**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KHOA HỌC MÁY TÍNH**

**TRỊNH HOÀNG MINH**

**PHAN VĂN TÂN**

**KHÓA LUẬN TỐT NGHIỆP**

**NHẬN DIỆN CẢM XÚC KHUÔN MẶT THEO THỜI GIAN THỰC BẰNG WEBCAM**

**CỬ NHÂN NGÀNH KHOA HỌC MÁY TÍNH**

**TP. HỒ CHÍ MINH, 2016**

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KHOA HỌC MÁY TÍNH**

**TRỊNH HOÀNG MINH - 12520638**

**PHAN VĂN TÂN - 12520971**

**KHÓA LUẬN TỐT NGHIỆP**

**NHẬN DIỆN CẢM XÚC KHUÔN MẶT THEO THỜI GIAN THỰC BẰNG WEBCAM**

**KỸ SƯ/ CỬ NHÂNNGÀNH <TÊN NGÀNH>**

**GIẢNG VIÊN HƯỚNG DẪN**

**TS. NGUYỄN THANH BÌNH**

**TP. HỒ CHÍ MINH, 2016**

DANH SÁCH HỘI ĐỒNG BẢO VỆ KHÓA LUẬN

Hội đồng chấm khóa luận tốt nghiệp, thành lập theo Quyết định số …………………… ngày ………………….. của Hiệu trưởng Trường Đại học Công nghệ Thông tin.

* 1. …………………………………………. – Chủ tịch.
  2. …………………………………………. – Thư ký.
  3. …………………………………………. – Ủy viên.
  4. …………………………………………. – Ủy viên.

MỤC LỤC

[Chương 1. TỔNG QUAN 2](#_Toc455013134)

[1.1. Đặt vấn đề 2](#_Toc455013135)

[1.2. Các nghiên cứu liên quan 2](#_Toc455013136)

[1.3. Mục tiêu đề tài, phạm vi bài toán 3](#_Toc455013137)

[Chương 2. HƯỚNG TIẾP CẬN GIẢI QUYẾT BÀI TOÁN 4](#_Toc455013138)

[2.1. Thuật toán SVM (Support Vector Machines) 4](#_Toc455013139)

[2.1.1. Tổng quan 4](#_Toc455013140)

[2.1.2. SVM tuyến tính (Linear SVM) 4](#_Toc455013141)

[2.1.2.1. Hard-margin 4](#_Toc455013142)

[2.1.2.2. Soft-margin 6](#_Toc455013143)

[2.1.3. Kernel method 6](#_Toc455013144)

[2.1.3.1. Định nghĩa 7](#_Toc455013145)

[2.1.3.2. Điều kiện Mercer 8](#_Toc455013146)

[2.1.3.3. Radial basis function (RBF) 8](#_Toc455013147)

[2.2. Cải tiến model 9](#_Toc455013148)

[2.3. Phương pháp tính toán feature 9](#_Toc455013149)

[Chương 3. TÊN CHƯƠNG 3 10](#_Toc455013150)

[3.1. Chủ đề cấp độ 2 10](#_Toc455013151)

[3.1.1. Chủ đề cấp độ 3 10](#_Toc455013152)

[3.1.1.1. Chủ đề cấp độ 4 10](#_Toc455013153)

[3.2. Chủ đề cấp độ 2 10](#_Toc455013154)

DANH MỤC TỪ VIẾT TẮT

LỜI NÓI ĐẦU

Cùng với sự phát triển không ngừng của Trí tuệ nhân tạo, con người đang hướng đến việc xây dựng những robot không chỉ có khả năng phân tích và dự đoán ưu việt hơn con người mà còn cần có khả năng giao tiếp tốt giữa các robot với nhau và với cả con người. Một trong những điểm thiết yếu để robot có thể giao tiếp với con người một cách thoải mái đó chính là việc biểu lộ cảm xúc khuôn mặt sao cho phù hợp với đối tượng giao tiếp. Hiện tại cũng có nhiều bài báo nghiên cứu về vấn đề này nhưng vẫn chưa giải quyết vấn đề một cách thỏa đáng.

Trong luận văn này, chúng tôi sẽ đưa ra một hướng tiếp cận mới nhằm giải quyết bài toán này: đó là thông qua các facial landmark ta sẽ thiết lập mô hình huấn luyện để nhận diện cảm xúc, từ đó tiến hành xây dựng ứng dụng nhận diện với mục tiêu là có thể thực thi thời gian thực.

