

KHẢO SÁT ĐỘ CHÍNH XÁC PHƯƠNG PHÁP PHÂN TÍCH THÀNH PHẦN CHÍNH ẢNH VỆ TINH ĐA PHỔ LANDSAT TRONG THEO DÕI BIẾN ĐỘNG ĐẤT NÔNG NGHIỆP KHU VỰC ĐỒNG BẰNG SÔNG HỒNG

ThS. NGUYỄN THỊ THUÝ HẠNH

Trường ĐH Tài nguyên Môi trường HN

1. Đặt vấn đề:

T_rước vấn đề công nghiệp hóa và đô thị hóa diện tích đất nông nghiệp ngày càng bị thu hẹp ảnh hưởng nghiêm trọng đến vấn đề an ninh lương thực quốc gia. Nghiên cứu sự biến động đất nông nghiệp, đặc biệt ở vùng đồng bằng Sông Hồng và Sông Cửu Long có ý nghĩa quan trọng vì đây là hai vụ lúa của đất nước. Khai thác dữ liệu vệ tinh đa thời gian sẽ cập nhật được thông tin về biến động đất nông nghiệp có tính chu kỳ và phủ trùm trên diện rộng, đồng thời đem lại hiệu quả kinh tế cao. Trong bài báo này chúng tôi xin giới thiệu kết quả khảo sát độ chính xác phương pháp phân tích thành phần chính (PCA- Principal Component Analysis) ảnh vệ tinh đa phổ LANDSAT trong theo dõi biến động đất nông nghiệp khu vực đồng bằng Sông Hồng.

2. Vài nét về đồng bằng Sông Hồng.

Đồng bằng sông Hồng có diện tích khoảng 500km². Điều kiện tự nhiên ở đây khá thuận lợi nên thu hút dân cư tập trung đông đúc. Giai đoạn gần đây, trước sức ép của vấn đề dân số, sự phát triển nhanh chóng của các khu công nghiệp, khu đô thị,... dẫn đến những chuyển đổi lớn về đời sống xã hội: ngành nghề đa dạng hơn, đời sống nhân dân được nâng cao hơn cả về vật chất lẫn tinh thần... Mặt khác, vấn đề tài nguyên thiên nhiên và môi trường chưa được giải quyết thỏa đáng, diện tích đất nông nghiệp bị lấy đi không hợp lý. Nghiên cứu

này sẽ cung cấp cho chúng ta tư liệu hữu ích, góp phần giải quyết những khó khăn trên.

Để phục vụ mục đích nghiên cứu, tác giả sử dụng ảnh vệ tinh Landsat TM chụp ngày 27/12/2001 và Landsat ETM+ chụp ngày 4/11/2007 [2] với 7 kênh phổ, kết hợp với tư liệu bản đồ (Bản đồ địa hình 1:50.000, Bản đồ hiện trạng sử dụng đất 1:50.000 [4]...), tài liệu về địa lý tự nhiên, số liệu thống kê sử dụng đất..., và đặc biệt là căn cứ vào kết quả kiểm tra thực địa tại khu vực thực nghiệm.

3. Cơ sở lý thuyết

Phương pháp phân tích thành phần chính (PCA) do Karl Pearson đề xuất năm 1901, được ứng dụng trong nhiều lĩnh vực: kinh tế-xã hội, vật lý... Đối với công nghệ viễn thám, do ảnh vệ tinh chụp ở các kênh phổ gần nhau thường có sự tương quan rất lớn, tức là thông tin bị trùng lặp nhiều nên ứng dụng PCA là hoàn toàn hợp lý. Thực chất, PCA cho phép chuyển đổi các giá trị độ xám của pixel về một hệ quy chiếu vuông góc mới, thuận lợi cho việc lưu trữ thông tin và nâng cao độ chính xác [1].

