

Metodologia para Indexação e Busca de Impressões Digitais através do uso de Função de Distância Métrica.

Evandro Jardini

Fundação Educacional de Fernandópolis
Departamento de Informática - eajardini@yahoo.com.br

Adilson Gonzaga

SEL/EESC-USP São Carlos
agonzaga@sc.usp.br

Caetano Traina Jr.

ICMC-USP São Carlos
caetano@icmc.usp.br

Abstract

To recognize and to authenticate individuals through characteristics biometrics comes turning usual in the day by day. The advantages of this authentication method on the traditional "tag/password" are countless. Besides, biometrics databases are growing more and more, influencing in the time of search of the information. This article presents a recognition system and indexation of fingerprints through function of metric distance.

1 Introdução

O reconhecimento e autenticação de um indivíduo através do uso de um autenticador e senha, apesar de muito utilizados, possui uma série de inconveniências como esquecimento, roubo, cópia indevida, perda, etc. Estes problemas sugerem a troca por meios mais seguros de autenticação. Entre os meios mais seguros encontra-se o uso de características biométricas. As características biométrica são divididas em dois grupos: i) as características comportamentais como gestos, voz, escrita manual, modo de andar e assinatura e ii) as características fisiológicas que incluem as impressão digital, face, geometria da mão, íris, veias da retina, voz e orelha. Estas características são chamadas de identificadores biométricos.

Um dos identificadores biométricos mais usados é a impressão digital. Isto ocorre por ser um identificador que é formado nos primeiros meses de vida do indivíduo e permanece inalterado no decorrer de sua vida; sua coleta é não intrusiva e fica registrada em objetos que forem tocados pela pessoa.

Uma vez que os bancos de impressões digitais são volumosos, há a necessidade de pesquisas ágeis e para isso podem ser usadas estruturas de indexação. Entre as estruturas de indexação tem-se as que utilizam acesso métrico, conhecidas como MAM (Método de Acesso Métrico), que utilizam funções de distâncias métricas como fator de indexação. Este artigo demonstra os resultados obtidos com uma metodologia original de identificação de impressões digitais baseado em uma função de distância métrica. Por ser uma metodologia baseada em função de distância, a mesma pode ser indexada em uma estrutura do tipo MAM.

2 Biometria e Identificadores Biométricos

A Biometria é um conjunto de métodos automatizados, com base em características comportamentais (gestos, voz, escrita manual, modo de andar e assinatura) e fisiológicas (impressão digital, face, geometria da mão, íris,

veias da retina, voz e orelha). Estas características são chamadas de identificadores biométricos [1]. Dentre os identificadores biométricos, o mais usado é a impressão digital. Trabalhos que utilizam outros identificadores podem ser encontrados em [2], [3] e [4]

A impressão digital é formada por sulcos presentes nos dedos. A parte alta dos sulcos é denominada de crista e a baixa de denominada de vale. Para o reconhecimento de impressões digitais, as cristas apresentam as características desejadas. Seguindo o fluxo das cristas nota-se a formação de dois pontos característicos usado para a identificação de indivíduos, as minúcias. Dos tipos existentes de minúcias, os dois mais utilizados para o reconhecimento de impressões digitais são: i) *minúcia do tipo cristas finais* e ii) *minúcia do tipo cristas bifurcadas*. A literatura de trabalhos biométricos realizados sobre impressões digitais é vasta. Frequentemente, as pesquisas recaem sobre as áreas de classificação de impressões digitais [5], detecção de minúcias [6] e processo de reconhecimento automático de impressões digitais [7].

3 Extração de Características e Indexação em Banco de Dados de Imagens

Nos Banco de Dados de Imagens (BDI), as imagens são armazenadas em grande volume. A quantidade pode chegar a *terabytes*. O acesso seqüencial a estas imagens compromete o desempenho do sistema tornando-o incapaz de ser usado para realização de tarefas no dia a dia-a-dia. Para diminuir o tempo de busca, estruturas de indexação são utilizadas. A indexação é feita em cima das características da própria imagem e exigem um pre-processamento para extração destas características. Uma vez extraídas as características, estas podem ser indexadas usando o Método de Acesso Métrico (MAM), usado para indexar dados complexos usando as distâncias entre os objetos. Para isto, os objetos devem estar contidos em um espaço métrico.

Um espaço métrico é uma coleção de objetos e uma função de distância definida sobre eles. Uma vez que um objeto pertence a um espaço métrico, pode-se utilizar estrutura de indexação (também denominada como árvore de busca) sobre eles para acelerar o processo de busca. Das diversas estruturas existentes de indexação pertencentes aos MAMs, tem-se a *SLIM-tree*. Esta estrutura foi criada por Traina *et al* [8]. A Slim-tree é uma árvore não é estática. Isto significa que permite que novos objetos sejam inseridos e retirados dinamicamente. Ela segue a mesma idéia de outras árvores métricas onde os dados são inseridos nas folhas e o balanceamento é feito pela altura.

