LUẬN VĂN CAO HỌC

LỜI CẢM ƠN

Em xin chân thành cảm ơn Thầy Trương Quốc Bảo đã tận tình giúp đỡ và hướng dẫn để em có thể hoàn thành luận văn một cách tốt nhất. Tuy trong thời gian thực hiện luận văn đã gặp phải rất nhiều khó khăn, nhưng với sự giúp đỡ của thầy Trương Quốc Bảo đã tạo cho em niềm tin và kiến thức để thực hiện luận văn.

Xin gửi lời cảm ơn trân quí đến nhà Trường Đại học Cần Thơ, khoa Sau đại học và Khoa Công nghệ thông tin và Truyền thông đã phối hợp tổ chức khóa đào tạo cao học ngành Hệ thống thông tin để em có cơ hội được học tập và nâng cao kiến thức, phát triển bản thân.

Con xin gửi lời cảm ơn đến gia đình, người thân và bạn bè, những người đã luôn ở bên cạnh là nguồn động lực để con có thể vượt qua những khó khăn và thực hiện tốt luận văn. Do sự hạn chế về mặt thời gian, không gian và kinh nghiệm nghiên cứu cá nhân nên luận văn còn nhiều thiếu sót, mong nhận được sự đóng góp và đánh giá của quý Thầy Cô và các bạn học viên.

MỤC LỤC

DANH MỤC CÁC CHỮ VIẾT TẮT VÀ KÝ HIỆU

|  |  |  |
| --- | --- | --- |
| **Từ viết tắt** | **Cụm từ đầy đủ** | **Ý nghĩa** |
| AAM | Active Appearance Model | Thuật toán thị giác máy tính so khớp mô hình thống kê của hình dạng đối tượng |
| ASM | Active Shape Model | Mô hình thống kê về hình dạng đối tượng |
| PCA | Principal Component Analysis | Phân tích thành phần chính |
| AUs | Action Units | Các đơn vị vận động |
| Open CV | Open Source Computer Vision Library | Thư viện về thị giác máy tính mã nguồn mở |
| SVM | Support Vector Machine | Máy học véc tơ hỗ trợ |
| ANNs | Artificial Neutral Networks | Mạng nơ ron nhân tạo |
| HOG | Histogram of oriented gradients | Bộ mô tả tính năng dùng để phát hiện đối tượng |
| BSD | Berkeley Software Distribution | Bản quyền phân phối phần mềm nguồn mở |
| DNA | Deoxyribonucleic Acid | Phân tử mang thông tin di truyền |
|  |  |  |

DANH MỤC BẢNG

DANH MỤC HÌNH

LỜI MỞ ĐẦU

Khi trò chuyện hoặc quan sát một người nào đó, chúng ta dễ dàng nhận ra cảm xúc hiện tại của họ. Họ vui khi nở một nụ cười trên môi và hai bên má nâng lên cao; họ cảm thấy khó chịu và nheo mày lại; hoặc miệng sẽ há thật to, mắt sẽ mở thật to khi nhận được một thông tin nào đó khiến học bất ngờ; và cảm xúc thông dụng nhất là trung tính, có nghĩa là không cảm xúc tương ứng với các cơ trên gương mặt luôn đặt ở trạng thái bình thường. Vậy làm như thế nào để máy tính cũng có khả năng này như là con người.

Trong giao tiếp, bên cạnh ngôn ngữ cơ thể thì gương mặt là một trong những kênh truyền thông phi ngôn ngữ quan trọng nhất. Ngoài việc biểu hiện cảm xúc, các cử chỉ trên gương mặt còn mang đến nhiều thông tin khác như truyền đạt một tín hiệu, thông điệp giao tiếp (nháy mắt, chớp mắt liên tục) hay là biểu hiện trong các trạng thái đau đớn, khó chịu; trường hợp bệnh lí, trầm cảm, rối loạn cảm xúc. Vì vậy, trong nhiều thế kỉ qua, vấn đề nghiên cứu để nhận dạng những cảm xúc thông qua biểu hiện gương mặt đã được thực hiện và không ngừng phát triển bởi các nhà khoa học.

Biểu cảm trên gương mặt là biểu hiện có thể nhìn thấy bằng mắt thường, hiểu được những trạng thái tình cảm, hoạt động nhận thức, tính cách và tâm lí của một người, dự đoán ý định của người đó, đóng góp hơn 55% hiệu quả trong hoạt động giao tiếp [Mehrabian]. Những biểu hiện gương mặt cùng với ngôn ngữ cơ thể giúp người nghe có thể hình dung thêm, hiểu thêm về ngữ cảnh đang nói và nắm bắt sâu hơn vấn đề, ý nghĩa và người nói muốn truyền đạt.

Do đó, biểu cảm gương mặt đóng một vai trò vô cùng quan trọng trong tương tác người – máy, đang được nghiên cứu và cải tiến để hỗ trợ cho những ứng dụng thực tế, giúp con người giải quyết các vấn đề dường như là không thể, như đánh giá độ trung thành của một nhân viên đối với công ty, tìm kiếm tội phạm hay các hệ thống hỗ trợ y tế.

1. Tình hình nghiên cứu hiện tại

Thế giới đang bước vào cuộc cách mạng công nghiệp 4.0 với những bước chuyển biến tích cực và mạnh mẽ của tất cả các ngành nghề trong mọi lĩnh vực của đời sống hiện đại. Cùng với đó là sự phát triển vượt bậc của khoa học và công nghệ, sự hỗ trợ thiết yếu của công nghệ thông tin trong việc tin học hóa và đơn giản hóa các quy trình làm việc của mọi lĩnh vực khác nhau. Để thực hiện được điều đó, cần có sự tương tác giữa người và máy để máy tính có thể hiểu và thực hiện những công việc tự động theo mong muốn của con người.