Về mặt toán học, PCA là kỹ thuật chuyển đổi tuyến tính giá trị độ xám của pixel ảnh từ không gian k chiều về không gian p chiều ($k > p$) được mô tả bởi phương trình:

$$z_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{ik}x_k$$

Trong đó: k - số kênh trên ảnh gốc

P - số kênh trên ảnh thành phần chính

x_i - giá trị độ xám của pixel trên ảnh gốc

z_i - giá trị độ xám của pixel trên ảnh thành phần chính

Hay:

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ \vdots \\ z_k \end{bmatrix} = \begin{bmatrix} a_{11}a_{12}\dots a_{1k} \\ a_{21}a_{22}\dots a_{2k} \\ \dots\dots\dots \\ a_{k1}a_{k2}\dots a_{kk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_k \end{bmatrix}$$

Viết lại dưới dạng ma trận: $Z = AX$ (1)

Để nhận ảnh thành phần chính cần phải giải hệ phương trình trên tìm các hệ số a_{ik} theo những điều kiện sau:

1. $\sum a_{ik}^2 = 1$

2. Giá trị phương sai của Z_i phải cực đại

3. Các đại lượng Z_i và Z_{i+1} phải độc lập tuyến tính với nhau

Quá trình tính toán được thực hiện như sau:

- Tính giá trị trung bình của các kênh ảnh

$$\mu_k = \frac{1}{n} \sum_{i=1}^n BV_{ik} \quad (2)$$

- Tính hiệp phương sai giữa các giá trị của pixel ở kênh k và kênh 1:

$$\sum_x = E \{ (X - \mu_x) (X - \mu_x) \}$$

$$\text{cov}_{kl} = \frac{\sum_{i=1}^n (BV_{ik} - \mu_k)(BV_{il} - \mu_l)}{n-1} \quad (3)$$

E: kỳ vọng toán học

BV_{ik} : giá trị độ xám pixel

- Lập ma trận hiệp phương sai giữa các giá trị của pixel ở dữ liệu ảnh mới:

$$\sum_z = A \sum_x A' \quad (4)$$

Để ảnh thành phần chính không tương quan thì \sum_z phải là ma trận chéo

Tìm ma trận A trong phương trình (1) cần phải xác định vector riêng và các giá trị riêng của \sum_x thông qua việc giải phương trình:

$$\left| \sum_x -\lambda I \right| = 0 \quad \text{trong đó } I \text{ - ma trận đơn vị}$$

Kết quả tạo ảnh thành phần chính không tương quan có \sum_z là ma trận chéo của các giá trị riêng của \sum_x :

$$\sum_z = A \sum_x A' = \begin{bmatrix} \lambda_{11} 0 \dots 0 \\ 0 \lambda_{22} \dots 0 \\ \dots\dots\dots \\ 0 \dots\dots \lambda_{kk} \end{bmatrix} \quad (5)$$

Các giá trị riêng rất quan trọng trong việc xác định thông tin chứa trong từng thành phần. Tổng các giá trị riêng cho ta tổng phương sai thể hiện trong toàn bộ ảnh thành phần chính. Phần trăm thông tin chứa trong từng thành phần chính được xác định bởi tỷ lệ phần trăm của tổng phương sai và được thể hiện bởi công thức [5]:

$$\%PC_p = (\text{Giá trị riêng } \lambda_p) / \text{Tổng giá trị riêng} (5)$$

4. Kết quả nghiên cứu

a. Tạo ảnh thành phần chính

(Xem bảng 1, 2)

Số liệu trên cho thấy: các kênh phổ của ảnh gốc có sự tương quan rất lớn:

Năm 2001: Kênh 1 và kênh 2 (90.9%), kênh 2 và kênh 3 (93.6%)...

Năm 2007: kênh 1 và kênh 2 (94.8%), kênh 4 và kênh 5 (92.7%)...

Tức là các kênh này sẽ không bổ sung thông tin cho nhau khi ta sử dụng để tổ hợp

màu.

