



LỌC ẢNH VỚI MÔ HÌNH TÚI TỪ TRỰC QUAN VÀ THUẬT TOÁN ARCX4-RMNB

Nguyễn Minh Trung¹, Nguyễn Ngọc Hưng² và Đỗ Thanh Nghi³

¹ Bộ môn Tin học Ứng dụng, Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ

² Trường Cao đẳng Cộng đồng Sóc Trăng

³ Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận: 14/09/2012

Ngày chấp nhận: 25/03/2013

Title:

Images Filtering With Bag-of-Visual-Words and ArcX4-RMNB

Từ khóa:

SIFT, BoVW, Arcx4-rMNB, SVM, C4.5

Keywords:

SIFT, BoVW, Arcx4-rMNB, SVM, C4.5

ABSTRACT

In this paper, we paper presents a novel approach to detect pornographic images based on the combination of the Scale-invariant feature transform method (SIFT), the bag-of-visual-words (BoVW) and the Arcx4 of random multinomial naive Bayes (Arcx4-rMNB). At the preprocessing step, we propose to use the Scale-invariant feature transform method (SIFT) which is locally based on the appearance of the object at particular interest points, invariant to image scale, rotation and also robust to changes in illumination, noise, occlusion. And then, the representation of the image that we use for classification is the bag-of-visual-words (BoVW), which is constructed from the local descriptors and the counting of the occurrence of visual words in a histogram like fashion. The pre-processing step brings out datasets with a very large number of dimensions. And then, we propose a new algorithm called Arcx4 of random multinomial naive Bayes (Arcx4-rMNB) that is suited for classifying very-high-dimensional datasets. We do setup experiment with two real datasets to evaluate performances. Our approach has achieved an accuracy of 91.75% for a small dataset and 87.93% for other large one.

TÓM TẮT

Trong bài này, chúng tôi giới thiệu một phương pháp mới phát hiện những ảnh khiêu dâm dựa trên sự kết hợp giữa phương pháp biểu diễn ảnh bằng các nét đặc trưng không đổi với những biến đổi tỉ lệ (Scale-invariant feature transform – SIFT), mô hình túi từ trực quan (the bag-of-visual-words (BoVW) và giải thuật ArcX4 của Bayes thơ ngẫu nhiên (the Arcx4 of random multinomial naive Bayes (Arcx4-rMNB)). Ở bước tiền xử lý, chúng tôi sử dụng phương pháp biểu diễn ảnh bằng các nét đặc trưng không đổi được thực hiện dựa trên đặc trưng cục bộ, không bị thay đổi trước những biến đổi tỉ lệ ảnh, tịnh tiến, phép quay, không bị thay đổi một phần đối với phép biến đổi hình học affine (thay đổi góc nhìn) và mạnh với những thay đổi về độ sáng, sự nhiễu và che khuất. Kế tiếp, mô hình túi từ trực quan được sử dụng để biểu diễn nội dung ảnh. Sau bước tiền xử lý, ảnh được biểu diễn bởi một véc-tơ có số chiều rất lớn, chúng tôi đề nghị một giải thuật mới ArcX4 của Bayes thơ ngẫu nhiên cho phép phân lớp hiệu quả dữ liệu có số chiều lớn. Để đánh giá hiệu quả của phương pháp đề xuất, chúng tôi thực nghiệm với tập dữ liệu thực tế và kết quả phương pháp của chúng tôi đạt được chính xác 91.75% cho tập dữ liệu nhỏ và 87.93% cho tập dữ liệu lớn.

1 GIỚI THIỆU

Sự bùng nổ của Internet giúp con người dễ dàng chia sẻ thông tin và tiếp cận kho tri thức của nhân loại. Bên cạnh đó, sách ảnh khiêu dâm cũng được phát tán nhanh chóng trên các trang Web có thể gây nguy hại cho trẻ em. Do đó việc bảo vệ trẻ em tránh tiếp xúc với hình ảnh khiêu dâm trên Internet là một vấn đề thực sự rất bức xúc hiện nay. Để giải quyết vấn đề này, các nhà nghiên cứu đề xuất các phương pháp phát hiện ảnh khiêu dâm bằng tiếp cận học từ nội dung ảnh.

Hiện nay, tiếp cận phát hiện ảnh khiêu dâm có hai hướng chính. Hướng thứ nhất là dựa trên phát hiện của các điểm màu da (skin color pixels), kết cấu da (skin texture), tổ chức đồ màu (color histogram) [7], [8], [6], [19], [24], mặt người (faces) [10] và hình dáng cơ thể (body shape) [22]. Hệ thống này sử dụng mạng nơ-ron (neural networks), máy học véc-tơ hỗ trợ (support vector machines) [21] hoặc rừng ngẫu nhiên (random forests) [3] để học phân lớp ảnh khiêu dâm. Tuy nhiên, hướng tiếp cận thứ nhất này không thu được độ chính xác cao do tiếp cận rút trích đặc trưng ảnh quá đơn giản. Gần đây, một hướng tiếp cận thứ hai [5], [12] dựa vào phương pháp biểu diễn ảnh bằng các nét đặc trưng không đổi với những biến đổi tỉ lệ (SIFT [13], [14]) và mô hình túi từ trực quan (BoVW) (được đề xuất đầu tiên bởi [1] cho phân lớp kết cấu). Một ảnh được biểu diễn bởi tập hợp túi từ trực quan được xây dựng bằng cách áp dụng một giải thuật gom nhóm lên các véc-tơ mô tả cục bộ SIFT. Giai đoạn tiền xử lý cho ra một tập dữ liệu với số chiều rất lớn (chẳng hạn 2000 chiều hoặc từ trực quan). Kế đến, phương pháp máy học véc-tơ hỗ trợ được sử dụng cho phân lớp tập dữ liệu số chiều lớn kiểu này. Hướng tiếp cận thứ hai này cho độ chính xác cao hơn hướng tiếp cận thứ nhất.

