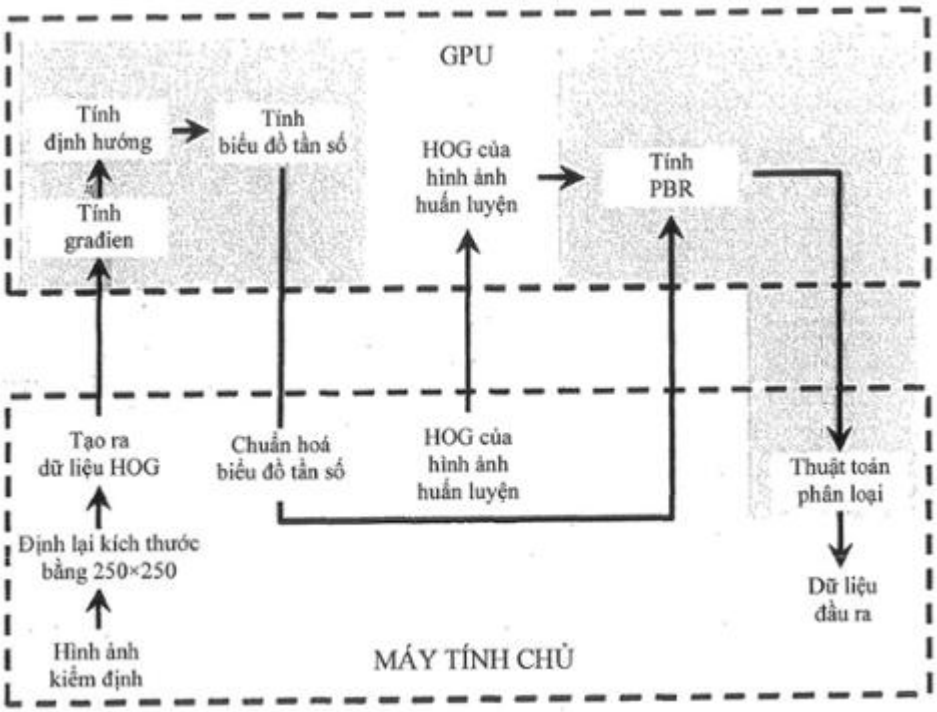
(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ
(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ
(11) **1-0032650**
(51)[7] **G06T 7/00**; G06F 19/24; G06K 9/78 (13) **B**

(21) 1-2017-01315 (22) 15/09/2015
(86) PCT/SG2015/050317 15/09/2015 (87) WO2016/043659 24/03/2016
(30) 62/050,414 15/09/2014 US
(45) 25/07/2022 412 (43) 26/06/2017 351A
(73) Temasek Life Sciences Laboratory Limited (SG)
1 Research Link, National University of Singapore, Singapore 117604, Singapore
(72) SWAMINATHAN, Muthukaruppan (SG); SJÖBLOM, Tobias (SE); CHEONG, Ian (SG); PILOTO, Obdulio (US).
(74) Công ty TNHH Tầm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) PHƯƠNG PHÁP VÀ HỆ THỐNG PHÂN LOẠI HÌNH ẢNH KỸ THUẬT SỐ

(57) Sáng chế đề cập đến hệ thống và phương pháp cải tiến để phân loại hình ảnh kỹ thuật số. Máy tính chủ có bộ xử lý được kết nối với bộ nhớ trên đó lưu trữ dữ liệu đặc trưng tham chiếu. Bộ xử lý đồ hoạ (*GPU: Graphics Processing Unit*) có bộ xử lý được kết nối với máy tính chủ và được tạo cấu hình để thu nhận, từ máy tính chủ, dữ liệu đặc trưng tương ứng với hình ảnh kỹ thuật số; để truy nhập, từ bộ nhớ, một hoặc nhiều dữ liệu đặc trưng tham chiếu; và xác định khoảng cách bán-mêtric dựa vào phân phối nhị thức-Poisson giữa dữ liệu đặc trưng thu được và một hoặc nhiều dữ liệu đặc trưng tham chiếu. Máy tính chủ được tạo cấu hình để phân loại hình ảnh kỹ thuật số bằng cách sử dụng khoảng cách bán-mêtric đã xác định.



Lĩnh vực kỹ thuật được đề cập

Nói chung, sáng chế đề cập đến hệ thống và phương pháp cải tiến để nhận dạng hình ảnh. Cụ thể hơn, sáng chế đề cập đến hệ thống và phương pháp nhận dạng mẫu trong hình ảnh kỹ thuật số. Cụ thể hơn nữa, sáng chế đề cập đến hệ thống và phương pháp thực hiện các chức năng phân loại và nhận dạng hình ảnh sử dụng độ đo khoảng cách bán mêtric mới được gọi là bán kính nhị thức-Poisson (*PBR: Poisson-Binomial Radius*) dựa trên phân phối nhị thức-Poisson.

Tình trạng kỹ thuật của sáng chế

Các phương pháp tự học của máy như phương pháp máy vectơ hỗ trợ (*SVM: Support Vector Machines*), phương pháp phân tích thành phần chính (*PCA: Principal Component Analysis*) và phương pháp k-phần tử lân cận gần nhất (*k-NN: k-Nearest Neighbors*) sử dụng các đại lượng đo khoảng cách để so sánh tính không đồng dạng tương đối giữa các điểm dữ liệu. Việc chọn đại lượng đo khoảng cách phù hợp là điều quan trọng chủ yếu. Các đại lượng đo khoảng cách được sử dụng rộng rãi nhất là tổng của các khoảng cách bình phương (khoảng cách L_2 hay khoảng cách Euclid) và tổng của các hiệu số tuyệt đối (khoảng cách L_1 hay khoảng cách Manhattan).

Câu hỏi đặt ra là sẽ sử dụng khoảng cách nào có thể được trả lời dựa vào quan điểm hợp lý cực đại (*ML: Maximum Likelihood*). Nói ngắn gọn là, khoảng cách L_2 được sử dụng cho các dữ liệu tuân theo phân phối Gauss độc lập và có cùng phân phối (*i.i.d.: independent and identically distributed*), trong khi khoảng cách L_1 được sử dụng cho các dữ liệu được phân bố Laplace i.i.d.. Xem các tài liệu [1], [2]. Do đó, khi phân phối dữ liệu liên quan là đã biết hoặc đã được đánh giá tốt, thì có thể xác định mêtric sẽ được sử dụng.

Vấn đề phát sinh khi phân phối xác suất cho các biến đầu vào là chưa biết hoặc không cùng loại. Lấy ví dụ về trường hợp thu nhận hình ảnh, hình ảnh được chụp bằng các camera kỹ thuật số hiện đại luôn bị nhiễu. Xem tài liệu [3]. Ví dụ, dữ liệu đầu ra của bộ cảm biến loại thiết bị ghép điện tích (*CCD: Charge-Coupled Device*) chứa nhiều thành phần nhiễu như nhiễu photon, nhiễu theo mẫu cố định (*FPN: Fixed-Pattern Noise*)

đi kèm với tín hiệu hữu ích. Xem tài liệu [4]. Ngoài ra, các hình ảnh có xu hướng bị nhiễu trong lúc khuếch đại và truyền tín hiệu. Xem tài liệu [5]. Một số loại nhiễu phổ biến nhất được nêu trong các tài liệu là nhiễu cộng, nhiễu xung hoặc nhiễu phụ thuộc vào tín hiệu. Tuy nhiên, loại nhiễu và mức nhiễu do các camera kỹ thuật số hiện đại tạo ra có xu hướng phụ thuộc vào các thông tin chi tiết cụ thể như tên nhãn hiệu và loại camera, ngoài các thông số cài đặt camera (khẩu độ, tốc độ cửa sập, chỉ số độ nhạy sáng ISO). Xem tài liệu [6]. Ngoài ra, chức năng chuyển đổi định dạng tệp hình ảnh và truyền tệp gây ra sự tổn hao của siêu dữ liệu có thể góp phần thêm vào vấn đề này. Thậm chí ngay cả ảnh chụp dường như không có nhiễu, thì ảnh đó vẫn có thể có các thành phần nhiễu không nhìn thấy rõ bằng mắt người. Xem tài liệu [7]. Vì các bộ mô tả đặc trưng phải chịu các nguồn nhiễu bất đồng nhất như vậy, do đó sẽ hợp lý khi giả định rằng các bộ mô tả đó có tính độc lập nhưng không cùng phân phối (*i.n.i.d.: independent but non-identically distributed*). Xem tài liệu [8].

Trong hầu hết các đại lượng đo khoảng cách luôn có giả định rằng các biến đầu vào có tính độc lập và có cùng phân phối (i.i.d.). Sự tiến bộ gần đây trong lĩnh vực phân tích dữ liệu giải trình tự sinh học và các lĩnh vực khác đã chứng minh rằng trong thực tế, dữ liệu đầu vào thường không tuân theo giả định i.i.d.. Lý do giải thích cho sự bất đồng này đã được nêu lên nhằm tìm ra các thuật toán dựa trên quyết định chính xác hơn.

Một vài xu hướng đã góp phần vào sự phát triển của các đại lượng đo khoảng cách bán mêtric. Xu hướng thứ nhất liên quan đến các tiền đề mà các đại lượng đo khoảng cách cần phải thoả mãn thì mới đủ điều kiện để trở thành các đại lượng đo khoảng cách. Đó là các tiền đề về tính không âm, tính đối xứng, tính phản xạ và bất đẳng thức tam giác. Các đại lượng đo không thoả mãn tiền đề bất đẳng thức tam giác theo định nghĩa được gọi là các khoảng cách bán mêtric.

Mặc dù các khoảng cách mêtric được sử dụng rộng rãi trong hầu hết các ứng dụng, nhưng có những lý do hợp lý để nghi ngờ về sự cần thiết của một số tiền đề, đặc biệt là tiền đề bất đẳng thức tam giác. Ví dụ, đã nhận thấy rằng tiền đề bất đẳng thức tam giác bị vi phạm ở mức có ý nghĩa thống kê khi các đối tượng là con người được yêu cầu phải thực hiện nhiệm vụ nhận dạng hình ảnh. Xem tài liệu [9]. Ví dụ khác, các điểm số khoảng cách được tạo ra bởi các thuật toán có hiệu quả tốt nhất để nhận dạng hình ảnh sử

dụng các tập dữ liệu Labelled Faces in the Wild (LFW) và Caltech101 cũng cho thấy là có vi phạm tiền đề bất đẳng thức tam giác. Xem tài liệu [10].

Một xu hướng khác liên quan đến không gian có số chiều quá lớn so với dữ liệu "curse of dimensionality". Khi số chiều của không gian đặc trưng tăng lên, thì tỷ số giữa các khoảng cách của phần tử lân cận gần nhất và phần tử lân cận xa nhất đến mẫu truy vấn cho trước bất kỳ có xu hướng hội tụ đến phần tử đơn vị với hầu hết các phân bố dữ liệu hợp lý và các hàm khoảng cách. Xem tài liệu [11]. Độ tương phản kém giữa các điểm dữ liệu có nghĩa rằng việc tìm kiếm phần tử lân cận gần nhất trong không gian có số chiều lớn sẽ là vô nghĩa. Do đó, khoảng cách bán mêtric L_p dạng phân số [12] được tạo ra như là một phương thức để bảo toàn độ tương phản. Nếu (x_i,y_i) là dãy vectơ ngẫu nhiên i.i.d., thì khoảng cách L_p được định nghĩa như sau:

$$L_p(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \quad (1)$$

Nếu lấy $p = 1$ thì thu được khoảng cách Manhattan và nếu lấy $p = 2$ thì thu được khoảng cách Euclid. Với các giá trị $p \in (0, 1)$, L_p sẽ là đại lượng đo khoảng cách L_p phân số.

Khi nghiên cứu so khớp mẫu gốc trong các hình ảnh khuôn mặt và các hình ảnh tổng hợp để so sánh khoảng cách L_p và khoảng cách L_2, đã đưa đến kết luận rằng các giá trị $p \in (0{,}25, 0{,}75)$ có hiệu quả tốt hơn với khoảng cách L_2 khi các hình ảnh bị suy giảm chất lượng do nhiễu và bị ẩn đi. Xem tài liệu [13]. Các nhóm khác cũng đã sử dụng khoảng cách L_p để so khớp các hình ảnh tổng hợp và các hình ảnh thực. Xem tài liệu [14]. Ý tưởng về việc sử dụng khoảng cách L_p để tìm kiếm hình ảnh dựa vào nội dung đã được Howarth et al [15] khai thác và các kết quả cho thấy rằng $p = 0{,}5$ có thể đạt được hiệu quả tìm kiếm cao hơn và luôn có hiệu quả tốt hơn với cả hai khoảng cách L_1 và L_2 chuẩn.

Các khoảng cách bán mêtric khác đáng kể là khoảng cách hàm bộ phận động (Dynamic Partial Function: DPF) [16], khoảng cách Jeffrey Divergence (JD) [17] và khoảng cách soạn thảo chuẩn hoá (Normalized Edit Distance: NED) [18].

Đến nay, chưa có đại lượng đo khoảng cách nào được chứng minh là có thể xử lý được các phân phối i.n.i.d. khi nhận dạng mẫu. Vì vậy, cần phải có các hệ thống và

phương pháp cải tiến để nhận dạng mẫu.

Tài liệu tham khảo

Các tài liệu công bố có sẵn công khai dưới đây được viện dẫn bằng số hiệu [#] trong phần mô tả trên đây và tạo thành một phần của sáng chế này. Các nội dung phù hợp trong các tài liệu được đưa vào trong sáng chế bằng cách viện dẫn, các nội dung đó phải được hiểu đúng theo ngữ cảnh và theo cách thức viện dẫn.

[1] N. Sebe, M. S. Lew, and D. P. Huijsmans, "Toward Improved Ranking Metrics", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 10, pp. 1132-1143, 2000.

[2] W. Dong, L. Huchuan, and Y. Ming-Hsuan, "Least Soft-Threshold Squares Tracking", in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 23-28 June 2013, pp. 2371-2378.

[3] G. Healey and R. Kondepudy, "Radiometric CCD Camera Calibration and Noise Estimation", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 16, no. 3, pp. 267-276, Mar 1994.

[4] J. R. Janesick, Scientific Charge-Coupled Devices. Bellingham, WA: SPIE, 2001.

[5] C.-H. Lin, J.-S. Tsai, and C.-T. Chiu, "Switching Bilateral Filter With a Texture/Noise Detector for Universal Noise Removal", Image Processing, IEEE Transactions on, vol. 19, no. 9, pp. 2307-2320, 2010.

[6] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman, "Automatic Estimation and Removal of Noise from a Single Image", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 299-314, 2008.

[7] N. Young and A. Evans, "Spatio-Temporal Attribute Morphology Filters for Noise Reduction in Image Sequences", in Proc. International Conference on Image Processing, vol. 1, 2003, pp. I-333-6.

[8] P. H. Westfall and K. S. S. Henning, Understanding Advanced Statistical Methods. Boca Raton, FL, USA: CRC Press, 2013.

[9] A. Tversky and I. Gati, "Similarity, Separability, and the Triangle Inequality", Psychological review, vol. 89, no. 2, p. 123, 1982.

[10] W. J. Scheirer, M. J.Wilber, M. Eckmann, and T. E. Boult, "Good Recognition is Non-Metric", Computing Research Repository, vol. abs/1302.4673, 2013.

[11] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When Is "Nearest Neighbor" Meaningful?" in Database Theory ICDT99, ser. Lecture Notes in Computer Science, C. Beeriand P. Buneman, Eds. Springer Berlin Heidelberg, 1999, vol. 1540, pp. 217-235.

[12] C. Aggarwal, A. Hinneburg, and D. Keim, "On the Surprising Behavior of Distance Metrics in High Dimensional Space", in Database Theory ICDT 2001, ser. Lecture Notes in Computer Science, J. Bussche and V. Vianu, Eds. Springer Berlin Heidelberg, 2001, vol. 1973, pp. 420-434.

[13] M. Donahue, D. Geiger, R. Hummel, and T.-L. Liu, "Sparse Representations for Image Decomposition with Occlusions", in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun 1996, pp. 7-12.

[14] D. W. Jacobs, D. Weinshall, and Y. Gdalyahu, "Classification with Nonmetric Distances: Image Retrieval and Class Representation", IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 6, pp. 583-600, 2000.

[15] P. Howarth and S. Rger, "Fractional Distance Measures for Content-Based Image Retrieval", in Advances in Information Retrieval, ser. Lecture Notes in Computer Science, D. Losada and J. Fernndez-Luna, Eds. Springer Berlin Heidelberg, 2005, vol. 3408, pp. 447-456.

[16] K.-S. Goh, B. Li, and E. Chang, "DynDex: A Dynamic and Non-metric Space Indexer", in Proc. Tenth ACM International Conference on Multimedia. New York, NY, USA: ACM, 2002, pp. 466-475.

[17] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann, "Empirical Evaluation of Dissimilarity Measures for Color and Texture", Computer Vision and Image Understanding, vol. 84, no. 1, pp. 25-43, 2001.

[18] A. Marzal and E. Vidal, "Computation of Normalized Edit Distance and

Applications", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 15, no. 9, pp. 926-932, 1993.

[19] L. Le Cam, "An approximation theorem for the poisson binomial distribution", Pacific Journal of Mathematics, vol. 10(4), pp. 1181-1197, 1960.

[20] H. Shen, N. Zamboni, M. Heinonen, and J. Rousu, "Metabolite Identification through Machine Learning - Tackling CASMI Challenge Using FingerID", Metabolites, vol. 3, no. 2, pp. 484-505, 2013.

[21] A. C. W. Lai, A. N. N. Ba, and A. M. Moses, "Predicting Kinase Substrates Using Conservation of Local Motif Density", Bioinformatics, vol. 28, no. 7, pp. 962-969, 2012.

[22] A. Niida, S. Imoto, T. Shimamura, and S. Miyano, "Statistical Model-Based Testing to Evaluate the Recurrence of Genomic Aberrations", Bioinformatics, vol. 28, no. 12, pp. 1115-1120, 2012.

[23] J.-B. Cazier, C. C. Holmes, and J. Broxholme, "GREVE: Genomic Recurrent Event ViEwer to Assist the Identification of Patterns Across Individual Cancer Samples", Bioinformatics, vol. 28, no. 22, pp. 2981-2982, 2012.

[24] H. Zhou, M. E. Sehl, J. S. Sinsheimer, and K. Lange, "Association Screening of Common and Rare Genetic Variants by Penalized Regression", Bioinformatics, vol. 26, no. 19, pp. 2375-2382, 2010.

[25] A. Wilm, P. P. K. Aw, D. Bertrand, G. H. T. Yeo, S. H. Ong, C. H. Wong, C. C. Khor, R. Petric, M. L. Hibberd, and N. Nagarajan, "LoFreq: a Sequence-Quality Aware, Ultra-Sensitive Variant Caller for Uncovering Cell-Population Heterogeneity from High-Throughput Sequencing Datasets", Nucleic Acids Research, vol. 40, no. 22, pp. 11189-11201, 2012.

[26] A. S. Macdonald, Encyclopedia of Actuarial Science, J. L. Teugels and B. Sundt, Eds. John Wiley & Sons, Ltd, Chichester, 2004.

[27] H. U. Gerber, "A Proof of the Schuette-Nesbitt Formula for Dependent Events", Actuarial Research Clearing House, vol. 1, pp. 9-10, 1979.

[28] Y. Hwang, J.-S. Kim, and I.-S. Kweon, "Difference-Based Image Noise Modeling Using Skellam Distribution", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 7, pp. 1329-1341, July 2012.

[29] J. Darroch, "On the Distribution of the Number of Successes in Independent Trials", The Annals of Mathematical Statistics, vol. 35, pp. 1317-1321, 1964.

