

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN CÔNG NGHỆ TRI THỨC**

BÙI TRUNG HẢI – PHẠM NGỌC TUẤN

**XÂY DỰNG ỨNG DỤNG TRỢ LÝ ẢO CHO MÁY
TÍNH SỬ DỤNG GOOGLE SPEECH API**

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN CNTT

TP. HCM, 2017

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN CÔNG NGHỆ TRI THỨC**

**BÙI TRUNG HẢI – 1312165
PHẠM NGỌC TUẤN – 1312669**

XÂY DỰNG ỨNG DỤNG TRỢ LÝ ẢO CHO MÁY TÍNH SỬ DỤNG GOOGLE SPEECH API

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN CNTT

**GIÁO VIÊN HƯỚNG DẪN
ThS. NGÔ MINH NHỰT**

KHÓA 2013 - 2017

[illegible]

TpHCM, ngày tháng năm
Giáo viên hướng dẫn
[Ký tên và ghi rõ họ tên]

This image shows a full page of white paper with horizontal dotted lines. The lines are evenly spaced and run across the width of the page, providing a guide for handwriting practice. There are no margins, text, or other markings on the page.

TpHCM, ngày tháng năm

Giáo viên phản biện

[Ký tên và ghi rõ họ tên]

LỜI CẢM ƠN

Trước tiên, em xin gửi lời tri ân sâu sắc đến Thầy Lê Hoài Bắc. Thầy đã rất tận tâm, nhiệt tình hướng dẫn và chỉ bảo em trong suốt quá trình thực hiện luận văn. Không có sự quan tâm, theo dõi chặt chẽ của Thầy chắc chắn em không thể hoàn thành luận văn này.

Em xin chân thành cảm ơn quý Thầy Cô khoa Công Nghệ Thông Tin - trường đại học Khoa Học Tự Nhiên, những người đã ân cần giảng dạy, xây dựng cho em một nền tảng kiến thức vững chắc.

Con xin cảm ơn ba mẹ đã sinh thành, nuôi dưỡng, và dạy dỗ để con có được thành quả như ngày hôm nay. Ba mẹ luôn là nguồn động viên, nguồn sức mạnh hết sức lớn lao mỗi khi con gặp khó khăn trong cuộc sống.

TP. Hồ Chí Minh, 3/2014

Trần Trung Kiên

ĐỀ CƯƠNG CHI TIẾT

| |
|--|
| Tên Đề Tài: Kết hợp đặc trưng màu-độ sâu cho nhận dạng hành động người trên video |
| Giáo viên hướng dẫn: TS. Lý Quốc Ngọc |
| Thời gian thực hiện: từ ngày 15/12/2012 đến ngày 15/07/2013 |
| Sinh viên thực hiện: Trần Duy Quang - 0912369 |
| Loại đề tài: Nghiên cứu và ứng dụng thị giác máy tính trong lĩnh vực rút trích thông tin hành động người trên video |

Nội Dung Đề Tài: Khảo sát, nghiên cứu, phân tích một số phương pháp trích chọn và biểu diễn đặc trưng màu-độ sâu nhằm mục đích áp dụng cho bài toán nhận dạng hành động người trên video. Bên cạnh đó, khóa luận đề xuất một mô hình trích chọn & kết hợp đặc trưng mới có thể ứng dụng cho nhiều lớp bài toán nhận dạng khác nhau.

Nội dung chi tiết của đề tài bao gồm:

- Nghiên cứu, khảo sát các hướng tiếp cận nhận dạng hành động người trong video
- Hai vấn đề quan tâm:
 - Thiết kế một đặc trưng giàu ngữ nghĩa, có khả năng biểu diễn cao và đặc thù cho dữ liệu độ sâu.
 - Mô hình trích chọn, kết hợp hiệu quả các kênh thông tin màu và độ sâu để biểu diễn cho hành động bất kỳ trên video.
- Mô hình nhận dạng hành động bao gồm các thành phần cơ bản:
 - Tiền xử lý dữ liệu
 - Trích chọn đặc trưng trên kênh thông tin màu-độ sâu
 - Ánh xạ các đặc trưng cấp thấp lên một cấp độ biểu diễn cao hơn
 - Học và hiệu chỉnh kết quả phân lớp
- Ứng dụng thử nghiệm:
 - Ứng dụng nhận dạng ngôn ngữ cử chỉ bàn tay và cánh tay.
 - Ứng dụng game tương tác-luyện trí nhớ thông qua cử chỉ.
 - Ứng dụng giám sát hành động người trên video.