Trong bài báo này, chúng tôi đề xuất sử dụng phương pháp biểu diễn ảnh bằng đặc trưng không đổi, mô hình túi từ trực quan. Tiếp đến, chúng tôi đề xuất một giải thuật học mới tên gọi là ArcX4 của Bayes thơ ngây ngẫu nhiên (Arcx4-rMNB) cho phân loại ảnh khiêu dâm. Giải thuật Arcx4-rMNB sử dụng phương

pháp Arcx4 [2] để xây dựng tuần tự k mô hình cơ sở Bayes thơ ngây ngẫu nhiên, mỗi phân lớp tập trung hầu hết các lỗi được tạo ra bởi các mô hình trước đó. Ngoài ra, chúng tôi đề nghị sử dụng các tập con đặc trưng ngẫu nhiên để ước tính xác suất các lớp của Bayes thơ ngây, ý tưởng này nhằm tăng khả năng chịu đựng nhiễu của Bayes thơ ngây. Vì vậy, ArcX4-rMNB có thể giải quyết tập dữ liệu với số chiều lớn (nhiều đặc trưng đầu vào với mỗi đặc trưng chỉ chứa đựng một lượng nhỏ thông tin cho phân lớp). Kết quả thực nghiệm trên hai tập dữ liệu ảnh thực chỉ rằng, đề xuất của chúng tôi đạt được 91,75% đối với tập dữ liệu nhỏ và 87,93% với tập dữ liệu lớn. Thuật giải ArcX4-rMNB thực hiện phân lớp tốt hơn so với những thuật giải khác bao gồm cây quyết định C4.5 [17], rừng ngẫu nhiên của (RF-CART [3]), AdaBoost của C4.5 [9]. So sánh với thuật giải SVM [21], thuật giải của chúng tôi tốt hơn SVM trong các tiêu chí đánh giá hiệu quả phát hiện ảnh khiêu dâm nhưng SVM cho kết quả tốt hơn đối với các tiêu chí đánh giá hiệu quả của ảnh bình thường.

Phần tiếp theo của bài báo được tổ chức như sau. Phần 2 trình bày phương pháp biểu diễn ảnh sử dụng SIFT và mô hình BoVW. Phần 3 trình bày ngắn gọn Bayes thơ ngây và thuật giải Arcx4-rMNB của chúng tôi đề xuất cho phân lớp tập dữ liệu với số chiều lớn. Kết quả thực nghiệm sẽ được trình bày ở phần 4 và kết luận ở phần 5.

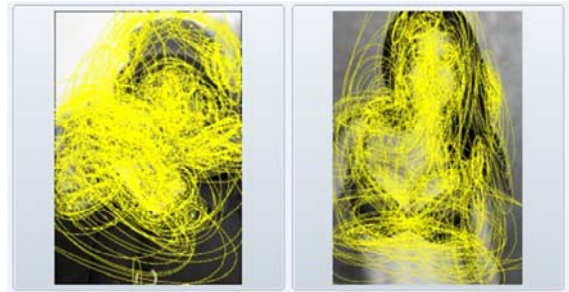
2 BIỂU DIỄN ĐẶC TRƯNG KHÔNG ĐỔI VÀ MÔ HÌNH TÚI TỪ TRỰC QUAN

Biểu diễn ảnh là một bước quan trọng trong phân loại ảnh. Bước này có ảnh hưởng rất lớn đến kết quả phân loại cuối cùng. Hai tiếp cận chính về biểu diễn ảnh hiện nay là: sử dụng nét đặc trưng toàn cục (global features) như véc-tơ bitmap, tổ chức đồ màu (color histogram) và sử dụng nét đặc trưng cục bộ (local features) như điểm đặc trưng, vùng đặc trưng để biểu diễn ảnh. Tiếp cận thứ nhất đơn giản nhưng lại không thật sự hiệu quả vì cách biểu diễn này không thích hợp với những biến đổi về góc nhìn, biến đổi tỉ lệ, phép quay, độ sáng, sự che khuất, sự biến dạng, sự xáo trộn của hình nền

và sự biến đổi trong nội bộ lớp. Ngược lại, tiếp cận thứ hai được đề nghị bởi [13], [14], lại rất mạnh với những thách thức này và đạt được hiệu quả cao trong phân loại ảnh, phát hiện ảnh và nhận dạng ảnh. Vì vậy, phương pháp của chúng tôi sử dụng các nét đặc trưng cục bộ để biểu diễn ảnh được chụp trong nhiều điều kiện khác nhau. Nghiên cứu của chúng tôi dựa trên một mô hình trong phân tích văn bản: mô hình túi từ (bag of words model). Để có thể áp dụng mô hình này lên ảnh, trước hết cần phải định nghĩa các “từ” cho ảnh (gọi là các từ trực quan hay visual words để phân biệt với các từ thông thường trong văn bản). Giai đoạn biểu diễn ảnh theo mô hình này bao gồm 3 bước chính: (i) phát hiện và biểu diễn các nét đặc trưng cục bộ, (ii) xây dựng từ điển các từ trực quan và (iii) biểu diễn ảnh dưới dạng véc-tơ tần xuất.

