

# THÔNG TIN LUẬN ÁN

Tên luận án: **Truy vấn hiệu quả thông tin thị giác từ dữ liệu lớn để phát triển môi trường thông minh**

Chuyên ngành: Khoa học máy tính

Mã số: 62.48.01.01

Họ tên NCS : Nguyễn Vinh Tiệp

Hướng dẫn khoa học: PGS. TS. Trần Minh Triết, PGS. TS. Dương Anh Đức

Cơ sở đào tạo : Trường Đại học Công nghệ thông tin – ĐHQG TP. HCM

## 1. TÓM TẮT

Luận án này tập trung đề xuất một số **phương pháp truy vấn hiệu quả** với các thể thức truy vấn khác nhau từ **kho dữ liệu lớn** các ảnh tĩnh hoặc video theo những nhu cầu tìm kiếm khác nhau của người dùng. Ứng với từng thể thức và đối tượng truy vấn, chúng tôi đã có các nghiên cứu sau:

**Truy vấn dựa trên ảnh mẫu**, người dùng có thể quan tâm đến (i) cảnh vật chứa trong toàn bộ ảnh mẫu, (ii) một đối tượng duy nhất chứa trong ảnh mẫu hoặc (iii) nhiều đối tượng thuộc nhiều ảnh mẫu khác nhau, cụ thể là người và địa điểm.

**Truy vấn bằng câu mô tả**: trong nhiều trường hợp, nội dung đi kèm do người đăng bài cung cấp không đầy đủ hoặc không đúng với những gì mà hình ảnh thể hiện. Mặt khác, câu truy vấn mô tả của người dùng đôi khi hàm chứa ngữ nghĩa phức tạp. Luận án này đề xuất hệ thống truy vấn trong đó khai thác các khía cạnh khác nhau của một tấm ảnh như: các đối tượng chính, quan hệ về ngữ cảnh giữa các đối tượng, các thuộc tính về địa điểm của ảnh và dữ liệu metadata.

Để minh họa cho các tiềm năng ứng dụng, chúng tôi xây dựng các hệ thống để minh họa cho ý tưởng tương tác và tiềm năng ứng dụng thực tế: hệ thống khuyến nghị gợi nhớ hình ảnh có liên quan trên mạng xã hội và hệ thống tìm kiếm bằng từ khoá trên tập dữ liệu video lớn.

Trong quá trình thực hiện luận án, NCS đã công bố 3 bài tạp chí: IJMDEM-2015 (ESCI), Informatica-2017 (Scopus), IJMIR-2019 (Scopus) và 4 bài hội thảo quốc tế : MMSP-2015, MMM-2017, SoICT-2016, MUM-2016.

## 2. CÁC ĐÓNG GÓP CHÍNH CỦA LUẬN ÁN

Luận án có những đóng góp chính sau:

1. Cấu trúc không gian trong chỉ mục ngược: đề xuất phương pháp kết hợp file chỉ mục ngược với cấu trúc kim tự tháp không gian (spatial pyramid) để vừa tăng tốc độ và giảm thời gian truy vấn.
2. Dung hợp phương pháp BOW và thuật toán phát hiện đối tượng: đề xuất phương pháp kiểm tra ràng buộc mới trong đó dung hợp phương pháp BOW và phương pháp phát hiện đối tượng. Đóng góp chính của chúng tôi trong đề xuất này là khai thác hiệu quả mối quan hệ về vị trí của các từ thị giác (visual word) với vị trí đề xuất đối tượng (object instance proposal) được ước lượng bởi bộ phát hiện đối tượng.
3. Học với đặc trưng cấp cao của mạng DNN: đề xuất phương pháp kết hợp đặc trưng học sâu với mô hình BOW và theo vết địa điểm (scene tracking) để tăng độ phủ của kết quả truy vấn.
4. Làm giàu ngữ nghĩa ảnh bằng các mạng CNN: đề xuất hệ thống khai thác các đối tượng chính, các thuộc tính liên quan đến không gian, quan hệ giữa các đối tượng trong ảnh và cả dữ liệu metadata cung cấp bởi người dùng để mô tả tất cả các khía cạnh của một frame ảnh.

