

# VỀ XỬ LÝ DỮ LIỆU VIDEO

**Đỗ Trung Tuấn**

*Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội*

**Lương Xuân Cường**

*Trường sĩ quan Thông tin, Nha Trang*

**Khun Piseth**

*Trường Đại học Bách khoa Hà Nội*

**Nguyễn Văn Tảo**

*Đại học Thái Nguyên*

## 1. Giới thiệu

Việc sử dụng các dữ liệu đa phương tiện là cần thiết trong xã hội hiện nay, nhưng cũng đặt ra những thách thức cho công nghệ cũng như cho công tác nghiên cứu. Trong số các dữ liệu đa phương tiện, dữ liệu video chưa được nghiên cứu để có các kĩ thuật cho phép thực hiện (i) các câu hỏi trên dữ liệu video; (ii) nhận dạng đoạn video. Để một phần đáp ứng nhu cầu về nghiên cứu trả lời cho câu hỏi vừa đặt ra, cần hình dung được các bước xử lí các dữ liệu video, mà tập trung vào (i) mô hình hoá các dữ liệu video; (ii) các phép toán đối với dữ liệu video, dựa trên mô hình được sử dụng; cho phép nhận dạng một số mẫu trong dữ liệu video, phục vụ công tác an toàn, hay cất giữ một số thông tin, phục vụ việc giấu dữ liệu; (iii) ngôn ngữ hỏi, nhằm trả lời các câu hỏi trên dữ liệu video; và (iv) thiết kế cơ sở dữ liệu video.

Tổng quan về những nghiên cứu liên quan đến xử lí dữ liệu đa phương tiện [3], người ta thấy các nhóm nghiên cứu, phòng thí nghiệm tập trung nhiều tại Châu Á, và triển khai những công việc liên quan đến tìm kiếm dữ liệu, chỉ số hoá, khai phá thông tin đa phương tiện/ đa hình thái có hoặc không có cấu trúc. Đặc biệt có nhiều hướng nghiên cứu liên quan đến lĩnh vực mà bài báo này cũng quan tâm, như (i) sử dụng các câu hỏi dữ liệu không gian/ thời gian; (ii) tổ chức dữ liệu đa phương tiện; (iii) tổ chức dữ liệu đa phương tiện theo bảng quan hệ; (iv) tự động hoá và xử lí video trong phạm vi hệ quản trị cơ sở dữ liệu.

## 2. Mô hình, phép xử lí, ngôn ngữ với dữ liệu video

Các dữ liệu video được thu nhận bằng các phương tiện ghi sẽ được thể hiện lại dưới dạng số hoá. Người ta dùng nhiều kĩ thuật nén, cho phép nén các tín hiệu video, để có thể sử dụng lại trong hệ thống số hoá, phù hợp với hiện trạng của truyền thông và mạng máy tính. Giải pháp xử lí khối lượng lớn dữ liệu trong hệ thống xử lí dữ liệu video là xác định khía cạnh ứng dụng, nhằm tổ chức dữ liệu video phù hợp với nhu cầu sử dụng. Những nhu cầu cơ thể là hệ thống tìm kiếm dữ

liệu video trong tổng thể hệ thống xử lý dữ liệu đa phương tiện [15], hay phục vụ cho yêu cầu cụ thể của đào tạo, huấn luyện [1], hay trong hệ thống kỹ thuật tự động giám sát [20, 24], hay hệ thống tìm những mẫu video tương tự [2, 11, 27].

### **2.1. Mô hình hoá**

Mô hình hoá các dữ liệu là yêu cầu trước tiên. Tiếp cận ngữ nghĩa [2, 28], tiếp cận tổng quát [13], thiết kế theo mẫu hình [18], mô hình hoá theo nhu cầu [9]... là một số tiếp cận cho phép mô hình hoá các dữ liệu video. Đi kèm với các tiếp cận mô hình hoá là kỹ thuật chỉ số hoá các dữ liệu video [6, 12, 16], tuy liên quan nhiều tới kỹ thuật vật lý, nhưng cần được tính đến khi mô hình hoá.