[30] J.-B. Baillon, R. Cominetti, and J. Vaisman, "A Sharp Uniform Bound for the Distribution of Sums of Bernoulli Trials", arXiv preprint arXiv:0arX.2350v4, 2013.

[31] V. N. Gudivada and V. V. Raghavan, "Content Based Image Retrieval Systems", Computer, vol. 28, no. 9, pp. 18-22, 1995.

[32] M. Arakeri and G. Ram Mohana Reddy, "An Intelligent Content-Based Image Retrieval System for Clinical Decision Support in Brain Tumor Diagnosis", International Journal of Multimedia Information Retrieval, vol. 2, no. 3, pp. 175-188, 2013.

[33] J. Kalpathy-Cramer and W. Hersh, "Automatic Image Modality Based Classification and Annotation to Improve Medical Image Retrieval", Stud Health Technol Inform, vol. 129, no. Pt 2, pp. 1334-8, 2007.

[34] B. Marshall, "Discovering Robustness Amongst CBIR Features", International Journal of Web & Semantic Technology (IJWesT), vol. 3, no. 2, pp. 19-31, April 2012.

[35] O. Boiman, E. Shechtman, and M. Irani, "In Defense of Nearest-Neighbor Based Image Classification", in Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2008, pp. 1-8.

[36] W. Zhang, J. Sun, and X. Tang, "Cat Head Detection - How to Effectively Exploit Shape and Texture Features", in Proc. Of European Conf. Computer Vision, 2008, pp. 802-816.

[37] Z. Weiwei, S. Jian, and T. Xiaoou, "From Tiger to Panda: Animal Head Detection", Image Processing, IEEE Transactions on, vol. 20, no. 6, pp. 1696-1708, 2011.

[38] T. Kozakaya, S. Ito, S. Kubota, and O. Yamaguchi, "Cat Face Detection with Two Heterogeneous Features", in Proc. IEEE International Conference on Image

Processing, 2009, pp. 1213-1216.

[39] H. Bo, "A Novel Features Design Method for Cat Head Detection", in Artificial Intelligence and Computational Intelligence, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, vol. 6319, ch. 47, pp. 397-405.

[40] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments", University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[41] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 886-893.

[42] Y. Mitani and Y. Hamamoto, "A Local Mean-Based Nonparametric Classifier", Pattern Recognition Letters, vol. 27, no. 10, pp. 1151-1159, 2006.

[43] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 4, pp. 743-761, 2012.

[44] A. Kembhavi, D. Harwood, and L. S. Davis, "Vehicle Detection Using Partial Least Squares", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 33, no. 6, pp. 1250-1265, 2011.

[45] M. Kaaniche and F. Brémond, "Recognizing Gestures by Learning Local Motion Signatures of HOG Descriptors", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 11, pp. 2247-2258, 2012.

[46] O. Ludwig, D. Delgado, V. Goncalves, and U. Nunes, "Trainable Classifier-Fusion Schemes: An Application to Pedestrian Detection", in Proc. 12th International IEEE Conference on Intelligent Transportation Systems, 2009, pp. 1-6.

[47] A. Klöckner, N. Pinto, Y. Lee, B. Catanzaro, P. Ivanov, and A. Fasih, "PyCUDA and PyOpenCL: A Scripting-Based Approach to GPU Run-Time Code Generation", Parallel Computing, vol. 38, no. 3, pp. 157-174, 2012.

[48] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing Multiple

Parameters for Support Vector Machines", Machine Learning, vol. 46, no. 1-3, pp. 131-159, 2002.

[49] F. Friedrichs and C. Igel, "Evolutionary Tuning of Multiple SVM Parameters", Neurocomputing, vol. 64, no. 0, pp. 107-117, 2005.

[50] S.-W. Lin, Z.-J. Lee, S.-C. Chen, and T.-Y. Tseng, "Parameter Determination of Support Vector Machine and Feature Selection Using Simulated Annealing Approach", Applied Soft Computing, vol. 8, no. 4, pp. 1505-1512, 2008.

[51] E. Fix and J. Hodges Jr, "Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties", USAF School of Aviation Medicine, Randolph Field, TX, Project 21-49-004, Rept. 4, Contract AF41 (128)-31, Tech. Rep., Feb. 1951.

[52] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. McLachian, A. Ng, B. Liu, P. Yu, Z.-H. Zhou, M. Steinbach, D. Hand, and D. Steinberg, "Top 10 Algorithms in Data Mining", Knowledge and Information Systems, vol. 14, no. 1, pp. 1-37, 2008.

[53] K. Fukunaga, Introduction to Statistical Pattern Recognition (2^{nd} ed.). San Diego, CA, USA: Academic Press Professional, Inc., 1990.

[54] A. K. Ghosh, "On Optimum Choice of k in Nearest Neighbor Classification", Computational Statistics & Data Analysis, vol. 50, no. 11, pp. 3113-3123, 2006.

[55] G. Toussaint, "Bibliography on Estimation of Misclassification", Information Theory, IEEE Transactions on, vol. 20, no. 4, pp. 472-479, 1974.

[56] B. Dasarathy, Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. Washington: IEEE Computer Society, 1991.

[57] V. Mnih, "CUDAMat: A CUDA-Based Matrix Class for Python", Technical Report UTML TR 2009-004, Department of Computer Science, University of Toronto, Tech. Rep., November 2009.

[58] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 1160-1165, 2003.

[59] M. Burge and W. Burger, "Ear Biometrics in Computer Vision", in Proc. 15th International Conference on Pattern Recognition, vol. 2, 2000, pp. 822-826 vol. 2.

[60] H. Galbally and A. Fierrez, "On the Vulnerability of Fingerprint Verification Systems to Fake Fingerprints Attacks", in Proc. 40th Annual IEEE International Carnahan Conferences Security Technology, USA, 2006, pp. 130-136.

[61] A. Iannarelli, Ear Identification. California: Paramount Publishing Company, 1989.

[62] H. Nejati, L. Zhang, T. Sim, E. Martinez-Marroquin, and G. Dong, "Wonder Ears: Identification of Identical Twins from Ear Images", in Proc. 21st International Conference on Pattern Recognition, Nov 2012, pp. 1201-1204.

[63] A. Kumar and C. Wu, "Automated Human Identification Using Ear Imaging", Pattern Recognition, vol. 45, no. 3, pp. 956-968, 2012.

[64] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior, "The Relation Between the ROC Curve and the CMC", in Proc. Fourth IEEE Workshop on Automatic Identification Advanced Technologies, Oct 2005, pp. 15-20.

[65] I. Steinwart, D. Hush, and C. Scovel, "Learning from Dependent Observations", Journal of Multivariate Analysis, vol. 100, no. 1, pp. 175-194, 2009.

[66] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support Vector Machines for Histogram-Based Image Classification", Neural Networks, IEEE Transactions on, vol. 10, no. 5, pp. 1055-1064, 1999.

[67] P. Brodatz, Textures: A Photographic Album for Artists and Designers. Dover Pubns, 1966.

[68] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh, "On the Significance of Real-World Conditions for Material Classification", in Computer Vision-ECCV 2004. Springer, 2004, pp. 253-266.

[69] Y. Xu, H. Ji, and C. Fermüller, "Viewpoint Invariant Texture Description using Fractal Analysis", International Journal of Computer Vision, vol. 83, no. 1, pp. 85-100, 2009.

[70] G. Kylberg, "The kylberg texture dataset v.1.0", Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, Uppsala, Sweden, External report (Blue series) 35, September 2011. [Online]. Available: http://www.cb.uu.se/~gustaf/texture/

[71] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories", in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2. IEEE, 2006, pp. 2169-2178.

[72] J. Wang, K. Markert, and M. Everingham, "Learning Models for Object Recognition from Natural Language Descriptions", in BMVC, vol. 1, 2009, p. 2.

[73] L. Sharan, R. Rosenholtz, and E. H. Adelson, "Accuracy and Speed of Material Categorization in Real-World Images", Journal of Vision, vol. 14, no. 10, 2014.

[74] O. J. O. Söderkvist, "Computer Vision Classification of Leaves from Swedish Trees", Master's thesis, Linköping University, SE-581 83 Linköping, Sweden, September 2001, liTH-ISY-EX-3132.

[75] L. Fei-Fei, R. Fergus, and P. Perona, "Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories", Computer Vision and Image Understanding, vol. 106, no. 1, pp. 59-70, 2007.

[76] X. Qi, R. Xiao, C.-G. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise Rotation Invariant Co-Occurrence Local Binary Pattern", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 36, no. 11, pp. 2199-2213, 2014.

[77] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope", International journal of computer vision, vol. 42, no. 3, pp. 145-175, 2001.

[78] L. Fei-Fei and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories", in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2. IEEE, 2005, pp. 524-531.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất hệ thống và phương pháp cải tiến để nâng cao hiệu quả nhận dạng hình ảnh.

Sáng chế đề xuất hệ thống và phương pháp nhận dạng mẫu sử dụng khoảng cách bán mêtric mới được gọi là bán kính nhị thức-Poisson (PBR) dựa trên phân phối nhị thức-Poisson. Sáng chế có nhiều ưu điểm không giới hạn. Ví dụ, sáng chế bao gồm khoảng cách bán mêtric mạnh để loại bỏ được giả định i.i.d. và tạo thành một phần của các bộ mô tả đặc trưng i.n.i.d. và lại còn có khả năng chịu được sự suy giảm chất lượng trong điều kiện có nhiễu. Ngoài ra, sáng chế nâng cao hiệu quả của chính thiết bị nhận dạng mẫu bằng cách giảm bớt khối lượng công việc xử lý và nâng cao hiệu quả.

Theo các khía cạnh của sáng chế, các hệ thống và phương pháp này phù hợp với các ứng dụng thời gian thực. Ví dụ, theo các phương án thực hiện sáng chế, các chức năng thực hiện được xử lý song song bằng bộ xử lý đồ hoạ (*GPU: Graphics Processing Unit*).

Theo các khía cạnh khác của sáng chế, một thuật toán phân loại mới được giới thiệu để đạt được độ chính xác cao khi phân loại mặc dù các tập hợp mẫu huấn luyện nhỏ. Theo các khía cạnh khác của sáng chế, thuật toán phân loại này có thể được suy rộng dễ dàng để xử lý nhiều loại hơn mà không cần có giai đoạn huấn luyện hoặc xác thực chéo để tối ưu hoá.

Theo các khía cạnh của sáng chế, đại lượng đo khoảng cách mới để nhận dạng mẫu là dựa trên phân phối nhị thức-Poisson, đại lượng đo khoảng cách này loại bỏ được giả định rằng các dữ liệu đầu vào có cùng phân phối. Các tác giả sáng chế đã kiểm định đại lượng đo mới này trong các thử nghiệm được mô tả trong sáng chế. Một thử nghiệm là nhiệm vụ phân loại nhị phân để phân biệt giữa các hình ảnh kỹ thuật số chụp người và mèo, và một thử nghiệm khác là nhận dạng các hình ảnh kỹ thuật số chụp tai được biên soạn từ hai thư viện hình ảnh. Trong hai thử nghiệm đó, hiệu quả của đại lượng đo này được so sánh với các đại lượng đo khoảng cách Euclid, Manhattan và L_p phân số. Việc trích xuất đặc trưng cho hai thử nghiệm đó được thực hiện bằng cách sử dụng cơ sở dữ liệu biểu đồ tần số gradien định hướng (*HOG: Histogram of Oriented Gradients*) được xử lý song song bằng GPU để thu giữ thông tin về hình dạng và kết cấu.

Các tác giả sáng chế đã chứng minh rằng sáng chế luôn có hiệu quả tốt hơn với các phương pháp nhận dạng mẫu sử dụng các đại lượng đo khoảng cách thuộc giải pháp đã biết nêu trên. Ngoài ra, các kết quả cho thấy rằng đại lượng đo khoảng cách được đề xuất theo sáng chế có thể nâng cao hiệu quả của các thuật toán tự học của máy.

Các khía cạnh của sáng chế đề xuất hệ thống phân loại hình ảnh. Hệ thống này bao gồm GPU để thực hiện chức năng tính toán các đặc trưng HOG cho hình ảnh thu được và so sánh các đặc trưng HOG đã tính với các đặc trưng HOG đã lưu trữ của các hình ảnh huấn luyện. Hệ thống này phân loại hình ảnh dựa trên hình ảnh huấn luyện phù hợp sát nhất dựa vào PBR.

Theo các khía cạnh của sáng chế, hệ thống phân loại hình ảnh này có thể được sử dụng để phân biệt các tế bào ung thư với các tế bào bình thường.

Theo các khía cạnh của sáng chế, hệ thống phân loại hình ảnh này có thể được sử dụng để so khớp các dấu vân tay.

Theo các khía cạnh của sáng chế, hệ thống phân loại hình ảnh này có thể được sử dụng để phát hiện các biến thể hiếm trong dữ liệu giải trình tự axit deoxyribonucleic (*ADN: deoxyribonucleic acid*) hoặc axit ribonucleic (*ARN: ribonucleic acid*).

Theo các khía cạnh của sáng chế, hệ thống phân loại hình ảnh này có thể được sử dụng để nhận dạng khuôn mặt.

Theo các khía cạnh của sáng chế, mẫu nhị phân địa phương đồng xuất hiện bất biến quay từng cặp (*PRICoLBP: Pairwise Rotation Invariant Co-occurrence Local Binary Pattern*) có thể được sử dụng như một phương án khác của HOG. Tương tự, nhân SVM có thể được sử dụng như một phương án khác của kNN.

Các ứng dụng và ưu điểm khác của các phương án thực hiện sáng chế được mô tả dưới đây dựa vào các hình vẽ kèm theo.

Mô tả vắn tắt các hình vẽ

Fig.1A và Fig.1B là các hình vẽ thể hiện hàm khối xác suất đầu ra để phân tích giải trình tự ADN.

Fig.2A và Fig.2B lần lượt là các hình ảnh làm ví dụ từ (a) tập dữ liệu LFW và

(b) tập dữ liệu mèo.

Fig.3A là sơ đồ khối thể hiện cấu trúc thực hiện làm ví dụ để nhận dạng hình ảnh theo phương án thực hiện sáng chế.

Fig.3B là sơ đồ khối thể hiện cấu trúc thực hiện làm ví dụ để phát hiện biến thể ADN hiếm theo phương án thực hiện sáng chế.

Fig.3C là lưu đồ cơ bản để thực hiện phương pháp nhận dạng hình ảnh theo các phương án thực hiện sáng chế.

Fig.4 là đồ thị biểu diễn độ chính xác phân loại dưới dạng hàm số phụ thuộc vào số lượng hình ảnh huấn luyện.

Fig.5 là đồ thị dạng thanh so sánh thời gian tính toán cho ứng dụng phân loại hình ảnh sử dụng các đại lượng đo khoảng cách khác nhau.

Fig.6A và Fig.6B lần lượt thể hiện các đường cong so khớp tích luỹ (*CMC: Cumulative Matching Curve*) cho các cơ sở dữ liệu (a) IIT Delhi I và (b) IIT Delhi II.

Fig.7A và Fig.7B lần lượt thể hiện mức độ ảnh hưởng của nhiễu đến hiệu quả nhận dạng hạng-một trên các cơ sở dữ liệu (a) IIT Delhi I và (b) IIT Delhi II.

Mô tả chi tiết sáng chế

Mặc dù sáng chế có thể được thực hiện theo nhiều phương án khác nhau, nhưng các phương án làm ví dụ được mô tả dưới đây dựa vào các hình vẽ nêu trên, nhằm mục đích giúp hiểu rõ về sáng chế, chỉ được coi là các ví dụ minh hoạ cho các nguyên lý của sáng chế và không được phép hiểu rằng các ví dụ đó nhằm mục đích giới hạn phạm vi của sáng chế ở các phương án ưu tiên được mô tả trong sáng chế và/hoặc được thể hiện trên các hình vẽ.

Phân phối nhị thức-Poisson được định nghĩa là hàm khối xác suất để có n sự kiện xảy ra thành công với các xác suất thành công có tính độc lập nhưng không cùng phân phối (p_1, …, p_N). Các sự kiện đó xảy ra trong không gian xác suất (Ω, F, P). Phân phối này là một yếu vị với giá trị trung bình μ là tổng của p_i, trong đó i tăng từ 1 đến N, và phương sai σ^2 là tổng của $(1-p_i)p_i$, trong đó i tăng từ 1 đến N.

Một trường hợp đặc biệt của phân phối này là phân phối nhị thức, trong đó p_i có

cùng một giá trị với mọi i. Phân phối nhị thức-Poisson có thể được sử dụng trong rất nhiều lĩnh vực như sinh học, nhận dạng hình ảnh, truy lục dữ liệu, kỹ thuật thông tin sinh học, và kỹ thuật. Tuy nhiên, thường tính gần đúng phân phối nhị thức-Poisson là phân phối Poisson. Cách tính gần đúng này chỉ hợp lệ khi các xác suất đầu vào là các giá trị nhỏ một cách rõ ràng so với các giới hạn dựa trên sai số được xác định theo định lý Le Cam [19] có công thức:

$$\sum_{n=0}^{\infty}\left|P(\Omega_n)-\frac{\lambda_N^n e^{-\lambda_N}}{n!}\right|<2\sum_{i=1}^{N}p_i^2 \qquad (2)$$

trong đó $P(\Omega_n)$ là xác suất của n lần thành công trong vùng phân phối nhị thức-Poisson và λ là thông số Poisson.

Phân phối nhị thức-Poisson đã được sử dụng ngày càng nhiều trong các ứng dụng nghiên cứu. Shen et al [20] đã phát triển thuật toán tự học của máy để nhận dạng sản phẩm chuyển hoá từ các cơ sở dữ liệu phân tử lớn như KEGG và PubChem. Vectơ dấu vân tay phân tử được xử lý như được phân phối nhị thức- Poisson và xác suất đỉnh thu được được sử dụng để tìm kiếm ứng viên. Tương tự, Lai et al [21] đã phát triển một mô hình thống kê để dự đoán các cơ chất kinaza dựa trên việc nhận biết vị trí phosphoryl hoá. Quan trọng là, xác suất quan sát thấy sự so khớp với các trình tự liên ứng được tính bằng cách sử dụng phân phối nhị thức-Poisson. Các nhóm khác [22], [23] đã sử dụng phân phối này để xác định sự sai lệch gen trong các mẫu khối u.

Vì xác suất của sự kiện sai lệch thay đổi theo các mẫu, nên vị trí của các bazơ trong trình tự ADN cá thể được coi là các thử nghiệm Bernoulli độc lập với các xác suất thành công không bằng nhau để xác định khả năng xuất hiện sự sai lệch gen ở mọi vị trí trong mỗi mẫu. Theo lập luận như vậy, các mô hình để phát hiện chính xác các biến thể hiếm [24], [25] đã được đề xuất.

Sáng chế tìm cách nâng cao độ chính xác trong phân tích giải trình tự ADN dựa vào các điểm số chất lượng giải trình tự, không kể những mục đích khác. Mỗi điểm số có sẵn cho từng bazơ trong một trình tự ADN được giải trình tự phản ánh xác suất để có giá trị đầu ra được phát hiện chính xác. Ví dụ, nếu có N kết quả đọc độc lập cho một vị trí cụ thể, thì phần mềm phân tích trình tự sẽ tạo ra điểm số chất lượng q_i cho mỗi kết quả đọc tại vị trí đó có tính đến xác suất xảy ra lỗi đọc. Xác suất mặc nhiên của kết quả đọc chính

xác được tính bằng công thức:

$$p_i = 1 - 10^{-\frac{q_i}{10}} \text{ trong đó } i \in [1, N] \quad (3)$$

Vì mức đồng nhất của từng vị trí được giải trình tự được gọi dựa vào nhiều kết quả đọc ở cùng một vị trí, đôi khi số kết quả đọc lên tới hàng nghìn, nên mỗi kết quả đọc là một sự kiện Bernoulli được xử lý và tìm cách xây dựng một phân phối xác suất cho mỗi vị trí được giải trình tự bằng cách sử dụng các điểm số chất lượng thích hợp cho vị trí đó. Các phương pháp có hiệu quả để tính toán phân bố xác suất này được nêu và mô tả dưới đây.

