**TỔNG QUAN**

Hiện nay, việc nhận diện cảm xúc thể hiện trên khuôn mặt con người bằng máy tính đang là 1 thử thách rất thú vị. Chúng tôi đưa ra 1 giải pháp mới về việc sử dụng các landmark của khuôn mặt để nhận diện cảm xúc thông qua những thể hiện trên khuôn mặt của con người. Khi ta phân tích được những biểu hiện cảm xúc trên khuôn mặt, điều đó sẽ rất hữu ích trong việc giao tiếp giữa người với người. Sau khi lần lượt thử nghiệm các phương pháp học như SVM, cây định danh, random forest chúng tôi đã tìm hiểu và nghiên cứu để đưa ra 1 tập các thuộc tính cho kết quả nhận diện với độ chính xác cao. Bên cạnh đó chúng tôi cũng sẽ thể hiện hướng giải quyết thông qua kết quả thu được từ những thí nghiệm cùng với 1 bản demo nhận diện cảm xúc thông qua stream webcam trực tiếp.

**Từ khóa:** machine learning, image processing, facial landmarks, facial emotions, SVM, decision tree, random forest

**GIỚI THIỆU**

Trong chương này, chúng ta sẽ bàn về bài toán nhận diện cảm xúc của con người thông qua biểu hiện trên khuôn mặt. Đầu tiên ta sẽ nói về động lực nghiên cứu của bài toán trên đồng thời tổng quát hóa bài toán. Toàn bộ đề cương của luận án nằm ở phần cuối chương.

1. *Động lực*

Khi con người giao tiếp với nhau, bên cạnh việc truyền đạt thông tin bằng miệng, các hình thức giao tiếp phi ngôn ngữ như thông qua cử chỉ, biểu hiện nét mặt mà con người có thể biểu hiện cảm xúc về những vấn đề trong việc giao tiếp. Với sự phát triển không ngừng của trí tuệ nhân tạo thì nhu cầu tương tác với máy tính để có thể đạt đến mức giao tiếp như giữa người với người là vô cùng thiết yếu. Một trong những cách thức để con người thể hiện cảm xúc rõ nhất chính là thông qua những biểu hiện thể hiện trên khuôn mặt. Vì vậy chúng tôi muốn xây dựng 1 hệ thống có thể nhận diện được cảm xúc của con người thông qua nét mặt với độ chính xác cao nhằm có thể ứng dụng về sau.

1. *Những đề tài liên quan*

Từ năm 1970 thì đã bắt đầu có những nghiên cứu về biểu cảm khuôn mặt của con người. Nhóm nghiên cứu của P. Ekman [1] đã tìm thấy những điểm chung trên nét mặt. Ông cho rằng ta có thể chia ra làm 6 loại gồm vui, buồn, giận dữ, sợ hãi, ngạc nhiên và thù ghét. Nghiên cứu của ông dựa trên nguồn dữ liệu đa dạng trên nhiều hình thức. Bên cạnh đó ông cũng đã thiết lập 1 hệ thống cho những biểu hiện trên khuôn mặt thông qua những chuyển động được tính theo số action units (AUs). Đó là 1 trong những điểm nhấn tạo cảm hứng cho những người đến sau. Qua những nguồn thu như phim ảnh, họ trích xuất những thuộc tính hữu ích từ khuôn mặt người từ đó phân loại thành các nhóm biểu hiện cảm xúc. Nhóm nghiên cứu của Aitor Azcarate [2] sử dụng Piecewise Bézier Volume Deformation (PBVD) tracker [3] được thiết kế bởi Tao và Huang đã được nâng cấp với bộ nhận diện khuôn mặt Haar để có thể tự động nhận diện khuôn mặt, từ đó sử dụng Naive Bayes và Tree-Augmented-Naive Bayes (TAN) để nhận diện khuôn mặt.

Riêng nhóm chúng tôi quyết định sẽ sử dụng hệ thống của V. Kazemi [4] để nhận diện khuôn mặt kèm landmark vì độ ưu việt của nó về thời gian xử lý. Từ đó nhóm sẽ nghiên cứu các landmark và đưa ra 1 tập các thuộc tính và tiến hành train và lọc ra các thuộc tính có độ tin cậy cao để từ đó rút ra được tập thuộc tính tối ưu để có thể nhận diện được cảm xúc thể hiện trên khuôn mặt 1 cách hiệu quả.

1. *Đề cương*

Ở chương sau ta sẽ đề cập về lý thuyết của những công nghệ mà ta sẽ sử dụng để giải quyết bài toán trên, sau đó là giải thích hướng giải quyết ở chương 4. Sau khi thiết kế và xây dựng hệ thống ta sẽ đánh giá kết quả của nó thông qua các thí nghiệm nhằm kiểm chứng độ chính xác của hướng tiếp cận mà ta đề cập đến và đi đến kết luận ở chương 5.