Thật vậy, để máy tính có thể giao tiếp với con người thì chúng ta cần có những phương pháp và kỹ thuật cụ thể, một trong số chúng là khả năng nhận dạng được cảm xúc của con người. Một trong những phương pháp vật lý và có hiệu quả để nhận dạng được cảm xúc của con người là thông qua cả biểu cảm trên gương mặt. Đây là chủ đề chính trong các nghiên cứu của các nhà khoa học ở nhiều năm trước đây, và hiện tại nó vẫn được xem là một đề tài hấp dẫn vì tính ứng dụng và độ phổ biến.

Các ứng dụng thực tế của việc nhận dạng cảm xúc là rất đa dạng và hữu ích. Trong đời sống, các ứng dụng di động nhận dạng được cảm xúc của người dùng để gán các biểu tượng cảm xúc tương ứng như snow, magic, polygram đang được giới trẻ rất ưa chuộng. Trong y học, các bác sĩ có thể theo dõi các thay đổi cảm xúc của bệnh nhân để đưa ra những chẩn đoán bệnh chính xác và điều trị bệnh hiệu quả với các ca thần kinh hay rối loạn cảm xúc. Trong thương mại, các tập đoàn, doanh nghiệp, nhà sản xuất có thể thu thập, phân tích và thống kê cảm xúc của khách hàng để đưa ra những quyết định kịp thời và đúng đắn nhằm tối ưu hóa lợi nhuận.

2. Tính cấp thiết của đề tài

Từ những ứng dụng thiết thực trên, việc nhận dạng cảm xúc của con người dựa vào biểu cảm trên gương mặt là một chủ đề rất hay, tuy không mới nhưng tính khoa học và thực tiễn cao. Các nhà nghiên cứu đi trước đã có những công trình nghiên cứu hay với các phương pháp nghiên cứu khác nhau đã cho kết quả và độ chính xác nhất định, đây cũng làm một động lực để các đề tài sau có cơ sở khoa học để so sánh và đánh giá, cải tiến hơn nữa hiệu quả mang lại.

Với sự phát triển và phổ biến của mạng xã hội và công nghệ thông tin như hiện nay, việc tạo ra các ứng dụng để hiểu được người dùng hơn cả người dùng hiểu chính họ là một điều tuyệt vời. Và việc nhận dạng cảm xúc của con người dựa trên gương mặt là một nền tảng cốt lõi cho các ứng dụng này.

3. Mục tiêu của đề tài

Mục tiêu của đề tài là tìm hiểu về các loại cảm xúc của con người, các đặc trung của gương mặt người và sự tương quan giữa cảm xúc và các đặc trưng đó. Nghiên cứu các kỹ thuật, phương pháp, thuật toán để thực hiện các công việc cụ thể trong toàn bộ quá trình hoạt động của hệ thống. Thông qua đó, thực hiện huấn luyện được một tập dữ liệu có khả năng nhận dạng được cảm xúc của con người thông qua gương mặt. Cuối cùng, xây dựng được một hệ thống nhận dạng cảm xúc con người dựa trên các hình ảnh đầu vào.

4. Đối tượng, phạm vi nghiên cứu của đề tài

Đối tượng nghiên cứu: phương pháp phân khúc ảnh và tiền xử lý hình ảnh dựa vào đặc trung HOG; các kỹ thuật nhận dạng và xác định vị trí của các thành phần trên gương mặt như AAM/ASM, PCA, AUs; mô hình máy học SVM và mạng nơ ron nhân tạo, thư viện OpenCV.

Phạm vi nghiên cứu: cơ sở lý thuyết và ứng dụng thực tế của các phương pháp, kỹ thuật hỗ trợ nhận dạng để áp dụng vào hệ thống; dữ liệu đầu vào là các ảnh tĩnh, hoặc video.

5. Phương pháp nghiên cứu

Nghiên cứu cơ sở lý thuyết của các kỹ thuật hỗ trợ nhận dạng cảm xúc dựa trên mặt người. Nghiên cứu và so sánh kết quả thực nghiệm với những đề tài cùng chủ đề đã được thực hiện trước đó. Xây dựng một chương trình thực tế để kiểm tra độ chính xác và khả năng mở rộng của dự án. Ứng dụng kết quả nghiên cứu vào một ngữ cảnh thực tế của đời sống xã hội.

6. Kết quả dự kiến

Tìm hiểu được các loại cảm xúc cơ bản của con người, kiến thức nền tảng về các phương pháp hỗ trợ nhận dạng cảm xúc dựa trên mặt người, xây dựng chương trình kiểm tra với kết quả chấp nhận được và ứng dụng vào một lĩnh vực cụ thể.

CHƯƠNG 1. CÁC LOẠI CẢM XÚC CỦA CON NGƯỜI VÀ ĐẶC TRƯNG TRÊN GƯƠNG MẶT NGƯỜI

* 1. Các trạng thái cảm xúc của con người

Cảm xúc có tác động rất lớn đến cuộc sống của con người. Cảm xúc là một thứ rất phức tạp, có thể thay đổi nhanh chóng, một người có thể có hơn một cảm xúc tại một thời điểm. Vì vậy, cảm xúc là gì và con người có tất cả bao nhiêu cảm xúc là các câu hỏi chưa có một đáp án thuyết phục.