Phương pháp sử dụng định lý Waring

Xác định $p_1, \ldots, p_N$ là mô tả các sự kiện độc lập nhưng không cùng phân phối xảy ra trong không gian xác suất (Ω, F, P). Z_k còn được xác định là tổng của toàn bộ k-tổ hợp duy nhất được lấy ra từ $p_1, \ldots, p_N$. Do đó, xét về mặt hình thức thì:

$$Z_k = \sum_{\substack{I \subset \{1,\ldots,N\} \\ |I|=k}} \bigcap_{i \in I} p_i, \qquad k \in \{0, \ldots, N\} \quad (4)$$

trong đó giao với tập hợp rỗng được gọi là Ω. Vì vậy, $Z_0 = 1$ và tổng tính trên tất cả các tập hợp con I có chỉ số 1, …, N chứa chính xác k phần tử. Ví dụ, nếu N = 3, thì

$$\begin{aligned} Z_1 &= p_1 + p_2 + p_3 \\ Z_2 &= p_1 \cdot p_2 + p_1 \cdot p_3 + p_2 \cdot p_3 \\ Z_3 &= p_1 \cdot p_2 \cdot p_3 \end{aligned} \quad (5)$$

Sau đó xác định P(n) với Z_k bằng cách chuẩn hoá cho tất cả các phép giao tập hợp được đếm dư bằng cách sử dụng định lý Waring [26], đây là trường hợp đặc biệt của công thức Schuette-Nesbitt [27].

$$P(\Omega_n) = \sum_{k=n}^{N} \binom{k}{n} Z_k (-1)^{k-n}, \qquad k \in \{0, \ldots, N\} \quad (6)$$

Định lý bao hàm-loại trừ được tính với n = 0. Phương tiện tính toán Z_k có thể mở rộng được mô tả trong thuật toán 1.

Thuật toán 1: Thuật toán Waring đệ quy

Input: $P = \{p_n \in R^m\}_{n=1}^N$: vectơ xác suất

Output: Tính các giá trị Z_k với một vectơ xác suất tuỳ ý

```
1   Thiết lập ma trận tam giác trên N×N là A^T = [a_{i,j}]
2   a_{1,*} ← P
3   for i = 2 to N do
4      k ← 0
5      for j = i to N do
6         k ← k + a_{i-1,j-1}
7         a_{i,j} ← k a_{1,j}
8      end for
9      Z_j ← Σ_{j=i}^{N} a_{i,j}
10  end for
```

Ưu điểm chính của thuật toán này là độ phức tạp thời gian có mức giảm tăng dần theo hàm mũ khi các giá trị N tăng lên. Kết quả đó phát sinh từ đặc điểm quy hoạch động của thuật toán này để nhóm các phép tính thành các khối để cực tiểu hoá phần dư. Cấu trúc đệ quy tự tương tự đó khiến cho việc tính toán có thể thực hiện được nhờ tránh được sự bùng nổ tổ hợp. Khi sử dụng thuật toán này, tổng số khối cần phải tính tăng theo N^2 và được mô tả bằng tổng số học N/2*(1+N).

Một ưu điểm nữa của thuật toán này là khả năng có thể tính được các phần tử của mỗi cột ở chế độ xử lý song song. Điều này có nghĩa là độ phức tạp thời gian giảm từ $O(N^2)$ ở chế độ không xử lý song song xuống chỉ còn là O(N) ở chế độ xử lý song song toàn phần. Có thể cải tiến thêm nữa bằng cách tính các phần tử ma trận theo hướng ngược, nhờ đó tạo ra phương pháp tiếp đôi để tính song song ma trận A^T. Phương pháp này được thực hiện bằng cách sử dụng đồng thời hai hàm đệ quy, $a_{i,N} = a_{1,N}(Z_{i-1} - a_{i-1,N})$ và $a_{i,j} = a_{1,j}.(a_{i,j+1}/a_{1,j+1} - a_{i-1,j})$, kết hợp với hàm đệ quy được xác định trong thuật toán 1. Các phương pháp nêu trên tạo thành phương thức có hiệu quả để tạo ra các hàm khối xác suất (*p.m.f.: probability mass functions*) kết hợp. Trường hợp N = 6 được biểu diễn dưới đây với dãy Z_k được nhân với các hệ số nhị thức thích hợp.

$$\begin{pmatrix} \binom{0}{0} & -\binom{1}{1} & \binom{2}{2} & -\binom{3}{3} & \binom{4}{4} & -\binom{5}{5} & \binom{6}{6} \\ 0 & \binom{1}{0} & -\binom{2}{1} & \binom{3}{2} & -\binom{4}{3} & \binom{5}{4} & -\binom{6}{5} \\ 0 & 0 & \binom{2}{0} & -\binom{3}{1} & \binom{4}{2} & -\binom{5}{3} & \binom{6}{4} \\ 0 & 0 & 0 & \binom{3}{0} & -\binom{4}{1} & \binom{5}{2} & -\binom{6}{3} \\ 0 & 0 & 0 & 0 & \binom{4}{0} & -\binom{5}{1} & \binom{6}{3} \\ 0 & 0 & 0 & 0 & 0 & \binom{5}{0} & -\binom{6}{1} \\ 0 & 0 & 0 & 0 & 0 & 0 & \binom{6}{0} \end{pmatrix} \bullet \begin{pmatrix} Z_0 \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \end{pmatrix} = \begin{pmatrix} P(0) \\ P(1) \\ P(2) \\ P(3) \\ P(4) \\ P(5) \\ P(6) \end{pmatrix}.$$

Hàm p.m.f. này có thể được tạo ra bằng cách sử dụng một phương pháp khác được mô tả dưới đây.

Phương pháp biến đổi Fourier nhanh

Bằng cách sử dụng các định nghĩa nêu trên, xác suất của một tổ hợp cụ thể bất kỳ ω có thể được viết dưới dạng tích tổ hợp của các sự kiện xuất hiện và các sự kiện không xuất hiện.

$$P(\omega) = \prod_{i \in I} p_i \prod_{i \in I^C} (1 - p_i) \tag{7}$$

Nếu Ω_n được xác định là không gian mẫu tương ứng của tất cả các tập hợp ghép cặp có thể có I và I^C thu được từ n lần xuất hiện và N–n lần không xuất hiện, thì

$$P(\Omega_n) = \sum_{j=1}^{\binom{N}{n}} \prod_{i \in I_j} p_i \prod_{i \in I_j^C} (1 - p_i) \tag{8}$$

Biểu thức nêu trên là trực quan vì nó là tổng của các xác suất của tất cả các tổ hợp có thể có của các sự kiện xuất hiện và các sự kiện không xuất hiện. Theo quan sát, có thể tạo ra đa thức để biểu diễn $P(\Omega_n)$ dưới dạng là các hệ số của đa thức bậc N.

$$\prod_{i=1}^{N} [p_i x + (1 - p_i)] = \sum_{n=0}^{N} P(\Omega_n) x^n \tag{9}$$

Sau đó các hệ số của đa thức nêu trên có thể được tìm ra dễ dàng bằng cách sử dụng

các thuật toán dựa vào phương pháp biến đổi Fourier rời rạc. Vectơ hệ số thích hợp có thể được tính theo cách có hiệu quả như sau:

$$C(x) = IFFT\left(\prod_{k=1}^{N} FFT(p_k \quad 1-p_k)\right) \tag{10}$$

Thực ra, các vectơ này có thể được thêm các số không dẫn đầu đến độ dài là luỹ thừa của hai và sau đó được xử lý lặp theo cặp bằng cách sử dụng thuật toán $IFFT^{-1}(FFT(a)\cdot FFT(b))$, trong đó a và b là cặp vectơ tuỳ ý bất kỳ. Khi thực hiện thuật toán biến đổi Fourier nhanh (*FFT: Fast Fourier Transform*) bằng GPU, nhiều dữ liệu đầu vào có thể được xử lý song song sử dụng một sơ đồ đơn giản có các dữ liệu đầu vào đan xen và các dữ liệu đầu ra được giải chập. Hàm số này trả về một danh sách gồm các bộ dữ liệu, trong đó bộ dữ liệu thứ i chứa phần tử thứ i từ mỗi chuỗi đối số hoặc các đối tượng lặp.

Giải trình tự ADN

Một ứng dụng quan trọng của sáng chế là phân tích các tập dữ liệu giải trình tự ADN thế hệ tiếp theo, trong đó hàng nghìn kết quả đọc phải được phân tích cho từng vị trí bazơ trong trình tự ADN. Nếu một vị trí bazơ cụ thể bị đột biến trong bệnh ung thư, thì việc phát hiện các biến thể như vậy sẽ là chẩn đoán lý tưởng. Thực tế, ADN đột biến thường trộn lẫn với ADN bình thường với tỷ lệ thấp và yêu cầu đặt ra là phải tính độ tin cậy thống kê với hai trạng thái mâu thuẫn được phát hiện ở cùng một vị trí bazơ. Việc này có thể được thực hiện bằng cách coi các trạng thái mâu thuẫn này là các sự kiện Bernoulli và thiết lập các hàm p.m.f. bằng cách sử dụng một trong hai phương pháp nêu trên. Ví dụ về dữ liệu đầu ra được thể hiện trên Fig.1A và Fig.1B.

Sau đó, các khoảng tin cậy tính được từ các hàm p.m.f. này sẽ cho phép quyết định là bằng chứng về trạng thái bazơ đột biến có đủ cao hơn ngưỡng có ý nghĩa hay không. Theo các khía cạnh của sáng chế, các nguyên lý tương tự có thể được áp dụng cho các ứng dụng nhận dạng mẫu, đặc biệt là các ứng dụng liên quan đến việc phân tích hình ảnh. Điều này có thể được chứng minh dựa trên thực tế là giá trị điểm ảnh có thể chỉ được coi là một biến ngẫu nhiên và nó không có giá trị thực, vì nó tuân theo các quy luật của vật lý lượng tử [28].

Khoảng cách bán mêtric gọi là bán kính nhị thức-Poisson

Việc tính các khoảng tin cậy cho mọi phép so sánh khoảng cách theo từng cặp là nặng về xử lý trong tập dữ liệu hình ảnh lớn. Để tránh chi phí này và để nâng cao hiệu quả, đại lượng đo khoảng cách có thể được xác định cho các bộ mô tả đặc trưng có tính độc lập nhưng không cùng phân phối như sau:

Định nghĩa: Cho hai vectơ đặc trưng N-chiều $X = (a_1, a_2, a_3, \ldots, a_N)$ và $Y = (b_1, b_2, b_3, \ldots, b_N)$ với $p_i = |a_i - b_i|$, khoảng cách giữa hai vectơ này là:

$$PB_m(X, Y) = P(\Omega_m)(N - m) \tag{11}$$

trong đó m là yếu vị và P(m) là xác suất đỉnh của phân phối. Darroch [29] trước đây đã chứng minh rằng yếu vị m có thể bị chặn như sau:

$$m = \begin{cases} n, & \text{nếu } n \le \mu \le n + \dfrac{1}{n+2} \\ n \text{ và/hoặc } n+1, & \text{nếu } n + \dfrac{1}{n+2} \le \mu \le n + 1 - \dfrac{1}{N-n+1} \\ n+1, & \text{nếu } n + 1 - \dfrac{1}{N-n+1} < \mu \le n+1 \end{cases}$$

trong đó $0 \le n \le N$. Điều này có nghĩa là m chênh lệch so với giá trị trung bình μ với lượng nhỏ hơn 1. Vì vậy, tuy yếu vị m là giá trị cực đại địa phương, nhưng giá trị này được tính gần đúng với giá trị trung bình μ. Điều này cho phép:

$$PB_\mu(X, Y) = P(\Omega_\mu)(N - \mu) \tag{12}$$

Có thể lọc thêm bằng cách xét độ nhọn vượt của phân phối nhị thức-Poisson được tính bằng công thức:

$$\text{Kurt}\left[\sum_{n=0}^{N} P(\Omega_n)\right] = \frac{1}{\sigma^2} - 6 \tag{13}$$

trong đó σ^2 là phương sai của hàm p.m.f.. Mối quan hệ ngược giữa giá trị cực đại của phân phối và σ^2 hàm ý mối quan hệ tương tự giữa $P(\Omega_n)$ và σ. Mối quan hệ ngược này cũng phù hợp với công trình của Baillon et al [30] thiết lập cận trên không đổi rõ nét dưới đây cho tổng của các thử nghiệm Bernoulli:

$$P(\Omega_n) \le \frac{\eta}{\sigma} \tag{14}$$

trong đó η là hằng số cận trên. Hàm ý của mối quan hệ ngược này là σ có thể được chấp nhận làm đại lượng thay thế của $P(\Omega_\mu)$, nhờ đó tránh được yêu cầu phải tạo ra hàm p.m.f. cho mỗi phép tính khoảng cách. Vì vậy, bán-mêtric dưới đây cho các bộ mô tả đặc trưng có tính độc lập và không cùng phân phối có thể được xác định.

Cho hai vectơ đặc trưng N-chiều $X = (a_1, a_2, a_3, \ldots, a_N)$ và $Y = (b_1, b_2, b_3, \ldots, b_N)$ với $p_i = |a_i - b_i|$, khoảng cách bán kính nhị thức-Poisson giữa hai vectơ này là:

$$PBR(X, Y) = \frac{\sigma}{N - \mu} = \frac{\sqrt{\sum_{i=1}^{N} p_i(1 - p_i)}}{N - \sum_{i=1}^{N} p_i} \tag{15}$$

PBR(X, Y) là bán-mêtric. Hàm d: X×X → [0, 1] là bán-mêtric trên tập hợp X nếu nó thỏa mãn các tính chất sau đây với {x, y} X: (1) tính không âm, d(x,y) >= 0; (2) tính đối xứng, d(x,y) = d(y,x); và 3) tính phản xạ, d(x,x) = 0. PBR là hàm không âm và thỏa mãn tính phản xạ. Vì chỉ sử dụng các giá trị tuyệt đối, nên PBR cũng thỏa mãn tính đối xứng. Xem bảng 4 dưới đây thể hiện là PBR và PB_μ là các đại lượng đo khoảng cách tương đương cho các mục đích ứng dụng thực tế.

Ứng dụng phân loại hình ảnh

Phân loại hình ảnh là một quy trình tự động hoá trên máy tính để gán một hình ảnh kỹ thuật số vào một lớp được chỉ định dựa vào kết quả phân tích nội dung kỹ thuật số của hình ảnh đó (ví dụ, phân tích dữ liệu điểm ảnh). Ứng dụng phổ biến nhất của các quy trình này là tìm kiếm hình ảnh, hoặc cụ thể hơn là, tìm kiếm hình ảnh theo nội dung (*CBIR: Content-Based Image Retrieval*). CBIR là quy trình tìm kiếm các hình ảnh so khớp giống hệt hoặc gần giống từ một hoặc nhiều kho chứa hình ảnh kỹ thuật số dựa vào các đặc trưng được trích xuất tự động từ hình ảnh truy vấn. Quy trình này có nhiều ứng dụng thực tế và hữu ích trong chẩn đoán y khoa, sở hữu trí tuệ, điều tra tội phạm, các hệ thống cảm biến từ xa và các hệ thống lưu trữ và quản lý hình ảnh. Xem tài liệu [31].

Các mục tiêu chính trong hệ thống CBIR bất kỳ là độ chính xác tìm kiếm cao và

độ phức tạp thấp khi tính toán (sáng chế cải thiện cả hai tiêu chí này). Việc thực hiện bước phân loại hình ảnh trước khi tìm kiếm hình ảnh có thể làm tăng độ chính xác tìm kiếm. Ngoài ra, độ phức tạp tính toán cũng có thể giảm bớt nhờ có bước này.

Gọi N_T là số lượng hình ảnh huấn luyện cho mỗi lớp, N_C là số lớp và N_D là số bộ mô tả đặc trưng cho mỗi hình ảnh. Độ phức tạp tính toán của hệ thống CBIR điển hình là $O(N_T \cdot N_C \cdot N_D + (N_T \cdot N_C) \cdot \log(N_T \cdot N_C))$. Xem tài liệu [34]. Ngược lại, việc bổ sung thêm bước phân loại trước sẽ làm cho độ phức tạp tính toán giảm xuống chỉ còn là $O(N_C \cdot N_D \cdot \log(N_T \cdot N_D)) + O(N_T \cdot N_D + N_T \cdot \log(N_T))$. Thuật ngữ thứ nhất đề cập đến quy trình phân loại trước hình ảnh sử dụng thuật toán phân loại lân cận gần nhất Naive-Bayes [35] và thuật ngữ thứ hai đề cập đến chính quy trình CBIR.

Để đưa ra quan điểm nào đó, xét trường hợp $N_T = 100$, $N_C = 10$ và $N_D = 150$. Độ phức tạp tính toán của hệ thống sau so với hệ thống trước nâng tốc độ xử lý lên nhanh gấp 7 lần. Vì vậy, quy trình phân loại trước hình ảnh sẽ nâng cao hiệu quả của hệ thống CBIR.

Việc phát hiện phần đầu và khuôn mặt của mèo đã thu hút sự quan tâm của các nhà nghiên cứu trong thời gian gần đây, phản ánh sự xuất hiện phổ biến của mèo trên mạng internet và chúng như là những người bạn của con người [36], [37], [38], [39]. Mèo là thách thức thú vị đối với việc nhận dạng mẫu. Mặc dù, có chung hình dạng khuôn mặt tương tự như con người, nhưng các thuật toán để phát hiện khuôn mặt của người không thể được áp dụng trực tiếp cho mèo vì có sự thay đổi lớn trong một lớp giữa các đặc trưng và cấu trúc khuôn mặt của mèo so với người. Sáng chế đề cập đến thuật toán phân loại dựa trên hệ thống PBR mà có thể phân biệt giữa hai nhóm này.

Tập dữ liệu hình ảnh Labelled Faces in the Wild (LFW) (Fig.2A) được tạo ra bởi các tác giả của tài liệu [40], và tập dữ liệu mèo (Fig.2B) được tạo ra bởi các tác giả của tài liệu [36]. Các tập dữ liệu này có 13.233 hình ảnh người và 9.997 hình ảnh mèo. Ví dụ, trong mỗi lớp, 70% hình ảnh được phân chia ngẫu nhiên để huấn luyện và 30% hình ảnh còn lại để kiểm định.