Khía cạnh thời gian, không gian của các dữ liệu video được quan tâm, nhằm đa dạng hoá các thuộc tính của đối tượng được thể hiện trong các đoạn băng video (gọi tắt là đối tượng video). Người ta có thể xem thời gian, không gian là thuộc tính của đối tượng video. Liên quan đến thuộc tính này là [4, 7, 14, 17, 25].

Việc mô hình hoá các dữ liệu video sẽ được tiếp nối bằng việc thiết kế cơ sở dữ liệu video, phục vụ cho nhu cầu chung, hay chỉ ít phục vụ nhu cầu của các trang tin trên mạng máy tính [15, 18]. Ngoài việc lựa chọn mô hình dữ liệu thích hợp, có nhóm nghiên cứu tập trung vào xây dựng cấu trúc dữ liệu hợp lý, cho phép tìm kiếm nhanh thông tin đa phương tiện [20]. Người ta thông kê thấy chưa có hệ thống tìm kiếm đa phương tiện nào đã dùng kỹ thuật hay cấu trúc dữ liệu để tăng tốc tìm kiếm. Các cấu trúc dữ liệu được nghiên cứu gồm cả các file nên theo nhiều chuẩn khác nhau.

### **2.2. Hỏi về dữ liệu video**

Trong hệ thống video hay chính xác là cơ sở dữ liệu video, người ta có nhu cầu hỏi, thông qua ngôn ngữ, cho phép (i) tìm dữ liệu đã thu nhận được [1, 9, 25, 26, 29]; (ii) cập nhật dữ liệu video. Mô hình dữ liệu video đã tính đến việc hỏi dữ liệu [6, 10, 13], nhằm thiết kế công cụ phù hợp với thể hiện của dữ liệu video.

Việc hỏi dữ liệu tuy thuộc nhiều vào mô hình dữ liệu video, tức tập các mô tả về các dữ liệu video cụ thể, mức độ lựa chọn là các đoạn video, khung hình, hay đối tượng xuất hiện trong những đoạn video; chẳng hạn tìm theo khung hình trong hệ thống dữ liệu lưu trữ theo nội dung [8]. Nhiều phương pháp cho phép phát hiện kết quả của tìm kiếm dữ liệu video, thông thường dựa trên (i) dữ liệu meta, như địa chỉ dữ liệu, hay đoạn văn bản mô tả đi kèm đoạn băng video; (ii) khía cạnh được trích trực tiếp từ các khung nhìn [5]. Những khía cạnh không gian, thời gian cũng tham gia vào quá trình hỏi dữ liệu.

### **3.3. Một số yêu cầu xử lý dữ liệu video**

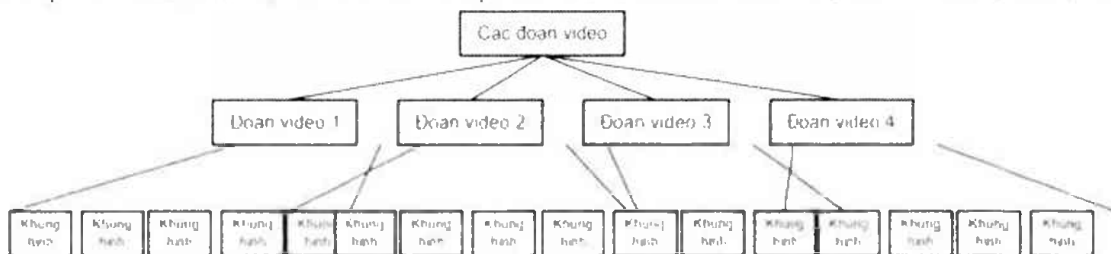
Quá trình thể hiện lại các dữ liệu video đã thu thập được, trích đoạn video, hay nén dữ liệu theo các chuẩn khác nhau là các yêu cầu khi mô hình hoá dữ liệu video, nhưng trong khuôn khổ bài báo, chỉ một số yêu cầu được đặt ra:

