

Relazione di tirocinio

Alessandro Bagnoli

651803

alessandro.bagnoli4@studio.unibo.it

31 ottobre 2015

1 ARGOMENTO, LUOGO, DATA E DURATA DELL'ATTIVITÀ

Il tirocinio curriculare previsto dal corso di laurea è stato svolto all'interno del *Biometric System Laboratory* (BioLab) del Dipartimento di Informatica - Scienza e Ingegneria (DISI) a Cesena. Le principali attività di ricerca svolte in questo laboratorio sono dedicate al riconoscimento di impronte digitali e del volto, e alla valutazione delle performance dei sistemi biometrici. I ricercatori che lavorano all'interno del BioLab sono organizzatori di una delle competizioni di riconoscimento di impronte più famose (FVC), svoltasi in più edizioni, e nel loro curriculum vantano la creazione e lo sviluppo di un software per la generazione di impronte digitali sintetiche (SFinGe).

L'attività di tirocinio è stata eseguita nel periodo compreso tra il 14 settembre 2015 e il 14 ottobre 2015, coprendo le 150 ore previste, e l'argomento principale è stato rappresentato appunto dalle impronte digitali con compiti e mansioni specifici che saranno descritti nel corso di questa relazione.

2 SCOPI, OBIETTIVI E MOTIVAZIONI

Ciò che mi ha spinto a svolgere un tirocinio interno piuttosto che in azienda è stato l'interesse che ho nutrito durante il corso di Fondamenti di Elaborazione di Immagini tenuto dal Prof. Raffaele Cappelli nel secondo semestre del terzo anno, e la mia curiosità nell'espandere e nel mettere in pratica le conoscenze acquisite dopo aver seguito questo corso e aver superato il relativo esame.

L'obiettivo principale del tirocinio era quello di migliorare le prestazioni di estrazione di minuzie da impronte digitali di bassa qualità. Un'impronta digitale è costituita da un insieme di linee, dette creste che scorrono in linee parallele, che a volte si intersecano oppure si interrompono, formando un disegno detto ridge pattern. A partire dal ridge pattern possono essere estratte ulteriori informazioni e le più importanti sono rappresentate dalle minuzie. Queste costituiscono un fattore importante per la discriminazione delle impronte infatti esse sono i punti in cui si può notare un comportamento anomalo delle creste. Le categorie principali di minuzie sono due: terminazioni e biforcazioni e possono essere osservate nell'immagine in Figura 2.1.

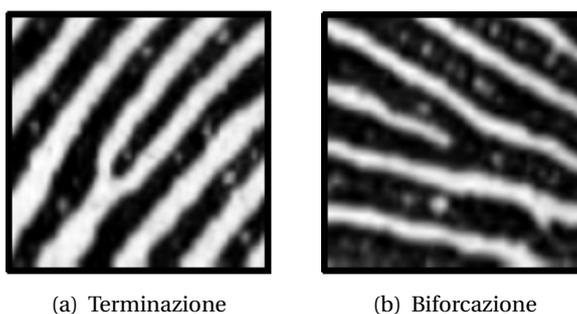


Figura 2.1: Le due principali categorie di minuzie

Le prestazioni degli algoritmi di estrazione delle minuzie e riconoscimento di impronte dipendono molto dalla qualità dell'immagine. Una parte non trascurabile delle immagini acquisite è di scarsa qualità per varie cause che vanno dalle condizioni della pelle (umida, secca o con tagli), ad una pressione non ideale del dito sul sensore o ad una bassa qualità intrinseca delle impronte di particolari categorie di persone (anziani, lavoratori manuali). Esempi di impronte di diverse qualità possono essere osservati in Figura 2.2. Esistono diversi tipi di degradazioni associate a immagini di impronte:

- Le creste non sono continue.
- Le creste parallele non sono ben separate a causa della presenza di rumore presente nell'immagine.
- Tagli, ferite e abrasioni.