Thực hiện phép phân tích thành phần chính, ta được ảnh mới hầu như không có sự tương quan giữa các kênh phổ (xem bảng 3, bảng 4 - Ma trận hiệp phương sai là ma trận chéo). (Xem bảng 5, hình 1)

Bảng 5 cho thấy: nếu chỉ sử dụng ba thành phần đầu tiên để tổ hợp màu khi phân loại ảnh đã cung cấp được 97.8% (năm 2001) và 97.5% (năm 2007) thông tin. Như vậy, ta đã thực hiện chuyển đổi ảnh gốc 7 kênh về ảnh thành phần chính chỉ có 3 kênh mà vẫn đảm bảo lượng thông tin không bị thay đổi đáng kể, tạo ảnh chứa thông tin chủ yếu dễ nhận biết hơn so với ảnh gốc, việc lưu trữ thông tin cũng thuận tiện hơn (tiết kiệm bộ nhớ).

b. Phân loại ảnh

Trên cơ sở ảnh thành phần chính (tổ hợp màu PC1-PC2-PC3) năm 2001 và năm 2007, ta tiến hành phân loại đa phổ

(Xem hình 2, 3)

Bảng 1: Ma trận tương quan giữa các kênh phổ của ảnh gốc năm 2001

| Correlation | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 | Band 7 |
|-------------|--------|--------|--------|--------|--------|--------|--------|
| Band 1 | 1 | 0.9092 | 0.8810 | 0.0236 | 0.4334 | 0.1492 | 0.5609 |
| Band 2 | 0.9092 | 1 | 0.9362 | 0.0670 | 0.4491 | 0.1831 | 0.5647 |
| Band 3 | 0.8810 | 0.9362 | 1 | 0.0507 | 0.5409 | 0.2834 | 0.6788 |
| Band 4 | 0.0236 | 0.0670 | 0.0507 | 1 | 0.6700 | 0.2877 | 0.4723 |
| Band 5 | 0.4334 | 0.4491 | 0.5409 | 0.6700 | 1 | 0.5637 | 0.9462 |
| Band 6 | 0.1492 | 0.1831 | 0.2834 | 0.2877 | 0.5637 | 1 | 0.5772 |
| Band 7 | 0.5609 | 0.5647 | 0.6788 | 0.4723 | 0.9462 | 0.5772 | 1 |

Bảng 2: Ma trận tương quan giữa các kênh phổ của ảnh gốc năm 2007

| Correlation | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 | Band 7 |
|-------------|--------|--------|---------|---------|--------|---------|--------|
| Band 1 | 1 | 0.9481 | 0.43119 | 0.5852 | 0.5568 | 0.6027 | 0.5771 |
| Band 2 | 0.9481 | 1 | 0.6250 | 0.4482 | 0.4412 | 0.4551 | 0.5471 |
| Band 3 | 0.4311 | 0.6250 | 1 | -0.0436 | 0.0263 | -0.0602 | 0.2553 |
| Band 4 | 0.5852 | 0.4482 | -0.0436 | 1 | 0.9269 | 0.8956 | 0.2081 |
| Band 5 | 0.5568 | 0.4412 | 0.0263 | 0.9269 | 1 | 0.9212 | 0.1998 |
| Band 6 | 0.6027 | 0.4551 | -0.0602 | 0.8956 | 0.9212 | 1 | 0.2510 |
| Band 7 | 0.5771 | 0.5471 | 0.2553 | 0.2081 | 0.1998 | 0.2510 | 1 |
| Band 8 | 0.5699 | 0.5403 | 0.2527 | 0.2059 | 0.1974 | 0.2478 | 0.9794 |