Ở bước đầu tiên, ảnh được đưa về dạng mức xám. Các điểm đặc trưng (Hình 1) được tính trên những ảnh này bằng cách sử dụng các giải thuật phát hiện điểm đặc trưng cục bộ (local feature detector) như là Harris-Affine, Hessian-Affine [16]. Những điểm đặc trưng này có thể là cực trị cục bộ của phép toán DoG (Difference of Gaussian) hoặc là cực đại của phép toán LoG (Laplace of Gaussian). Sau đó, vùng xung quanh các điểm đặc trưng được xác định và mô tả bằng các véc-tơ mô tả cục bộ. Véc-tơ mô tả

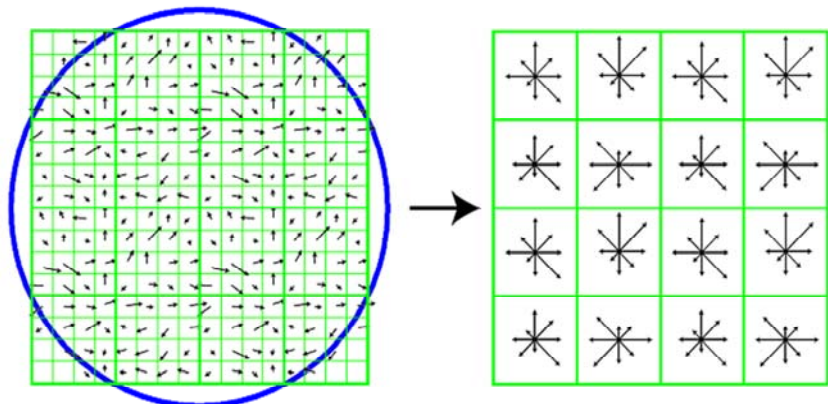
SIFT [13], [14] được đánh giá rất cao bởi giới chuyên môn trong việc biểu diễn các vùng xung quanh điểm đặc trưng bởi vì nó không đổi đối với những biến đổi tỉ lệ, tịnh tiến, phép quay, và không đổi một phần đối với những thay đổi về góc nhìn, đồng thời nó cũng rất mạnh với những thay đổi về độ sáng, sự che khuất, nhiễu.



Hình 1: Các điểm đặc trưng được phát hiện bởi thuật toán Hessian-Affine

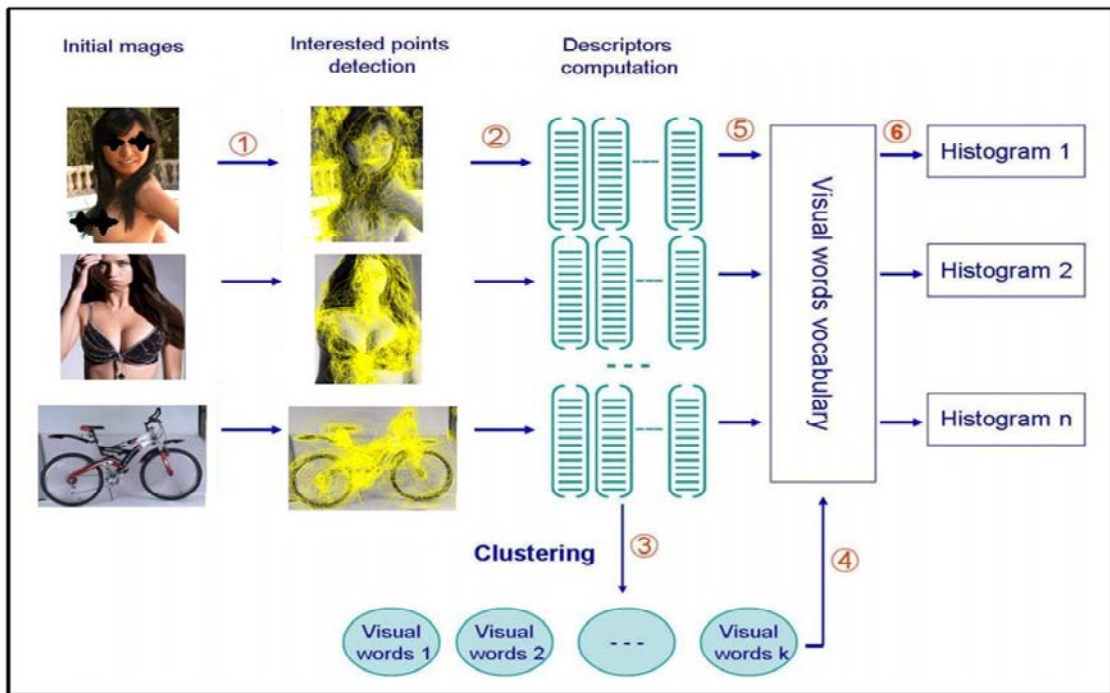
Hình 2 minh họa một ví dụ của véc-tơ mô tả SIFT được xây dựng từ vùng cục bộ xung quanh một điểm đặc trưng. Mỗi véc-tơ mô tả là một ma trận 4x4 các tổ chức đồ. Mỗi tổ chức đồ có 8 khoảng tương ứng với 8 hướng. Do đó, mỗi véc-tơ mô tả SIFT là một véc-tơ 4x4x8=128 chiều. Lúc này, mỗi ảnh được biểu diễn bởi một tập các véc-tơ mô tả SIFT.