Vào thế kỷ IV trước công nguyên, Aristole đã nhận định có 14 loại cảm xúc cơ bản nhất, bao gồm: hài lòng, tử tế, tranh đua, ganh tị, đáng thương, câm phẫn, sợ hãi, bình tĩnh, tự tin, thù địch, giận dữ, xấu hổ, bằng hữu và vô liêm sĩ. Những năm gần đây, sáu là con số đại diện cho số lượng cảm xúc cơ bản nhất của con người được các nhà tâm lý học công bố, gồm có: vui vẻ, buồn bã, ngạc nhiên, sợ hãi, ghê tởm, giận dữ.

Ở một nhận định khác, TS. Rachael Jack của đại học Glasgow, Vương quốc Anh cho rằng con người chỉ có bốn loại cảm xúc cơ bản khi mà những biểu hiện của các cơ trên gương mặt trong hai cảm xúc sợ hãi và ngạc nhiên là như nhau, tương tự với giận dữ và ghê tởm. Thật sự là hai loại cảm xúc trong từng cặp ở trên có sự giống nhau trong quá trình vận động các nhóm cơ mặt để hình thành cảm xúc, chúng chỉ khác nhau khi được thể hiện đầy đủ và xong xuôi, vì thực chất chúng là những cảm xúc khác nhau.

Đề tài được thực hiện nhận dạng sáu loại cảm xúc cơ bản nhất của con người (vui vẻ, buồn bã, ngạc nhiên, sợ hãi, ghê tởm, giận dữ).

* 1. Đặc trưng của gương mặt người

Gương mặt là một phần của cơ thể người, bộ phận trung tâm để bộ lộ cảm xúc và là phương tiện truyền tải cảm xúc giữa người với người. Gương mặt người có các thành phần đặc trưng giống nhau như chân mày, mắt, mũi, miệng, v.v. nhưng mỗi người lại có một gương mặt khác nhau và là duy nhất. Do đó, gương mặt là đặc trưng tốt nhất để phân biệt một người với những người khác.

Gương mặt là nơi biểu hiện của những cảm xúc. Một nụ cười thể hiện cho một niềm vui, một cái chau mày đồng nghĩa với việc không tán thành hay khó chịu. Chính vì thế, nhận dạng cảm xúc của người đối diện qua gương mặt có một vai trò quan trọng trong giao tiếp. Con người có thể đọc được cảm xúc của người khác nhờ biểu cảm trên gương mặt của người đó, từ đó dự đoán được khả năng xảy ra của các hành vi tiếp theo.

Cơ mặt đóng một vai trò nổi bật trong việc mô tả cảm xúc con người, cùng với các đặc trưng khác trên gương mặt mang đến sự đa dạng trong sự biểu lộ nhiều cảm xúc khác nhau.

* 1. Nhận dạng cảm xúc dựa trên mặt người

Gương mặt là nơi cảm xúc được bộc lộ rõ nhất và dễ dàng nhận thấy nhất. Chính vì thế các biểu hiện trên gương mặt người được sử dụng như là một nền tảng cốt lõi để nhận dạng cảm xúc con người.

Tương ứng với từng cảm xúc riêng biệt là những biểu hiện khác nhau trên gương mặt. Như vậy, với một loại cảm xúc nhất định sẽ có một tập các biểu hiện nhất định của các nhóm cơ và các đặc trưng khác trên gương mặt.

Đối với một người, việc quan sát gương mặt và nhận dạng cảm xúc của một người khác là một điều rất dễ dàng. Vậy đối với máy tính thì như thế nào. Chúng cần được học, được huấn luyện để có được khả năng này như con người.

Chương 2. CÁC PHƯƠNG PHÁP GIÚP NHẬN DẠNG CẢM XÚC DỰA TRÊN GƯƠNG MẶT

* 1. Phương pháp dựa trên đặc trưng của gương mặt

Sử dụng phương pháp phân tích thành phần chính PCA. Phương pháp này sẽ lấy ra được các thành phần chính của gương mặt (thể hiện bằng các eigenvectors) trong tập ảnh huấn luyện và tạo ra không gian mặt. Tiếp tục sử dụng các thuật toán máy học để huấn luyện tạo thành các tập dữ liệu huấn luyện là các lớp tướng ứng với các loại cảm xúc cơ bản.

(Hình ảnh)

PCA là một công cụ mạnh mẽ cho việc xác định hình dạng của gương mặt, phân tích các thành phần trên gương mặt.

* 1. Phương pháp sử dụng các đơn vị vận động trên gương mặt

Cảm xúc được xác định dựa trên sự chuyển động của các đơn vị trên khuôn mặt, được gọi là các action units. Có tất cả 64 action unit tương ứng với 64 biểu hiện khác nhau của các nhóm cơ trên gương mặt [Ekman và Friesen 1978]. Việc xác định cảm xúc chỉ đơn giản là việc xác định có bao nhiêu action unit cùng xuất hiện trên gương mặt tại một thời điểm, và sự kết hợp của chúng cho ra một cảm xúc duy nhất. Ví dụ: cảm xúc vui là kết quả của sự kết hợp hai action unit 6 (má nâng lên) và 12 (góc ở mép môi đưa lên cao), cảm xúc buồn gồm có các action unit 1 (vầng trán nâng lên), 4 (chân mày hạ xuống) và 15 (góc ở mép môi đưa hạ xuống) biểu hiện đồng thời. [ (1)]

Ưu điểm

Nhược điểm của phương pháp này là mang tính chủ quan cao và vấn đề thời gian thực hiện

* 1. Sử dụng mô hình AAM kết hợp tương quan điểm.

Đầu tiên, mô hình AAM được sử dụng để phát hiện vùng mặt, sau đó dựa vào sự phân bố hiện tại của các landmark và tỉ lệ giữa chúng để biết được tương ứng với loại cảm xúc nào. Ý tưởng chính là: Nếu có thể xác định được chính xác tọa độ các điểm trên khuôn mặt thì có thể dựa vào tương quan các điểm đó để nhận dạng cảm xúc.