Theo các khía cạnh của sáng chế, hệ thống phân loại hình ảnh được thể hiện trên Fig.3A có khả năng thực hiện chức năng phân loại hình ảnh như được mô tả trong sáng chế (xem quy trình cơ bản được thể hiện trên Fig.3C). Như được thể hiện trên hình vẽ, hệ

thống này có bộ xử lý đồ hoạ kết nối với hệ thống máy tính chủ (ví dụ, bộ xử lý trung tâm (*CPU: Central Processing Unit*)), để truy nhập bộ nhớ (không được thể hiện trên hình vẽ). Như được thể hiện trên Fig.3A, máy tính chủ có khả năng truy nhập các hình ảnh đã lưu trữ hoặc dữ liệu hình ảnh đối với các hình ảnh huấn luyện. Như được mô tả dưới đây, mỗi hình ảnh được định lại kích thước bằng kích thước tiêu chuẩn (ví dụ 250×250 điểm ảnh), kích thước tiêu chuẩn này có thể được chọn dựa vào ứng dụng cụ thể. Sau khi định lại kích thước, cơ sở dữ liệu biểu đồ tần số građien định hướng (*HOG: Histogram of Oriented Gradients*) được sử dụng để trích xuất đặc trưng. Dữ liệu HOG có thể được lưu trữ và truy nhập đối với mỗi hình ảnh huấn luyện sao cho hình ảnh đó không cần phải được tạo ra (hoặc tạo lại) khi đang thực hiện. Hình ảnh cần phân loại (trên Fig.3A, hình ảnh kiểm định) nhận được ở máy tính chủ từ nguồn cung cấp hình ảnh như bộ nhớ, mạng hoặc hệ thống chụp ảnh (camera, máy quét hoặc thiết bị chụp ảnh khác). Hình ảnh này được định lại kích thước bằng kích thước tiêu chuẩn. Sau khi định lại kích thước, cơ sở dữ liệu biểu đồ tần số građien định hướng (HOG) được sử dụng để trích xuất đặc trưng.

Dữ liệu HOG được nhập vào GPU để xử lý tiếp. Sự định hướng được tính và biểu đồ tần số được tạo ra. Biểu đồ tần số được chuẩn hoá (như được thể hiện trên hình vẽ, bằng máy tính chủ). Quy trình tính PBR được thực hiện trên cả dữ liệu hình ảnh huấn luyện lẫn dữ liệu hình ảnh kiểm định. Đương nhiên là, quy trình tính PBR có thể được thực hiện trước đối với các hình ảnh huấn luyện và các kết quả được lưu trữ. Cuối cùng, các so sánh được thực hiện để phân loại hình ảnh bằng cách tìm đặc trưng khớp nhất sử dụng các kết quả PBR. Ví dụ, thuật toán 2 (dưới đây) có thể được sử dụng.

Trong một ví dụ, phiên bản được xử lý song song bằng GPU của cơ sở dữ liệu biểu đồ tần số građien định hướng (HOG) [41] được sử dụng để trích xuất đặc trưng. Thuật toán phân loại có tên là k-phần tử lân cận gần nhất dựa vào giá trị trung bình địa phương thích ứng (*ALMkNN: Adaptive Local Mean-based k-Nearest Neighbors*) được sử dụng, thuật toán này là một dạng sửa đổi của thuật toán phân loại không thông số dựa vào giá trị trung bình địa phương được sử dụng trong tài liệu Mitani et al [42]. Thuật toán ALMkNN được thực hiện một phần bằng GPU.

Việc trích xuất đặc trưng của cơ sở dữ liệu HOG có thể được thực hiện bằng GPU sử dụng khung kiến trúc thiết bị thống nhất tính toán (*CUDA: Compute Unified Device*

Architecture) của NVIDIA. Các đặc trưng của cơ sở dữ liệu HOG được mô tả sớm nhất bởi Navneet Dalai và Bill Triggs [41] như là phương tiện tách diện mạo và hình dạng bằng cách biểu diễn phân phối không gian của các gradien trong hình ảnh. Thuật toán này được áp dụng để phát hiện người đi bộ [43], phát hiện phương tiện giao thông [44] và nhận biết cử động [45]. Theo một phương án thực hiện sáng chế, thuật toán cải biến HOG-hình chữ nhật (*R-HOG: Rectangular-HOG*) [46] được sử dụng như được mô tả dưới đây.

Theo khía cạnh khác, sáng chế đề xuất hệ thống và phương pháp giải trình tự ADN, như phát hiện các biến thể hiếm, ví dụ, trong trường hợp sinh thiết khối u. Vectơ X của các xác suất chất lượng giải trình tự là từ mẫu ADN đầu vào ở một vị trí bazơ duy nhất có độ sâu giải trình tự d_x sao cho $X = (x_1, x_2, x_3, \ldots, x_{dx})$, và vectơ tương tự Y là từ mẫu ADN tham chiếu được giải trình tự đến độ sâu d_y sao cho $Y = (y_1, y_2, y_3, \ldots, y_{dy})$. Giá trị trung bình ($\mu$) và độ lệch chuẩn ($\sigma$) của hai vectơ này được tính như sau:

$$\mu_Y = \frac{1}{d_y}\sum_{i=1}^{d_y} y_i \qquad \mu_X = \frac{1}{d_x}\sum_{i=1}^{d_x} x_i$$

$$\sigma_Y = \sqrt{\sum_{i-1}^{d_y}(1-y_i)y_i} \qquad \sigma_X = \sqrt{\sum_{i-1}^{d_x}(1-x_i)x_i}$$

Để so sánh vectơ X với vectơ Y, PBR_{seq} có thể được định nghĩa là khoảng cách giữa vectơ X và vectơ Y như sau:

$$PBR_{seq}(X,Y) = \frac{\sigma_X \sigma_Y}{|\mu_X - \mu_Y|}.$$

Giá trị PBR_{seq} nhỏ biểu thị khả năng mẫu khối u cao hơn. Nhằm mục đích phân loại, ngưỡng T đơn giản có thể được xác định sao cho mẫu X được phân loại là khối u nếu $PBR_{seq} \leq T$, và ngược lại thì mẫu đó được phân loại là mẫu bình thường.

Như được thể hiện trên Fig.3B, sáng chế đề xuất hệ thống giải trình tự ADN. Như được thể hiện trên Fig.3B, hệ thống này tương tự như hệ thống được thể hiện trên Fig.3A, nhưng hệ thống này thực hiện phương pháp trên đây để giải trình tự ADN bằng cách sử dụng dữ liệu vectơ. Như được thể hiện trên hình vẽ, vectơ điểm số chất lượng đầu vào như đã nêu trên được biến đổi thành vectơ xác suất đầu vào, bước biến đổi này có thể được thực hiện bằng máy tính chủ. Các vectơ xác suất tham chiếu có thể được cung cấp

từ trước hoặc được tính bằng máy tính chủ và được cung cấp cho GPU. GPU được tạo cấu hình để thu nhận hai vectơ xác suất này và tính khoảng cách PBR_{seq} giữa các vectơ đầu vào và các vectơ tham chiếu. Khoảng cách đó được sử dụng để phân loại trình tự ADN và máy tính chủ xuất ra thông tin chỉ báo về lớp được gán.

Tính građien

Cho hình ảnh đầu vào I(x,y), các đạo hàm không gian 1-chiều (1-D) $I_x(x,y)$ và $I_y(x,y)$ có thể được tính bằng cách áp dụng các bộ lọc građien theo các hướng x và y. Độ lớn građien Mag(x,y) và định hướng (x,y) của mỗi điểm ảnh có thể được tính bằng cách sử dụng các biểu thức như sau:

$$\mathrm{Mag}(x,y) = \sqrt{I_x(x,y)^2 + I_y(x,y)^2} \tag{16}$$

$$\theta(x,y) = \tan^{-1}(I_y(x,y)/I_x(x,y)) \tag{17}$$

Tích lũy các biểu đồ tần số

Các biểu đồ tần số có thể được tạo ra bằng cách tích luỹ độ lớn građien của mỗi điểm ảnh vào các hộp định hướng tương ứng trên các vùng không gian cục bộ được gọi là các ô. Để giảm bớt sự ảnh hưởng của độ sáng và độ tương phản, các biểu đồ tần số được chuẩn hoá trên toàn bộ hình ảnh. Cuối cùng, bộ mô tả HOG được tạo ra bằng cách ghép các biểu đồ tần số đã chuẩn hoá của tất cả các ô thành một vectơ.

Trong một ví dụ, thuật toán HOG nêu trên được thực hiện bằng cách sử dụng bộ công cụ PyCUDA [47] phiên bản 2012.1 và phiên bản 5.0 của bộ công cụ NVIDIA CUDA và chạy trên cạc đồ hoạ GeForce GTX 560 Ti. Mỗi hình ảnh được định lại kích thước bằng 250×250 (62.500 điểm ảnh) sau đó được phân chia đều thành 25 ô, mỗi ô có kích thước 50×50 điểm ảnh. Để lập địa chỉ cho 62.500 điểm ảnh, thì 65.536 chuỗi được tạo ra trong GPU có 32×32 chuỗi cho mỗi khối ảnh và 8×8 khối ảnh cho mỗi lưới. Sau khi phân định bộ nhớ trong cả máy tính chủ lẫn GPU, nhân được khởi động.

Độ lớn građien, sự định hướng và biểu đồ tần số có thể được tính sau khi biểu đồ tần số được truyền đến máy tính chủ, ở đó quy trình chuẩn hoá cho toàn bộ hình ảnh được thực hiện.

Môđun phân loại

Các thuật toán phân loại có thể có thông số hoặc không có thông số. Các thuật toán phân loại có thông số giả định một phân phối thống kê cho mỗi lớp, phân phối thống kê đó thường là phân phối chuẩn. Dữ liệu huấn luyện chỉ được sử dụng để thiết lập mô hình phân loại và sau đó được loại bỏ hoàn toàn. Vì vậy, các thuật toán này được gọi là thuật toán phân loại theo mô hình hoặc thuật toán phân loại hăng hái. Trong phép so sánh, các thuật toán phân loại không có thông số không đưa ra giả định về phân phối xác suất cho dữ liệu, việc phân loại bộ dữ liệu kiểm định chỉ dựa vào dữ liệu huấn luyện đã lưu trữ và vì vậy nên còn được gọi là thuật toán phân loại theo trường hợp hoặc thuật toán phân loại lười. Ví dụ nguyên mẫu của thuật toán phân loại có thông số là thuật toán máy vectơ hỗ trợ (SVM), thuật toán này cần có giai đoạn huấn luyện tăng cường của các thông số trong thuật toán phân loại [48], [49], [50] và ngược lại, một trong số các thuật toán phân loại không có thông số nổi tiếng nhất là thuật toán phân loại k-phần tử lân cận gần nhất (kNN).

Thuật toán kNN [51] đã được sử dụng rộng rãi trong các bài toán nhận dạng mẫu vì nó đơn giản và hiệu quả. Ngoài ra, thuật toán này được coi là một trong mười thuật toán tốt nhất để truy lục dữ liệu [52]. Thuật toán kNN gán cho mỗi mẫu truy vấn một lớp kết hợp với nhãn lớp chính của k-phần tử lân cận gần nhất của nó trong tập dữ liệu huấn luyện. Trong các bài toán phân loại nhị phân (hai lớp), giá trị của k thường là số lẻ để tránh tình trạng các số lượng biểu quyết ngang nhau. Mặc dù thuật toán kNN có một số ưu điểm như có khả năng xử lý số lượng lớp rất lớn, tránh được tình trạng quá khớp và không có giai đoạn huấn luyện, nhưng thuật toán này lại có ba nhược điểm lớn: (1) thời gian tính toán, (2) sự ảnh hưởng của các giá trị ngoại lệ [53] và (3) cần phải chọn giá trị cho k [54].

Vấn đề thứ nhất, độ phức tạp thời gian, phát sinh trong khi tính các khoảng cách giữa tập dữ liệu huấn luyện và mẫu truy vấn, đặc biệt là khi kích thước của tập dữ liệu huấn luyện là rất lớn. Vấn đề này có thể được giải quyết bằng cách thực hiện song song thuật toán kNN, làm giảm độ phức tạp thời gian chỉ còn là hằng số O(1). Điều này gần giống với các phương án thực hiện thay thế khác như các cây tìm kiếm là O(logN) về thời gian. Vấn đề thứ hai liên quan đến sự ảnh hưởng của các giá trị ngoại lệ. Để giải

quyết vấn đề này, có thể sử dụng thuật toán tập trung vào các phần tử lân cận địa phương. Tuy nhiên, thuật toán loại này, được gọi là thuật toán kNN dựa vào giá trị trung bình địa phương (*LMkNN: Local Mean kNN*), vẫn còn tồn tại vấn đề là cần phải chọn giá trị cho k. Hầu hết thời gian, k được chọn bằng các kỹ thuật kiểm chứng chéo [55].

Tuy nhiên, các kỹ thuật đó tốn nhiều thời gian và có nguy cơ quá khớp. Vì vậy, sáng chế sử dụng thuật toán trong đó k được chọn thích ứng, nhờ vậy tránh được yêu cầu cần phải có giá trị k cố định. Để thiết lập cận trên cho k, có thể sử dụng quy tắc kinh nghiệm thông thường, cận trên này bằng căn bậc hai của N, trong đó N là tổng số trường hợp huấn luyện thuộc tập hợp T [56]. Thuật toán này được gọi là thuật toán LMkNN thích ứng (Adaptive LMKNN) hay ALMkNN.

Các bước làm việc trong thuật toán phân loại này được mô tả trong thuật toán 2.

Thuật toán 2: Thuật toán ALMkNN

Input: Mẫu truy vấn: x; $T = \{x_n \in R^m\}_{n=1}^N$: tập hợp kiểm định TS; $c_1, c_2, \ldots, c_M$: M nhãn lớp; k_{min}: số phần tử lân cận gần nhất tối thiểu; k_{max}: số phần tử lân cận gần nhất tối đa; LB: cận dưới; UB: cận trên

Output: Gán nhãn lớp của mẫu truy vấn cho vectơ trung bình địa phương gần nhất trong số các lớp

```
1  Tính các khoảng cách giữa x và tất cả các x_i thuộc tập hợp T
2  Chọn phần tử lân cận gần nhất k_min thuộc tập hợp T, ký hiệu là T_{k_min}(x)
3  Xác định số lượng phần tử lân cận trong tập hợp T_{k_min}(x) cho mỗi lớp c_i
4  if tất cả các phần tử trong tập hợp T_{k_min}(x) chỉ biểu diễn một lớp c then
5     Gán x cho lớp c
6  else
7     while k_min ≤ k_max do
8        count ← 0
9        Tính vectơ trung bình địa phương cho mỗi lớp c_i thuộc tập hợp T_{k_min}(x)
10       Tính các khoảng cách d_i giữa x và mỗi vectơ trung bình địa phương
11       Sắp xếp d_i theo thứ tự tăng dần để thu được d1 < d2 < d3 < ... < dM
12       Tính tỷ lệ phần trăm thay đổi giữa hai khoảng cách gần nhất d1 và d2
13       if tỷ lệ phần trăm thay đổi> LB + ((UB – LB)/(k_max – k_min) then
14          Gán x cho lớp c[d1]
15          break while
16       else
17          k_min ← k_min + 1
18          count ← count + 1
19       end if
20    end while
21    else
22       Tính vectơ trung bình địa phương cho mỗi lớp c_i thuộc tập hợp T_{k_min}(x)
23       Tính các khoảng cách d_i giữa x và mỗi vectơ trung bình địa phương
24       Gán x cho lớp c_i với vectơ trung bình địa phương gần nhất
25 end if
```

Với 16.261 trường hợp huấn luyện (N thuộc tập hợp T), các giới hạn về số lượng phần tử lân cận, k_{min} và k_{max}, có thể được xác định lần lượt bằng 20 và 127 (số nguyên lớn nhất nhỏ hơn hoặc bằng $\sqrt{N}$). Cận dưới (*LB: Lower Bound*) và cận trên (*UB: Upper Bound*) có thể được xác định để ra quyết định lần lượt bằng 2% và 50%. Bước thứ nhất để tính khoảng cách có thể được thực hiện bằng GPU sử dụng công cụ CUDAMat [57]. Các bước còn lại của thuật toán này được thực hiện bằng bộ xử lý CPU (máy tính chủ). Không có giai đoạn huấn luyện và các bộ mô tả HOG của các hình ảnh huấn luyện được lưu trữ trong bộ nhớ.

Hiệu quả phân loại

Bằng cách sử dụng thuật toán ALMkNN để làm khung, các đại lượng đo khoảng cách khác nhau được đánh giá cạnh nhau, đó là PBR, $L_{0.1}$, $L_{0.5}$, L_1 và L_2. Độ chính xác phân loại theo sáng chế được tính trung bình cho sáu lần chạy kiểm chứng việc lấy mẫu con ngẫu nhiên lặp lại và các kết quả này được thể hiện trên Fig.4. Điều đáng chú ý là, PBR và L_1 có độ chính xác gần giống nhau, có hiệu quả tốt hơn đáng kể so với các đại lượng đo khoảng cách khác. Khoảng cách Euclid có thể có hiệu quả tốt hơn một chút với tập dữ liệu huấn luyện nhỏ nhưng lại nhanh chóng giảm sút hiệu quả khi số lượng hình ảnh huấn luyện tăng lên.

Sự ảnh hưởng của nhiễu

Để kiểm tra PBR có khả năng chống được sự suy giảm do nhiễu tốt hơn so với các đại lượng đo khoảng cách khác hay không, các hình ảnh huấn luyện và các hình ảnh kiểm định đều được gây hỏng bằng nhiễu dạng muối tiêu có mật độ tăng, d. Ở giá trị d = 0, PBR có hiệu quả tốt hơn đáng kể so với tất cả các đại lượng đo khoảng cách trừ L_1. Tuy nhiên, phù hợp với giả thiết của sáng chế, PBR có hiệu quả tốt hơn đáng kể so với tất cả các đại lượng đo khoảng cách kể cả L_1 ngay cả khi mức nhiễu nhỏ nhất được đưa vào (d = 0,05) (Bảng 1).

Bảng 1: So sánh diện tích dưới đường cong thu được với 5 phương pháp

Mức nhiễu (d)	$L_{0.1}$	$L_{0.5}$	L_1	L_2	PBR
0,00	0,9758*	0,9899*	0,9916	0,9892*	0,9918
0,05	0,9061*	0,9246*	0,9295*	0,9224*	0,9336
0,25	0,8348*	0,8543*	0,8572*	0,8443*	0,8607

0,50	0,7938*	0,8229*	0,8326*	0,8292*	0,8364

Diện tích dưới đường cong (*AUC: Area Under the Curve*) cho mỗi phương pháp được tính trung bình cho 6 lần chạy độc lập để kiểm chứng việc lấy mẫu con ngẫu nhiên lặp lại. Kiểm định dấu và hạng Wilcoxon với độ tin cậy 95% được dùng để so sánh các phương pháp khác với phương pháp PBR. Các phương pháp có hiệu quả kém hơn đáng kể so với phương pháp PBR được đánh dấu *. Giá trị AUC cao nhất với mỗi mức nhiễu được in đậm.

Thời gian tính toán

Thời gian tính toán được đo trên hệ thống máy tính cá nhân (*PC: Personal Computer*) dùng bộ xử lý 64-bit Intel Core i5-3470 CPU ở tốc độ 3,20GHz 12GB RAM chạy hệ điều hành Ubuntu 12.04 LTS.

Bảng 2: Thời gian tính toán trung bình để xử lý hình ảnh 250×250 điểm ảnh

	CPU (ms)	GPU (ms)	Tăng tốc
Chụp hình ảnh	0,007	-	-
HoG	17,19	5,73	3
Tính PBR	31,7	12,11	2,6
Thuật toán phân loại	1,37	-	-
Tổng	50,26	19,21	2,6

Nhìn vào bảng 2, có thể thấy rằng trường hợp xử lý bằng GPU theo sáng chế nhanh hơn cỡ khoảng 2,6 lần so với trường hợp xử lý thuần tuý bằng bộ xử lý CPU. Sự tăng tốc này làm cho phương pháp PBR gần như ngang bằng với các phương pháp L_1 và L_2. Thời gian tính toán còn được giảm xuống hơn nữa khi đưa vào thuật toán phân loại trung bình gần nhất (*NMC: Nearest Mean Classifier*) (Thuật toán 3) làm bước được thực hiện trước thuật toán phân loại ALMkNN. Độ tin cậy (*CM: Confidence Measure*) bằng 20% được sử dụng, có nghĩa là kết quả NMC được sử dụng để phân loại khi độ tương phản giữa các khoảng cách đến trọng tâm vượt quá 20%.