Đánh giá độ chính xác kết quả phân loại trên ảnh thành phần chính (Xem bảng 6, 7)

c. Xác định biến động

Bảng 8: Thống kê diện tích biến động giữa năm 2001 và năm 2007

| STT | Lớp đối tượng | Diện tích (km^2) | Phần trăm (%) |
|-----|---------------|-----------------------------|---------------|
| 1 | Rừng thưa | -3.40 | -0.74 |
| 2 | Lúa | -112.29 | -24.26 |
| 3 | Màu | -35.62 | -7.7 |
| 4 | Nước | -2.69 | -0.58 |
| 5 | Dân cư | 154.02 | 33.29 |

Trong bảng 8:

Các giá trị mang dấu (+) thể hiện sự gia tăng của các đối tượng. Ví dụ từ năm 2001 đến năm 2007, diện tích khu dân cư tăng 154.02km^2 (tăng 33.29%).

Các giá trị mang dấu (-) thể hiện sự biến mất (giảm) của các đối tượng. Ví dụ từ năm 2001 đến năm 2007, diện tích rừng thưa giảm 3.40km^2 (giảm 0.74%), diện tích đất trồng lúa giảm 112.29km^2 (giảm 24.26%).

Đo đạc - Bản đồ và các ngành liên quan

Bảng 3: Ma trận hiệp phương sai của ảnh thành phần chính năm 2001

| | | | | | | | |
|---------|---------|--------|-------|-------|-------|---|-------|
| 643.915 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 110.309 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 35.187 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 8.033 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 4.855 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 4.116 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.028 |

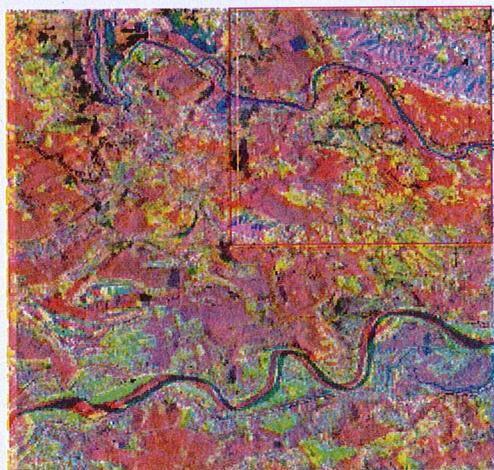
Bảng 4: Ma trận hiệp phương sai của ảnh thành phần chính năm 2007

| | | | | | | | |
|----------|---------|--------|--------|--------|--------|---|-------|
| 1439.919 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 311.415 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 91.170 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 19.418 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 12.518 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 11.058 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3.809 |

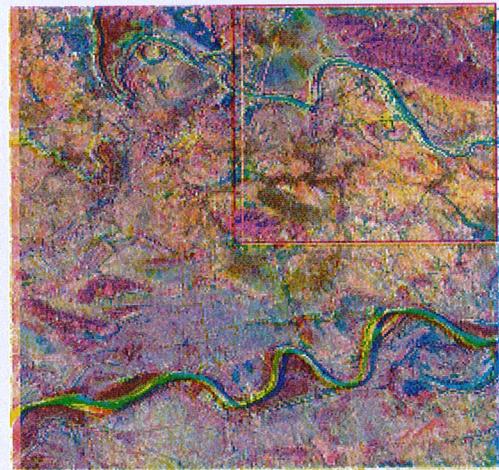
Bảng 5: Phần trăm thông tin chứa trong từng thành phần chính

| Năm | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
|------|----------|----------|----------|----------|----------|----------|----------|
| 2001 | 0.797475 | 0.136615 | 0.043578 | 0.009949 | 0.006013 | 0.005098 | 0.001273 |
| 2007 | 0.762054 | 0.164811 | 0.04825 | 0.010277 | 0.006625 | 0.005852 | 0.002016 |