4 Metodologia do Trabalho

Neste trabalho é apresentado um sistema de identificação de impressões digitais (ID), denominado de SI-IDM (*Sistema de Identificação de Impressões Digitais Métrico*), que utiliza uma inovadora função de distância métrica responsável pelo cálculo da diferença entre as impressões.

O SIIDM é um repositório indexado de vetores de características de ID. A indexação dos vetores é feita através da MAM *SLIM-Tree*. O funcionamento do SIIDM é mostra no diagrama de blocos da figura 1. Nela nota-se duas fases, **Fase I** e **Fase 2**. A Fase 1 é responsável pela extração das características das impressões digitais e o armazenamento do vetor destas características. Os vetores são armazenados de forma indexada.

O vetor de característica de uma ID tem o formato representado pela figura 2. O vetor é formado por duas listas principais. A primeira lista (representa pela letra M) contém as minúcias detectadas na ID. A segunda lista (representada pela letra V) contém a vizinhança de uma minúcia. Cada posição de lista é formada por uma estrutura representando uma minúcia. O número de minúcias detectadas em uma ID não é exato. Para este trabalho foram consideradas as IDs com mais de 10 minúcias e no máximo 100. Para cada minúcia são considerados as seguintes

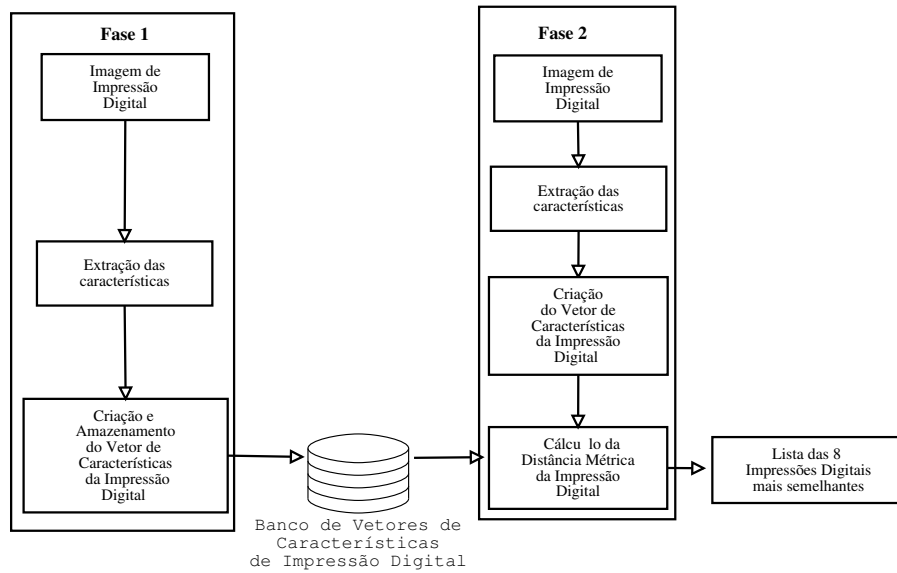


Figura 1: Diagrama de blocos do SIIDM

características: as coordenadas X e Y da minúcia, a direção da minúcia e uma lista de minúcias vizinhas contendo também suas coordenadas, direções e a quantidade de cristas entre a minúcia e sua vizinha.

Todo trabalho de armazenamento em disco dos vetores do SIIDM é feito pela *SLIM-Tree*. A mesma já possui uma série rotinas de manipulação de objetos, tanto em disco quanto em memória. Uma vez criado o banco de vetores de características, o SIIDM está pronto para o processo de identificação de IDs. Vale lembrar que novos vetores de IDs podem ser acrescentados no banco de dados.

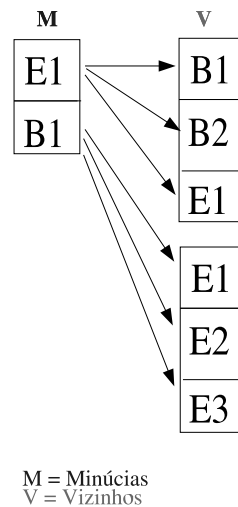


Figura 2: Formato do vetor de característica

Na fase 2 tem-se o processo de identificação de IDs. É usado uma ID de entrada e esta é comparada com as IDs previamente cadastradas no banco de dados pela fase 1. Para cada ID de entrada, suas características são extraídas e o vetor de característica é criado. O vetor, então, é comparado com os do banco de dados. O resultado é uma lista contendo de 1 a 8 IDs mais parecidas com a ID de entrada. A comparação é feita através de busca de padrões semelhantes. Segundo a lista representada na figura 2, para cada minúcia existe uma lista de minúcias vizinhas. Existem dois padrões que são extraídos destes dados: as distâncias entre as minúcias e o triângulo formado por três minúcias. Os padrões encontrados nas minúcias da ID de busca e seus vizinhos, são procurados nas minúcias das IDs armazenadas no banco de dados.