1. Tìm trong những đoạn video những đối tượng đáng lưu ý. Các đối tượng này đã có mẫu, dưới dạng dữ liệu video, dữ liệu meta. Việc so sánh trực tiếp đoạn video mẫu với những dữ liệu video đã có là không thể thực hiện được, do khối lượng tính toán, kích thước của dữ liệu, và khả năng của máy tính. Do vậy tiếp cận tìm theo dữ liệu meta là hợp lý. Tuy nhiên do đối tượng mẫu không được mô tả chi tiết, lệ thuộc nhiều vào thuộc tính về thời gian và không gian, nên người ta đã dùng các kĩ thuật tìm theo mức độ tương tự vừa phải [5, 27], kĩ thuật phân tách sơ bộ các khung hình [2, 19]. Những kết quả nghiên cứu thu được cũng có thể áp dụng cho các cơ sở dữ liệu đa phương tiện về âm thanh, đặc biệt tự động nhận thức một số âm mẫu trong những dữ liệu âm thu được, hoặc đang được xử lí trên trang tin của mạng máy tính; hoặc cho cơ sở dữ liệu siêu văn bản đối với bộ chữ không giống bộ chữ Latinh.

2. Dựa vào mô hình dữ liệu video, người ta không thay đổi những khía cạnh chính của các đối tượng được video thể hiện, mà thay đổi những khía cạnh không quan trọng theo hướng mang thông tin. Thông tin này có thể là thông tin cần truyền đi kèm đoạn video, hoặc thông tin về mã hoá của bản thân dữ liệu video. Việc tìm kiếm dữ liệu video theo hành động, hành vi [21] và màu sắc [28] cũng là một hướng nghiên cứu, nhằm phát hiện các đối tượng qua những khía cạnh tương tự. Việc tìm đối tượng đa phương tiện có thể theo (i) con trỏ, ứng với các yêu cầu xác định chính xác đối tượng; (ii) câu hỏi về phạm vi, ứng với yêu cầu đánh giá gần đúng, cho phép xác định một loạt các đối tượng gắn với đối tượng cần tìm; và (iii) gắn sát với k tiêu chí tìm, ứng với yêu cầu xác định các đối tượng đáp ứng nhiều tiêu chí tìm kiếm, tức là k khía cạnh của đối tượng tìm kiếm.

### 3. Đề xuất hệ thống xử lí dữ liệu video

Bài báo giới thiệu hệ thống xử lí dữ liệu video, với các chức năng từ (i) thu thập dữ liệu video, đến (ii) xử lí dữ liệu video theo hai nhu cầu vừa nêu trên.



Hình 1. Phân rã các đoạn video

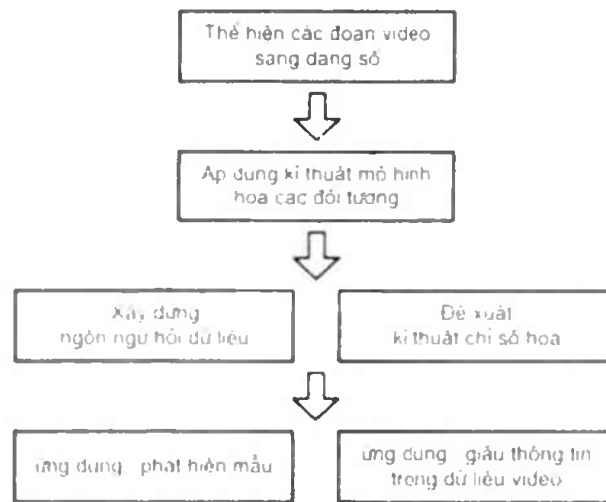
Mô hình hoá được thực hiện theo :

#### 3.1. Các dữ liệu video được tách thành (i) các đoạn video; và (ii) các khung hình

Các đoạn băng video dưới dạng số hoá, mô tả cùng một chủ đề quan tâm được chia nhỏ tuỳ theo nội dung mà các đối tượng thể hiện. Các đối tượng nằm trong nhiều đoạn video; mỗi đoạn gồm nhiều khung hình. Hầu hết các hệ thống xử lí dữ

liệu video đều phân rã các đoạn video như vậy. Một số khung hình có thể thuộc nhiều đoạn video.