# TỔNG QUAN

## Đặt vấn đề

Trong khoa học Trí tuệ nhân tạo đặc biệt là robot, việc tăng cường khả năng giao tiếp với con người của robot là hệ quả tất yếu. Một robot với khả năng giao tiếp giống người thật sẽ giúp cho công việc giữa người và máy có thể thuận tiện và suôn sẻ hơn. Để đạt được mục đích đó, ta cần phải tạo điều kiện giúp cho robot phân tích được những biểu lộ cảm xúc của con người thông qua giọng nói, sóng não, nét biểu lộ trên khuôn mặt, v.v... Một trong những hướng tiếp cận thuận tiện nhất đó chính là thông qua những thể hiện trên khuôn mặt. Để đạt được mục tiêu nói trên, ta phải đảm bảo việc nhận diện cảm xúc phải có được độ chính xác cao và tốc độ nhận diện chấp nhận được. Nhóm chúng em đề xuất ra 1 hướng tiếp cận mới nhằm giải quyết vấn đề đó.

## Các nghiên cứu liên quan

Theo như nghiên cứu của Pantic & Rothkrantz [0] thì nhóm ông cho rằng để giải quyết bài toán nhận diện cảm xúc khuôn mặt ta cần phải trải qua 3 quá trình: nhận diện khuôn mặt, trích xuất dữ liệu từ khuôn mặt thu được và cuối cùng là phân loại dạng cảm xúc.

Những nghiên cứu từ trước đến nay đều mặc định sẵn rằng trong 1 ảnh hoặc chuỗi ảnh đang được xử lý thì ta có được 1 số thông tin về vị trí của frontal face. Nhóm nghiên cứu của Viola & Jones [1] đã sử dụng thuật toán Adaboost để thực hiện tìm kiếm vét cạn thông qua các sub-window với các kích cỡ khác nhau của ảnh gốc nhằm xác định được vị trí chính xác của khuôn mặt. Một điểm chưa mạnh của nghiên cứu trên là tốc độ xử lý khá chậm (15FPS). V.Kazemi và J.Sullivan [2] đã đưa ra được 1 hướng tiếp cận tốt hơn, đó chính là sử dụng cây hồi quy để xác định được vị trí chính xác của khuôn mặt kèm với các landmark với tốc độ xử lý cao.

Về vấn đề trích xuất dữ liệu, nhóm của Cohn [3] đầu tiên khoanh vùng các điểm feature trong frame ảnh đầu tiên của chuỗi ảnh bằng tay, sau đó sử dụng hierarchical optical flow để track chuyển động của các cửa sổ nhỏ xung quanh các điểm đó xuyên suốt các frame. Những vector chuyển vị tính được cho mỗi landmark giữa frame khởi đầu và đỉnh điểm thể hiện thông tin thu được về biểu cảm khuôn mặt.

Công đoạn cuối cùng của quá trình nhận diện cảm xúc khuôn mặt đó là việc phân loại cảm xúc thông qua một cách sắp xếp cụ thể. Đa số các công trình nghiên cứu hiện nay đều chia cảm xúc khuôn mặt thành 6 loại chính: giận dữ, ghê tởm, sợ hãi, vui vẻ, buồn rầu và ngạc nhiên theo như nhóm nghiên cứu của P. Ekman [4]. Nhóm của ông đã thiết lập hệ thống Facial Action Coding System (FACS) cho những biểu hiện trên khuôn mặt thông qua 1 bộ 44 cử chỉ khuôn mặt được gọi là Action Units (AUs).

## Mục tiêu đề tài, phạm vi bài toán

Nhóm sẽ sử dụng thuật toán nhận diện facial landmark của V.Kazemi và J.Sullivan [2] để tối ưu hóa thời gian nhận diện các facial landmark. Sau khi tiền xử lý landmark thu được nhóm sẽ tính toán 1 set gồm 105 feature kèm với các label. Trong nghiên cứu này thì nhóm không dựa vào FACS mà chỉ nhận dạng frontal face với 3 dạng cảm xúc cơ bản: tích cực (positive), tiêu cực (negative) và bình thường (neutral). Sau khi thử nghiệm qua các model máy học thì nhóm quyết định sử dụng thuật toán SVM để phân loại vì tính hiệu quả và tốc độ xử lý của nó phù hợp với mục tiêu của đề tài.

# HƯỚNG TIẾP CẬN GIẢI QUYẾT BÀI TOÁN

## Thuật toán SVM (Support Vector Machines)

### Tổng quan

Support Vector Machines (SVM) là một kĩ thuật học có giám sát được dùng để phân lớp hoặc phân tích hồi quy. Bắt nguồn từ lý thuyết học thống kê (Statistical Learning Theory) được phát triển bởi nhóm nghiên cứu của V. Vapnik [5], SVM được xây dựng dựa trên nguyên tắc giảm thiểu rủi ro (Structural Risk Minimization).