Ví dụ một cách chi tiết. Xem hình bên dưới:

(Chèn hình vào đây)

Gọi:

R1 = (Độ rộng miệng)/(Độ rộng mặt)

R2 = (Chiều cao mắt)/(Chiều cao mũi)

R3 = (Khoảng cách hai chân mày)/(Khoảng cách hai mắt)

*Với R1:* Khi gương mặt ở trạng thái bình thường, độ rộng miệng luôn nhỏ hơn độ rộng gương mặt, nên R1 luôn nhỏ hơn 1. Khi cười, miệng của chúng ta sẽ rộng hơn bình thường và độ rộng miệng tăng lên, độ rộng gương mặt vẫn giữ nguyên, kéo theo R1 sẽ tăng lên.

*Với R2:* Khi gương mặt ở trạng thái bình thường, chiều cao mắt luôn nhỏ hơn chiều cao của mũi, nên R2 luôn nhỏ hơn 1. Khi chúng ta nhạc nhiên vì một điều gì đó, mắt chúng ta sẽ mở to và rộng hơn, đồng thời là miệng há to làm cho độ rộng miệng nhỏ lại. Điều đó làm cho R2 tăng lên và R1 giảm đi.

*Với R3:* Khi gương mặt ở trạng thái bình thường, khoảng chắc giữa hai chân mày và khoảng cách giữa hai mắt là tương đương nhau, nên R3 xấp xỉ 1. Khi chúng ta giận dữ hoặc khó chịu với một điều gì đó, chúng ta thường nheo mày lại và làm cho khoảng cách giữa hai chân mày thu ngắn lại, kéo theo R3 giảm.

Để chi tiết hơn, các trường hợp được miêu tả cụ thể trong bảng bên dưới (đây chỉ là một số trường hợp đại diện cho thuật toán):

|  |  |  |  |
| --- | --- | --- | --- |
| **Cảm xúc** | **R1** | **R2** | **R3** |
| Trung tính | *Bình thường* | *Bình thường* | *Bình thường* |
| Vui vẻ | **Tăng** | *Bình thường* | Giảm |
| Ngạc nhiên | *Bình thường* | **Tăng** | *Bình thường* |
| Giận dữ | **Tăng** | *Bình thường* | Giảm |

Phương pháp này có ưu điểm rất lớn khi được sử dụng trong nhận dạng cảm xúc thời gian thực vì có tốc độ xử lý nhanh, thường được dùng trong nhận dạng cảm xúc người dùng qua video hay người dùng thực tế. Ít bị ảnh hưởng của yếu tố nền khuôn mặt lên kết quả nhận dạng. Việc nhận dạng có thể được thực hiện trong điều kiện gương mặt chịu ảnh hưởng của các phép biến hình (như phép tịnh tiến, tỉ lệ, xoay).

Một hạn chế của phương pháp này là việc xác định được ngưỡng tỉ lệ để nhận định đó là cảm xúc nào.

Chương 3. CƠ SỞ LÝ THUYẾT

3.1 Principal Component Analysis

PCA - Phân tích thành phần chính – là một trong những phương pháp phân tích dữ liệu nhiều biến đơn giản nhất, là một kỹ thuật dùng để làm nổi bật sự thay đổi và đưa ra các mô hình chính yếu trong tập dữ liệu, mô hình hóa tập dữ liệu, giúp cho việc phân tích và khám phá tập dữ liệu dễ dàng và chính xác.

PCA là một kỹ thuật có tính ứng dụng cao trong việc nhận dạng gương mặt, phổ biến để phát hiện mẫu của dữ liệu nhiều chiều.

Các tính năng chính của PCA:

Giảm số chiều của dữ liệu quan sát. Được sử dụng khi các mẫu và thông tin trong dữ liệu khó nhận ra trong không gian đa chiều.

PCA xây dựng một không gian mới với trục tọa độ mới có số chiều ít hơn, nhưng có khả năng thể hiện dữ liệu tương đương hoặc tốt hơn không gian cũ, nghĩa là đảm bảo độ biến thiên của dữ liệu trên mỗi chiều mới.

Các trục tọa độ trong không gian mới là tổ hợp tuyến tính của các trục tọa độ ở không gian cũ.

Trong không gian mới, các thông tin của dữ liệu có thể được phát hiện ở một khía cạnh khác mà ở không gian của không tìm thấy được, vì các liên kết tìm ẩn của dữ liệu có thể được khai phá.

Một số khái niệm toán học quan trọng được sử dụng trong PCA: độ lệch chuẩn (standard deviation), phương sai (variance), hiệp phương sai (covariation), giá trị riêng (aigenvalue) và vector riêng (eigenvector).

3.1.1 Độ lệch chuẩn

Độ lệch chuẩn là một đại lượng dùng để đo khoảng cách giữa các phần tử trong tập dữ liệu. Độ lệch chuẩn còn được hiểu là khoảng cách trung bình từ trung bình mẫu (X ngang) đến các điểm của dữ liệu.

Giả sử ta có tập dữ liệu X = [x1, x2, ..., x­­n]. Vậy trung bình mẫu X ngang là một giá trị nằm giữa của tập X:

X ngang =

Công thức tính độ lệch chuẩn (s):

s =

3.1.2 Phương sai

Phương sai là bình phương của độ lệch chuẩn. Phương sai là một đại lượng khác dùng để biểu diễn dữ liệu: đo khoảng cách giữa các phần tử trong tập dữ liệu.