Thuật toán 3: Thuật toán trung tâm gần nhất và ALMkNN

Input: Mẫu truy vấn: x; $T = \{x_n \in R^m\}_{n=1}^N$: tập hợp kiểm định TS; $c_1, c_2, \ldots, c_M$: M nhãn lớp; k_{min}: số phần tử lân cận gần nhất tối thiểu; k_{max}: số phần tử lân cận gần nhất tối đa; LB: cận dưới; UB: cận trên; CM: độ tin cậy

Output: Gán nhãn lớp của mẫu truy vấn cho vectơ trung bình địa phương gần nhất trong số các lớp

1 Tính vectơ trung bình địa phương cho mỗi lớp c_i thuộc tập hợp T
2 Tính các khoảng cách d_i giữa x và mỗi vectơ trung bình địa phương sử dụng đại lượng đo khoảng cách Manhattan chuẩn hoá
3 Sắp xếp d_i theo thứ tự tăng dần để thu được $d_1 < d_2 < d_3 < \ldots < d_M$
4 Tính tỷ lệ phần trăm thay đổi giữa hai khoảng cách gần nhất d_1 và d_2
5 if tỷ lệ phần trăm thay đổi > CM then
6 Gán x cho lớp $c[d_1]$
7 else
8 Gán x cho lớp c_i sử dụng thuật toán phân loại ALMkNN
9 end if

Các kết quả về độ chính xác đều hoàn toàn giống nhau nhưng thời gian tính toán được cải thiện đáng kể, như được thể hiện trên Fig.5.

Ứng dụng công nghệ sinh trắc tai

Công nghệ sinh trắc liên quan đến các phương pháp tự động hoá để xác minh đặc điểm nhận dạng của một cá nhân sử dụng các đặc trưng có thể là đặc trưng sinh lý hoặc hành vi. Lĩnh vực sinh trắc học tự động hoá đã đạt được những tiến bộ đáng kể trong thập kỷ qua với công nghệ sinh trắc khuôn mặt, dấu vân tay và mống mắt đã nổi lên như những phương thức ứng dụng thông dụng nhất. Không có một phương thức sinh trắc học nào mà không có nhược điểm. Ví dụ, công nghệ sinh trắc khuôn mặt đã được nghiên cứu rộng rãi và nhưng vẫn có thể thất bại trong các điều kiện dưới mức tối ưu [58], [59].

Trong khi đó, các dấu vân tay về lý thuyết là đủ độ phức tạp để dùng làm ký số duy nhất, nhưng trên thực tế, công nghệ sinh trắc dấu vân tay không chống giả mạo được, vì hệ thống dễ bị tấn công bởi các dấu vân tay giả làm từ gelatin, silicon và latec [60]. Công nghệ sinh trắc mống mắt đã được chứng minh là có độ chính xác cao và đáng tin cậy

nhưng hiệu quả của nó bị suy giảm mạnh trong điều kiện ánh sáng yếu, di chuyển mục tiêu, sự lão hoá, khép một phần mí mắt và nhạy với hoàn cảnh thu nhận. Điều này đã thúc đẩy việc nghiên cứu về các đặc điểm khác có thể khắc phục được các vấn đề liên quan đến công nghệ sinh trắc đã được thiết lập. Một trong những đặc điểm mới đó, công nghệ sinh trắc tai, đã nhận được sự quan tâm ngày càng tăng vì nhiều lý do.

1) Khác với khuôn mặt và mống mắt, hình dạng của tai khá là bất biến ở độ tuổi thiếu niên và trưởng thành. Mọi sự thay đổi thường xuất hiện trước độ tuổi lên 8 và sau độ tuổi 70 [61].

2) Không đòi hỏi phải có môi trường được điều chỉnh để chụp ảnh tai vì khung cảnh chụp ảnh lấy chuẩn từ phía bên cạnh của khuôn mặt.

3) Công nghệ sinh trắc tai có thể phân biệt được giữa những người sinh đôi giống nhau về mặt di truyền trong khi công nghệ sinh trắc khuôn mặt không làm được điều đó [62].

4) Tai có sự phân bố đồng đều hơn về màu sắc và ít thay đổi theo những biểu hiện trên khuôn mặt.

Theo các khía cạnh của sáng chế, hệ thống nhận dạng tai được tạo ra dựa vào các đặc điểm HOG và PBR, theo phần mô tả nêu trên. Các cơ sở dữ liệu đã được sử dụng là cơ sở dữ liệu IIT Delhi Ear I và II [63]. Có 125 đối tượng và 493 hình ảnh trong cơ sở dữ liệu IIT Delhi DB I; 221 đối tượng và 793 hình ảnh trong cơ sở dữ liệu IIT Delhi DB II.

Hình ảnh kiểm định cho từng đối tượng trong cả hai cơ sở dữ liệu được chọn ngẫu nhiên và các hình ảnh còn lại được dùng để huấn luyện.

Cấu trúc phân tích sinh trắc

Có ba bước chính trong hệ thống nhận dạng tai: (1) xử lý trước, (2) trích xuất đặc trưng và (3) so khớp mẫu gốc. Bước hiệu chỉnh biểu đồ tần số có thể được dùng làm bước xử lý trước. Bước trích xuất đặc trưng có thể giống như đã mô tả trên đây. Theo các khía cạnh của sáng chế, môđun so khớp có thể tìm hình ảnh phù hợp sát nhất trong số các hình ảnh huấn luyện. Các hình ảnh trong các cơ sở dữ liệu này có kích thước 50×180 điểm ảnh và được định lại kích thước bằng 50×50 điểm ảnh.

Hiệu quả nhận dạng

Hiệu quả được đánh giá bằng cách sử dụng độ chính xác nhận dạng hạng-một. Kết quả nhận dạng được tính trung bình cho mười lần chạy. Giá trị trung bình và độ lệch chuẩn của tỷ lệ nhận dạng hạng-một cho tất cả các đại lượng đo khoảng cách được thể hiện trong bảng 3.

Bảng 3: Hiệu quả nhận dạng hạng-một trên các cơ sở dữ liệu IIT Delhi

	IIT Delhi I	IIT Delhi II
PBR	92,9 ± 1,7	93,0 ± 1,2
$L_{0.1}$	90,7 ± 2,4	90,3 ± 1,3
$L_{0.5}$	92,7 ± 1,9	93,0 ± 1,1
L_1	92,6 ± 1,7	92,9 ± 1,2
L_2	89,3 ± 1,6	89,8 ± 1,5

Các đường cong so khớp tích luỹ (*CMC: Cumulative Match Curve*) được dùng để xác định hiệu quả của các hệ thống nhận dạng sinh trắc và đã được chứng minh là có liên quan trực tiếp đến đường cong đặc trưng hoạt động của thiết bị thu tín hiệu (*ROC: Receiver Operating Characteristic curve*) trong ngữ cảnh kiểm tra hiệu quả [64]. Do đó, các đường đặc trưng CMC cho tất cả các đại lượng đo cũng được thể hiện trên Fig.6.

Sự ảnh hưởng của nhiễu

Trong thử nghiệm, sáng chế được áp dụng cho các hình ảnh huấn luyện và các hình ảnh kiểm định được tạo nhiễu kiểu muối tiêu có mật độ tăng, d. Các kết quả so sánh được thể hiện trên Fig.7A và Fig.7B. Có thể thấy rằng tất cả các đại lượng đo khoảng cách trừ L_2 đều chịu được nhiễu với L_2 có hiệu quả bị giảm rõ rệt khi tăng mật độ nhiễu d.

Sự tương quan giữa PB_μ và các đại lượng đo khoảng cách

Bằng cách xếp hạng các hình ảnh được so khớp theo các đại lượng đo khoảng cách khác nhau với hình ảnh kiểm định đã xác định, thì sự tương quan giữa PB_μ và các đại lượng đo khác (tức là PBR, $L_{0.1}$, $L_{0.5}$, L_1 và L_2) được xác định. Các kết quả thể hiện trong bảng 4 cho thấy rằng PBR và PB_μ có hệ số tương quan cao và thứ hạng gần như giống nhau giữa hai đại lượng đo khoảng cách đó. Điều này phù hợp với nhận định PB_μ và PBR

là các đại lượng đo khoảng cách tương đương gần đúng.

Bảng 4: Hệ số tương quan theo hạng của Spearman
giữa PB_μ và các đại lượng đo khoảng cách khác đối với một hình ảnh kiểm định

	PB_μ
PBR	0,9995
$L_{0.1}$	0,8452
$L_{0.5}$	0,9887
L_1	0,9889
L_2	0,9738

Phân loại hình ảnh dựa vào nhân

PBR là đại lượng đo khoảng cách chấp nhận các dữ liệu đầu vào khác nhau (PRICoLBP, HOG) và còn hoạt động trong các khung tự học của máy khác nhau (kNN, nhân SVM).

Mặc dù các phương pháp SVM (Support Vector Machines) đòi hỏi dữ liệu đầu vào phải có tính độc lập và có cùng phân phối, nhưng các phương pháp này được áp dụng thành công trong các trường hợp phi i.i.d., như nhận dạng tiếng nói, chẩn đoán hệ thống v.v. [65]. Do đó, khung SVM có thể được sử dụng để mô tả hiệu quả của khoảng cách PBR trong ứng dụng phân loại hình ảnh. Để đưa PBR vào trong khung SVM, thì phải sử dụng dạng khái quát hoá của các nhân RBF như sau [66]:

$$K_{d-RBF}(X, Y) = e^{-\rho d(X,Y)}$$

trong đó ρ là tham số tỷ lệ thu được bằng cách sử dụng kiểm chứng chéo và $d(X, Y)$ là khoảng cách giữa hai biểu đồ tần số X và Y. Khoảng cách có thể được định nghĩa bằng cách sử dụng một dạng sửa đổi nhỏ của PBR như sau:

Định nghĩa: Cho hai vectơ đặc trưng N-chiều $X = (a_1, a_2, a_3, \ldots, a_N)$ và $Y = (b_1, b_2, b_3, \ldots, b_N)$ với $p_i = a_i \cdot \ln(2a_i/(a_i+b_i)) + b_i \cdot \ln(2 \cdot b_i/(a_i+b_i))$, khoảng cách giữa hai vectơ này là:

$$d(X,Y) = \frac{\sum_{i=1}^{N} p_i(1-p_i)}{N - \sum_{i=1}^{N} p_i}.$$

Nhân PBR có thể thu được bằng cách thế d(X,Y) vào khung SVM.

Các thử nghiệm

Hiệu quả của nhân khoảng cách PBR được đánh giá trong sáu ứng dụng khác nhau như sau: phân loại kết cấu, phân loại quang cảnh, nhận dạng loài, nhận dạng vật liệu, nhận dạng lá cây và nhận dạng đối tượng. Các tập dữ liệu kết cấu là Brodatz [67], KTH-TIPS [68], UMD [69] và Kylberg [70]. Ứng dụng phân loại quang cảnh dựa vào tập dữ liệu Scene-15 [71]. Đối với các nhiệm vụ nhận dạng, các tập dữ liệu Leeds Butterfly [72], FMD [73], Swedish Leaf [74] và Caltech-101 [75] được sử dụng. Đối với cả hai nhiệm vụ phân loại và nhận dạng, sự phụ thuộc của hiệu quả vào số lượng hình ảnh huấn luyện cho mỗi lớp được đánh giá. Trong mỗi tập dữ liệu, n hình ảnh huấn luyện được chọn ngẫu nhiên, và các hình ảnh còn lại là các hình ảnh kiểm định, ngoại trừ tập dữ liệu Caltech-101 trong đó số lượng hình ảnh kiểm định được giới hạn ở giá trị 50 hình ảnh cho mỗi lớp. Tất cả các thử nghiệm được lặp lại một trăm lần với các tập dữ liệu kết cấu và mười lần với các tập dữ liệu khác. Với mỗi lần chạy, độ chính xác trung bình cho mỗi loại được tính. Kết quả thu được từ các lần chạy riêng biệt được sử dụng để báo cáo giá trị trung bình và độ lệch chuẩn như là các kết quả cuối cùng. Chỉ sử dụng các giá trị cường độ theo thang độ xám cho tất cả các tập dữ liệu, kể cả khi có sẵn hình ảnh màu.

Việc phân loại thành nhiều lớp được thực hiện bằng cách sử dụng kỹ thuật một-đối với-phần còn lại. Với mỗi tập dữ liệu, các siêu thông số của SVM như C và gama được chọn bằng kiểm chứng chéo trong tập dữ liệu huấn luyện với:

$$C \in \left[2^{-2}, 2^{18}\right]$$

và

$$\text{gama} \in \left[2^{-4}, 2^{10}\right] \text{ (cỡ bước bằng 2).}$$

Hiện nay, đặc điểm của mẫu nhị phân địa phương đồng xuất hiện bất biến quay từng cặp (*PRICoLBP: Pairwise Rotation Invariant Co-occurrence Local Binary Pattern*)

đã được chứng minh là có hiệu quả và hữu ích với rất nhiều ứng dụng [76]. Các thuộc tính có ý nghĩa của đặc điểm này là tính bất biến quay và sự bắt hiệu quả thông tin đồng xuất hiện trong ngữ cảnh không gian. Vì vậy, đặc điểm này được sử dụng trong các thử nghiệm.

Phân loại kết cấu

Bộ sưu tập Brodatz là tập dữ liệu kết cấu định chuẩn phổ biến có 111 lớp kết cấu khác nhau. Mỗi lớp có một hình ảnh được phân chia ra thành chín hình ảnh con không chồng lên nhau.

Tập dữ liệu KTH-TIPS có 10 lớp kết cấu với 81 hình ảnh cho mỗi lớp. Các hình ảnh này có sự thay đổi lớn trong lớp vì chúng được chụp ở chín cỡ theo ba hướng độ sáng khác nhau và với ba tư thế khác nhau.

Tập dữ liệu kết cấu UMD có 25 loại với 40 mẫu cho mỗi loại. Các hình ảnh không có định cỡ, không có đăng ký được chụp trong điều kiện có sự thay đổi đáng kể về góc ngắm và cỡ cùng với sự chênh lệch đáng kể về độ tương phản.

Tập dữ liệu Kylberg có 28 lớp kết cấu của 160 mẫu duy nhất cho mỗi lớp. Các lớp đồng nhất về cỡ, độ sáng và hướng. Phiên bản 'không có' bổ sung kết cấu quay của tập dữ liệu này được sử dụng.

Cấu hình mẫu gốc 2_a của dữ liệu PRICoLBP được sử dụng, cấu hình này tạo ra vectơ đặc trưng 1.180-chiều với mọi tập dữ liệu. Các kết quả thử nghiệm được thể hiện trong bảng 5, bảng 6, bảng 7 và bảng 8 lần lượt cho các tập dữ liệu Brodatz, KTH-TIPS, UMD và Kylberg. Từ các kết quả đó, có thể thấy rằng phương pháp PBR luôn có hiệu quả tốt hơn so với các phương pháp khác khi số lượng hình ảnh huấn luyện ít và có độ lệch chuẩn nhỏ hơn cùng với tốc độ phân loại cao hơn so với các đại lượng đo khoảng cách khác.

Bảng 5: Kết quả phân loại kết cấu (tỷ lệ phần trăm) trên tập dữ liệu Brodatz

Phương pháp	Các hình ảnh huấn luyện cho mỗi lớp	
	2	3
PBR	95,9 ± 0,6	96,8 ± 0,5
BD	95,6 ± 0,9	96,6 ± 0,5
JD	95,7 ± 0,7	96,6 ± 0,6
χ^2	95,5 ± 1,0	96,6 ± 0,5
L_1	93,1 ± 1,1	94,9 ± 0,8
L_2	89,2 ± 1,4	92,3 ± 1,0
$L_{0.5}$	92,9 ± 0,8	94,4 ± 0,8
L_1-BRD	93,1 ± 1,1	95,0 ± 0,9
HI	93,1 ± 1,0	95,0 ± 0,8
Hellinger	94,2 ± 0,9	95,8 ± 0,6

Bảng 6: Kết quả phân loại kết cấu (tỷ lệ phần trăm) trên tập dữ liệu KTH-TIPS

Phương pháp	Các hình ảnh huấn luyện cho mỗi lớp			
	10	20	30	40
PBR	87,7 ± 2,1	94,3 ± 1,6	97,3 ± 1,2	98,4 ± 1,0
BD	87,2 ± 2,4	94,1 ± 1,8	97,2 ± 1,2	98,3 ± 1,0
JD	87,1 ± 2,5	94,1 ± 1,8	97,1 ± 1,2	98,4 ± 1,0
χ^2	86,5 ± 2,6	94,0 ± 1,9	97,1 ± 1,1	98,5 ± 1,1
L_1	83,3 ± 2,7	92,3 ± 2,0	95,9 ± 1,3	97,6 ± 1,1
L_2	79,9 ± 2,8	89,9 ± 2,1	94,4 ± 1,5	96,5 ± 1,2
$L_{0.5}$	74,9 ± 3,0	80,6 ± 2,1	83,6 ± 2,0	84,4 ± 1,7
L_1-BRD	83,4 ± 2,7	92,3 ± 2,0	95,9 ± 1,3	97,6 ± 1,1
HI	83,7 ± 2,9	92,4 ± 1,9	95,8 ± 1,2	97,5 ± 1,2
Hellinger	84,6 ± 2,8	92,8 ± 1,9	96,3 ± 1,2	97,8 ± 1,1

Bảng 7: Kết quả phân loại kết cấu (tỷ lệ phần trăm) trên tập dữ liệu UMD

Phương pháp	Các hình ảnh huấn luyện cho mỗi lớp			
	5	10	15	20
PBR	90,4 ± 2,0	95,6 ± 1,2	97,4 ± 0,9	98,4 ± 0,8
BD	90,1 ± 2,4	95,2 ± 1,2	97,1 ± 1,0	98,3 ± 0,7
JD	90,0 ± 2,3	95,5 ± 1,4	97,3 ± 1,0	98,3 ± 0,8
χ^2	89,8 ± 2,5	95,5 ± 1,4	97,3 ± 1,0	98,3 ± 0,8
L_1	86,7 ± 2,6	93,5 ± 1,5	95,8 ± 1,0	97,1 ± 0,9
L_2	80,3 ± 2,5	89,7 ± 1,5	93,1 ± 1,3	95,0 ± 1,2
$L_{0.5}$	84,9 ± 1,7	90,7 ± 1,3	92,8 ± 1,2	94,3 ± 1,1
L_1-BRD	86,7 ± 2,5	93,5 ± 1,5	95,8 ± 1,0	97,1 ± 0,9
HI	86,6 ± 2,6	93,4 ± 1,5	95,8 ± 1,0	97,0 ± 0,9
Hellinger	86,9 ± 2,4	93,3 ± 1,5	95,7 ± 1,1	97,0 ± 1,0

Bảng 8: Kết quả phân loại kết cấu (tỷ lệ phần trăm) trên tập dữ liệu Kylberg

Phương pháp	Các hình ảnh huấn luyện cho mỗi lớp			
	2	3	4	5
PBR	85,0 ± 2,8	90,7 ± 1,9	93,6 ± 1,7	95,8 ± 1,2
BD	83,9 ± 3,6	89,8 ± 2,5	93,1 ± 2,1	95,5 ± 1,5
JD	83,9 ± 2,6	89,8 ± 2,7	93,1 ± 1,9	95,5 ± 1,4
χ^2	83,1 ± 3,3	89,6 ± 2,4	92,5 ± 2,4	95,0 ± 1,6
L_1	78,5 ± 2,8	84,1 ± 2,8	88,2 ± 2,1	90,4 ± 1,8
L_2	73,6 ± 3,1	79,8 ± 3,0	83,9 ± 2,8	87,4 ± 2,3
$L_{0.5}$	76,3 ± 4,6	82,9 ± 2,0	85,9 ± 1,9	88,3 ± 1,5
L_1-BRD	78,5 ± 2,8	84,1 ± 2,8	88,1 ± 2,1	90,4 ± 1,8
HI	78,5 ± 2,4	84,4 ± 2,4	88,0 ± 2,1	90,4 ± 1,8
Hellinger	80,5 ± 2,5	86,0 ± 2,0	89,2 ± 1,9	91,2 ± 1,7

Nhận dạng lá cây

Tập dữ liệu Swedish Leaf có 15 loài cây khác nhau ở Thuỵ Điển với 75 hình ảnh cho mỗi loài. Các hình ảnh này có sự đồng dạng lớn giữa các lớp và có sự khác nhau lớn về hình học và trắc quang trong lớp. Tác giả sáng chế sử dụng cấu hình của dữ liệu PRICoLBP giống như đối với các tập dữ liệu kết cấu. Cần lưu ý rằng, tác giả sáng chế không sử dụng thông tin có trước về sơ đồ bố trí không gian của lá. Các kết quả thử nghiệm được thể hiện trong bảng 9. Có thể nhận thấy rằng PBR tạo ra các kết quả chính xác hơn so với các đại lượng đo khoảng cách khác.