Cada padrão encontrado, é acrescido uma pontuação. Essa pontuação vai até um limite máximo. Então, quanto mais próximo for a pontuação obtida entre as IDs, mais semelhantes elas são.

4.1 Resultados

Para os testes, foi utilizado um banco de ID usado na *Fingerprint Verification Competition* (FVC)¹ do ano de 2000 e que está disponível no livro *Handbook of Fingerprint Recognition* [9]. O banco contém 800 IDs. Nele existem 100 classes de IDs e para cada classe, existem 8 IDs. As imagens de uma uma classes possui diferenças quanto rotação e posicionamento. Isto pode ser observado pela figura 3. Nota-se que na figura 3(a), o delta (representado por um círculo em negrito) esta posicionado mais abaixo na imagem. Já na figura 3(b), o delta esta posicionado mais a esquerda na imagem.



(a) Delta deslocado mais a baixo da imagem

(b) Delta deslocado mais a esquerda da imagem

Figura 3: Imagens de impressões digitais de uma mesma classe.

Foram realizadas duas classes de testes. Uma para determinar a precisão da função de distância e outra para verificar o tempo de busca entre acesso sequencial e acesso indexado.

Para a primeira classe de teste foi traçado a curva *Precision x Recall* (PxR). É importante ressaltar que não foi encontrada na literatura uma curva PxR indicando a precisão de uma função de distância para a identificação de IDs, sendo este trabalho o primeiro. O que se encontra são curvas *Receiver Operating Characteristic* (ROC). Posteriormente também será calculado a curva ROC para este trabalho.

¹O FVC é uma competição internacional para avaliar sistemas de reconhecimento de impressões digitais. Disponível em <http://bias.csr.unibo.it/fvc2000>

A figura 4(a) demonstra a curva PxR para a ID da figura 3(a). Nota-se pelo gráfico, que para esta classe de impressão digital, que a função de distância do SIIDM obteve um resultado muito bom. Chegando ao acerto das 8 imagens da classe. Entretanto, a base possui imagens de impressões digitais com baixa qualidade ou com grande variação de posicionamento. Isto implica sensivelmente na tarefa de identificação. O gráfico PxR da base inteira pode ser observado na figura 4(b).

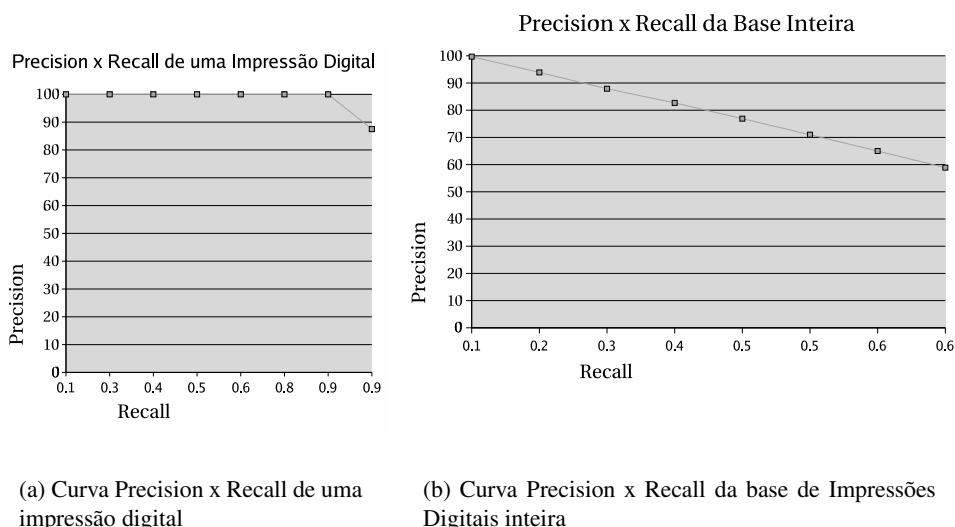


Figura 4: Curvas *Precision x Recall* .