Với kĩ thuật SVM không gian dữ liệu đầu vào sẽ được ánh xạ vào không gian đặc trưng, và từ đó ta sẽ xác định được siêu phẳng phân chia tối ưu.

### SVM tuyến tính (Linear SVM)

Ta có tập S gồm *e* các mẫu học

với một vector đầu vào *n* chiều thuộc 2 lớp với nhãn ứng với *yi = 1* hoặc *yi = -1*.

Đối với bài toán phân chia dữ liệu tuyến tính, ta cần tìm siêu phẳng maximum- margin có thể tách riêng nhóm điểm có *yi = 1* với *yi = -1*, tức là tìm siêu phẳng sao cho khoảng cách giữa nó và vector gần nhất từ cả 2 nhóm là lớn nhất.

#### Hard-margin

Với điều kiện data để học có thể phân chia tuyến tính, ta chọn 2 siêu phẳng song song phân chia 2 lớp sao cho khoảng cách giữa chúng là lớn nhất. Miền tạo bởi 2 siêu phẳng đó ta gọi là margin, siêu phẳng maximum-margin là siêu phẳng nằm giữa chúng.

Ta có thể mô tả 2 siêu phẳng ấy bằng phương trình:

và

Trong hình học giải tích, khoảng cách giữa 2 siêu phẳng trên là . Do đó để cực đại hóa khoảng cách giữa 2 siêu phẳng trên ta cần phần giảm thiểu . Đồng thời để các điểm không rơi vào trong margin, ta thêm ràng buộc rằng với mọi *i* ta có

nếu *yi = 1*

hoặc

nếu *yi = -1*

Những ràng buộc này đảm bảo các điểm đều nằm đúng vào đúng vùng class của nó. Ta có thể viết lại như sau:

với mọi (1)

Tổng hợp lại ta có bài toán tối ưu như sau:

“Tìm sao cho thỏa với mọi ”

Vector pháp tuyến xác định chiều của siêu phẳng, còn giá trị ngưỡng *b* xác định khoảng cách giữa siêu phẳng và gốc tọa độ. Từ 2 kết quả trên ta thu được bộ phân loại . Thêm một điểm nữa ta cần lưu ý, maximum-margin được xác định bởi những nằm kề nó. Chúng được gọi là các support vectors.



*Hình 2.1 Siêu phẳng maximum-margin và các margin của 1 bộ SVM được học từ các mẫu được gán nhãn thành 2 lớp. Những mẫu nằm trên margin được gọi là support vectors*

#### Soft-margin

Trong trường hợp dữ liệu không thể được phân chia tuyến tính, chúng ta sử dụng hàm *hinge loss*

Nếu thỏa điều kiện (1) thì kết quả trả về của hàm này bằng 0. Ngược lại thì giá trị của nó sẽ tỉ lệ thuận với khoảng cách từ margin. Từ đó ta cần phải giảm thiểu

với λ thể hiện tradeoff kích thước của margin nhằm đảm bảo thỏa điều kiện (1). Với λ thích hợp, soft-margin SVM cũng giống như hard-margin SVM nếu dữ liệu đầu vào có thể phân loại tuyến tính, ngược lại nó vẫn có thể học được 1 cách phân loại khả thi khác.

### Kernel method

Được sử dụng rộng rãi trong lĩnh vực máy học, các kĩ thuật kernel methods có thể mở rộng các thuật toán học, trong bài viết này là SVM có thể phân loại được trường hợp dữ liệu không thể phân loại tuyến tính được. Ở hình 2.2 nêu lên cho ta 1 ví dụ về việc mọi siêu phẳng đều chứa cả 2 class. Với việc sử dụng 1 hàm vô cùng phức tạp để phân loại nó là khả thi, nhưng thay vào đó ta chỉ cần dùng 1 hàm ánh xạ Ф biến không gian đầu vào *X* thành 1 không gian ℍ với số chiều lớn hơn và có thể phân loại tuyến tính được.

 (a) (b)

*Hình 2.2 Trường hợp không thể phân chia tuyến tính với 2 class* ⚫ *và* ◼*. Với hình (a) ta không tìm được siêu phẳng nào có thể tách riêng 2 class được. Thay vào đó ta sử dụng 1 ánh xạ không tuyến tính ở (b)*

Trong thực tế, số chiều của ℍ có thể vô cùng lớn. Cụ thể là trong bài toán phân loại tài liệu, với bộ từ điển chỉ với 100,000 từ, số chiều của không gian feature ℍ đạt đến 1015, điều đó khiến cho việc tính toán vô cùng tốn kém. Để giải quyết vấn đề trên ta dùng *kernel methods* vốn dựa trên *kernels* hoặc *hàm kernels*.