Hình 1: Ảnh thành phần chính



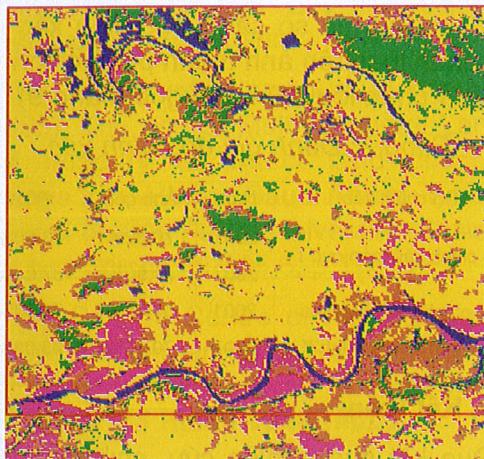
Ảnh thành phần chính năm 2001 (tổ hợp màu PC1-PC2-PC3)



Ảnh thành phần chính năm 2007 (tổ hợp màu PC1-PC2-PC3)

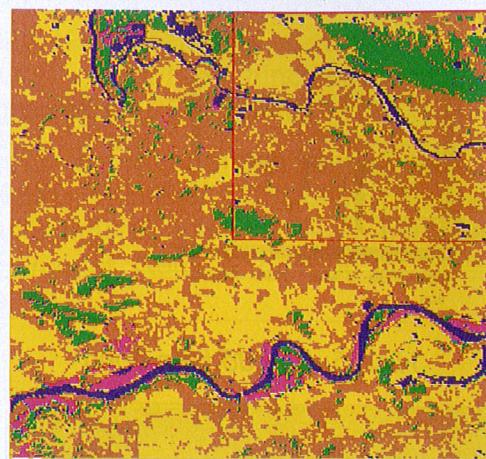
Đo đạc - Bản đồ và các ngành liên quan

Hình 2: Ảnh phân loại năm 2001



| | |
|-----------|--------|
| Rừng thưa | Màu |
| Nước | Lúa |
| | Dân cư |

Hình 3: Ảnh phân loại năm 2007



| | |
|-----------|--------|
| Rừng thưa | Màu |
| Nước | Lúa |
| | Dân cư |

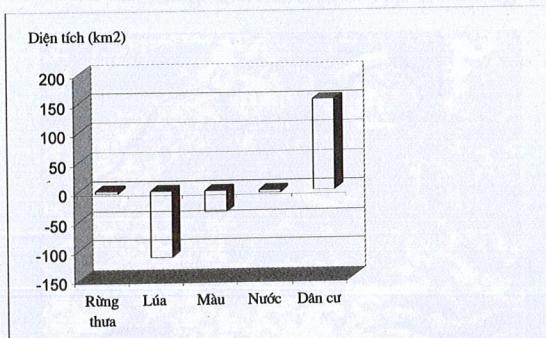
Bảng 6: Độ chính xác kết quả phân loại năm 2001

| Loại thực vật | % Rừng thưa | Loại giải đoán | | | | | Tổng hàng |
|------------------------------|----------------|----------------|-------|--------|--------|-------|--------------|
| | | Lúa | Màu | Nước | Dân cư | | |
| Rừng thưa | 97.70 | 0.00 | 1.17 | 5.62 | 17.92 | 35.30 | |
| Lúa | 0.08 | 94.76 | 5.57 | 24.72 | 22.74 | 32.35 | |
| Màu | 0.24 | 5.24 | 91.20 | 0.00 | 4.62 | 10.13 | |
| Nước | 0.00 | 0.00 | 0.00 | 63.15 | 0.00 | 6.86 | |
| Dân cư | 1.98 | 0.00 | 2.05 | 6.52 | 54.72 | 15.36 | |
| Tổng cột | 100 | 100 | 100 | 100 | 100 | 100 | |
| Độ tin cậy của nhà sản xuất | 97.70 | 94.76 | 91.20 | 63.15 | 54.72 | | |
| Độ tin cậy của người sử dụng | 85.73 | 72.38 | 74.94 | 100.00 | 90.30 | | |