Công thức tính phương sai (s2):

s2 =

3.1.3 Hiệp phương sai

Hai khái niệm độ lệch chuẩn và phương sai được sử dụng để biểu diễn dữ liệu một chiều, nhưng dữ liệu trong thực tế có nhiều hơn một chiều và có sự liên hệ với nhau mật thiết. Do đó, đại lượng hiệp phương sai ra đời để có thể tính toán và biểu diễn được dữ liệu đa chiều.

Hiệp phương sai thực chất chỉ tính toán được sự biến thiên của hai chiều dữ liệu. Nên ta có thể tính từng cặp chiều dữ liệu cho toàn bộ chiều của tập dữ liệu.

Công thức tính hiệp phương sai:

3.1.4 Vector riêng

3.1.5 Giá trị riêng

3.1.6 Các bước thực hiện cơ bản của PCA

Bước 1: Lấy dữ liệu đầu vào. Tính trung bình mẫu, hay còn gọi là vector kỳ vọng của toàn bộ dữ liệu.

Công thức:

Bước 2: Trừ mỗi điểm dữ liệu đi một lượng vector kỳ vọng của toàn bộ dữ liệu. Xét chiều dữ liệu ở chiều x đều có một giá trị trung bình mẫu. Thực hiện trừ lần lượt các giá trị chiều x cho trung bình mẫu.

Công thức:

Bước 3: Tính ma trận hiệp phương sai.

Công thức:

Bước 4: Tính các vector riêng và giá trị riêng của ma trận hiệp phương sai. Sắp xếp kết quả theo thứ tự giảm dần của giá trị riêng.

Bước 5: Chọn ra các thành phần chính. Chọn ra K vector riêng ứng với K giá trị riêng lớn nhất để xây dựng ma trận UK có các cột tạo thành một hệ trực giao. K vector này còn được gọi là các thành phần chính, tạo thành không gian con gần với phân bố của dữ liệu ban đầu đã chuẩn hóa. Tùy vào số lượng thành phần chính yêu cầu, lấy lần lượt các thành phần (các vector riêng) tương ứng có các giá trị riêng cao nhất.

Bước 6: Chiếu dữ liệu ban đầu đã chuẩn hóa xuống không gian con tìm được. Dữ liệu mới chính là tọa độ của các điểm trong không gian mới.

3.1.7 Một số hạn chế của PCA

Chỉ làm việc với dữ liệu numeric.

Độ chính xác còn phụ thuộc nhiều vào điệu kiện chiếu sáng.

Nhạy cảm với các điểm outlier/extreme. Khi góc mặt nghiêng hay quá xa với webcam hay công cụ ghi hình có thể sẽ cho ra kết quả nhận dạng sai.

Do PCA hoàn toàn dựa trên các biến đổi tuyến tính, nên không phù hợp với các mô hình phi tuyến.

Để đạt được độ chính xác cao hơn, cần sử dụng nhiều hình ảnh để huấn luyên, kéo theo tốc độ xử lý chậm.

Không phù hợp với các ứng dụng đòi hỏi xứ lý thời gian thực.

3.2 Mô hình ASM/AAM

3.3 Đặc trưng HOG

HOG (Histogram of Oriented Gradients) là một bộ mô tả tính năng được sử dụng để phát hiện đối tượng trong thị giác máy tính và xử lý ảnh. HOG được tính toán trên một lưới dày đặc các ô và chuẩn hóa sự tương phản giữa các khối để nâng cao độ chính xác. HOG được dùng chủ yếu để phát hiện và mô tả hình dạng của một đối tượng trong ảnh. Bài toán tính toán HOG thông thường gồm năm bước chính:

Bước 1: Chuẩn hóa hình ảnh.

Bước 2: Tính gradient theo x và y.

Bước 3: Thống kê thành phần cùng trọng số trong mỗi ô.

Bước 4: Chuẩn hóa các khối.

Bước 5: Thu thập tất cả các biểu đồ cường độ gradient định hướng để tạo ra vector tính năng cuối cùng.

3.4 Open CV

OpenCV (Open Source Computer Vision Library) là một thư viện nguồn mở của thị giác máy tính, xử lý ảnh và máy học, gồm các hàm chức năng tính toán được tích hợp sẵn, được phân phối dưới giấy phép BSD. OpenCV được thiết kế để tính toán hiệu quả và tập trung chủ yếu vào các ứng dụng thời gian thực.

OpenCV được sử dụng rộng rãi trong các ứng dụng như: kiểm tra và giám sát tự động, robot và xe hơi tự hành, phân tích hình ảnh y tế, tìm kiếm và phục hồi hình ảnh hay video, thực tế ảo, và nhiều ứng dụng khác.

OpenCV được bắt đầu phát triển tại Intel vào năm 1999 bởi Gary Bradsky và phát hành phiên bản đầu tiên năm 2000. OpenCV hỗ trợ đa dạng ngôn ngữ lập trình như C++, Python, Java, v.v. và đa nền tảng bao gồm Windows, Linux, OS X, Android và iOS. (2).