Bảng 9: Kết quả nhận dạng (tỷ lệ phần trăm) trên tập dữ liệu Swedish Leaf

Phương pháp	Các hình ảnh huấn luyện cho mỗi lớp		
	5	10	25
PBR	95,7 ± 1,0	97,7 ± 0,7	99,7 ± 0,2
BD	94,5 ± 1,7	97,2 ± 1,1	99,6 ± 0,3
JD	93,9 ± 1,9	97,2 ± 0,7	99,6 ± 0,3
χ^2	94,0 ± 1,9	97,2 ± 0,9	99,6 ± 0,3
L_1	92,2 ± 1,5	95,8 ± 1,2	98,8 ± 0,6
L_2	85,8 ± 3,5	91,7 ± 1,8	97,2 ± 0,9
$L_{0.5}$	88,7 ± 1,5	92,2 ± 1,0	94,8 ± 0,6
L_1-BRD	92,2 ± 1,5	95,8 ± 1,2	98,8 ± 0,6
HI	91,6 ± 1,6	95,6 ± 1,3	98,8 ± 0,6
Hellinger	93,3 ± 1,4	96,7 ± 0,9	99,2 ± 0,5

Nhận dạng vật liệu

Cơ sở dữ liệu Flickr Material Database (FMD) là tập dữ liệu định chuẩn theo yêu cầu được công bố gần đây để nhận dạng vật liệu. Các hình ảnh trong cơ sở dữ liệu này được chọn thủ công từ các hình ảnh Flickr và mỗi hình ảnh thuộc một trong số 10 loại vật liệu thông thường, bao gồm vải, lá cây, thủy tinh, da, kim loại, giấy, nhựa, đá, nước và gỗ. Mỗi loại có 100 hình ảnh (50 hình ảnh chụp cận cảnh và 50 hình ảnh chụp cả đối tượng), các hình ảnh này chụp sự thay đổi hình thức bên ngoài của các vật liệu trong thế giới thực.

Vì vậy, các hình ảnh này có sự thay đổi lớn trong lớp và các điều kiện độ sáng khác nhau. Trên thực tế, các hình ảnh được kết hợp với các mạng che từng phần để mô tả vị trí của đối tượng. Các mạng che đó có thể được sử dụng để tách ra mẫu PRICoLBP chỉ từ các vùng đối tượng. Cụ thể là, cấu hình 6-mẫu gốc có thể được sử dụng cho dữ liệu PRICoLBP tạo ra vectơ đặc trưng 3.540 chiều.

Bảng 10 thể hiện sự phụ thuộc của tỷ lệ nhận dạng vào số lượng hình ảnh huấn luyện cho mỗi lớp của tập dữ liệu FMD. Có thể nhận thấy rằng, nhân PBR có hiệu quả tốt nhất, sau đó đến Bhattacharyya Distance và Jeffrey Divergence.

Bảng 10: Các kết quả thử nghiệm (tỷ lệ phần trăm) trên tập dữ liệu FMD

Phương pháp	Các hình ảnh huấn luyện cho mỗi lớp					
	5	10	20	30	40	50
PBR	30,3 ± 3,9	39,8 ± 2,1	47,4 ± 1,8	51,5 ± 1,5	55,8 ± 1,7	57,3 ± 2,2
BD	28,6 ± 5,7	38,5 ± 2,2	46,7 ± 1,8	50,6 ± 2,0	54,7 ± 1,6	57,1 ± 2,1
JD	27,3 ± 3,9	38,6 ± 2,2	46,6 ± 2,0	51,0 ± 2,4	55,4 ± 1,9	57,0 ± 2,0
χ^2	27,4 ± 4,2	38,0 ± 1,7	45,9 ± 2,0	50,2 ± 2,4	54,5 ± 1,4	56,4 ± 1,8
L_1	25,3 ± 4,9	35,1 ± 1,7	41,3 ± 2,1	43,6 ± 2,0	47,5 ± 1,6	51,3 ± 1,9
L_2	22,0 ± 3,4	28,2 ± 2,9	34,3 ± 1,2	36,6 ± 1,0	39,9 ± 1,4	42,5 ± 2,2
$L_{0.5}$	16,3 ± 2,4	18,2 ± 1,4	21,3 ± 1,3	21,3 ± 1,4	23,9 ± 1,3	23,9 ± 1,8
L_1-BRD	25,1 ± 4,9	34,5 ± 2,3	41,1 ± 2,4	43,6 ± 2,0	47,5 ± 1,6	51,2 ± 1,9
HI	27,1 ± 4,1	35,0 ± 2,0	42,0 ± 1,4	44,3 ± 2,0	47,8 ± 1,7	51,3 ± 1,7
Hellinger	28,5 ± 3,4	35,7 ± 1,8	42,4 ± 1,1	44,5 ± 2,0	49,2 ± 1,3	51,9 ± 1,6

Lưu ý rằng trong bảng 11, nhân PBR có hiệu quả tốt nhất trong 5 loại trên tổng số 10 loại, so với các nhân đại lượng đo khoảng cách khác.

Bảng 11: Độ chính xác theo từng loại (tỷ lệ phần trăm) trên tập dữ liệu FMD

Loại	PBR	BD	JD	χ^2	HI
Vải	46,6	47,0	46,6	46,2	39,2
Lá cây	86,2	85,2	84,2	83,0	78,2
Thủy tinh	59,2	59,4	61,0	58,6	42,8
Da	52,2	52,8	52,6	50,8	54,8
Kim loại	28,6	28,0	29,2	30,2	24,2
Giấy	47,6	45,4	46,6	46,2	46,6
Nhựa	52,8	51,8	50,6	49,8	48,4
Đá	69,2	71,2	68,8	70,2	63,0
Nước	70,0	69,8	69,8	69,6	61,2
Gỗ	61,0	60,8	60,4	59,6	54,2

Phân loại quang cảnh

Tập dữ liệu Scene-15 có tổng số 4.485 hình ảnh, tập dữ liệu này là sự kết hợp của một số tập dữ liệu trước [71], [77], [78]. Mỗi hình ảnh trong tập dữ liệu này thuộc về một trong số 15 loại, bao gồm phòng ngủ, ngoại ô, công nghiệp, nhà bếp, phòng khách, bờ biển, rừng, đường cao tốc, trong thành phố, núi, vùng lộ thiên, đường phố, cao ốc, văn phòng và cửa hàng. Số lượng hình ảnh cho mỗi loại thay đổi từ 210 đến 410. Các hình ảnh này có độ phân giải khác nhau, vì vậy nên các hình ảnh này được định lại kích thước sao cho có kích thước tối thiểu là 256 điểm ảnh (trong khi vẫn giữ nguyên cỡ ảnh).

Cấu hình 2_a mẫu gốc của dữ liệu PRICoLBP được sử dụng, nhưng có hai cỡ (bán kính của các phần tử lân cận: 1,2). Do đó, số chiều của vectơ đặc trưng là 2.360. Bảng 12 thể hiện các kết quả phân loại của các phương pháp khác nhau cho số lượng hình ảnh huấn luyện thay đổi. Đã quan sát thấy rằng PBR có hiệu quả tốt nhất với số lượng hình ảnh huấn luyện ít hơn và có hiệu quả tương đương với 100 hình ảnh huấn luyện cho mỗi lớp.

Bảng 12: Kết quả phân loại (tỷ lệ phần trăm) trên tập dữ liệu Scene-15

Phương pháp	Các hình ảnh huấn luyện cho mỗi lớp			
	10	20	30	100
PBR	62,4 ± 1,9	69,6 ± 1,4	72,4 ± 1,0	79,7 ± 0,5
BD	61,3 ± 2,0	69,1 ± 1,5	72,0 ± 1,1	79,9 ± 0,4
JD	61,4 ± 1,7	69,1 ± 1,4	71,7 ± 1,0	79,9 ± 0,6
χ^2	61,6 ± 1,3	68,6 ± 1,8	71,9 ± 1,1	79,9 ± 0,7
L_1	58,1 ± 1,2	65,6 ± 0,6	69,1 ± 0,8	77,4 ± 0,5
L_2	50,2 ± 2,7	58,2 ± 1,9	62,4 ± 1,7	73,0 ± 0,6
$L_{0.5}$	43,3 ± 1,5	48,4 ± 1,6	50,7 ± 2,0	52,9 ± 5,6
L_1-BRD	58,2 ± 1,1	65,6 ± 0,6	69,1 ± 0,8	77,4 ± 0,5
HI	58,1 ± 1,0	65,2 ± 1,2	69,1 ± 0,8	77,3 ± 0,6
Hellinger	59,4 ± 1,4	66,8 ± 1,0	69,9 ± 1,0	78,6 ± 0,7

Nhận dạng đối tượng

Tập dữ liệu Caltech-101 là tập dữ liệu định chuẩn quan trọng để nhận dạng đối tượng. Tập dữ liệu này có 9.144 hình ảnh với 102 lớp (101 lớp đa dạng và một lớp nền). Số lượng hình ảnh cho mỗi lớp thay đổi từ 31 đến 800. Các hình ảnh này có sự thay đổi lớn trong lớp và chúng cũng có sự thay đổi về kích thước. Vì vậy, các hình ảnh này được định lại kích thước sao cho có kích thước nhỏ nhất là 256 điểm ảnh (trong khi vẫn giữ nguyên cỡ ảnh). Cấu hình 6-mẫu gốc của dữ liệu PRICoLBP được sử dụng cùng với hai cỡ (bán kính của các phần tử lân cận: 1,2), cấu hình này tạo ra vectơ đặc trưng 7.080 chiều.

Bảng 13 thể hiện độ chính xác nhận dạng của các phương pháp khác nhau với số lượng hình ảnh huấn luyện thay đổi. Có thể thấy rằng các kết quả thu được của nhân khoảng cách PBR tương đương với các kết quả thu được của các nhân dựa vào các đại lượng đo khoảng cách khác.

Bảng 13: Kết quả nhận dạng (tỷ lệ phần trăm) trên tập dữ liệu Caltech-101

Phương pháp	Các hình ảnh huấn luyện cho mỗi lớp		
	10	20	30
PBR	36,42	44,08	49,16
BD	36,27	43,88	48,73
JD	36,33	43,98	49,15
χ^2	36,22	44,03	49,10
L_1	31,71	39,00	44,18
L_2	24,58	31,27	35,65
L_1-BRD	31,70	39,00	44,18
HI	31,64	38,96	44,07
Hellinger	32,99	40,54	45,88

Nhận dạng loài

Tập dữ liệu Leeds Butterfly có tổng cộng 832 hình ảnh cho 10 loại (loài) bướm. Số lượng hình ảnh trong mỗi loại nằm trong khoảng từ 55 đến 100. Các hình ảnh này khác nhau về độ sáng, tư thế và hướng. Các hình ảnh này được định lại kích thước sao cho có kích thước tối thiểu bằng 256 điểm ảnh (trong khi vẫn giữ nguyên cỡ ảnh). Cấu hình của dữ liệu PRICoLBP giống như cấu hình dùng cho các tập dữ liệu kết cấu được sử dụng. Bảng 14 thể hiện độ chính xác nhận dạng của các phương pháp khác nhau trên tập dữ liệu Leeds Butterfly với số lượng hình ảnh huấn luyện thay đổi. Có thể nhận thấy rằng nhân PBR có hiệu quả tương đương so với các nhân dựa vào các đại lượng đo khoảng cách khác.

Bảng 14: Kết quả nhận dạng (tỷ lệ phần trăm) trên tập dữ liệu Leeds Butterfly

Phương pháp	Các hình ảnh huấn luyện cho mỗi lớp			
	10	20	30	40
PBR	90,8 ± 1,6	94,5 ± 0,8	95,7 ± 0,7	97,5 ± 0,9
BD	90,2 ± 1,6	94,0 ± 1,1	95,0 ± 0,6	97,0 ± 0,8
JD	89,7 ± 2,4	94,4 ± 0,8	95,5 ± 0,6	97,2 ± 1,0
χ^2	89,7 ± 2,4	94,4 ± 1,2	95,6 ± 0,8	97,5 ± 0,8
L_1	88,3 ± 1,3	92,9 ± 1,2	94,9 ± 0,9	96,6 ± 1,4
L_2	82,6 ± 2,3	88,5 ± 1,1	91,6 ± 1,3	93,7 ± 1,9
$L_{0.5}$	68,9 ± 2,3	73,4 ± 2,2	74,7 ± 2,2	77,1 ± 1,8
L_1-BRD	88,3 ± 1,2	92,9 ± 1,2	94,9 ± 0,9	96,6 ± 1,4
HI	88,0 ± 2,4	93,0 ± 1,0	95,0 ± 0,9	96,7 ± 1,4
Hellinger	89,1 ± 1,2	93,2 ± 1,2	94,4 ± 1,2	96,8 ± 1,2

Do đó, một số phương án ưu tiên của sáng chế đã được mô tả chi tiết trên đây tham chiếu đến các hình vẽ. Theo các khía cạnh của sáng chế, các hệ thống và phương pháp được đề xuất có thể nâng cao hiệu quả tính toán, tốc độ và độ chính xác cho các hệ thống nhận dạng hình ảnh. Các ứng dụng của sáng chế bao gồm các hệ thống y khoa như máy chẩn đoán y khoa, máy giải trình tự ADN, rôbot phẫu thuật, và các hệ thống chụp ảnh khác. Các ứng dụng khác của sáng chế có thể bao gồm máy để xác minh ký số sinh trắc, các hệ thống điều tra tội phạm, như hệ thống nhận dạng dấu vân tay hoặc hệ thống nhận dạng khuôn mặt. Người có hiểu biết trung bình trong lĩnh vực kỹ thuật này có thể tìm ra các ứng dụng mới và hữu ích khác của sáng chế.

Mặc dù sáng chế đã được mô tả dựa vào các phương án ưu tiên nêu trên, nhưng người có hiểu biết trung bình trong lĩnh vực kỹ thuật này phải hiểu rõ là nhiều phương án sửa đổi, cải biến và thay đổi có thể được tạo ra dựa trên các phương án được mô tả trong sáng chế mà vẫn nằm trong phạm vi của sáng chế.

Ví dụ, người dùng có thể được phân loại theo đặc trưng của người dùng, và việc so khớp có thể chỉ giới hạn ở những người dùng có đặc trưng người dùng xác định.

YÊU CẦU BẢO HỘ

1. Phương pháp phân loại hình ảnh kỹ thuật số được thực hiện bằng máy tính, trong đó phương pháp này bao gồm các bước:

thu nhận, từ máy tính chủ, dữ liệu đặc trưng tương ứng với hình ảnh kỹ thuật số;

xác định, bằng bộ xử lý đồ hoạ, khoảng cách bán-mêtric dựa vào phân phối nhị thức-Poisson giữa dữ liệu đặc trưng thu được và một hoặc nhiều dữ liệu đặc trưng tham chiếu được lưu trữ trong bộ nhớ của máy tính chủ; và

phân loại hình ảnh kỹ thuật số bằng cách sử dụng khoảng cách bán-mêtric đã xác định.

2. Phương pháp theo điểm 1, trong đó khoảng cách bán-mêtric là bán kính nhị thức-Poisson (*PBR: Poisson-Binomial Radius*).

3. Phương pháp theo điểm 1, trong đó bước phân loại hình ảnh kỹ thuật số bao gồm sử dụng thuật toán phân loại máy vectơ hỗ trợ (*SVM: Support Vector Machines*).

4. Phương pháp theo điểm 1, trong đó bước phân loại hình ảnh kỹ thuật số bao gồm sử dụng thuật toán phân loại k-phần tử lân cận gần nhất (*kNN: k-Nearest Neighbors*).

5. Phương pháp theo điểm 4, trong đó thuật toán phân loại kNN là thuật toán phân loại k-phần tử lân cận gần nhất dựa vào giá trị trung bình địa phương thích ứng (*ALMkNN: Adaptive Local Mean-based k-Nearest Neighbors*), trong đó giá trị của k-phần tử lân cận gần nhất (k) được chọn thích ứng.

6. Phương pháp theo điểm 5, trong đó giá trị thích ứng của k-phần tử lân cận gần nhất không vượt quá căn bậc hai của số lượng dữ liệu của một hoặc nhiều dữ liệu đặc trưng tham chiếu.

7. Phương pháp theo điểm 1, trong đó dữ liệu đặc trưng thu được và một hoặc nhiều dữ liệu đặc trưng tham chiếu bao gồm dữ liệu mẫu nhị phân địa phương đồng xuất hiện bất biến quay từng cặp (*PRICoLBP: Pairwise Rotation Invariant Co-occurrence Local Binary Pattern*).

8. Phương pháp theo điểm 1, trong đó dữ liệu đặc trưng thu được và một hoặc nhiều dữ liệu đặc trưng tham chiếu bao gồm dữ liệu biểu đồ tần số gradien định hướng (*HOG: Histogram of Oriented Gradients*).

9. Phương pháp theo điểm 1, trong đó dữ liệu đặc trưng thu được bao gồm vectơ đặc trưng N-chiều X sao cho $X = (a_1 \dots a_N)$ và dữ liệu đặc trưng tham chiếu bao gồm vectơ đặc trưng N-chiều Y sao cho $Y = (b_1 \dots b_N)$, và trong đó bước xác định khoảng cách bán-mêtric (PBR(X, Y)) bao gồm bước tính:

$$PBR(X,Y) = \frac{\sigma}{N - \mu} = \frac{\sqrt{\sum_{i=1}^{N} p_i(1-p_i)}}{N - \sum_{i=1}^{N} p_i},$$

trong đó N là số nguyên lớn hơn 0,

$\sigma = \sqrt{\sum_{i=1}^{N} p_i(1-p_i)}$ là độ lệch chuẩn của vectơ $(p_1 \dots p_N)$,

$\mu = \sum_{i=1}^{N} p_i$ là giá trị trung bình của vectơ $(p_1 \dots p_N)$, và

p_i là $|a_i - b_i|$.