Việc tách các đoạn video căn cứ vào (i) xuất hiện hay ra khỏi đoạn video của đối tượng đang xét; chẳng hạn người đang theo dõi ra khỏi đoạn video và một số yếu tố khác sẽ dùng làm dấu hiệu tách đoạn video; (ii) âm thanh đột ngột thay đổi; âm thanh tự nhiên, lời nói của đối tượng, âm nhạc thay đổi thường phù hợp với sự thay đổi nội dung; do vậy có thể dùng làm dấu hiệu ngắt đoạn video; (iii) môi trường thay đổi; dấu hiệu này sử dụng thông tin từ dữ liệu meta.



Hình 2. Chức năng hệ thống xử lý dữ liệu video

**3.2.** Các đối tượng được các đoạn video thể hiện sẽ được thể hiện lại trong cơ sở dữ liệu cố điển, chẳng hạn theo mô hình dữ liệu quan hệ; một số khía cạnh về đối tượng, tức thuộc tính của đối tượng, sẽ được dùng như dữ liệu meta, sử dụng trong việc chỉ số hoá khung hình, hay thông tin cho phép tách đoạn video;

Thí dụ hệ thống cần lưu trữ dữ liệu về một số người cần được theo dõi. Các dữ liệu meta gắn với người cho phép đánh dấu khung hình hay cả đoạn video. Các thuộc tính được quản lý theo mô hình cơ sở dữ liệu quan hệ:

*Người (tên, tỉ lệ mắt xuất hiện, độ sáng so với độ sáng ngưỡng, đã có hình mẫu hay không, khoảng thời gian,...)*

Họ và tên	Tỉ lệ mắt (%)	Độ sáng (%)	Hình mẫu	T1	T2
Nguyễn Văn Ba	65	80	Có	12	34
Trần Văn Nam	70	70	Không	15	65
Phạm Thị Toàn	45	70	Có	34	68

Hình 3. Thí dụ về thuộc tính của đối tượng tham gia trong đoạn video

Tuy nhiên, để nhận thức dễ dàng các đối tượng của thế giới thực, hệ thống cần để ra các *kiểu dữ liệu video*. Điều này sẽ làm tăng độ phức tạp của hệ thống xử lý dữ liệu video đang xử lý một số bài toán riêng, đặc thù.

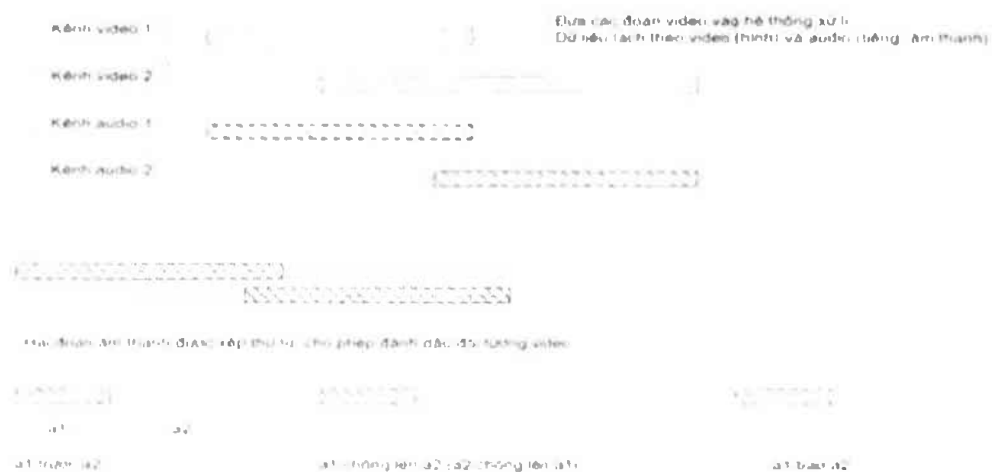
**3.3.** Khía cạnh không gian của các đối tượng được thể hiện qua vị trí của dữ liệu trong khung cảnh chung của đoạn video. Khía cạnh này mang thông tin về quan hệ giữa các đối tượng trong cùng đoạn video. Thứ tự về không gian có thể được tính đến:

Trên màn thể hiện các đối tượng video, hệ thống xác định vị trí tương đối của một đối tượng so với các dữ liệu khác. Các đối tượng khác có thể xuất hiện đồng thời hay không đồng thời. Trường hợp xuất hiện không đồng thời trên màn hiện gắn với khái niệm đối tượng ảo.