Hình 2: Đặc trưng cục bộ SIFT được tính toán từ vùng xung quanh điểm đặc biệt (vòng tròn): gradient của ảnh (trái), véc-tơ mô tả (phải)



Bước kế tiếp là thiết lập các từ trực quan từ các mô tả cục bộ đã được tính ở bước trước. Thuật giải k-means [15] được thực hiện trên các véc-tơ mô tả để phân các véc-tơ SIFT thành vào các nhóm (cluster) và mỗi cluster tương ứng với một từ trực quan. Tập các cluster này tạo thành

một từ điển. Sau cùng, mỗi véc-tơ mô tả trong ảnh sẽ được gán vào cluster gần nhất (khoảng cách mỗi véc-tơ đến các tâm điểm của các cluster đại diện đã được định nghĩa trước đó). Tiếp theo, một ảnh sẽ được biểu diễn bằng tần số của các từ trực quan trong ảnh. Hình 3 mô tả các bước tạo mô hình BoVW biểu diễn các ảnh.



Hình 3: Tạo mô hình BoVW để biểu diễn ảnh

Giai đoạn tiền xử lý cho một tập dữ liệu với số chiều rất lớn (ví dụ, 3000 từ trực quan với nhiều đặc trưng đầu vào với mỗi đặc trưng chỉ chứa ít thông tin cho phân lớp).

3 THUẬT GIẢI ARCX4-RMNB

Việc biểu diễn ảnh bằng túi từ trực quan đưa vấn đề phân lớp ảnh về bài toán phân loại văn bản sử dụng mô hình túi từ. Có nhiều thuật toán máy học cho bài toán này [20], một giải thuật đơn giản Bayes thơ ngây Multinomial Naïve Bayes (MNB [11]) thường được sử dụng để phân lớp văn bản.

3.1 Giải thuật Multinomial Naïve Bayes (MNB)

Phương pháp ước lượng xác suất khi phân lớp một văn bản của mô hình MNB được trình bày tóm tắt như sau. Giả sử C là tập hợp các lớp của văn bản. Tập các từ vựng của văn bản có kích thước là N . Khi có một văn bản mới đến là t_i thì mô hình MNB gán lớp cho t_i sao cho ước lượng xác suất để t_i thuộc vào một lớp c_i là lớn nhất hay là tìm giá trị lớn nhất của $Pr(c|t_i)$. Ước lượng xác suất $Pr(c|t_i)$ được tính như sau:

$$Pr(c | t_i) = \frac{Pr(c)Pr(t_i | c)}{Pr(t_i)} \quad c \in C \quad (1)$$

Trong công thức (1), xác suất $Pr(c)$ được tính bằng tổng số văn bản của lớp c chia cho tổng số văn bản của tất cả các lớp. Trong tính toán tìm giá trị lớn nhất của $Pr(c|t_i)$, người ta có thể bỏ qua $Pr(t_i)$ do nó không đổi khi ước lượng xác suất của từng lớp.

Xác suất $Pr(t_i|c)$ được tính bằng công thức (2) như sau:

$$Pr(t_i | c) = \left(\sum_n f_{ni}\right)! \prod_n \frac{Pr(w_n | c)^{f_{ni}}}{f_{ni}!} \quad (2)$$

Trong công thức (2), f_{ni} là tần suất từ thứ n trong t_i và $Pr(w_n|c)$ là xác suất của từ thứ n khi cho trước lớp c . $Pr(w_n|c)$ có thể được ước lượng bằng cách lấy tần suất từ thứ n trong tất cả các văn bản của lớp c chia cho tổng số tần suất của các từ vựng trong các văn bản của lớp c . Hơn nữa, $\left(\sum_n f_{ni}\right)!$ và $\prod_n f_{ni}!$ trong công thức (2) có thể thay bằng hằng số chuẩn hóa α mà không làm thay đổi kết quả. Việc ước lượng xác suất

$Pr(t_i|c)$ của công thức (2) được tính bằng công thức (3) như sau:

$$Pr(t_i | c) = \alpha \prod_n Pr(w_n | c)^{f_{ni}} \quad (3)$$

Mặc dù cách tính của MNB đơn giản, nhưng lại cho hiệu quả khá tốt khi phân lớp dữ liệu văn bản [11]. Tuy nhiên, khi dữ liệu có số chiều quá lớn, mô hình MNB đơn không còn hiệu quả khi so sánh với mô hình máy học SVM [21] hay các kỹ thuật tập hợp mô hình [2], [3] và [9].