**SƠ LƯỢC LÝ THUYẾT**

1. *Facial landmark*
2. *SVM*
3. *Cây định danh*
4. *Random forest*

**HƯỚNG GIẢI QUYẾT VẤN ĐỀ**

1. *Thu thập dữ liệu*

Do lớp bài toán này xuất hiện cũng đã khá lâu nên lượng dữ liệu có sẵn cũng khá lớn và vô cùng đa dạng nhưng đa phần chúng đều được thu trong những điều kiện nghiêm ngặt. Nhằm hướng đến việc ứng dụng trong đời sống, phần lớn nguồn dữ liệu nhóm thu thập được là từ cuộc thi 300-W In The Wild [5][6][7] bao gồm 4 tập dữ liệu: HELEN, AFW, IBUG, LFPW được đánh dấu sẵn vị trí của 68 điểm landmark vì độ tự nhiên của chúng.

**LFPW:** Tập Labeled Face Parts in-the-wild (LFPW) chứa 1287 ảnh lấy từ google.com, flickr.com, yahoo.com. Những tấm ảnh này được chụp với nhiều tư thế, biểu cảm trong nhiều điều kiện ánh sáng và khung cảnh khác nhau.

**HELEN:** Tập Helen chứa 2330 ảnh thu thập từ trang flickr.com. Tập này chứa nhiều ảnh có độ phân giải cao, đôi khi lớn hơn 500x500 pixel.

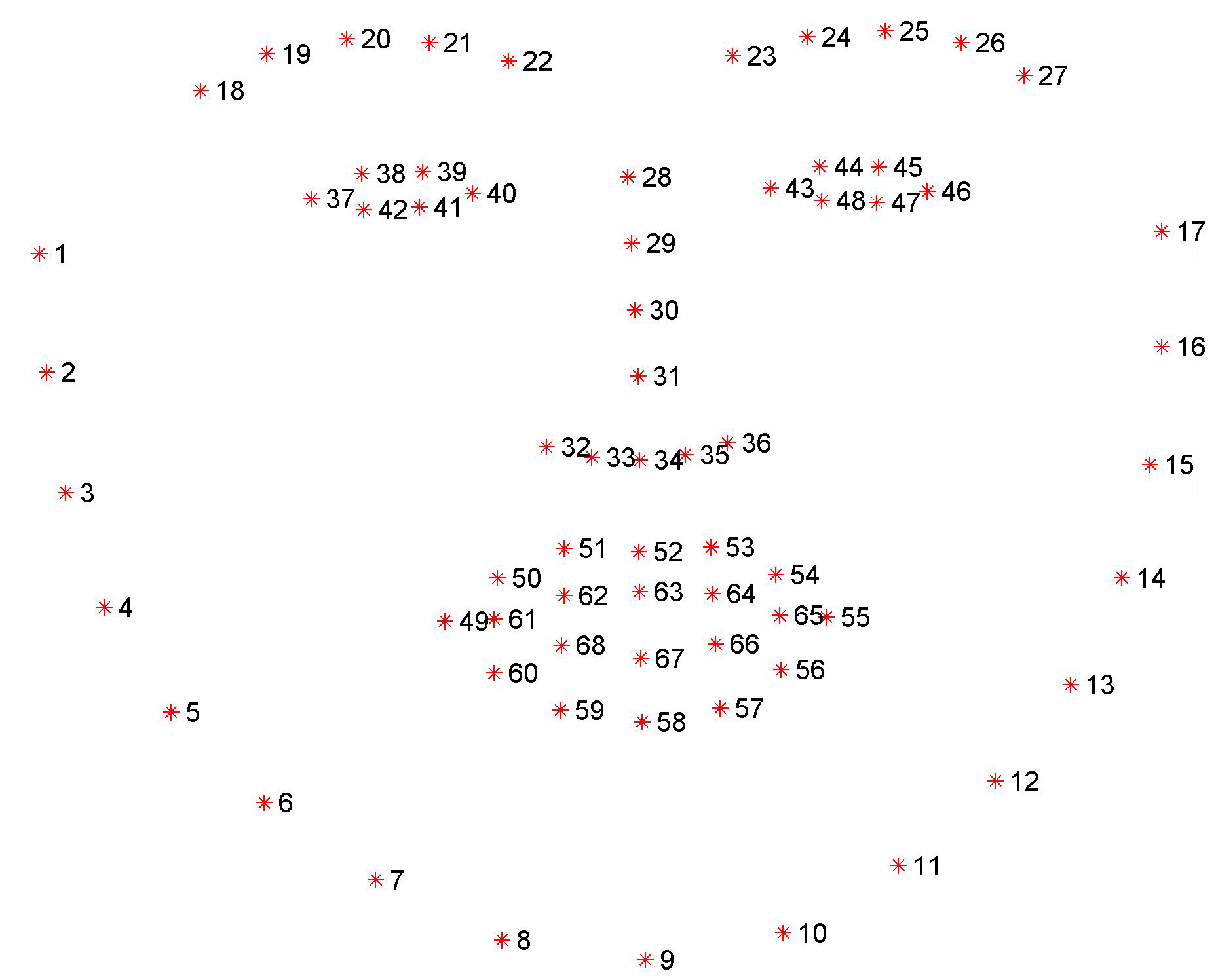
**AFW:** Tập Annotated Faces in-the-wild (AFW) chứa 250 ảnh với 468 khuôn mặt. Điểm đặc biệt của tập này là mỗi ảnh chứa dến 6 khuôn mặt.