Figura 2.2: Impronte di diverse qualità

3 PRESENTAZIONE GENERALE DELLE ATTIVITÀ SVOLTE

Le attività principali da svolgere nel BioLab si sono presentate come segue:

- Etichettatura manuale delle minuzie presenti in uno o più database contenente impronte digitali di bassa qualità utilizzando un'applicazione creata ad hoc
- Analisi degli errori su tali database creati da algoritmi di estrazione delle minuzie allo stato dell'arte
- Sviluppo di tecniche che consentano di migliorare le prestazioni dell'estrazione di minuzie in impronte digitali di bassa qualità

Nel capitolo seguente verranno descritte una per una queste tre diverse fasi di lavoro

4 METODO SEGUITO NELLE ATTIVITÀ

Prima di partire con l'etichettatura manuale delle minuzie, mi è stato altamente consigliato di leggere una parte del libro *Handbook of Fingerprint Recognition* per avere una maggiore cognizione di quello che sarebbe stato l'argomento chiave del tirocinio. Si è rivelata una lettura propedeutica e indispensabile per capire i concetti principali che ruotano attorno al mondo della fingerprint recognition.

In seguito sono stati selezionati insieme al Prof. Cappelli alcuni database di immagini di impronte in cui avrei dovuto etichettare manualmente le minuzie. I database comprendevano immagini di impronte di tutte le qualità, e per selezionare solo quelle di bassa qualità è stata utilizzata una semplice console application che mi separasse le immagini in base al loro indice di qualità, che da ora in avanti chiamerò NFIQ. Ogni database comprendeva oltre alle immagini anche dei file con estensione .ist che sta per ISO template. Questi file hanno lo stesso nome dell'immagine a cui sono associati e contengono in forma binaria i dati sulle minuzie estratte dall'immagine cui fanno riferimento. Le minuzie contenute negli ISO template dei database di analisi sono state estratte con uno degli algoritmi partecipanti all'FVC chiamato P088, che spicca per prestazioni migliori rispetto all'estrattore sviluppato all'interno del BioLab.

Una volta separate le immagini e i relativi ISO template in base al loro NFIQ è partita la fase di etichettatura manuale sulle immagini di qualità più bassa (NFIQ 4 e NFIQ 5) utilizzando un applicativo sviluppato sempre all'interno del BioLab chiamato *Minutiae Template Modifier*. Dato che non si partiva da zero con l'etichettatura ma, come detto prima, da dei template già prodotti, il mio compito era quello di correggere tali template, eliminando le false minuzie, posizionarne altre dove non sono state individuate, o correggere la tipologia di minuzia segnalata (biforcazione al posto di una terminazione o viceversa). Questa è stata forse la fase più noiosa del tirocinio poichè molto ripetitiva, ma indispensabile per i passi successivi.

Dopo aver corretto un discreto numero di template (precisamente 100), si è provveduto a estrarre le minuzie dalle stesse immagini utilizzando l'estrattore BioLab ed è iniziata la fase di analisi degli errori che questo algoritmo compie su tali immagini. Per questo compito è stato utilizzato un altro applicativo, *ISO Minutiae Template Quality Analyzer*, che confronta i template corretti manualmente (Ground Truth) con i template generati dagli algoritmi BioLab e P088, mostrando il numero di minuzie false e non trovate che ha generato un algoritmo confrontandolo direttamente col Ground Truth, insieme ad altre statistiche. Durante l'analisi parallela BioLab / Ground Truth effettuata impronta per impronta tramite l'applicativo si è notato come la maggior parte degli errori (soprattutto minuzie false) fossero presenti in impronte con molti tagli cutanei, impronte eccessivamente secche dove le creste erano poco distinguibili e in alcuni casi più rari dai pori della pelle.

Le immagini NFIQ 4 e NFIQ 5 estratte dal database di analisi erano spesso dello stesso dito, si è quindi implementato un piccolo programma in grado di calcolare i valori percentuali del *False Match Rate 100 (FMR100)* ed *Equal Error Rate (EER)* partendo da due file di testo, uno contenente i match genuine e l'altro i match impostor. Il False Match Rate rappresenta la probabilità che un algoritmo di riconoscimento classifichi come autentico (genuine) un confronto tra due impronte che in realtà sono di dita diverse (impostor) mentre il False Non Match Rate (FNMR) rappresenta la probabilità che tale algoritmo classifichi come impostor

un confronto che in realtà è di tipo genuino. L'EER denota invece il punto in cui, alla soglia t , l'FMR e l'FNMR assumono lo stesso valore.