Đầu vào:

- m phần tử dữ liệu : $\{(x_i, y_i)\}_{i=1, m}$ với $x_i \in R^n$ và $y_i \in \{1, -1\}$
- số bước lặp T

Huấn luyện:

- ▶ khởi động trọng số của m phần tử dữ liệu $Dist_1(j)$ cho $j = 1$ tới m thực hiện $Dist_1(j) = 1/m$

- ▶ cho $i = 1$ tới T thực hiện (lặp T bước)
 - lấy mẫu S_i phần tử dựa trên trọng số $Dist_i$
 - học mô hình Bayes thơ ngây ngẫu nhiên h_i từ tập mẫu S_i $h_i = rMNB(S_i)$
 - tính lại lỗi dự đoán của từng phần tử x_j khi sử dụng các bộ phân lớp được xây dựng trước đó

$$\epsilon_j = \sum_{t=1}^i h_t(x_j) \neq y_j$$

- cập nhật lại trọng số của m phần tử dữ liệu cho $j = 1$ tới m thực hiện

$$Dist_{i+1}(j) = (1 + \epsilon_j^4)/fac_i \quad \text{với } fac_i = \sum_{j=1}^m (1 + \epsilon_j^4)$$

- ▶ trả về tập T mô hình Bayes thơ ngây ngẫu nhiên $\{h_i\}_{i=1, T}$

Phân lớp:

- ▶ phân lớp phần tử x : bình chọn số đông của $\{h_i(x)\}_{i=1, T}$

Giải thuật 1: ArcX4 Bayes thơ ngây ngẫu nhiên

Boosting (AdaBoost [9]) là kỹ thuật áp dụng một tập các bộ phân lớp yếu (weak learner) để nâng cao hiệu quả của các bộ phân lớp này bằng cách giảm bias và variance. Trong cùng thời điểm [2] cũng đề xuất lớp các giải thuật Arcing nhằm giảm cả bias và variance. Theo Breiman, Boosting là một dạng trong lớp giải thuật Arcing. Trong đó có giải thuật ArcX4 cho kết quả tương tự như AdaBoost.

3.2 Giải thuật ArcX4-rMNB

Trong [2], Breiman đã nghiên cứu phân tích hiệu quả giải thuật học dựa trên cơ sở của hai thành phần lỗi là bias và variance. Trong đó thành phần lỗi bias là lỗi của mô hình học và variance là lỗi do tính biến thiên của mô hình so với tính ngẫu nhiên của các mẫu dữ liệu học. Trong nghiên cứu kết hợp nhiều mô hình phân loại yếu thành tập hợp các mô hình phân loại để cho tính chính xác cao hơn so với chỉ một mô hình đơn.

mỗi bước học, các trọng số này sẽ được cập nhật lại (tăng trọng số cho các phần tử bị phân lớp sai). Ở bước thứ i , ta lấy tập mẫu S_i trên tập dữ liệu và xây dựng mô hình h_i từ tập mẫu S_i . Lặp lại quá trình này sau T bước, ta sẽ được T mô hình cơ sở, kết hợp các mô hình cơ sở này lại ta sẽ có được một bộ phân lớp mạnh. ArcX4 rất dễ cài đặt và cho kết quả tốt trong thực tế.

ArcX4 thường dùng giải thuật cơ sở là cây quyết định C4.5 [15]. Các nghiên cứu của [25] chỉ ra rằng Boosting và Arcing mặc dù cho kết quả tốt trong thực tế nhưng thường bị học vệt khi tăng số bước lặp vượt qua một ngưỡng nào đó. Để khắc phục nhược điểm này, [26] đề xuất sử dụng mô hình cơ sở cây phải đơn giản (cây có kích thước không quá 8 nút), khi đó số bước lặp tăng cao vẫn đảm bảo rằng Boosting và Arcing không bị tình trạng học vệt.

Xuất phát từ ý tưởng của [26] kết hợp với ý tưởng được đề xuất bởi [3] cho cây ngẫu nhiên. Chúng tôi đã đề xuất xây dựng giải thuật Bayes thơ ngây ngẫu nhiên. Thay vì giải thuật MNB sử dụng toàn bộ tập các thuộc tính (chiều) để dự báo nhãn thì Bayes thơ ngây ngẫu nhiên (rMNB) chỉ sử dụng tập con các thuộc tính được lấy ngẫu nhiên từ tập thuộc tính ban đầu. Việc ước lượng xác suất $Pr(t_i|c)$ của công thức (3) được tính bằng công thức (4) bằng cách thay thế n từ vệt bởi n' từ vệt lấy ngẫu nhiên từ n từ vệt.

$$Pr(t_i | c) = \alpha \prod_{n'} Pr(w_{n'} | c)^{f_{n'i}} \quad (4)$$

Trong các nghiên cứu thực nghiệm cho thấy chỉ cần lấy tập con, n' thuộc tính khoảng căn bậc 2 của tổng số n thuộc tính ($n' = \sqrt{n}$). Mặc dù mô hình đơn của rMNB thì không mạnh do quá đơn giản, nhưng khi áp dụng kỹ thuật ArcX4 để xây dựng tập hợp các mô hình rMNB thì hiệu quả của giải thuật được cải thiện rất nhiều. So với mô hình ArcX4 sử dụng MNB với tập đầy đủ các thuộc tính thì giải thuật ArcX4-rMNB đơn giản, nhanh hơn, hiệu quả phân lớp tốt hơn do khả năng chịu đựng nhiễu cao hơn.