### 3. NHỮNG VẤN ĐỀ CÒN BỎ NGỎ CẦN TIẾP TỤC NGHIÊN CỨU

Dưới đây là một số hướng phát triển cho một số thể thức truy vấn:

**Đối tượng nhỏ ít đặc trưng:** khai thác cấu trúc chỉ mục ngược cho bài toán phát hiện đối tượng để tăng tốc độ truy vấn.

**Nhóm đối tượng:** tích hợp cấu trúc chỉ mục ngược và chiến lược nhánh cận trong việc lưu trữ đặc trưng gương mặt người rút trích từ kho dữ liệu ảnh để tăng tốc độ truy vấn.

**Truy vấn dựa trên ngữ nghĩa:** kết hợp với các công cụ tìm kiếm hình ảnh dựa trên văn bản hiện nay để học online và cho kết quả gần với dữ liệu gán nhãn trước đó.

CÁN BỘ HƯỚNG DẪN

NGHIÊN CỨU SINH

DƯƠNG ANH ĐỨC

TRẦN MINH TRIẾT

NGUYỄN VINH TIỆP

# THESIS INFORMATION

Title: **Efficient Visual Information Retrieval from Large-Scale Data for Smart Environment Development**

Major: Computer Science

Major Code: 62.48.01.01

PhD student : NGUYEN Vinh Tiep

Advisors: Assos. Prof. TRAN Minh Triet, Assos. Prof. Duong Anh Duc

University : University of Information Technology, VNU-HCMC

## 1. ABSTRACT

This thesis propose some **efficient retrieval methods** with different modalities from **large-scale datasets** of images and videos for many user demands. User can use image examples of an object or a description in text to query. Corresponding to each type of query, we contribute the following methods.

**For visual example based retrieval**, the user may be interested in (i) a scene in the entire query image, (ii) a single object contained in the query image or (iii) multiple objects belonging to multiple image examples, e.g. a person and a place.

**Query by text description:** in many cases, the content provided by the its owner is incomplete or incorrect. On the other hand, the user's query description sometimes contains complex semantics. This thesis proposes a semantic based retrieval system which exploits various aspects of an image such as: main objects, relations between objects, location attributes and metadata.

To illustrate potential applications, we build systems to demonstrate potential applications: the recommendation system for images reminiscence on social networks and description based image retrieval on a large-scale video dataset.

During the thesis, we have published 3 journals indexed by Scopus: IJMDEM-2015, Informatica-2017, IJMIR-2019 and 4 international conference papers: MMSP-2015, MMM-2017, SoICT-2016, MUM -2016.

## 2. THE MAIN CONTRIBUTIONS OF THE THESIS ARE AS FOLLOWS:

1. Inverted index with spatial structure embedding: the system proposes a method to combine inverted file with the spatial pyramid structure to improve both accuracy and query time.

2. Combining BOW model and object detection algorithm: the system proposes a new method of spatial verification in which BOW and object detection method are leveraged. Our main contribution is to effectively exploit the relationship of visual words with the position proposed by a detector.

3. Learning with the high-level features of DNN networks: the system proposes to combine deep feature with the BOW model and scene tracking method to increase the recall of the system.

4. Enriching semantics with CNN networks: the system proposes to exploit main objects, spatial-related attributes, relationships between objects in images and metadata data provided by the content owners to describe all aspects of an image.

### **3. OPEN PROBLEMS**

Here are some future works:

**For fairy-textured and small objects:** exploiting the inverted index structure for object detection problems to speed up the system.

**Compound query:** integrating the inverted index structure and pruning strategy in storing the facial features extracted from the images in the database to speed up the system.

**Semantic-based retrieval:** combining existing text-based image search tools to learn online.

### **ADVISORS**

### **PHD STUDENT**

TRAN MINH TRIET

DUONG ANH DUC

NGUYEN VINH TIEP