Tramite lo stesso programma si sono elencate e in seguito analizzate le coppie di impronte che contribuivano agli errori a FMR100 e si sono effettuati calcoli statistici sulla media e sulla distribuzione del numero di minuzie estratte da queste coppie di impronte e le restanti. I calcoli effettuati hanno portato a notare che le impronte da cui erano state estratte un numero maggiore di una certa soglia (tra 140 e 150) erano tutte responsabili degli errori a FMR100.

In seguito a questa scoperta, si è deciso di effettuare prove al fine di diminuire il numero di minuzie che l'algoritmo estrae, per verificare se gli indici FMR100 e EER fossero calati di conseguenza. I tentativi effettuati durante le ultime ore di tirocinio rimaste a disposizione hanno riguardato la modifica di alcuni parametri che utilizza l'estrattore BioLab. Per verificare mano a mano il risultato delle modifiche si è utilizzata l'applicazione *Minutiae Extraction Lab*. Grazie alla modifica di alcune combinazioni di parametri dell'estrattore è stato possibile apprezzare un discreto miglioramento delle performance sul database preso in analisi.

5 MATERIALI E STRUMENTI UTILIZZATI

Per la scrittura e l'esecuzione di piccole console application come quelle per l'estrazione di minuzie con l'algoritmo BioLab, o per il calcolo degli errori EER e FMR100 è stato utilizzato l'IDE Microsoft Visual Studio e il linguaggio C#. Nelle Figure 5.1, 5.2, e 5.3 vengono mostrati gli applicativi con interfaccia grafica utilizzati durante lo svolgimento delle attività sopra elencate.

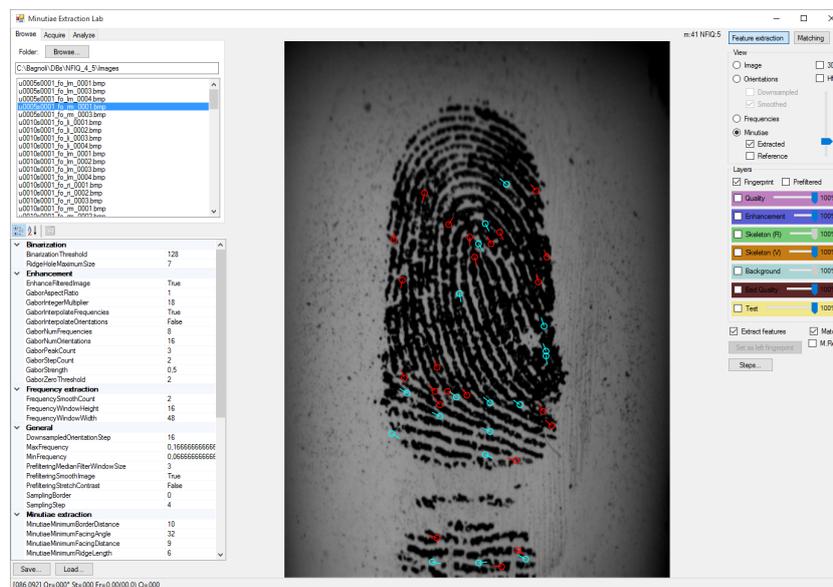


Figura 5.1: Minutiae Extraction Lab. Applicativo che estrae le minuzie con l'algoritmo BioLab su immagini fornite in input. Permette di modificare tutti i parametri che l'algoritmo utilizza, di mostrare lo scheletro delle creste e delle valli, etc.