4 KẾT QUẢ THỰC NGHIỆM

Chúng tôi tiến hành đánh giá hiệu quả của phương pháp đề xuất (mô hình túi từ trực quan SIFT-BoVW và giải thuật Arcx4-rMNB) cho phát hiện ảnh khiêu dâm. Chúng tôi muốn so sánh hiệu quả của giải thuật Arcx4-rMNB với các giải thuật học khác, bao gồm MNB [11], SVM [21], cây quyết định C4.5 [17], AdaBoost của C4.5 (AdaBoost-C4.5 [9]) và rừng ngẫu nhiên (RF-CART [3]). Chúng tôi tiến hành cài đặt giải 2 giải thuật MNB và ArcX4-rMNB bằng ngôn ngữ lập trình C/C++. Các giải thuật khác đã có trong các thư viện phần mềm miễn phí, như LibSVM [5], thư viện Weka [22] có sẵn các giải thuật C4.5, RF-CART, AdaBoost-C4.5 và kNN.

4.1 Kết quả thực nghiệm

Chúng tôi tiến hành thực nghiệm trên hai tập dữ liệu thực. Tập dữ liệu thứ nhất có chứa 1414 ảnh và tập dữ liệu lớn thứ 2 có 14971 ảnh trong hai lớp (khiêu dâm và không khiêu dâm - thường) được mô tả ở bảng 1. Ảnh không khiêu dâm được thu thập từ các hình ảnh video quảng cáo đồ lót và người mẫu triển lãm xe hơi, các ảnh khác trên internet.

Bảng 1: Tập dữ liệu ảnh

STT	Tập dữ liệu	Ảnh khiêu dâm	Ảnh thường
1	Tập nhỏ	484	930
2	Tập lớn	6944	8027

Để biểu diễn ảnh bằng mô hình túi từ trực quan, chúng tôi sử dụng giải thuật phát hiện điểm đặc trưng cục bộ Hessian Affine của [16] để rút trích các véc-tơ mô tả SIFT. Sau đó, thực thi giải thuật k -means [15] để gom nhóm các véc-tơ mô tả SIFT vào 3000 clusters tương ứng với 3000 từ trực quan. Giai đoạn tiền xử lý tạo ra hai tập (bảng) dữ liệu với 1414 và 1971 phần tử với 3000 chiều tương ứng. Chúng tôi cố gắng thay đổi số clusters (từ trực quan từ 1000 đến 50000) để tìm các kết quả thực nghiệm tốt nhất. Cuối cùng, chúng tôi thu được độ chính xác ổn định với 3000 từ trực quan.

Để so sánh hiệu suất của các thuật toán phân lớp, chúng tôi tiến hành phân tích hiệu quả của các thuật toán phân lớp dựa trên các tiêu chí sau:

$$TP\ Rate = Recall = tp/(tp+fn)$$

$$TN\ Rate = tn/(tn+fp)$$

$$Precision = tp/(tp+fp)$$

$$F1\text{-Measure} = \frac{2 * Precision * Recall}{Precision + Recall}$$

$$Accuracy = (tp + tn)/(tp+fp+tn+fn)$$

Với các thông số để tạo thành các tiêu chí:

tp: số ảnh khiêu dâm được mô hình phân lớp là khiêu dâm

fp: số ảnh bình thường bị mô hình phân lớp là khiêu dâm

fn: số ảnh khiêu dâm bị mô hình phân lớp là bình thường

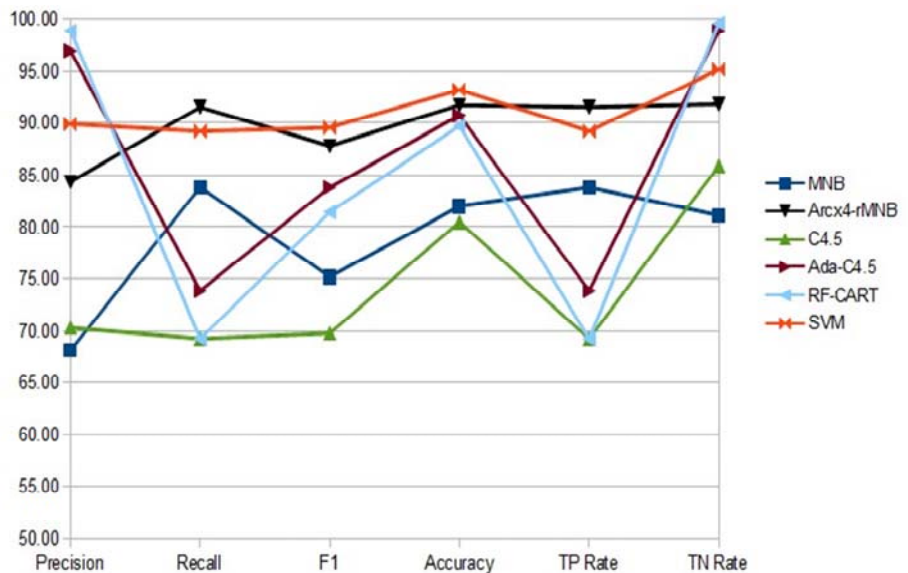
tn: số ảnh bình thường được mô hình phân lớp là bình thường