*Hình 4.* Xác định thuộc tính không gian của các đối tượng video

**3.4.** Khía cạnh thời gian của các đối tượng được thể hiện qua (i) mốc thời gian tương đối, so với mốc của các đoạn video đang xét; (ii) thứ tự thời gian trước - sau giữa các đối tượng cùng được thể hiện trong dữ liệu video;



*Hình 5.* Khía cạnh thời gian của các đối tượng video

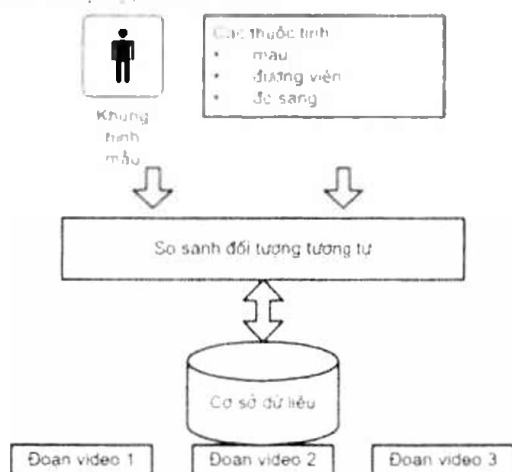
Việc xét thuộc tính thời gian của các đối tượng yêu cầu hệ thống sử dụng khoảng thời gian ( $T1$ ,  $T2$ ) đối với mỗi đối tượng; trong đó  $T1$  là mốc thời gian xuất hiện,  $T2$  là mốc thời gian biến mất.

3.5. Các thuộc tính của các đối tượng được thể hiện qua bộ  $(a_1, a_2, \dots, a_n)$ , với bộ trọng số  $(w_1, w_2, \dots, w_n)$ , trong đó  $w_i$  cho biết độ quan trọng của thuộc tính  $a_i$  đối với đối tượng.

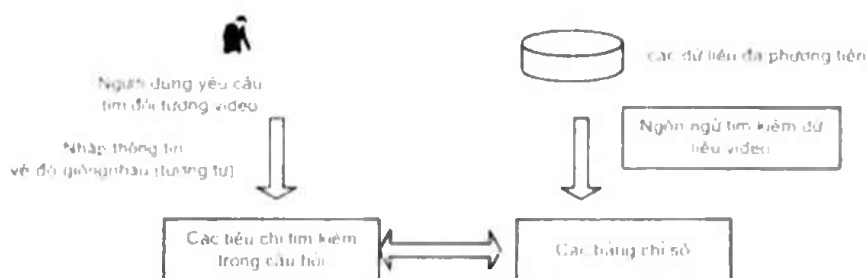
Người dùng có thể đặt câu hỏi với điều kiện tìm (i) hoặc là khung hình mẫu, (ii) hoặc thông tin về khung hình mẫu. Kết quả tìm kiếm là khung hình, hay đoạn video có đối tượng "tương tự" với đối tượng cần tìm. Tính tương tự cần được xác định qua độ đo trong không gian các thuộc tính.

Đối tượng A, B được thể hiện trong đoạn video được đặc trưng bằng bộ  $(a_1, a_2, \dots, a_n)$ ,  $(b_1, b_2, \dots, b_n)$ , với trọng số  $(w_1, w_2, \dots, w_n)$ . Mô hình dữ liệu video cần xác định các hàm  $d_i$  cho phép tính độ sai khác giữa thuộc tính  $a_i$  và  $b_i$ . Khoảng cách giữa hai đối tượng A và B được tính theo  $d(A, B) = \sum d_i(a_i, b_i) * w_i$ .

Đối tượng A được coi là tương tự B nếu  $d(A, B) < \alpha$ . Hệ số  $\alpha$  được xác định tùy theo ứng dụng, theo loại đối tượng cần tìm.

