. Multiresolution covariation among Landsat and AVHRR vegetation indices. In Quattrochi, D.A., Goodchild, M.F. (editors) **Scale in Remote Sensing and GIS**, chapter3, p.73-91, 1st edition, CRC, Boca Raton, FL, USA, january, 1997.

Gan, R.B., Wang, J.G. Distribution-based CFAR detectors in SAR images **Journal of Systems Engineering and Electronics**, 17 (4): p. 717-721, dec, 2006.

He, H.S., Ventura, S.J., Mladenoff, D.J. Effects of spatial aggregation approaches on classified satellite imagery **International Journal of Geographical Information Science**, 16 (1): p.93-109, jan, 2002.

Ritter, G. X., Wilson, J., Davidson, J. Image Algebra An Overview, **Computer Vision, Graphics and Image Processing**, 49, p.297-331, 1990.

Takeyama, M. Building spatial models within GIS through geo-algebra **Transactions in GIS**, 2, p.245–256, 1997.

Takeyama, M.; Couclelis, H. Map dynamics: integrating cellular automata and GIS through Geo-Algebra, **International Journal of Geographical Information Science**, 11, p. 73-91, 1991.

Tomlin, D. **Geographic Information Systems and Cartographic Modeling**. NJ: Prentice Hall, Englewood Cliffs, 1990. 249 p.