Chúng tôi đề xuất sử dụng nghi thức kiểm tra chéo (hold-out) được áp dụng để đánh giá hiệu quả của các giải thuật phân lớp. Cách làm như sau: lặp lại 3 lần thực nghiệm, mỗi lần thực nghiệm lấy ngẫu nhiên 2/3 tập dữ liệu làm tập

huấn luyện mô hình và 1/3 dữ liệu còn lại dùng làm tập kiểm tra kết quả phân lớp. Chúng tôi sử dụng tập huấn luyện để điều chỉnh các thông số cho các thuật toán Arcx4-rMNB, AdaBoost-C4.5, RF-CART, LibSVM thu được độ chính xác tốt nhất trong từng giai đoạn. Kết quả cuối cùng là trung bình của 3 lần thực thi như vừa mô tả. Chúng tôi cố gắng sử dụng các hàm nhân (kernel function) của thuật toán SVM gồm hàm đa thức bậc d , Radial Basis Function (hàm nhân RBF của hai điểm dữ liệu x_i, x_j là $K[i, j] = \exp(-\gamma \|x_i - x_j\|^2)$). Giải thuật SVM sử dụng hàm nhân RBF (với $\gamma=0.0002$) cho kết quả tốt nhất. Arcx4-rMNB học 200 phân lớp yếu (rMNB) sử dụng 300 chiều ngẫu nhiên để ước lượng các xác suất phân lớp. RF-CART xây dựng 200 cây sử dụng 300 chiều ngẫu nhiên để thực hiện phân hoạch tại các nút. AdaBoost-C4.5 cũng thực hiện phân lớp sử dụng 200 cây.

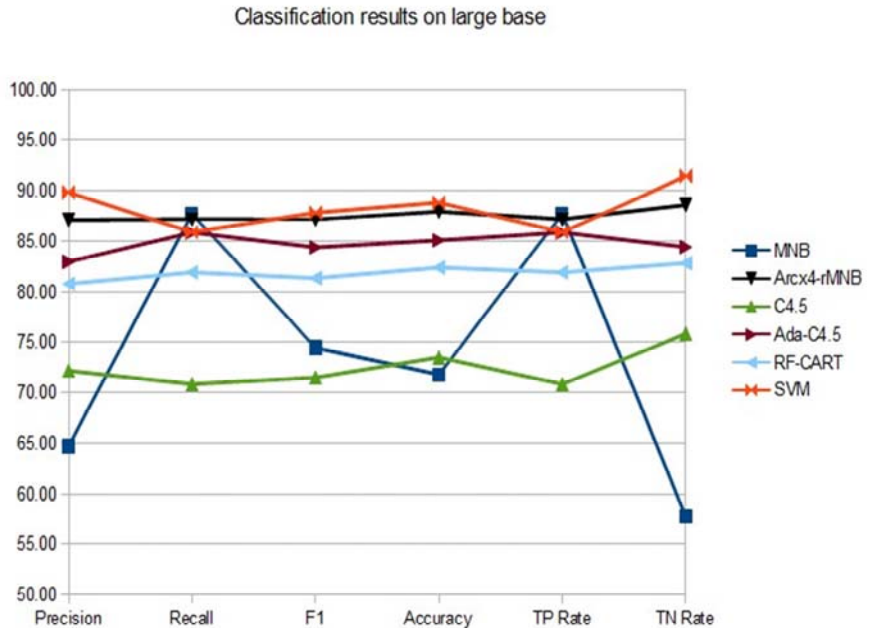
Kết quả thu được từ các giải thuật được hiển thị trực quan trong 2 đồ thị như trình bày trong hình 4 (tập dữ liệu nhỏ) và hình 5 (tập dữ liệu lớn). Chúng tôi tiến hành so sánh và phân tích kết quả phân lớp dựa trên các tiêu chí Precision, Recall, F1, Accuracy, tỉ lệ TP và TN.

Classification results on small base



Hình 4: Kết quả phân lớp trên tập dữ liệu nhỏ

Hình 5: Kết quả phân lớp trên tập dữ liệu lớn



Khi so sánh kết quả thu được bởi giải thuật Arcx4-rMNB với các thuật giải khác. Giải thuật ArcX4-rMNB thực hiện hiệu quả hơn giải thuật MNB [11], C4.5 [17], RF-CART [3], AdaBoost of C4.5 [9] qua các tiêu chí đánh giá hiệu quả phát hiện ảnh khiêu dâm như F1- measure và Accuracy

Kết quả thực nghiệm cũng cho thấy giải thuật ArcX4-rMNB phát hiện ảnh khiêu dâm chính xác hơn LibSVM nhưng ngược lại LibSVM cho kết quả tốt hơn ArcX4-rMNB đối với các ảnh bình thường. Khi xét các tiêu chí khác thì ArcX4-rMNB có kết quả tương đương với LibSVM. Việc ArcX4-rMNB có kết quả tốt khi so sánh với LibSVM là thành công lớn bởi vì LibSVM được xem là giải thuật nhận dạng rất chính xác hiện nay.

Với các kết quả phân lớp này, chúng tôi tin rằng giải thuật ArcX4-rMNB thực hiện tốt cho vấn đề phát hiện ảnh khiêu dâm khi so sánh với các giải thuật như MNB, SVM, cây quyết định C4.5, AdaBoost-C4.5 và rừng ngẫu nhiên RF-CART.