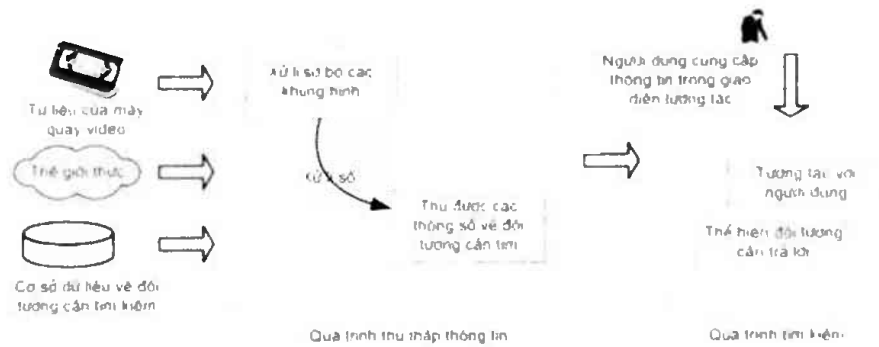


Hình 6. Kiến trúc phần so sánh độ tương tự giữa các đối tượng



Hình 7. Các tiêu chí tìm kiếm của người dùng cần được khớp với dữ liệu đã được chỉ số hoá

Việc tìm kiếm các đối tượng thông qua các khía cạnh tương tự được thực hiện trong cơ sở dữ liệu theo các quá trình (i) thu thập thông tin số từ các máy quay, cơ sở dữ liệu... (ii) ngôn ngữ xử lý dữ liệu video sẽ tìm theo các tiêu chí, vừa tương tác với người dùng. Trong quá trình này, yêu cầu đưa ra kết quả ở dạng thực, đồ họa là cần thiết đối với ứng dụng đa phương tiện.



Hình 8. Quá trình tìm kiếm đối tượng theo tiêu chí nhập vào

#### 4. Kết luận, trao đổi

Bài báo đã đặt ra yêu cầu về xử lý dữ liệu video, dựa trên các dữ liệu số hoá của các đoạn video. Việc tìm kiếm đoạn video hay một khung hình cần được thực hiện theo ngôn ngữ, được xác định phù hợp với mô hình dữ liệu video. Tiêu chuẩn tìm được viết dưới dạng biểu thức logic của các dữ liệu meta. Với công cụ hỏi dữ liệu, nhiệm vụ tìm đối tượng X để đảm bảo an ninh được giải quyết.

Thuật toán tìm dữ liệu video và tối ưu hoá việc tìm kiếm có thể hiệu quả hơn trên tổ chức vật lý phù hợp, được chỉ số hoá tốt. Bài báo này chưa đề cập đến khía cạnh này.

#### Tài liệu tham khảo

1. Antonio Bililoni, Ricardo Galli, *Content based retrieval video system for educational purpose*, [dmuabegps.tub.es](http://dmuabegps.tub.es), 2000.
2. Ardizzone E. et al., *A real-time neural approach to scene cut detection*, Univ. Palermo, Italy, 2001.
3. Basu Sankar et al., Introduction to the special issue on multimedia database, *IEEE trans. on Multimedia*, Vol. 4, N. 2, June 2002, p.141-145.
4. Bo Huang et al., *A Spacio - temporal object model and query language*, Univ. of Keele, UK ST5 5GB, [bohuang@cs.keele.ac.uk](mailto:bohuang@cs.keele.ac.uk), 2000.
5. Cheung S. S., A. Zakhor, Efficient video similarity measurement and search, *Hughes research lab and california digital media innovation (DiMI) program*, D97-03, 2001.
6. Elisa Bertino et al., An access control model for video database systems, *National Science Foundation N. 9972883-EIA*, 2000.
7. Etievent E., F. Lebourgeois, J. M. Jolion, *Assisted video sequences indexing: shot detection and motion analysis based on interest points*, INSA Lyon, French, [ciik.univ-lyon1.fr](http://ciik.univ-lyon1.fr), 2001.