3.5 Máy học SVM

3.5.1. Giới thiệu SVM

SVM (Support Vector Machines) là một máy học vector hỗ trợ, được Vapnik nghiên cứu từ những năm 1965, đến những năm 1990 thì giải thuật chính thức được phát triển mạnh, trở thành công cụ hữu hiệu và phổ biến của lĩnh vực máy học, nhận dạng và khai phá dữ liệu. SVM đã được áp dụng thành công trong rất nhiều lĩnh vực như nhận dạng gương mặt người, phân loại văn bản, phân loại bệnh, … Bằng việc kết hợp với phương pháp hàm nhân, SVM cung cấp các mô hình hiệu quả và chính xác cho các vấn đề phân lớp, hồi quy tuyến tính và phi tuyến trong thực tế. Giải thuật SVM nhận đầu vào là một hàm nhân (kernel function) sẽ tạo ra một mô hình mới mà không cần đến bất kỳ sự thay đổi nào từ mã chương trình. Giải thuật học dẫn đến việc giải bài toán quy hoạch toàn phương, luôn có kết quả tối ưu toàn cục. SVM là một trong những giải thuật quan trọng của khai mỏ dữ liệu. (3)

Trong kỹ thuật SVM, không gian dữ liệu nhận vào ban đầu sẽ được ánh xạ và không gian đặc trưng, và trong không gian đặc trưng này thì mặt siêu phẳng phân chia dữ liệu tối ưu sẽ được xác định.

3.5.2. Giới thiệu về phân lớp dữ liệu

Phân lớp dữ liệu là một kỹ thuật quan trọng trong khai phá dữ liệu và được sử dụng rộng rãi nhất bên cạnh kỹ thuật hồi quy.

Mục đích: Để dự đoán và gán nhãn phân lớp cho các bộ dữ liệu mới hoặc mẫu mới. Với đầu vào là một tập các mẫu dữ liệu huấn luyện có riêng từng nhãn phân lớp và đầu ra là một bộ phân lớp dựa trên tập huấn luyện hoặc nhãn phân lớp. Phân lowpsd dữ liệu dựa trên tập huấn luyện và các gái trị trong một thuộc tính phân lớp và dùng nó để xác định lớp cho dữ liệu mới. (4 p. 8)

Kỹ thuật phần lớp dữ liệu gồm hai bước cơ bản:

Bước 1: Xây dựng mô hình tự tập huấn luyện

Dữ liệu huấn luyện

Bộ phân lớp (Mô hình)

Các thuật toán phân lớp

Bước 2: Kiểm tra tính đúng đắn của mô hình và sử dụng mô hình để phân lớp dữ liệu mới. (4)

Phân lớp cho những đối tượng mới hoặc chưa được phân lớp.

Đánh giá độ chính xác của mô hình: Lớp đã biết của một mẫu dữ liệu đem kiểm tra được so sánh với kết quả thu được từ mô hình. Tỉ lệ chính xác được tính bằng phần trăm các mẫu dữ liệu được phân lớp đúng bởi mô hình trong số các lần kiểm tra. (4 p. 9)

Bộ phân lớp (Mô hình)

Dữ liệu kiểm tra

Dữ liệu chưa phân lớp

3.5.3. Sử dụng SVM trong phân lớp dữ liệu

SVM rất hiệu quả trong việc giải quyết các bài toán với dữ liệu có số chiều quan sát lớn, như ảnh của dữ liệu gen, tế báo, ADN.

SVM giải quyết vấn đề overfitting rất tốt (dữ liệu có nhiễu và tách rời nhóm, hoặc số lượng dữ liệu huấn luyện quá ít).

Tốc độ phân lớp nhanh, hiệu suất tổng hợp tốt và hiệu năng tính toán cao. (4 p. 9)

3.5.4 Ứng dụng SVM vào đề tài

SVM được sử dụng để phân lớp cảm xúc, phân ra sáu loại cảm xúc riêng biệt với từng tiêu chí khác nhau, để tạo thành một tập huấn luyện. Sau đó, dự vào tập huấn luyện này để nhận dạng hình ảnh mới có cảm xúc gì.

3.6 Mạng nơ-ron nhân tạo

Mạng nơ-ron nhân tạo (ANNs – Artificial Neural Networks) là một họ phương pháp tính toán tổng quát mô hình hóa hoạt động của hệ thần kinh con người. Là một mạng phức tạp kết nối các đơn vị tính toán lại với nhau, trong đó mỗi đơn vị tính toán là một nơ-ron nhân tạo, có thể có nhiều đầu vào, như chỉ có một đầu ra duy nhất cuối cùng.

Mạng nơ-ron nhân tạo là một giải thuật học có giám sát. Mạng nơ-ron nhân tạo là một mô hình tính toán được xây dựng mô phỏng theo mạng nơ-ron sinh học, bao gồm một nhóm các nơ-ron nhân tạo (nút) nối với nhau, và thông tin được xử lý bằng cách truyền theo các kết nối và tính giá trị mới tại các nút. Mạng nơ-ron nhân tọa được ứng dụng nhiều trong lĩnh vực nhận dạng, phân lớp, điều khiển và dự báo. (3 p. 19)

Mạng nơ-ron nhân tạo là một trong những kỹ thuật xử lý dữ liệu hiện đại, cho phép lấy được lượng thông tin tối đa từ dữ liệu như: nhận dạng, phân loại, dự báo, xây dựng mô hình, nghiên cứu về suy nghĩ của con người và cách để tạo ra trí thông minh nhân tạo. (3 p. 20)

Do giải thuật đơn giản, mạng nơ-ron nhân tạo được cài đặt khá dễ dàng trên hệ thống nhúng. Tuy nhiên, quá trình huấn luyện mạng nơ-ron tốn nhiều thời gian, do phải huấn luyện nhiều lần vì kế quả thu được chỉ là tối ưu cục bộ. Một khó khăn khác nữa là mạng nơ-ron chỉ làm việc với dữ liệu số, vì thế cần phải có một bước tiền xử lý nếu dữ liệu không phải là số. Ngoài ra, kết quả của mạng nơ-ron không dễ hiểu chút nào, rất khó để giải thích kết quả của mạng nơ-ron với dữ liệu đầu vào được cho. Dù đã có nhiều nghiên cứu về kiến trúc mạng nơ-ron, nhưng vấn đề thiết kế mạng nơ-ron để phù hợp với từng ứng dụng là một chủ đề được đang quan tâm và nghiên cứu. (3 p. 47)