#### Định nghĩa

Hàm được gọi là 1 kernel của *X*

Ta định nghĩa 1 hàm kernel *K* sao cho với 2 điểm bất kì bằng với inner product của vector Ф(*x*) và Ф(*y*):

cho ánh xạ đến 1 không gian Hilbert được gọi là không gian feature. Điểm mạnh của kernel *K* là độ hiệu quả của nó: ta có thể tính được *K* với hiệu quả cao hơn so với việc tính toán Ф và inner product trong ℍ. Bên cạnh đó ta không cần định nghĩa tường minh hoặc tính toán 1 hàm ánh xạ Ф. Ta có thể chọn kernel *K* tùy ý miễn tồn tại hàm Ф, nghĩa là *K* phải thỏa mãn điều kiện Mercer.

#### Điều kiện Mercer

Với là 1 tập compact và hàm là 1 hàm liên tục và đối xứng. Ta có [K admits a uniformly convergent expansion of the form]

với *an > 0* khi và chỉ khi với bất kì hàm bình phương khả tích thỏa:

#### Radial basis function (RBF)

Với mọi hàng số , một *Gaussian kernal* hoặc *Radial basis function (RBF)* là kernel *K* định nghĩa trong tập ℝN bởi:

Gaussians kernels nằm trong số những kernel được sử dụng rộng rãi nhất trong các ứng dụng. Trong quá trình huấn luyện model học ta cần quan tâm đến 2 thông số quan trọng nhất chính là và C. Thông số gamma cho ta biết ảnh hưởng của 1 training sample với xung quanh nó, hiểu đơn giản nó là nghịch đảo của bán kính của các sample được lựa chọn làm support vector của model. Gamma tradeoff giữa tốc độ xử lý và độ hiệu quả, với giá trị càng cao thì tốc độ xử lý càng nhanh nhưng chưa chắc hiệu quả và ngược lại. Còn với thông số C tradeoff giữa việc phân loại nhầm giữa các sample và độ phức tạp của hàm kernel. C thấp sẽ cho kernel với độ phức tạp thấp trong khi C cao dễ dẫn đến hiện tượng overfit.

## Cải tiến model

Như đã đề cập ở trên, vấn đề mới nảy sinh chính là tìm cặp (gamma,C) sao cho model thu được là tốt nhất. Ta sử dụng kĩ thuật k-fold cross-validation.

Với k-fold cross-validation, ta chia tập training thành k tập con cùng kích cỡ. Từng tập con sẽ lần lượt được dùng để test trong khi các tập còn lại được dùng để train. Vì thế mỗi phần của training data đều được dùng để dự đoán 1 lần nên độ chính xác của cross-validation cũng chính là độ chính xác của lượng data được dự đoán chính xác. Việc sử dụng cross-validation sẽ giúp ta hạn chế được vấn đề overfit của model. Nghiên cứu của Chih-Wei Hsu, Chih-Chung Chang, và Chih-Jen [6] đã gợi ý bằng cách dùng dãy các số cấp số nhân tăng tiến để tìm ra thông số tốt nhất (VD: )

## Phương pháp tính toán feature

[NO DATA]

# XÂY DỰNG ỨNG DỤNG

## Hệ điều hành và ngôn ngữ lập trình

Ứng dụng của nhóm sẽ chạy trên hệ điều hành Ubuntu là 1 nhánh con của hệ điều hành Linux. Điểm mạnh của Ubuntu chính là độ tin cậy cao và lượng thư viện opensource là vô cùng dồi dào. Không đặt nặng UX hơn hiệu suất và lại mở hoàn toàn, Ubuntu là 1 lựa chọn hợp lí cho các lập trình viên.

Ứng dụng chính của nhóm sẽ được lập trình trên nền C++ trong khi SVM model sẽ được train riêng bằng 1 Python script khác.

## Thư viện sử dụng

### Thư viện dlib

Là một thư viện mở được release kèm Boost Software License, dlib là 1 thư viện C++ cross-platform miễn phí được sử dụng rộng rãi. Từ năm 2002 đến nay, dlib được phát triển trong nhiều lĩnh vực như mạng, thread, GUI, máy học, v.v... Lí do chính nhóm chọn thư viện dlib là vì dlib đã implement được thuật toán nhận diện facial landmark với hiệu quả cao, rất cần thiết cho việc xử lý real-time thông qua webcam.

### Thư viện libSVM

Thư viện libSVM cũng là 1 thư viện mở rất phổ biến được phát triển trên nền C++ và port sang các ngôn ngữ khác. Vì khả năng cross-platform của libSVM nên nhóm chọn sử dụng thư viện này để hỗ trợ cho việc training model thuận tiện hơn.

## Hướng tiếp cận

## Kết quả thử nghiệm

# KẾT LUẬN

TÀI LIỆU THAM KHẢO

Theo chuẩn IEEE