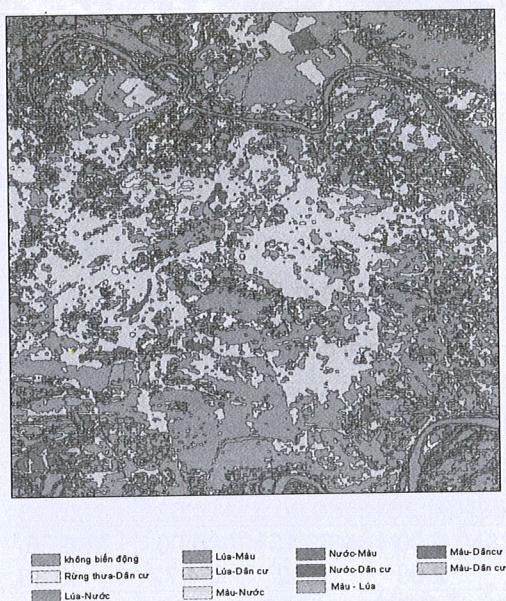
Bảng 7: Độ chính xác kết quả phân loại năm 2007

| Loại thực vật | % Rừng thưa | Loại giải đoán | | | | | Tổng hàng |
|------------------------------|----------------|----------------|-------|------|--------|-------|-----------|
| | | Lúa | Màu | Nước | Dân cư | | |
| Rừng thưa | 83.49 | 0 | 4.69 | 0 | 10.12 | 28.64 | |
| Lúa | 0.16 | 90.91 | 0 | 0 | 8.29 | 24.61 | |
| Màu | 10.32 | | 93.55 | | 1.83 | 11.43 | |
| Nước | 0 | 0 | 0 | 100 | 0 | 10.86 | |
| Dân cư | 6.03 | 9.09 | 1.76 | 0 | 79.77 | 24.46 | |
| Tổng cột | 100 | 100 | 100 | 100 | 100 | | |
| Độ tin cậy của nhà sản xuất | 83.49 | 90.91 | 93.55 | 100 | 79.77 | | |
| Độ tin cậy của người sử dụng | 89.68 | 91.27 | 68.16 | 100 | 82.63 | | |

Hình 4: Biểu đồ biến động



Hình 5: Bản đồ biến động



Từ kết quả biến động các lớp đối tượng giai đoạn 2001 - 2007 (bảng 8) ta thấy: diện tích đất nông nghiệp bị thu hẹp 151km², vốn đất này bị lấy bởi các dự án: Dự án khu Công nghệ thông tin Bắc Ninh với tổng diện tích 50ha (khởi công xây dựng năm 2004), Khu đô thị dịch vụ sân golf Vân Trung Bắc Giang,... Như vậy, trung bình mỗi năm diện tích đất nông nghiệp bị thu hẹp 25km². Điều hình tại xã Đại Đồng - Tiên Du - Bắc Ninh, diện tích đất nông nghiệp bị thu hồi chiếm 70.08% đã làm cho bình quân lương thực/người giảm từ 449kg xuống còn 96.4kg.

d. *Dánh giá độ chính xác phương pháp PCA*

Để phục vụ việc khảo sát độ chính xác phương pháp PCA trên ảnh vệ tinh đa phổ Landsat, tác giả so sánh độ chính xác kết quả phân loại trên ảnh thành phần chính với kết quả phân loại trên ảnh gốc (bảng 9)

Bảng 9: Đánh giá độ chính xác

| Độ chính xác | Năm | Ảnh gốc | PCA |
|---|------|---------|--------|
| Độ chính xác toàn cục T | 2001 | 79.32% | 83.50% |
| | 2007 | 81.90% | 87.01% |
| Độ chính xác phân loại cho toàn ảnh vệ tinh K | 2001 | 68.05% | 75.93% |
| | 2007 | 70.12% | 83.08% |

5. Kết luận

Qua kết quả nghiên cứu, chúng ta thấy ưu điểm của kỹ thuật PCA như sau:

- Nén nội dung thông tin của dữ liệu ảnh có nhiều kênh phổ sang ảnh chỉ có 3 kênh (từ 7 kênh trở thành 3 kênh), tạo điều kiện thuận lợi cho việc lưu trữ và giải đoán ảnh (hiện ảnh tổ hợp màu ba kênh phổ ứng với ba màu cơ bản PC1-PC2-PC3).