CHƯƠNG 4. QUY TRÌNH THỰC HIỆN

4.1 Sơ đồ tổng quan các thành phần chính của hệ thống nhận dạng biểu cảm gương mặt

(Vẽ sơ đồ tổng quan của hệ thống, đây là sơ đồ tạm thời, chưa chính xác lắm)

Trích chọn đặc trưng HOG

Phát hiện vùng mặt

Phân đoạn ảnh

Ảnh đầu vào

Hiển thị kết quả

Phân lớp, nhận dạng

4.2 Các nghiên cứu liên quan

4.3 Định hướng giải quyết của luận văn

Trong thực tế, để dự đoán được cảm xúc hiện tại của một người, đầu tiên chúng ta cần nhìn vào gượng mặt của người đó. Cụ thể hơn, chúng ta quan sát các thành phần chính trên gương mặt như miệng, mắt, má, chân mày,…. Trong quá trình này, bộ não đang thu thập dữ liệu để nhận dạng cảm xúc. Và như một cách tự nhiên, hay đã được học ở một nơi nào đó, chúng ta biết được với các biểu hiện như thế nào của các thành phần trên gương mặt sẽ thể hiện cảm xúc gì tương ứng.

Quá trình nhận dạng cảm xúc gương mặt trên máy tính cũng được thực hiện tương tự như những gì diễn ra trong bộ não, gồm ba bước chính:

**Bước 1**: Thu thập dữ liệu. Dữ liệu là hình ảnh của các gương mặt người với những biểu cảm cần nhận dạng.

**Bước 2**: Huấn luyện bộ nhận dạng. Thực hiện công việc dạy cho máy tính biết với những biểu hiện như thế nào của các thành phần chính trên gương mặt thì sẽ tương ứng với những cảm xúc gì.

**Bước 3**: Nhận dạng cảm xúc gương mặt: Kiểm tra sự chính xác của bộ nhận dạng vừa huấn luyện bằng việc tiến hành nhận dạng cảm xúc của những dữ liệu mới.

4.4 Quy trình thực hiện luận văn

\*\*\* Dữ liệu huấn luyện:

- Dữ liệu được dùng để huấn luyện là tập dữ liệu Kohn-Kanade, JAFFE (Japanese Femal Facial Expression)

+ **Cohn-Kanade** (CK và CK+): Là cơ sở dữ liệu biểu cảm gương mặt dựa trên các đơn vị vận động gương mặt, được dùng cho nghiên cứu trong tự động phân tích và tổng hợp hình ảnh gương mặt, và cho các lĩnh vực học về cảm xúc. Cohn-Kanade hiện tại có hai phiên bản:

*Phiên bản 1:* **CK**. Là phiên bản đầu tiên, bao gồm 486 chuỗi hình ảnh cảm xúc gương mặt của 97 đối tượng. Mỗi chuỗi hình ảnh cảm xúc bắt đầu với một cảm xúc trung tính và tăng dần để đạt đến cảm xúc mạnh mẽ nhất. Từng hình ảnh cảm xúc trong mỗi chuỗi được mã hóa và gán nhãn. (5)

*Phiên bản 2:* **CK+**. Chứa 593 chuỗi hình ảnh (327 chuỗi có các nhãn cảm xúc riêng biệt), mỗi hình ảnh thể hiện một trong tám loại cảm xúc: vui vẻ, buồn bã, ngạc nhiên, giận dữ, sợ hãi, kinh tởm, trung tính và kinh thường. Các hình ảnh hầu hết là ảnh mức xám với độ phân giải quy định là 640x490 pixels.

+ **JAFFE**: Kho dữ liệu chứa 213 hình ảnh của 7 biểu cảm gương mặt (vui vẻ, buồn bã, ngạc nhiên, giận dữ, sợ hãi, kinh tởm và trung tính) được thể hiện bởi 10 người mẫu nữ đến từ Nhật Bản. Mỗi hình ảnh được đánh giá trên 6 tính từ cảm xúc bởi 60 người Nhật. (6)

\*\*\* Môi trường và công cụ thực nghiệm:

Ngôn ngữ lập trình: Python

Thư viện tích hợp: OpenCV, Numpy, Matplotlib

Hệ điều hành: Windows

\*\*\* Các mô-đun cần thiết:

**- cv2**: mô-đun OpenCV sử dụng cho ngôn ngữ lập trình Python, được dùng để phát hiện và nhận dạng gương mặt. Để sử dụng mô-đun cv2 trong Python, ta thực hiện chèn cv2 vào mã chương trình như một thư viện bằng dòng lệnh: “*import cv2*”.

**- os**: là một mô-đun của Python, cung cấp một cách linh hoạt để sử dụng các chức năng phục thuộc vào hệ điều hành. Ví dụ như mô-đun os.path được sử dụng để thao tác với các đường dẫn hệ thống, như là đường dẫn trỏ tới file hình ảnh đầu vào. Tôi sử dụng mô-đun này để đọc thư mục huấn luyện và tên file. Để sử dụng, chèn dòng lệnh “*import os*” và mã chương trình.

**- numpy**: là một mô-đun tuyệt vời của Python, hỗ trợ cho việc tính toán trên các ma trận một cách dễ dàng. Nó có các hàm mạnh mẽ được tích hợp để xử lý các mảng nhiều chiều, chuyển đổi các danh sách trong Python thành các mảng nunpy phục vụ cho việc tính toán và nhận dạng. Sử dụng dòng lệnh dưới đây: “*import numpy as np*”.