**IBUG:** Tập IBUG được nhóm Intelligent Behaviour Understanding Group phát hành bao gồm 135 ảnh có độ phân giải cao.

Bên cạnh đó nhóm cũng đã tự đi thu thập thêm khoảng 600 ảnh frontal face được chia ra thành 3 loại cảm xúc cơ bản: positive, neutral và negative. Tổng cộng có khoảng 4000 ảnh được thu thập để sử dụng trong nghiên cứu này.

1. *Tiền xử lý dữ liệu*

Với những ảnh chưa có file đánh dấu vị trí của các landmark, ta áp dụng thuật toán nhận diện landmark của khuôn mặt thu được file đánh dấu vị trí landmark có dạng như sau



Hình 1. 68 landmark trên khuôn mặt dùng trong file đánh dấu

Sau khi có được file đánh dấu, ta xác nhận lại bằng cách vẽ lại các điểm từ file lên ảnh gốc. Để tiện cho việc nhận diện cảm xúc khuôn mặt, ta tiến hành xoay ảnh sao cho 2 điểm 32 và 36 song song với trục hoành rồi crop lấy khuôn mặt và resize lại ảnh với độ rộng 100 pixel. Ta tính toán lại vị trí của các điểm landmark và lưu lại thành file đánh dấu mới kèm với ảnh đã được chỉnh sửa.



Hình 2. Ảnh trước và sau khi được chỉnh sửa

Với tập ảnh thu được sau khi đã chỉnh sửa, nhóm tiến hành đánh nhãn cảm xúc cho ảnh với 3 loại cảm xúc cơ bản là positive, neutral và negative.

1. *Xây dựng training model*



Hình 3. Quá trình xây dựng training model

Từ file đánh dấu 68 landmark thu được từ ảnh đã qua xử lý, chúng tôi đưa nó vào module tính toán để tạo ra các thuộc mới và lưu lại thành file .csv. Sau cùng ta đưa file vào module SVM để học và lưu lại thành file model để có thể tái sử dụng trong app chính.

1. *Thiết kế app*



Hình 4. Mô hình của app

Từ webcam của laptop ta trích xuất từng frame rồi đem đi xác định vị trí các landmark, từ đó ta đi tính giá trị của các thuộc tính thu được từ vị trí các landmark. Với model đã train sẵn từ trước ta chỉ cần load lại vào module chính của app rồi predict để thu ra được kết quả.

**REFERENCE**

[1] P. Ekman. **Strong evidence for universals in facial expressions**. Psychol. Bull., 115(2): 268–287, 1994.

[2] Azcarate, Aitor, et al. **Automatic facial emotion recognition**. Universiteit van Amsterdam (2005).

[3] H. Tao, T.S. Huang. **Connected vibrations: a modal analysis approach to non-rigid motion tracking**. Proc. IEEE Conf. on CVPR, 735–740, 1998.

[4] Kazemi, Vahid and Sullivan, Josephine. **One Millisecond Face Alignment with an Ensemble of Regression Trees**. In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)

[5] C. Sagonas, E. Antonakos, G, Tzimiropoulos, S. Zafeiriou, M. Pantic. **300 faces In-the-wild challenge: Database and results**. Image and Vision Computing (IMAVIS), Special Issue on Facial Landmark Localisation "In-The-Wild". 2016

[6] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. **A semi-automatic methodology for facial landmark annotation**. Proceedings of IEEE Int’l Conf. Computer Vision and Pattern Recognition (CVPR-W), 5th Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2013). Oregon, USA, June 2013

[7] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. **300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge**. Proceedings of IEEE Int’l Conf. on Computer Vision (ICCV-W), 300 Faces in-the-Wild Challenge (300-W). Sydney, Australia, December 2013