- Thực hiện phân loại trên ảnh thành phần chính (không có nhiễu và tương quan giữa các biến) giúp nâng cao đáng kể độ chính xác so với việc sử dụng ảnh gốc (bảng 9).

- Mặc dù chỉ sử dụng ảnh vệ tinh có độ phân giải trung bình (Landsat) nhưng bằng phương pháp phân tích thành phần chính, tác giả đã xác định được biến động đất nông nghiệp khu vực Đồng bằng sông Hồng với độ chính xác cao. Điều đó cho thấy hiệu quả kinh tế cũng như sự phù hợp của phương pháp được sử dụng để nghiên cứu. Đây sẽ là công cụ hữu ích để các nhà quản lý giám sát được nguồn tài nguyên quý giá-đất nông nghiệp, từ đó có chính sách sử dụng đất hợp lý, đảm bảo vấn đề an ninh lương thực quốc gia.

TÀI LIỆU THAM KHẢO

[1]. www.gov.com

[2]. <http://www.glcf.umiacs.edu/>

[3]. Lê Văn Trung (2005), Cơ sở viễn thám, Nhà xuất bản Đại học Quốc gia Thành phố Hồ Chí Minh.

[4]. Aldakheel Y and Abdulrahman, AH (2005), The use of Multi-Temporal Landsat TM Imagery to detect land cover/Use change in Al-Hassa, Saudi Arabia, Scientific Journal of King Faisal University (Basic and Applied Sciences), Vol 6, p.111-126.

[5]. Canada Centre of Remote Sensing (1998), Fundamentals of Remote Sensing

[6]. Pham Thi Mai Thi (2008), "Change detection of Multi Temporal Remote sensing Data using Principal Components Case study: Pimpri Chinchwad Municipal Corporation (PCMC) India".

[7]. S.E.Ingebritsen and R.J.P Lyon (1985), "Principal Component Analysis of Multitemporal Image pairs", International Journal of Remote sensing 6, pp. 687-696.○

ỨNG DỤNG GIS TRONG CÔNG TÁC.....

(Tiếp theo trang 46)

Summary

APPLICATION OF GIS IN ASSIGNING STATE-REGULATED PRICES TO LAND PARCELS

Ass. Prof. Dr. Tran Trong Duc

Department of Geomatics Engineering

Hochiminh city University of Technology

Land pricing is a routine yet important task in the state management of land parcels. Price of land parcels is the basis to determine land use tax and any other financial obligations concerning land use. Land prices are separately regulated for urban and rural regions. Then in each region, land prices are set specifically for each main street. At each main street, prices are set for four different position categories. Land position is determined based on the distance from a land parcel to the main street that the land parcel belongs to. In current practice, there are two methods to determine such a distance. One calculates the distance by accumulating walking distances along actual roads to access a particular land parcel from the main street. The other method uses the shortest distance from the land parcel to its main street. In each province or city the number of land parcels to be priced is huge therefore it is a terribly time consuming task if its administrative staffs have to manually measure distances, categorize position of land parcels, and assign prices accordingly. In addition, manually assigning position category to each land parcel is prone to produce error. Aiming to reduce these burdens for government staffs, a computational model is designed to determine position category of each land parcel based on its spatial relationship with the road and alley allowing access to the parcel. The model is built based on a selected combination of fundamental spatial analysis in GIS including intersection, buffering, distance measuring... Steps in the model are converted into functions in a software developed based on ArcGIS Engine and programming languages. The model is tested using information on land parcels of Long Thanh District, Dong Nai Province of Viet Nam. The experiment results demonstrate the model's effectiveness and efficiency in quickly and accurately pricing parcels at the test site.○