4.1 Chuẩn bị dữ liệu huấn luyện

- Tiền đề được đặt ra rất đơn giản là càng nhiều hỉnh ảnh được sử dụng để huấn luyện thì hiệu quả tập huấn luyện càng tốt.

- Thực hiện tổ chức thư mực như sau: có một thư mục cha tên là “training-data”. Thư mục này chứa sáu thư mục con tương ứng là sáu cảm xúc của con người: vui vẻ, buồn bã, ngạc nhiên, giận dữ, sợ hãi, kinh tởm. Mỗi thư mục con sẽ chứa hình ảnh gương mặt của con người với biểu lộ cảm xúc tương ứng với tên thư mục con. Các hình ảnh được đánh số từ 1 cho đến số lượng hình ảnh có trong thư mục. Được mô tả cụ thể như hình bên dưới:

training-data

|----------------happiness

| |------ 1.jpg

| |------ 2.jpg

| |------ …

| |------ n.jpg

|----------------sadness

| |------ 1.jpg

| |------ 2.jpg

| |------ …

| |------ n.jpg

|----------------surprise

| |------ 1.jpg

| |------ 2.jpg

| |------ …

| |------ n.jpg

|----------------anger

| |------ 1.jpg

| |------ 2.jpg

| |------ …

| |------ n.jpg

|----------------fear

| |------ 1.jpg

| |------ 2.jpg

| |------ …

| |------ n.jpg

|----------------disgust

| |------ 1.jpg

| |------ 2.jpg

| |------ …

| |------ n.jpg

- Tiếp theo, chúng ta cần có một thư mục “test-data” chứa các hình ảnh sẽ được sử dụng để kiểm tra độ chính xác của chương trình nhận dạng cảm xúc sau khi đã huấn luyện thành công.

- Chi tiết hơn, quá trình chuẩn bị dữ liệu gồm những bước sau:

+ **Bước 1**: Đọc các thư mục con chứa ảnh và các landmarks tương ứng của nó.

+ **Bước 2**:

+ **Bước 3**: Đọc tất cả hình ảnh của gương mặt người, và áp dụng bộ nhận dạng cảm xúc với từng hình ảnh.

+ **Bước 4**:

CHƯƠNG 5. CÀI ĐẶT VÀ THỰC NGHIỆM

5.1 Yêu cầu phần cứng và phần mềm

5.2 Giao diện chương trình

5.3 Kiểm thử và kết quả

Ảnh tĩnh

Ảnh từ camera máy tính

CHƯỚNG 6. ĐÁNH GIÁ VÀ KẾT LUẬN

6.1 Đánh giá kết quả đạt được

(Kèm theo các so sánh các phương pháp)

6.2 Kết luận

6.3 Thách thức trong nhận dạng cảm xúc dựa trên mặt người

Nhận dạng cảm xúc con người dựa vào các biểu cảm trên gương mặt là một bài toán khó và mang tính tương đối, vì cảm xúc của con người là vô cùng đa dạng. Ngoài sáu cảm xúc cơ bản, con người còn có rất nhiều cảm xúc khác với sự sai khác không rõ rệt và rất khó phân biệt.

Bên cạnh đó, phương pháp nhận dạng cảm xúc dựa trên gương mặt mang tính tương đối là do chúng ta chỉ dựa vào các nhóm cơ mặt và vị trí các thành phần trên gương mặt để xác định cảm xúc, và cảm xúc này có thể chưa thật sự đúng đắn.

Một khó khăn khác về môi trường thực hiện, hệ thống sẽ cho các kết quả không tốt với độ chính xác thấp nếu ảnh đầu vào có chất lượng kém, điều kiện ánh sáng không tốt, ảnh mờ hay vùng gương mặt khó nhận dạng, kích thước quá nhỏ. (7)

6.4 Hướng phát triển

Cải tiến để có thể nhận dạng thêm được nhiều loại cảm xúc phức tạp hơn của con người.

Nghiên cứu và áp dụng hệ thống vào một ứng dụng thực tế và hữu ích.

Cải tiến độ chính xác của đề tài với nhiều dữ liệu huấn luyện hơn.

TÀI LIỆU THAM KHẢO

1. **Wikipedia.** Facial Action Coding System. *Wikipedia.* [Online] April 05, 2018. https://en.wikipedia.org/wiki/Facial\_Action\_Coding\_System.

2. **Team, OpenCV Dev.** OpenCV Documentation. *Introduction to OpenCV - Python Tutorials.* [Online] Novemver 10, 2014. [Cited: April 27, 2018.] https://docs.opencv.org/3.0-beta/doc/py\_tutorials/py\_setup/py\_intro/py\_intro.html#intro.

3. **Đỗ, Nghị Thanh and Phạm, Khang Nguyên.** *Giáo trình Nguyên Lý Máy Học.* Cần Thơ : Đại học Cần Thơ, 2012.

4. *Tìm hiểu về Support Vector Machine cho bài toán phân lớp quan điểm.* **Phạm, Sơn Văn.** Hải Phòng : s.n., 2012.

5. **Jeffrey, Cohn.** Affect Analysis Group. *Cohn-Kanade AU-Coded Expression Database.* [Online] Facial expression database. [Cited: 5 23, 2018.]

6. **Michael, Lyons, Miyuki , Kamachi and Jiro, Gyoba.** The Japnese Female Facial Expression (JAFFE) Database. [Online] Psychology Department, Kyushu University. [Cited: 5 23, 2018.] http://www.kasrl.org/jaffe.html.

7. *Nghiên cứu nhận dạng biểu cảm mặt người trong tương tác người máy.* **Nguyễn, Vân Thị Thanh.** Hải phòng : s.n., 2016.