5 KẾT LUẬN VÀ ĐỀ XUẤT

Chúng tôi vừa trình bày một phương pháp mới đạt được hiệu quả cao trong việc phân lớp các ảnh khiêu dâm. Ý tưởng chính xuất phát từ

kết hợp biểu diễn ảnh bằng mô hình túi từ trực quan và giải thuật ArcX4-rMNB. Ở bước tiền xử lý, thực hiện rút trích đặc trưng không đổi SIFT từ ảnh. Bước kế tiếp là thiết lập các từ trực quan từ các mô tả cục bộ đã được tính dựa vào thuật giải *k*-means. Sau bước tiền xử lý, ảnh được biểu diễn bởi một véc-tơ có số chiều rất lớn, chúng tôi đề xuất một giải thuật mới ArcX4-rMNB cho phép phân lớp hiệu quả dữ liệu có số chiều lớn. Kết quả thực nghiệm trên tập dữ liệu thực tế cho thấy phương pháp của chúng tôi đề xuất đạt được chính xác 91.75% cho tập dữ liệu nhỏ và 87.93% cho tập dữ liệu lớn. Giải thuật ArcX4-rMNB thực hiện tốt hơn các thuật giải cây quyết định C4.5 [17], AdaBoost của C4.5 (AdaBoost-C4.5 [9]) và rừng ngẫu nhiên (RF-CART [3]). So sánh hiệu quả với các giải thuật như MNB, SVM, cây quyết định C4.5, AdaBoost-C4.5 và rừng ngẫu nhiên RF-CART cũng cho thấy rằng giải thuật ArcX4-rMNB thực hiện tốt cho vấn đề phát hiện ảnh khiêu dâm.

Trong tương lai, chúng tôi dự định ứng dụng phương pháp đề xuất vào vấn đề phân lớp video khiêu dâm. Bên cạnh đó, chúng tôi cũng muốn tăng tốc quá trình xây dựng mô hình học ArcX4-rMNB bằng việc xây dựng giải thuật song song.

TÀI LIỆU THAM KHẢO

1. Bosch, A., Zisserman, A., Munoz, X.: Scene classification via pLSA. In: Proceedings of the European Conference on Computer Vision, pp. 517–530 (2006).
2. Breiman, L.: Arcing classifiers. *The annals of statistics* 26(3), 801–849 (1998)
3. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001).
4. Chang, C.C., Lin, C.J.: LIBSVM – a library for support vector machines (2001). <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
5. Deselaers, T., Pimenidis, L., Ney, H.: Bag-of-visual-words models for adult image classification and filtering. In: Proceeding of The 19th International Conference on Pattern Recognition, pp. 1–4 (2008).
6. Duan, L., Cui, G., Gao, W., Zhang, H.: Adult image detection method base-on skin colormodel and support vector machine. In: Proceeding of The 5th Asian Conference on Computer Vision, pp. 797–800 (2002).
7. Fleck, M., Forsyth, D., Bregler, C.: Finding naked people. In: Proceedings of the European Conference on Computer Vision, vol. 2, pp. 592–602 (1996).
8. Forsyth, D., Fleck, M.: Identifying nude pictures. In: Proceedings of the IEEE Workshop on the Applications of Computer Vision, pp. 103–108 (1996)
9. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Computational Learning Theory: Proceedings of the Second European Conference, pp. 23–37 (1995).
10. Jeong, C., Kim, J., Hong, K.: Appearance-based nude image detection. In: Proceedings of The 17th International Conference on Pattern Recognition, pp. 467–470 (2004).
11. Lewis, D., Gale, W.: A sequential algorithm for training text classifiers. In: Proceedings of SIGIR (1994).
12. Lopes, A., Avila, S., Peixoto, A., Oliveira, R., Coelho, M., Araujo, A.: Nude detection in videousing bag-of-visual-feature. In: Proceedings of The 22th Brazilian Symposium on Computer Graphics and Image Processing, pp. 224–231 (2009).
13. Lowe, D.: Object recognition from local scale invariant features. In: Proceedings of the 7th International Conference on Computer Vision, pp. 1150–1157 (1999).
14. Lowe, D.: Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision* pp. 91–110 (2004).
15. MacQueen, J.: Some methods for classification and analysis of multivariate observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, *University of California Press* Vol.1, pp. 281–297 (1967).
16. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. *International Journal of Computer Vision* 60(1), 63–86 (2004)
17. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA (1993).
18. Van Rijsbergen, C.V.: *Information Retrieval*. Butterworth (1979).
19. Schettini, R., Brambilla, C., Cusano, C., Ciocca, G.: On the detection of pornographic digital images. In: Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, pp. 2105–2113 (2003).
20. Sebastiani, F.: Machine learning in automated text categorization. *ACM Computing Surveys* 34(1), 1–47 (1999).
21. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer-Verlag (1995).
22. Wang, Y., Wang, W., Gao, W.: Research on the discrimination of pornographic and bikiniimages. In: Proceedings of the Seventh IEEE International Symposium on Multimedia, pp. 558–564 (2005).
23. Witten, I., Frank, E.: *DataMining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann (2005).
24. Zheng, H., Daoudi, M.: Blocking adult images based on statistical skin detection. *Electronic Letters on Computer Vision and Image Analysis* 4(2), 1–14 (2004).
25. Grove, A.J. and Schuurmans, D.: Boosting in the limit: Maximizing the margin of learned ensembles. In Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), pp. 692–699 (1998).
26. Friedman, J., Hastie, T. and Tibshirani, R.: Response to Mease and Wyner, Evidence Contrary to the Statistical View of Boosting. *Journal Machine Learning Research* Vol. 9, pp. 175–180 (2008).