8. Fushikida K. et al., A representative - video - frame selection method for a content - based video-query-agent system, *IEICE Trans. Inf. & Syst.*, Vol E83-D, N. 6 June 2000, p. 1274-1281.
9. Giuseppe Amato et al., An approach to a content-based retrieval of multimedia data, *Project Hermes*, N.9141, 2000, p. 1-31, Kluwer Academic Publishers, Boston.
10. Gulrukh Ahanger et al., *Video query formulation*, Dept. of Electrical, computer and systems engineering, Boston Univ., Boston, Massachusetts 02215, USA, 2000.
11. Heinrichs A. et al., *Image Indexing and content-based search using pre-attentive similarities*, poster in Proc. Fourth, 2000.
12. Hjelsvold R. et al., *Digital video archives*, *Hybrid multimedia*, 1994.
13. Hjelsvold R., R. Midtstraum, Modelling and querying video data, *Proc. of 20<sup>th</sup> VLDB Conf.*, Santiago Chile, 9p, 1994.
14. Hjelsvold R., R. Midtstraum, O. Sandsta, A temporal foundation of video databases, *Norwegian Institute of Technology*, 2001.
15. Howard D. Waetlar, Multi-document summarization and visualization in the informedia digital video library, *New Information Technology 2001 Conference*, Beijing, May 29-31, 2001.
16. Ichiro Ide et al., An automatic video indexing method based on shot classification, *Systems and Computers in Japan*, Vol. 32, N. 9(2001), p. 32-41.
17. John Z. Li et al., *Modeling of video spacial relationships in an object database Management Systems*, CITR (Canada Institute for Telecommunications Research), 2000.
18. Jon Olav Hauglid et al., *Video database on WWW*, 1998.
19. Li Zhao et al., Content-based retrieval of video shot using the improved nearest feature line method, *Microsoft research China*, 2001.
20. Lu Guojun, Techniques and data structures for efficient multimedia retrieval based on similarity, *IEEE trans. on Multimedia*, Vol. 4, N. 3, Sept., 2002, p.372-384.
21. Manabu Kuramoto, et al., Video database retrieval based on gestures and its application, *IEEE trans. on Multimedia*, Vol. 4, N. 4, Dec. 2002, p. 500-508
22. Ngo Chong-Wah, et al., On clustering and retrieval of video shots through temporal slides analysis, *IEEE trans. on Multimedia*, Vol. 4, N. 4, Dec. 2002, p.446-458.
23. Niels Haering et al., *A Framework for the design of event detectors*, 2001



24. Niels Hacring, et al., *Detecting hunts in wildlife videos*, 2001.
25. Petkovic M., W. Jonker, *Content-based retrieval of spatio-temporal video events*, milan@cs.utwente.nl, 2000.
26. Shermann S. M. Chan, Yi Wu, A hybrid approach to video retrieval in a generic video management and application processing framework, *Project N. 7001073*, HongKong, 2000.
27. Tan Y., S. R. Kulkarni, P. J. Ramadge, *A framework for measuring video similarity and its application to video query by example*, New Jersey Center for Multimedia Research, and the national Science Foundation NYI Grant IRI-9456745, 1998.
28. Theo Gevers, Image segmentation and similarity of color-texture objects, *IEEE trans. on Multimedia*, Vol. 4, N. 4, Dec. 2002, p. 509-516.
29. Zhuang Y. et al., *Applying semantic association to support content-based video retrieval*, Dept. of Computer Science, P. R. China, Army Research Laboratory, N. DAAL01-96-2-0003, 2002.

## VIDEO DATA MANIPULATION

**Do Trung Tuan**

*College of Science, VNU*

**Luong Xuan Cuong**

*Military Communicationa Institute, Nha Trang*

**Khun Piseth**

*Hanoi University of Technology*

**Nguyen Van Tao**

*Uninversity of Thai Nguyen*

One among research for data manipulation in certain big information systems, especially on communication networks is research on multimedia information. Video data is interesting in research; in Vietnam it is in special laboratory on information technology, in research development projects, in national informatics-electronics-telecommunication program. The article aims at presenting to general architecture of the systems allowing to modeling, in order to manipulate multimedia information, video data particularly; and to propose some problems in safe environment, linguistics processing and information hiding on video. Wavelet transform and it's application on image compression