A segunda classe de testes refere-se aos tempos gastos na busca das IDs. Este trabalho propõe a indexação das IDs através da MAM *SLIM-Tree* e compara a busca indexada com a busca seqüencial. Os testes foram realizados da seguinte maneira: Como há 8 IDs para cada classe, mediu-se o tempo de busca para recuperar uma ID, em seguida duas IDs, assim sucessivamente até que as 8 IDs sejam recuperadas. Este processo foi realizado para um banco indexado e outro sem indexação. Para realização dos testes, foi utilizado um computador com processador AMD *Athlon* 2.0 ghz com sistema operacional Windows 2000. Os resultados podem ser observados nas figuras 5(a) e 5(b).

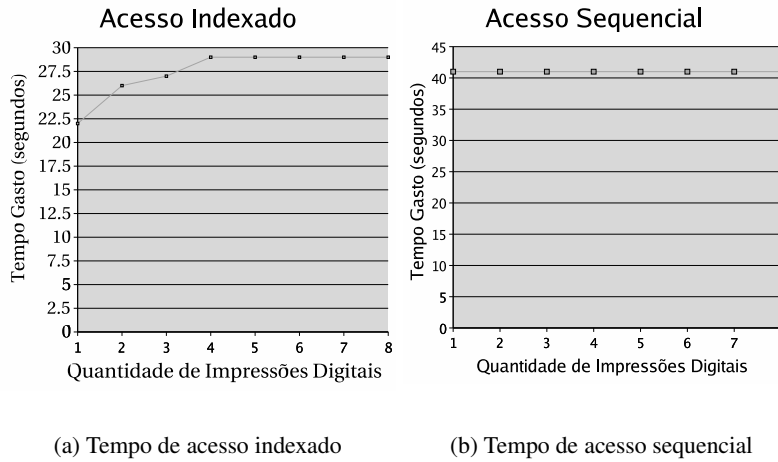


Figura 5: Tempo de acesso indexado e sequencial.

Na figura 5(a) mostra o gráfico de tempo de busca em segundos para uma determinada classe de ID. Nota-se que o tempo para a recuperação de uma ID levou 27 segundos, para duas levou 31 segundos e sendo que para as oito o tempo gasto foi de 36 segundos.

Na figura 5(b) tem-se o gráfico de tempo de busca em segundos para acesso sequencial para a mesma classe de ID usada para o teste indexado. Pelo gráfico percebe-se que não há variação no tempo de busca independente da quantidade de impressões digitais buscadas. Além disso, o tempo gasto para recuperação das oito IDs foi de 41 segundos, maior que o tempo gasto com o banco de dados indexado. Isto demonstra que o uso da MAM *SLIM-Tree* é vantajoso para banco de dados de IDs.

5 Conclusão

O artigo demonstrou o SIIDM (Sistema de Identificação de Impressões Digitais Métrico), que é um sistema de identificação de impressões digitais que utiliza uma inovadora função de distância Métrica. A identificação de impressões digitais não é tarefa fácil devido a posicionamento dos dedos, qualidade da imagem, etc. Os resultados foram apresentados através de gráficos *Precision x Recall* (PxR). Como não existem gráficos PxR para este tipo de trabalho, fica difícil fazer um comparativo deste sistema com outros. O SIIDM trabalha com um banco de dados de impressões digitais indexado através da MAM *SLIM-Tree*. Através dos testes, verificou-se que o uso desta MAM possibilitou obter ganhos no tempo de busca em relação ao acesso sequencial.

Referências

- [1] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki, "An introduction to evaluating biometric systems," *IEEE Computer*, pp. 56–63 February, 2000.
- [2] W. Boles and B. Boashash, "A human identification technique using images of the iris and wavelet transform," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 1185–1188, 1998.

- [3] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterisation of human faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 831–835, December, 1990.
- [4] S. J. Vaughan-Nichols, "Voice authentication speaks to the marketplace," *IEEE Computer*, March, 2004.
- [5] D. Maio and D. Maltoni, "A structural approach to fingerprint classification," *IEEE Proceedings of ICPR'96*, pp. 578–585, 1996.
- [6] F. Zhao and X. Tang, "Preprocessing for skeleton-based fingerprint minutiae extration," *CISST Internation Conference*, pp. 742–745, 2002.
- [7] A. Jain, L. Hong, S. Pankanti, and R. Bolle, "An identity authentication system using fingerprints," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1365–1388, september, 1997.
- [8] C. TRAINA Jr., A. Traina, B. Seeger, and C. Faloutsos, "Slim-Trees: High performance metric trees minimizing overlap between nodes," *Lecture Notes in Computer Science*, vol. 1777, pp. 51–65, 2000.
- [9] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. New York: Springer, 1 ed., 2003.