10. Phương pháp theo điểm 1, trong đó hình ảnh kỹ thuật số bao gồm thông tin tương ứng với trình tự ADN hoặc ARN, và dữ liệu đặc trưng thu được bao gồm vectơ X của các xác suất chất lượng giải trình tự cho mẫu ADN hoặc ARN thứ nhất có độ sâu giải trình tự d_x sao cho $X = (x_1 \dots x_{dx})$ và dữ liệu đặc trưng tham chiếu bao gồm vectơ Y của các xác suất giải trình tự cho mẫu ADN hoặc ARN tham chiếu có độ sâu giải trình tự d_y sao cho $Y = (y_1 \dots y_{dy})$, và trong đó bước xác định khoảng cách bán-mêtric (PBR_{seq}) bao gồm bước tính:

$$PBR_{seq}(X,Y) = \frac{\sigma_X \sigma_Y}{|\mu_X - \mu_Y|},$$

trong đó μ_X là giá trị trung bình của vectơ X,

μ_Y là giá trị trung bình của vectơ Y,

σ_X là độ lệch chuẩn cho vectơ X, và

σ_Y là độ lệch chuẩn cho vectơ Y.

11. Phương pháp theo điểm 10, trong đó bước phân loại hình ảnh kỹ thuật số bao gồm các bước:

xác định khoảng cách bán-mêtric (PBR_{seq}) có lớn hơn giá trị ngưỡng hay không, và

phân loại trình tự ADN hoặc ARN là khối u hay bình thường dựa vào kết quả xác định khoảng cách bán-mêtric (PBR_{seq}) có lớn hơn giá trị ngưỡng hay không.

12. Phương pháp theo điểm 10, trong đó bước phân loại hình ảnh kỹ thuật số bao gồm phát hiện các đột biến hiếm trong trình tự ADN hoặc ARN.

13. Phương pháp theo điểm 1, trong đó phương pháp này còn bao gồm bước:

xác định dữ liệu đặc trưng tham chiếu phù hợp sát nhất trong số một hoặc nhiều dữ liệu đặc trưng tham chiếu.

14. Phương pháp theo điểm 13, trong đó phương pháp này còn bao gồm bước:

nhận dạng người dựa vào dữ liệu đặc trưng tham chiếu phù hợp sát nhất đã xác định, trong đó hình ảnh kỹ thuật số bao gồm ít nhất một trong số các loại hình ảnh: tai, khuôn mặt, dấu vân tay, và mống mắt.

15. Hệ thống phân loại hình ảnh kỹ thuật số bao gồm:

máy tính chủ có bộ xử lý, trong đó máy tính chủ được kết nối với bộ nhớ lưu trữ một hoặc nhiều dữ liệu đặc trưng tham chiếu; và

bộ xử lý đồ hoạ (*GPU: Graphics Processing Unit*) có bộ xử lý,

trong đó GPU được kết nối với máy tính chủ và được tạo cấu hình để:

thu nhận, từ máy tính chủ, dữ liệu đặc trưng tương ứng với hình ảnh kỹ thuật số;

truy nhập, từ bộ nhớ, một hoặc nhiều dữ liệu đặc trưng tham chiếu;

xác định khoảng cách bán-mêtric dựa vào phân phối nhị thức-Poisson giữa dữ

liệu đặc trưng thu được và một hoặc nhiều dữ liệu đặc trưng tham chiếu,

trong đó máy tính chủ được tạo cấu hình để:

phân loại hình ảnh kỹ thuật số bằng cách sử dụng khoảng cách bán-mêtric đã xác định.

16. Hệ thống theo điểm 15, trong đó khoảng cách bán-mêtric là bán kính nhị thức-Poisson (PBR).

17. Hệ thống theo điểm 15, trong đó máy tính chủ còn được tạo cấu hình để phân loại hình ảnh kỹ thuật số bằng cách sử dụng thuật toán phân loại máy vectơ hỗ trợ (SVM).

18. Hệ thống theo điểm 15, trong đó máy tính chủ còn được tạo cấu hình để phân loại hình ảnh kỹ thuật số bằng cách sử dụng thuật toán phân loại k-phần tử lân cận gần nhất (kNN).

19. Hệ thống theo điểm 18, trong đó thuật toán phân loại kNN là thuật toán phân loại k-phần tử lân cận gần nhất dựa vào giá trị trung bình địa phương thích ứng (ALMkNN), trong đó giá trị của k-phần tử lân cận gần nhất (k) được chọn thích ứng.

20. Hệ thống theo điểm 19, trong đó giá trị thích ứng của k-phần tử lân cận gần nhất (k) không vượt quá căn bậc hai của số lượng dữ liệu của một hoặc nhiều dữ liệu đặc trưng tham chiếu.

21. Hệ thống theo điểm 15, trong đó dữ liệu đặc trưng thu được và một hoặc nhiều dữ liệu đặc trưng tham chiếu bao gồm dữ liệu mẫu nhị phân địa phương đồng xuất hiện bất biến quay từng cặp (PRICoLBP).

22. Hệ thống theo điểm 15, trong đó dữ liệu đặc trưng thu được và một hoặc nhiều dữ liệu đặc trưng tham chiếu bao gồm dữ liệu biểu đồ tần số gradien định hướng (HOG).

23. Hệ thống theo điểm 15, trong đó dữ liệu đặc trưng này bao gồm vectơ đặc trưng N-chiều X sao cho $X = (a_1 \dots a_N)$ và dữ liệu đặc trưng tham chiếu bao gồm vectơ đặc trưng N-chiều Y sao cho $Y = (b_1 \dots b_N)$, và trong đó GPU còn được tạo cấu hình để tính:

$$PBR(X,Y) = \frac{\sigma}{N-\mu} = \frac{\sqrt{\sum_{i=1}^{N} p_i(1-p_i)}}{N-\sum_{i=1}^{N} p_i},$$

trong đó PBR(X, Y) là khoảng cách bán kính nhị thức-Poisson (PBR) giữa vectơ X và vectơ Y,

N là số nguyên lớn hơn 0,

$\sigma = \sqrt{\sum_{i=1}^{N} p_i(1-p_i)}$ là độ lệch chuẩn của vectơ ($p_1 \ldots p_N$),

$\mu = \sum_{i=1}^{N} p_i$ là giá trị trung bình của vectơ ($p_1 \ldots p_N$), và

p_i là $|a_i - b_i|$.

24. Hệ thống theo điểm 15, trong đó hình ảnh kỹ thuật số bao gồm thông tin tương ứng với trình tự ADN hoặc ARN, và dữ liệu đặc trưng thu được bao gồm vectơ X của các xác suất chất lượng giải trình tự cho mẫu ADN hoặc ARN thứ nhất có độ sâu giải trình tự d_x sao cho $X = (x_1 \ldots x_{dx})$ và dữ liệu đặc trưng tham chiếu bao gồm vectơ Y của các xác suất giải trình tự cho mẫu ADN hoặc ARN tham chiếu có độ sâu giải trình tự d_y sao cho $Y = (y_1 \ldots y_{dy})$, và trong đó GPU còn được tạo cấu hình để xác định khoảng cách bán-mêtric (PBR_{seq}), trong đó bước xác định (PBR_{seq}) bao gồm bước tính:

$$PBR_{seq}(X,Y) = \frac{\sigma_X \sigma_Y}{|\mu_X - \mu_Y|},$$

trong đó $PBR_{seq}(X, Y)$ là khoảng cách bán kính nhị thức-Poisson (PBR) giữa vectơ X và vectơ Y,

μ_X là giá trị trung bình cho vectơ X,

μ_Y là giá trị trung bình cho vectơ Y,

σ_X là độ lệch chuẩn cho vectơ X, và

σ_Y là độ lệch chuẩn cho vectơ Y.

25. Hệ thống theo điểm 24, trong đó máy tính chủ còn được tạo cấu hình để:

xác định khoảng cách bán-mêtric (PBR_{seq}) có lớn hơn giá trị ngưỡng hay không, và

phân loại trình tự ADN hoặc ARN là khối u hay bình thường dựa vào kết quả xác định khoảng cách bán-mêtric (PBR_{seq}) có lớn hơn giá trị ngưỡng hay không.

26. Hệ thống theo điểm 24, trong đó máy tính chủ còn được tạo cấu hình để:

phát hiện các biến thể hiếm trong trình tự ADN hoặc ARN.

27. Hệ thống theo điểm 15, trong đó máy tính chủ còn được tạo cấu hình để:

xác định dữ liệu đặc trưng tham chiếu phù hợp sát nhất trong số một hoặc nhiều dữ liệu đặc trưng tham chiếu.

28. Hệ thống theo điểm 27, trong đó máy tính chủ còn được tạo cấu hình để:

nhận dạng người dựa vào dữ liệu đặc trưng tham chiếu phù hợp sát nhất đã xác định, trong đó hình ảnh kỹ thuật số bao gồm ít nhất một trong số các loại hình ảnh: tai, khuôn mặt, dấu vân tay, và mống mắt.

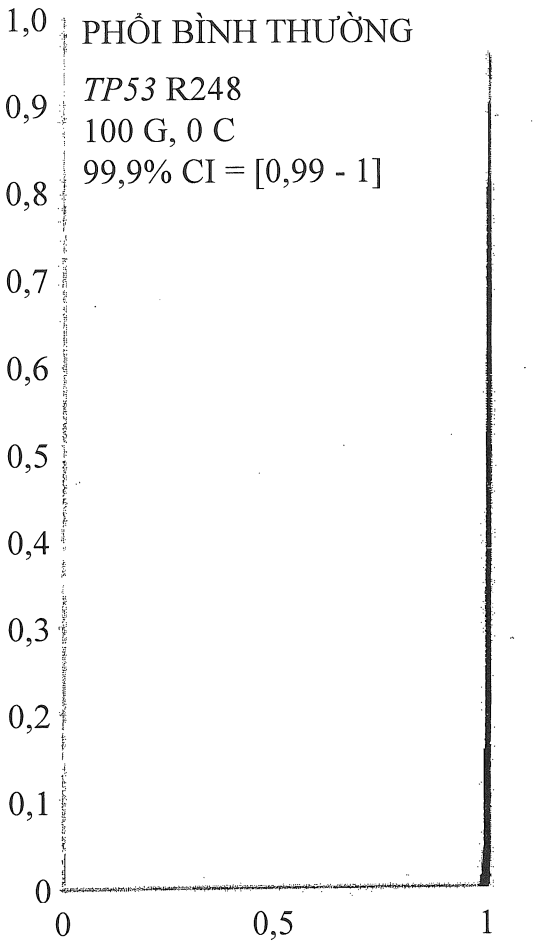

PHỔI BÌNH THƯỜNG
TP53 R248
100 G, 0 C
99,9% CI = [0,99 - 1]
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
0,5
1

FIG. 1A

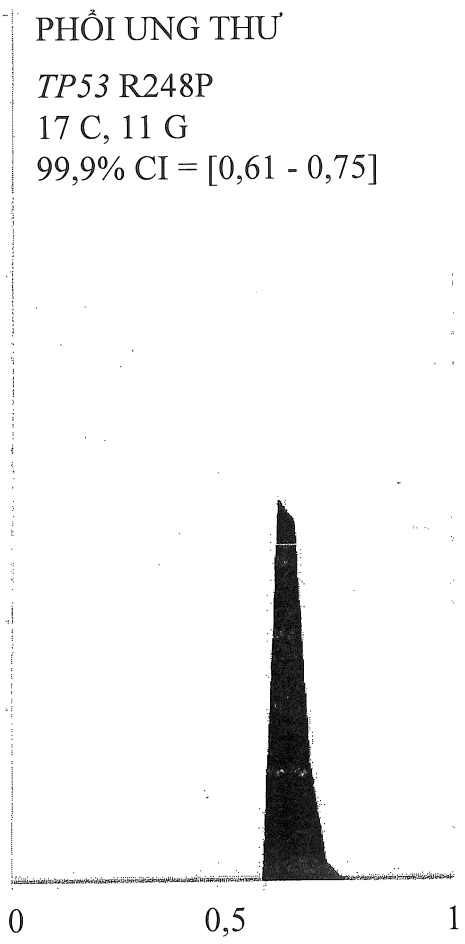

PHỔI UNG THƯ
TP53 R248P
17 C, 11 G
99,9% CI = [0,61 - 0,75]
0
0,5
1

FIG. 1B

FIG. 2A

FIG. 2B

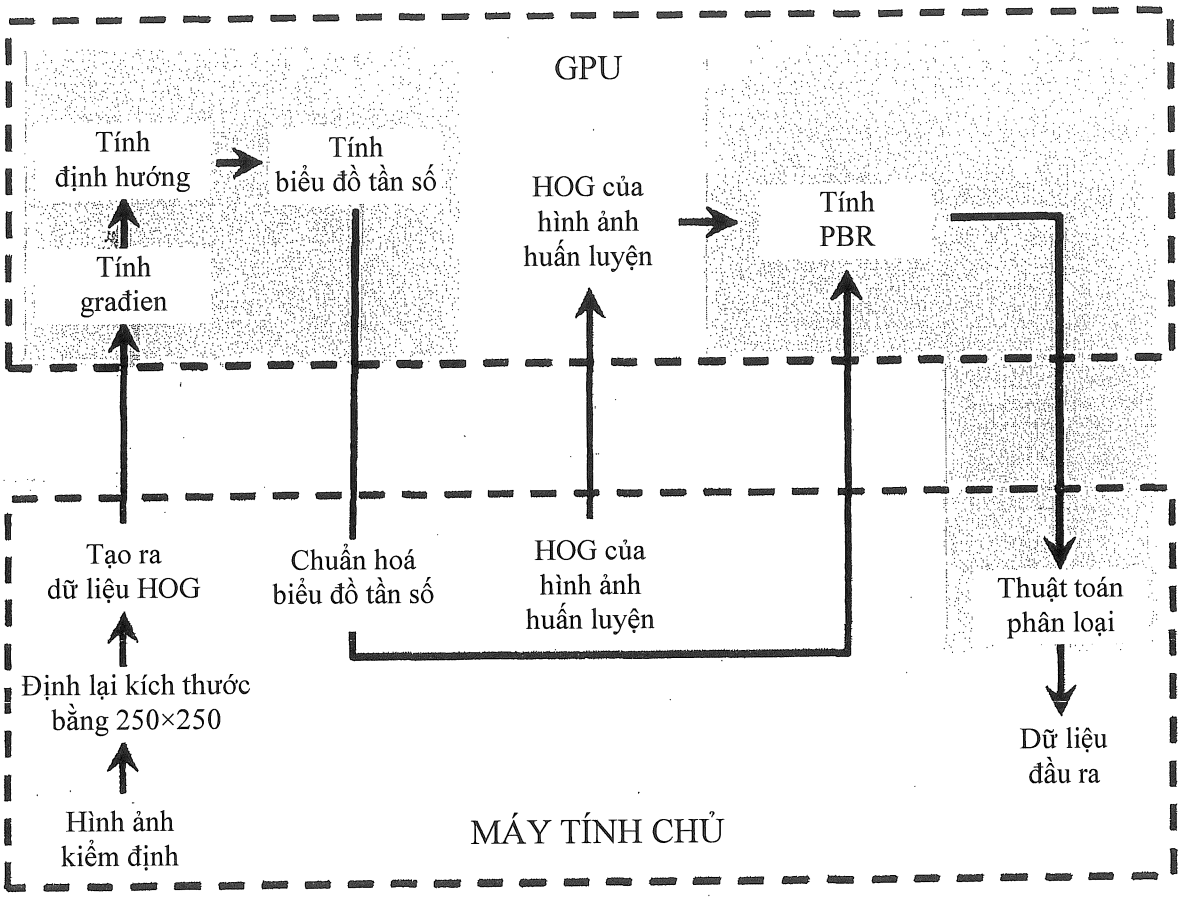

GPU
Tính
định hướng
Tính
biểu đồ tần số
HOG của
hình ảnh
huấn luyện
Tính
PBR
Tính
gradien
Tạo ra
dữ liệu HOG
Chuẩn hoá
biểu đồ tần số
HOG của
hình ảnh
huấn luyện
Thuật toán
phân loại
Định lại kích thước
bằng 250×250
Dữ liệu
đầu ra
Hình ảnh
kiểm định
MÁY TÍNH CHỦ

FIG. 3A

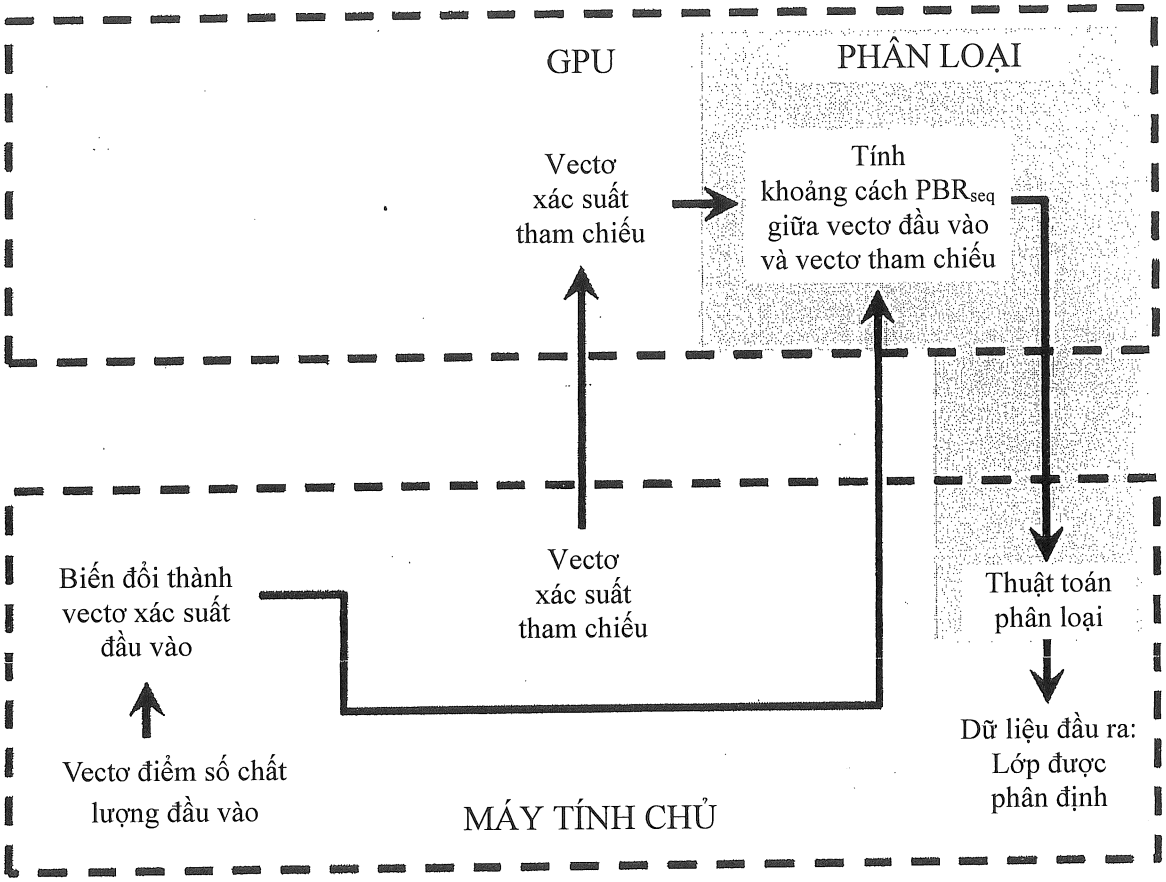

GPU
PHÂN LOẠI
Vectơ
xác suất
tham chiếu
Tính
khoảng cách PBR$_{seq}$
giữa vectơ đầu vào
và vectơ tham chiếu
Vectơ
xác suất
tham chiếu
Biến đổi thành
vectơ xác suất
đầu vào
Thuật toán
phân loại
Vectơ điểm số chất
lượng đầu vào
Dữ liệu đầu ra:
Lớp được
phân định
MÁY TÍNH CHỦ