Kế Hoạch Thực Hiện:

- 15/12/2012 – 14/01/2013: Khảo sát về các hướng tiếp cận cho bài toán nhận dạng hành động người, cử chỉ bàn tay, tập trung vào các nghiên cứu liên quan đến áp dụng đặc trưng không gian-thời gian.
- 15/01/2013 – 14/02/2013: Tìm hiểu các thuật toán trích chọn và biểu diễn đặc trưng.
- 15/02/2013 – 14/03/2013: Nghiên cứu và thiết kế mô hình trích chọn đặc trưng có ngữ nghĩa, đặc thù cho hành động người xét 2 kênh dữ liệu màu-độ sâu.
- 15/03/2013 – 14/04/2013: Tìm hiểu và xây dựng mô hình kết hợp, biểu diễn hiệu quả thông tin đặc trưng đã trích chọn.
- 15/04/2013 – 14/05/2013: Tiến hành thực nghiệm, kiểm định hiệu năng của mô hình đề xuất trên nhiều tập dữ liệu chuẩn khác nhau.
- 15/05/2013 – 14/06/2013: Phát triển ứng dụng nhận dạng hành động-ngôn ngữ cử chỉ của người trong thời gian thực. Viết paper nộp cho hội nghị RIVF2013.
- 15/06/2013 – 07/07/2013: Hoàn thiện luận văn & viết paper nộp cho hội nghị ICCAIS2013.

Xác nhận của GVHD**Ngày 08 tháng 07 năm 2013**
SV Thực hiện

MỤC LỤC

| | |
|---|-----------|
| LỜI CẢM ƠN | i |
| ĐỀ CƯƠNG CHI TIẾT | ii |
| MỤC LỤC | iv |
| DANH MỤC HÌNH ẢNH | v |
| DANH MỤC BẢNG | vi |
| Chương 1 Mở Đầu | 1 |
| Chương 2 Tổng Quan | 2 |
| Chương 3 Nghiên Cứu Thực Nghiệm Hoặc Lý Thuyết | 3 |
| Chương 4 Trình Bày, Đánh Giá, Bàn Luận Về Các Kết Quả | 4 |
| Chương 5 Kết Luận và Hướng Phát Triển | 5 |
| 5.1 Kết luận | 5 |
| 5.2 Hướng phát triển | 5 |
| TÀI LIỆU THAM KHẢO | 6 |
| Phụ Lục: Các Công Trình Đã Công Bố | 8 |

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG

Chương 1

Mở Đầu

Chương 2

Tổng Quan

Chương 3

Nghiên Cứu Thực Nghiệm Hoặc Lý Thuyết

Chương 4

Trình Bày, Đánh Giá, Bàn Luận Về Các Kết Quả

Chương 5

Kết Luận và Hướng Phát Triển

5.1 Kết luận

5.2 Hướng phát triển

TÀI LIỆU THAM KHẢO

- [1] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” 2013.
- [2] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio, “Theano: a CPU and GPU math expression compiler,” in *Proceedings of the Python for Scientific Computing Conference (SciPy)*, Jun. 2010, oral Presentation.
- [3] A. Coates, “Demystifying unsupervised feature learning,” Ph.D. dissertation, Stanford University, 2012.
- [4] A. Coates, A. Y. Ng, and H. Lee, “An analysis of single-layer networks in unsupervised feature learning,” in *International Conference on Artificial Intelligence and Statistics*, 2011, pp. 215–223.
- [5] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier networks,” in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. JMLR W&CP Volume*, vol. 15, 2011, pp. 315–323.
- [6] I. Goodfellow, H. Lee, Q. V. Le, A. Saxe, and A. Y. Ng, “Measuring invariances in deep networks,” in *Advances in neural information processing systems*, 2009, pp. 646–654.
- [7] Q. V. Le, A. Karpenko, J. Ngiam, and A. Y. Ng, “Ica with reconstruction cost for efficient overcomplete feature learning,” in *NIPS*, 2011, pp. 1017–1025.
- [8] Y. LeCun, “The MNIST database,” <http://yann.lecun.com/exdb/mnist/>.

- [9] H. Lee, A. Battle, R. Raina, and A. Ng, “Efficient sparse coding algorithms,” in *Advances in neural information processing systems*, 2006, pp. 801–808.
- [10] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807–814.
- [11] B. A. Olshausen *et al.*, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
- [12] S. Rifai, G. Mesnil, P. Vincent, X. Muller, Y. Bengio, Y. Dauphin, and X. Glorot, “Higher order contractive auto-encoder,” *Machine Learning and Knowledge Discovery in Databases*, pp. 645–660, 2011.
- [13] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, “Contractive auto-encoders: Explicit invariance during feature extraction,” in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 833–840.
- [14] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1096–1103.
- [15] M. D. Zeiler, “Hierarchical convolutional deep learning in computer vision,” Ph.D. dissertation, New York University, 2014.
- [16] M. Zeiler, M. Ranzato, R. Monga, M. Mao, K. Yang, Q. Le, P. Nguyen, A. Senior, V. Vanhoucke, J. Dean *et al.*, “On rectified linear units for speech processing.” ICASSP, 2013.

Phụ Lục: Các Công Trình Đã Công Bố

Hội nghị quốc tế:

- **K. Tran** and B. Le, “Demystifying Sparse Rectified Auto-Encoders,” in *Proceedings of the Fourth Symposium on Information and Communication Technology*, ser. SoICT’13. New York, NY, USA: ACM, 2013, pp. 101–107. [Online]. Available: <http://doi.acm.org/10.1145/2542050.2542065>