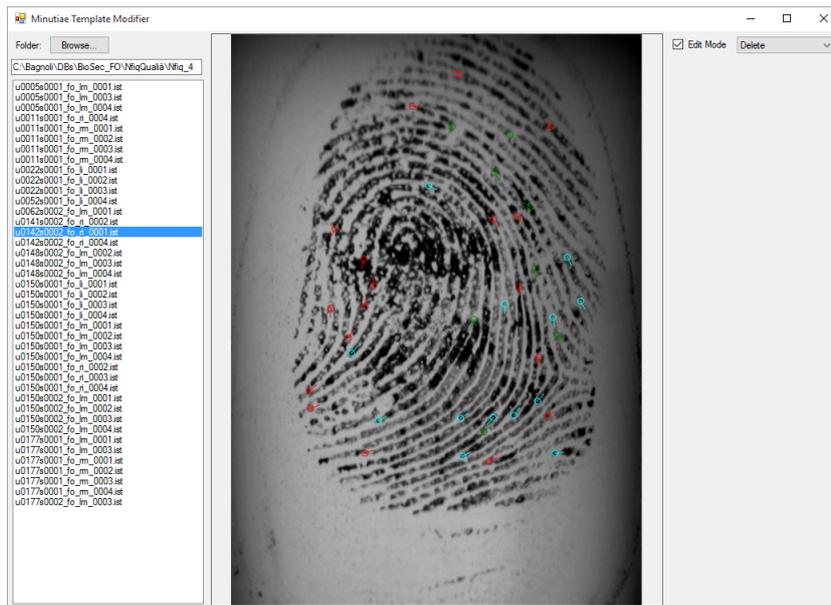


Figura 5.2: Minutiae Template Modifier. Applicativo che permette di modificare gli ISO templates con aggiunta, rimozione o modifica delle posizioni delle minuzie.

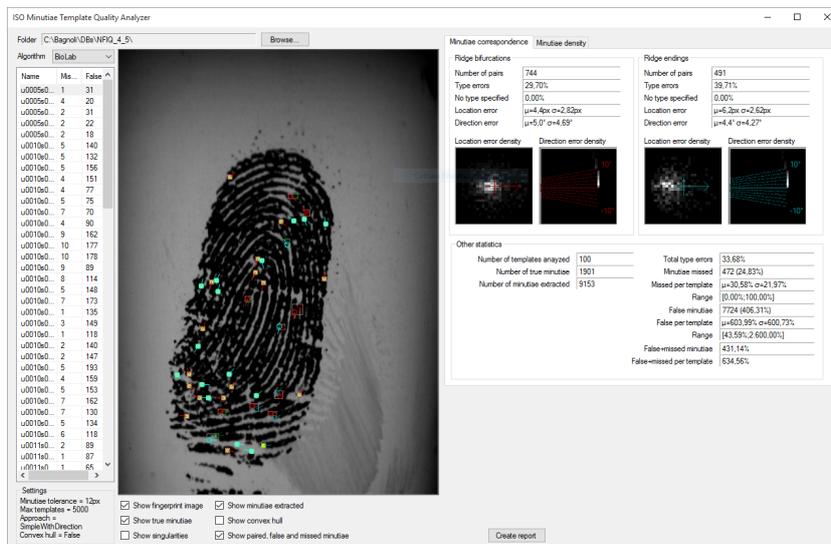


Figura 5.3: ISO Minutiae Template Quality Analyzer. Applicativo che confronta i risultati di un algoritmo a scelta tra BioLab e P088 con il Ground Truth mostrando statistiche sugli errori di estrazione.

6 CONOSCENZE AFFINATE O ACQUISITE

Non avevo mai avuto esperienze o conoscenze pregresse nel campo della biometria, quindi quello delle impronte digitali era un concetto piuttosto astratto per me prima di questa esperienza, la quale mi ha dato la possibilità di imparare molto in questo campo. L'utilizzo di Visual Studio e di C# mi ha permesso di consolidare le scarse conoscenze che avevo dell'intera piattaforma .NET.

7 SINTESI DEI RISULTATI RAGGIUNTI E VALIDITÀ DELL'ESPERIENZA IN RAPPORTO ALLE ASPETTATIVE

I risultati ottenuti durante l'arco delle 150 ore sono in linea con le aspettative e l'obiettivo iniziale di migliorare le prestazioni dell'algoritmo di estrazione è stato in parte centrato. Tuttavia, il lavoro svolto e le prove effettuate non sono finite, e lasciano presupporre che ci siano altri margini di miglioramento, anche provando approcci diversi da quelli utilizzati durante il tirocinio.

8 CONCLUSIONI

In conclusione, posso ritenermi soddisfatto di questa esperienza. L'ambiente di lavoro è risultato gradevole e la costante disponibilità del Prof. Cappelli, tutor del tirocinio, e del Dottor Ferrara, sempre presente nel laboratorio, mi ha permesso di svolgere le mie mansioni al meglio e senza intoppi. Visto l'interesse che ho nei confronti di questa materia, ho deciso di continuare le attività svolte durante il tirocinio e di integrarle nel mio lavoro di tesi.