FIG. 3B

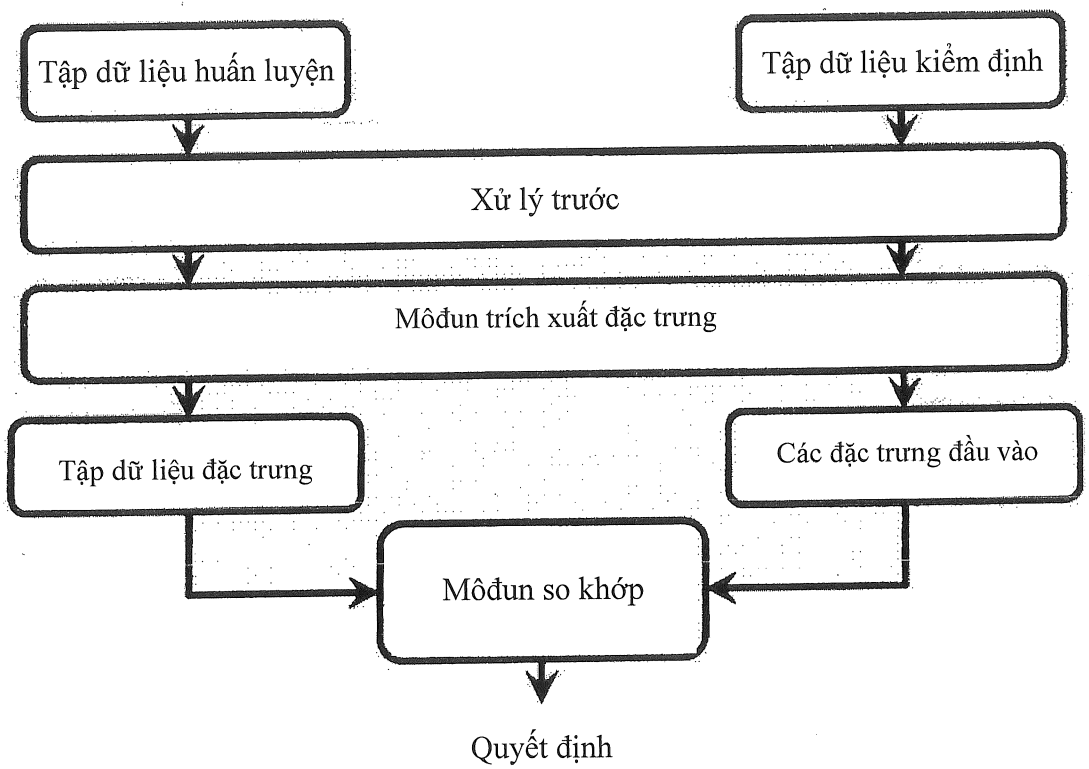

Tập dữ liệu huấn luyện
Tập dữ liệu kiểm định
Xử lý trước
Môđun trích xuất đặc trưng
Tập dữ liệu đặc trưng
Các đặc trưng đầu vào
Môđun so khớp
Quyết định

FIG. 3C

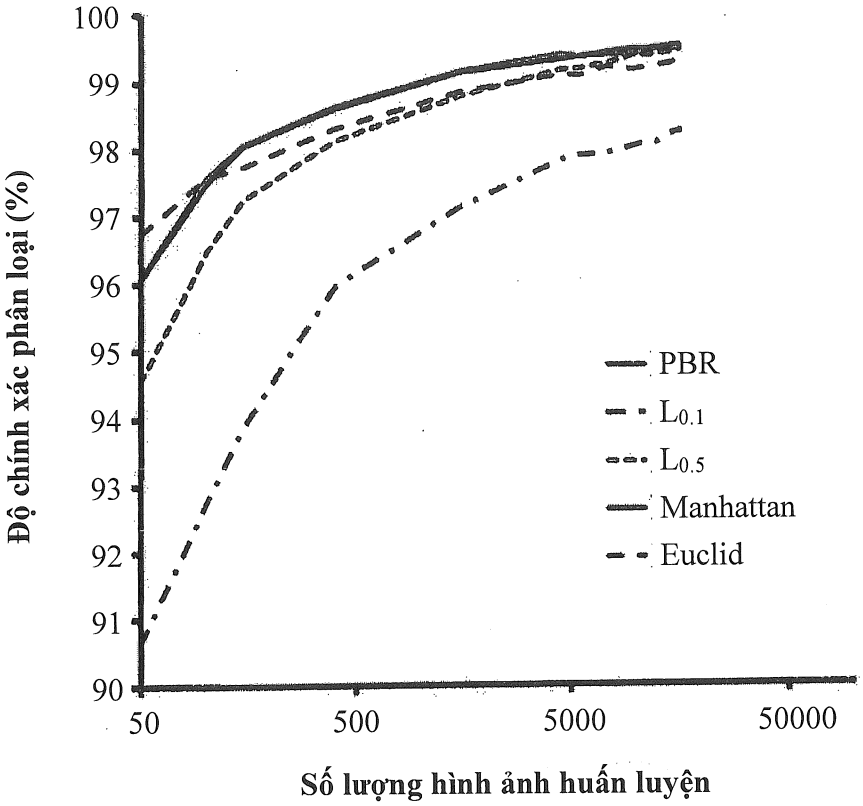

100
99
98
97
96
95
94
93
92
91
90
Độ chính xác phân loại (%)
50
500
5000
50000
Số lượng hình ảnh huấn luyện
PBR
L0.1
L0.5
Manhattan
Euclid

FIG. 4

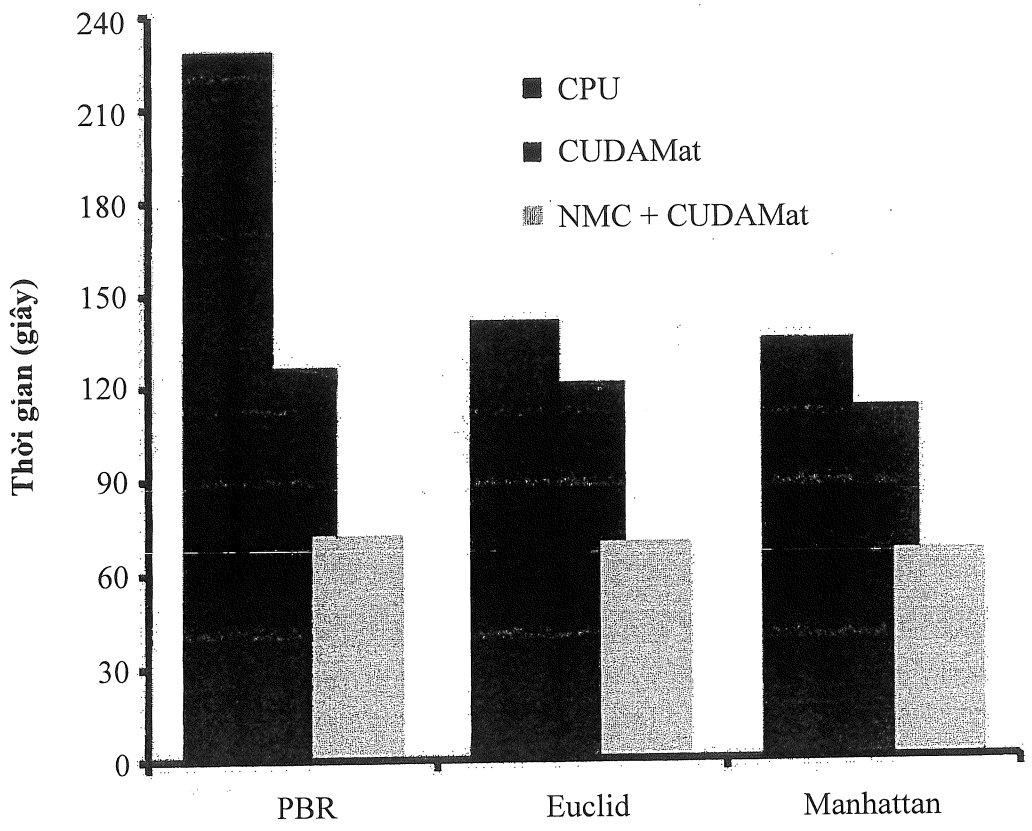

240
210
180
150
120
90
60
30
0
Thời gian (giây)
CPU
CUDAMat
NMC + CUDAMat
PBR
Euclid
Manhattan

FIG. 5

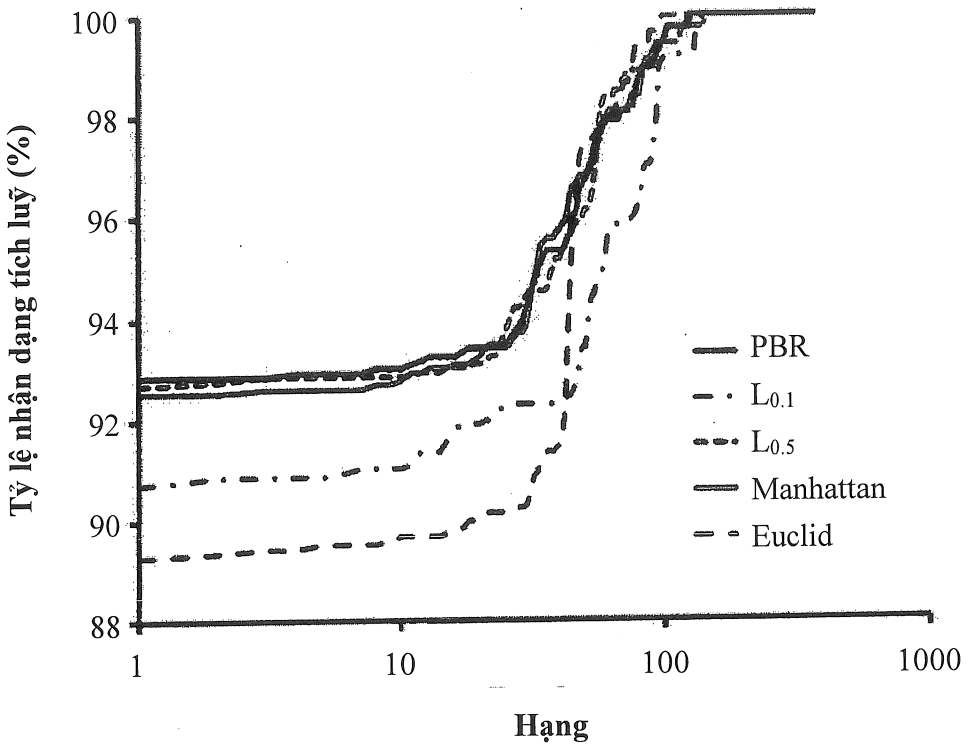

100
98
96
94
92
90
88
Tỷ lệ nhận dạng tích luỹ (%)
1
10
100
1000
Hạng
PBR
L0.1
L0.5
Manhattan
Euclid

FIG. 6A

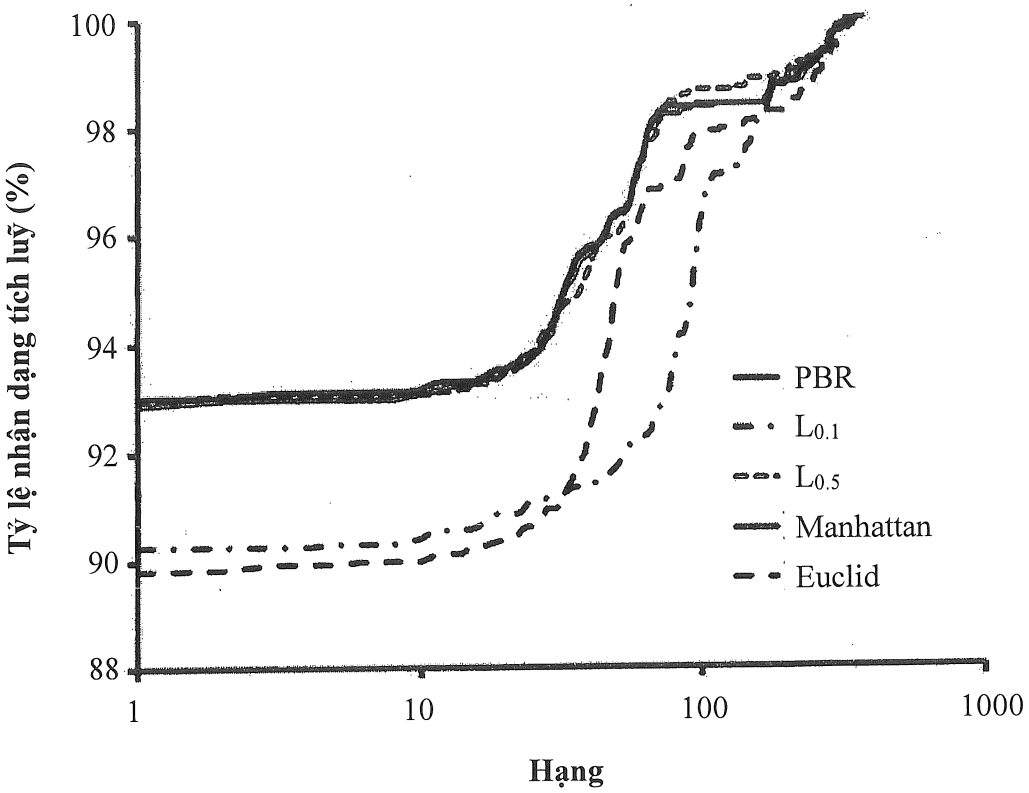

100
98
96
94
92
90
88
Tỷ lệ nhận dạng tích luỹ (%)
1
10
100
1000
Hạng
PBR
L0.1
L0.5
Manhattan
Euclid

FIG. 6B

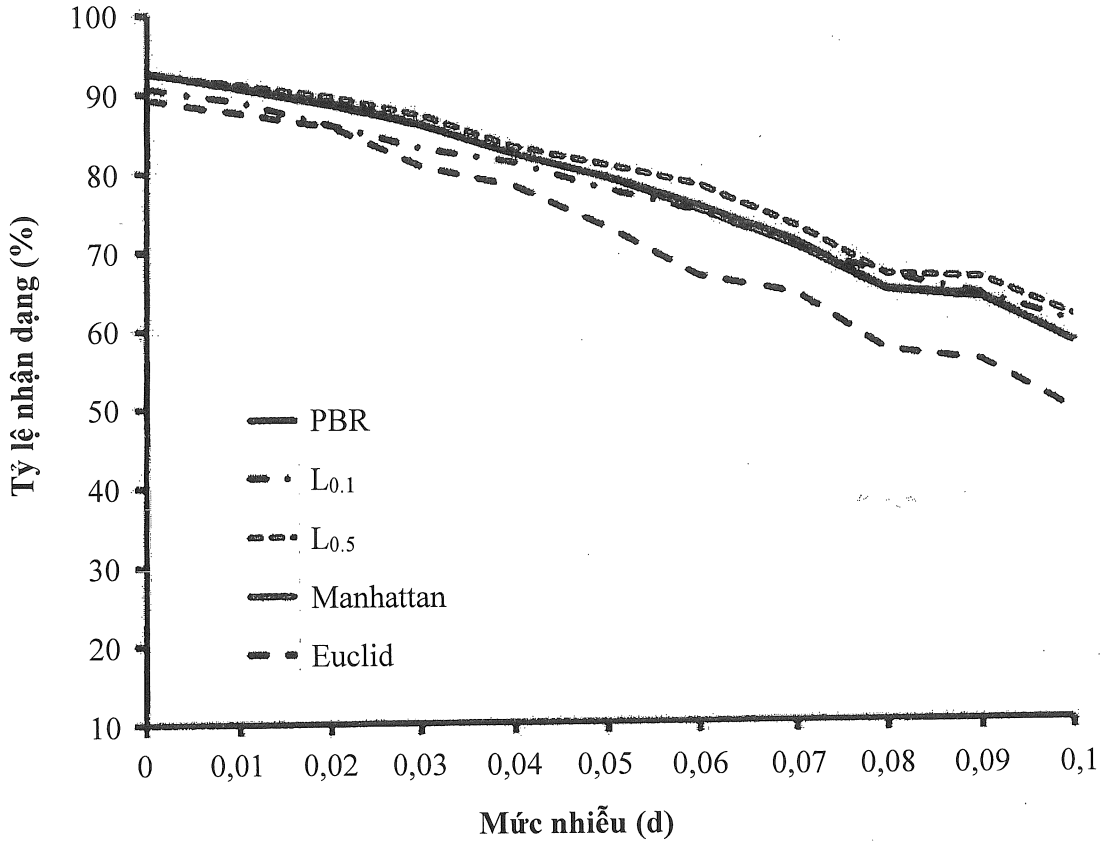

100
90
80
70
60
50
40
30
20
10
Tỷ lệ nhận dạng (%)
PBR
L0.1
L0.5
Manhattan
Euclid
0
0,01
0,02
0,03
0,04
0,05
0,06
0,07
0,08
0,09
0,1
Mức nhiễu (d)

FIG. 7A

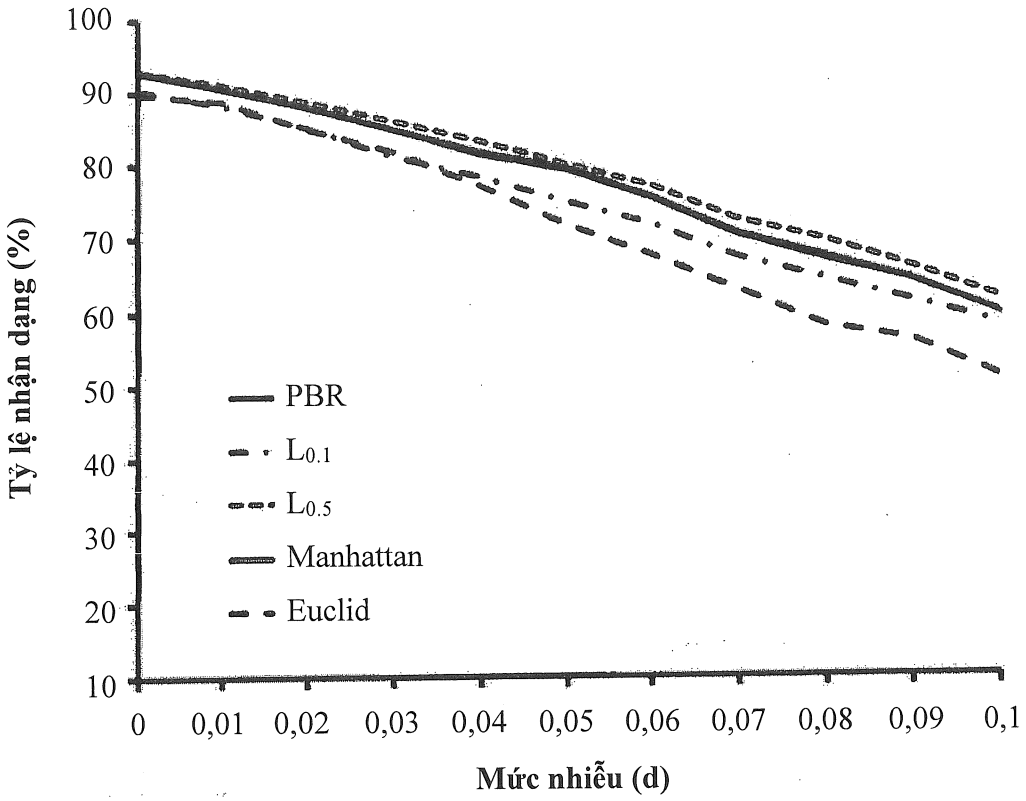

100
90
80
70
60
50
40
30
20
10
Tỷ lệ nhận dạng (%)
PBR
L0.1
L0.5
Manhattan
Euclid
0
0,01
0,02
0,03
0,04
0,05
0,06
0,07
0,08
0,09
0,1
Mức nhiễu (d)

FIG. 7B